The Paretian Heritage*

John S. Chipman
Department of Economics
University of Minnesota
Minneapolis, MN. 55455

First published in the *Revue européenne des sciences sociales et Cahiers Vilfredo Pareto*, Tome XIV, 1976, No. 37, pp. 65–173. Reprinted in *Vilfredo Pareto: Critical Assessments of Leading Economists* (edited by John Cunningham Wood and Michael Mclure), London and New York: Routledge, 1999, Volume II, pp. 258–265.

*Opening lecture to commemorate the fiftieth anniversary of Vilfredo Pareto's death, Canadian Economic Association, Kingston, Ont., Canada, June 1, 1973. The present work is a revised and expanded version of the original lecture which was prepared while I was a Fellow at the Center for Advanced Study in Behavioral Sciences, Stanford, California, 1972-73. I wish to thank the Center for the assistance which made this work possible, as well as to thank G. H. Bousquet, H. Scott Gordon, Piet Tommissen, and, most particularly, Paul Samuelson, for valuable comments on the earlier draft; they are, of course, not be held responsible for any defects that remain.

Revised, January 13, 1976

Corrected, January 22, 2010

Contents

Chapter 1

Introduction

It has become increasingly unfashionable to pay any attention to one's forebears; whereas, a generation ago, it was considered a necessary ritual to cite classical authorities even if what they had to say might not be very pertinent, today it is generally assumed that there is very little to be gained from such probing into the past. This view is certainly quite understandable, for a number of reasons. In the first place, scientific economics has been progressing at such an extraordinary rate that the literature of the past twenty years, I would estimate, easily exceeds in quantity the entire literature that preceded it. Secondly, while the recent literature is, Lord knows, not all of unimpeachable quality, the general level of sophistication has certainly also risen. Thirdly, much of classical thought can be considered to have been assimilated into the corpus of received knowledge. Finally, interpretation of the classics is an increasingly difficult thing to do; both their economics and their mathematics bear the same relation to modern economics and mathematics as do ancient Latin and Greek to their modern counterparts.

With respect to this final reason, Pareto is certainly no exception. To interpret his work is no easy task: it requires at least some knowledge of the extraordinary events that took place in Italy in the 1880s and 90s; it requires at least some knowledge of the work of Pareto's contemporaries, notably Walras, Edgeworth, and Fisher; it requires, of course, a knowledge of the languages in which Pareto wrote, namely French, German, and, above all, Italian; and it requires a knowledge of the notations, conventions, and practices of 19th century mathematics, many of which, thankfully for the sake of progress, have long since fallen out of use.

With respect to the first three of the reasons I have listed, on the other hand, Pareto is surely exceptional. First of all, he is perhaps the most prolific economist who has ever lived, with an output exceeding the combined writings of David Ricardo, John Stuart Mill, and John Maynard Keynes; the bibliography of his writings contained in the third volume of his letters to Pantaleoni [195, pp. 476–542] occupies 67 pages, and is still not complete. Secondly, most of Pareto's writings are on a high intellectual level, and their total scope — covering mathematical economics, statistics, economic history, sociology, political science, and scientific method, as well as actuarial science, physics, and pure mathematics — is nothing short of staggering. Thirdly, unlike Ricardo, Mill, Jevons, and Marshall, Pareto never formed a part of the mainstream of economic tradition. He himself had few students of any note;[1] few of his contemporaries were able to follow his analyses; and what influence his work has had seems to have been the result of a few sporadic but notable rediscoveries and developments of fragments of it, among which we may mention those of Barone [18], Bowley [32], Slutsky [223], Schultz [215], Al1en [7], Hicks and Allen [115], Georgescu-Roegen [98], Bergson [23], Little [133], Koopmans [128], and Mandelbrot [140].

One of the unfortunate consequences of the relative inaccessibility of Pareto's work has been the sheer waste and duplication of effort that have resulted. In the important developments that took place in welfare economics between 1934 and 1951 on the part of Lerner [130], Lange [129], and Arrow [13], the principle of what we now call "Pareto optimality"[2] was fully developed, but without reference to Pareto; not until the publication of Little's work [133] did Pareto finally get his due. But the Compensation Principle continued to be credited to Kaldor [121] and Hicks [112], and to this day it is not generally recognized that it a1so is due to Pareto, who introduced it simultaneously with the Pareto Principle in a fundamental article published in 1894 [157]. Likewise, when Bergson [23] introduced the concept of a social welfare function, he gave generous credit to Pareto for the principle that society may be deemed better off if one person is benefited by a measure and no one harmed, i.e. the Pareto Principle;[3] and Bergson's concept was developed in significant ways by Samuelson [208, 211]. However, neither Bergson nor Samuelson was aware that Pareto had himself essentially developed the concept of a social welfare function in 1913 [189].

Examples such as these are quite disturbing. While they serve to strengthen our admiration for Pareto's contributions, they serve at the same time to underscore the futility of his efforts. As proof of the importance of his concepts we can offer the fact that in large part they had to be and were rediscovered independently after being neglected for over half a century. This leads one to speculate about a possible intellectual determinism: good ideas are bound to be developed anyway by somebody else if not by ourselves; and the only justification for our efforts is to hasten the process, and in some cases — as in Pareto's — possibly not even this can be claimed. Perhaps this is too pessimistic an

[1]Pareto nevertheless attracted a band of loyal followers who founded what may be considered a Paretian school in Italy, starting around the second decade of the twentieth century. Among the more prominent of these may be mentioned Bresciani-Turroni, Amoroso, Vinci, and D'Addario. As proof, however, of the assertion in the text, the contributions of this school — which include some valuable developments of Pareto's work on income distribution — have to a large extent suffered the same neglect among English-speaking economists as has Pareto's work itself. Since these contributions may be considered part of the Paretian heritage, some of the more noteworthy among them — dealing with the theory of income distribution — will be discussed in Chapter 4 below.

[2]This term appears to have become current only following the publication of Little's work [133, p. 89], where the term "Pareto 'optimum'" was introduced.

[3]This term was introduced by Arrow in the second edition of [14].

assessment. In any case the point has I hope been made that the process of absorbing new ideas into the corpus of economic knowledge has been a very inefficient one; and that the study of the history of economic doctrines is a legitimate and appropriate, even if diminishing, part of the general enterprise of pushing our subject forward.

In what follows I shall try to make an assessment of Pareto's contributions to economic thought, as well as to trace the interesting evolution of his own thinking. This will be done under four headings: utility theory; welfare economics and international trade; population and income distribution; and time series analysis and methods of interpolation.

Chapter 2

Utility Theory

2.1 The law of demand

In his very earliest contribution to utility theory in 1892 [154], Pareto already displayed concern for the empirical content of utility theory, one of whose objectives was to yield testable hypotheses concerning demand. And the natural question to investigate at the very beginning was the "law of demand," i.e., the proposition that the quantity demanded of a commodity is a decreasing function of its price, other prices being held constant. Pareto was dissatisfied with the Marshallian explanation which was based on the hypothesis which today we would call additive separability of the utility function, that is, the hypothesis that the marginal utility of a commodity is a function of the quantity consumed of that commodity alone, combined with the hypothesis of diminishing marginal utility, and combined further with the hypothesis of constant marginal utility of income. Pareto was ready as a first approximation to accept the first two (additive separability and diminishing marginal utility) but not the third (constant marginal utility of income); given the first two he was able to show [154, Part 2] that the third, if interpreted to mean that the marginal utility of income is independent of prices, implied that marginal utility would have to have the form $\varphi_i(x_i) = a\alpha_i/x_i$ hence the utility function would be log-linear — a result that was rediscovered independently by Samuelson [207] fifty years later.[1] Pareto's proof may be restated as follows [154, Part 2, pp.

[1] See also Samuelson [208, pp. 189–202]. Samuelson [207] actually established a stronger result, not resting on Marshall's assumption of additive separability of the utility function, namely that if the marginal utility of income is independent of prices then the income elasticity of demand must be unitary, hence the utility function must be expressible as a linear function of the logarithm of a function which is homogeneous of first degree. Alternative proofs of this result have since been presented by Wilson [253] and Katzner [124]. Like Pareto before him, Samuelson [207] showed that it is impossible (given the usual assumptions concerning the individual's preference map) for the marginal utility of income to be at the same time independent of both prices and income. Wilson [251] and Samuelson [207] also examined the consequences of an alternative interpretation of constancy of the marginal utility of income; see footnote 2 below.

493–4]. Let $p = (p_1, p_2, \ldots, p_n)$ be the price vector, let I be income, and let $x_i = h_i(p, I)$ be the demand function for the ith commodity, assumed to maximize the utility function $U(x) = U(x_1, x_2, \ldots, x_n)$ subject to the budget constraint $\sum_{i=1}^n p_i x_i = I$.[2] By the separability assumption, the marginal utility of commodity i depends only on the amount of commodity i consumed, i.e., $U_i(x_i) \equiv \partial U(x)/\partial x_i = \varphi_i(x_i)$, where φ_i is a decreasing function by virtue of the assumption of diminishing marginal utility, and therefore has an inverse, φ_i^{-1}. Defining the marginal utility of income by

$$\mu(p, I) = \frac{\varphi_i[h_i(p, I)]}{p_i} \quad (i = 1, 2, \ldots, n), \tag{2.1}$$

and assuming it to depend only on income, i.e.,

$$\mu(p, I) = m(I), \tag{2.2}$$

it follows from the budget identity that

$$I = \sum_{i=1}^n p_i h_i(p, I) = \sum_{i=1}^n p_i \varphi_i^{-1}(p_i m(I)).$$

This equation can be satisfied for all prices p_i only if each term in the sum is independent of the corresponding price, i.e.,

$$p_i \varphi_i^{-1}(p_i m(I)) = A_i(I) \quad (i = 1, 2. \ldots, n),$$

whence from (2.1) and (2.2) it follows that

$$h_i(p, I) = \varphi_i^{-1}(p_i m(I)) = \frac{A_i(I)}{p_i}. \tag{2.3}$$

[2] That this is equivalent to Pareto's formulation may not be apparent on the surface, since Pareto wrote the budget constraint in a form equivalent to $x_1 = \sum_{i=2}^n p_i x_i$ (cf. [154, Part 2, p. 493, formula (5)], where x_1 is the quantity of an "instrumental good" (*bene instrumentale*) which serves as money or unit of account). Pareto went out of his way, however, to emphasize that such an instrumental good did not enter the consumer's utility function; in his words (p. 490): "If one is dealing with an economic good which is not enjoyed directly, but which is used only to produce other economic goods which are consumed, that is, if one is dealing with a good which is *solely* an *instrumental* good, this good does not have a final degree of utility of its own ..., but its degree of utility is equal simply to the common value of the degrees of utility of the goods which are obtained with it." In his subsequent more extended and, unfortunately, also more ambiguous treatment in the *Manuel* [185, Appendix, §§52–9, pp. 579–588], Pareto described commodity 1 above merely as "the commodity whose price is unity, that is to say, money." It is understandable, then, that Wilson [251], basing himself on the treatment in the *Manuel*, interpreted "constancy of the marginal utility of money" to mean constancy, with respect to variations in other prices and income, of the marginal utility of some directly enjoyed good which is chosen as numéraire. He then showed that this could follow if the utility function were chosen to be of the form $U(x) = ax_1 + \psi(x_2, \ldots, x_n)$ — a form which Samuelson [207, p. 85] subsequently showed to be necessary for this result. In a more recent paper, Samuelson [213, pp. 1278–9] has reverted to this second interpretation of constancy of the marginal utility of income, and has also attributed it to Auspitz and Lieben, who employed a utility function of this form in their analysis [17, Appendix II, §2, pp. 470–477], where x_1 (their η) was, however, definitely interpreted as "money" or "cash" (*Bargeld* [17, p. 452]), rather than as Walras's *numéraire* or even Pareto's *bene instrumentale*; that is, in the Auspitz-Lieben formulation one has to assume that fiat money is an argument of the utility function.

With the budget identity this yields $\sum_{i=1}^n A_i(I) = I$ whence, defining $\alpha_i = A_i(1)$ and making use of the fact that h_i is homogeneous of degree 0 in prices and income, we have

$$x_i = h_i(p, I) = \frac{\alpha_i I}{p_i} \quad (i = 1, 2, \ldots, n), \tag{2.4}$$

where $\sum_{i=1}^n \alpha_i = 1$ from the budget identity. Now, observing from (2.2)that $\mu(p_i/I, I) = \mu(p, I)$, we see from (2.1), (2.2), and (2.4) that[3]

$$\varphi\left(\frac{\alpha_i I}{p_i}\right) = p_i m(I), \quad \text{or} \quad \varphi(x_i) = \left(\frac{\alpha_i I}{x_i}\right) m(I); \tag{2.5}$$

and since the left side of the second equation of (2.5) depends on x_i only, so must the right side, hence we must have

$$Im(I) = \text{constant} = m(1) \equiv a,$$

whence the marginal-utility functions must have the form

$$\varphi_i(x_i) = a\frac{\alpha_i}{x_i} \quad (i = 1, 2, \ldots, n), \tag{2.6}$$

and the marginal utility of income is $m(I) = a/I$.

In a previous article [153, pp. 225–6] Pareto had already asserted that the form (2.6) was necessary and sufficient for the marginal utility of income to be independent of prices, but he had only demonstrated the sufficiency. And in an earlier part of his 1892–3 treatise [154, Part 1, p. 413] he had observed that the integral of (2.6) was a linear function of the logarithm of x_i.[4] He remarked further that (2.6) "is a form which is not encountered because it ought to be excluded a priori." He did not, however, write out explicitly the log-linear expression for the total utility function $U(x) = a(\sum_{i=1}^n \alpha_i \log x_i) + b$, which today is very familiar in the form of the logarithm of the "Cobb-Douglas" utility function $B \cdot (\prod_{i=1}^n x_i^{\alpha_i})^a$.[5]

Pareto's discussion of the special form (2.6) of the marginal utility function led him into a lengthy discussion of the St. Petersburg paradox, since

[3] This step in the development is not contained in Pareto's argument — which assumed income to be fixed.

[4] He also remarked that $x_i = 0$ could not be taken as the lower limit of integration, since this would lead to an infinite integral — a remark subsequently made by Samuelson [207, p. 90n] in criticism of Marshall's consumer's surplus concept. Pareto's observation incidentally proves, if proof was necessary, that he could not have been guilty of the elementary error Edgeworth [73] subsequently accused him of in connection with his income distribution formula — see §4.5 below.

[5] He could hardly have been unaware that this was the implied form of the utility function, however, since in his treatise [154, Part 3, p. 152] he had displayed the explicit formula for the total utility function in the general case of independent utilities (additive separability), and had also [154, Part 5, p. 290] considered as an illustration the case in which the marginal utilities had the form $\varphi_i(x) = c_i/(\alpha_i + x_i)$, $i = 1, 2$, and had displayed the corresponding total utility function $U(x_1, x_2) = c_0 + c_1 \log(\alpha_1 + x_1) + c_2 \log(\alpha_2 + x_2)$.

Bernoulli [24] had introduced the logarithmic utility function for his proposed solution; Pareto suggested as an alternative form the right, concave portion of the cumulative normal distribution function [154, Part 4, pp. 15–20]. While he, unfortunately, did not comment explicitly on the fact that this function enjoyed the important property of boundedness, nevertheless his insight correctly anticipated that of Karl Menger [146] 42 years later.

Rejecting the form (2.6) as being much too special, Pareto proceeded to show [154, Part 3, pp. 121–4] that the "law of demand" could be proved on the basis of the first two hypotheses alone (additive separability and diminishing marginal utility).[6] His proof, which was straightforward and elegant, proceeded as follows. Differentiating (2.1) with respect to p_j one obtains

$$\frac{\partial\mu(p,I)}{\partial p_j} = \frac{1}{p_i}\varphi_i'[h_i(p,I)]\frac{\partial h_i(p,I)}{\partial p_j} - \delta_{ij}\frac{1}{p_i^2}\varphi_i[h_i(p,I)],$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 0$ if $i \neq j$ and $= 1$ if $i = j$). It follows that

$$\frac{\partial h_i(p,I)}{\partial p_j} = \frac{p_i}{\varphi_i'[h_i(p,I)]}\frac{\partial\mu(p,I)}{\partial p_j} + \delta_{ij}\frac{\varphi_i[h_i(p,I)]}{p_i\varphi_i'[h_i(p,I)]}. \tag{2.7}$$

From this it is immediate that if $\partial\mu/\partial p_j = 0$ then $\partial h_j/\partial/p_j < 0$; however, Pareto showed that the result followed without assuming μ to be independent of prices. Multiplying (2.7) by p_i and summing the resulting equations for $i = 1, 2, \ldots, n$ one obtains, taking account of the fact that differentiation of the budget identity $\sum_{i=1}^n p_i h_i(p,I) = I$ with respect to p_j yields

$$h_j(p,I) + \sum_{i=1}^n p_i \frac{\partial}{\partial p_j} h_i(p,I) = 0$$

the following expression for $\partial\mu/\partial p_j$:

$$\frac{\partial\mu(p,I)}{\partial p_j} = -\frac{h_j(p,I) + \varphi_j[h_j(p,I)]/\varphi_j'[h_j(p,I)]}{\sum_{i=1}^n \{p_i^2/\varphi_i'[h_i(p,I)]\}}. \tag{2.8}$$

This equation provides a second proof of Pareto's proposition that constancy of the marginal utility of income (with respect to prices) implies that marginal utility must have the form (2.6); this follows simply from integrating the differential equation

$$\frac{d\log\varphi_j(x_j)}{d\log x_j} = \frac{x\varphi_j'(x_j)}{\varphi_j(x_j)} = -1$$

[6] An obvious generalization of this result, involving the assumption of additive separability as between the commodity in question and the remaining commodities, was derived by Wilson [250]. Actually, Pareto's theorem can be proved in much more elementary fashion by observing from the condition $\varphi_i(x_i)/\varphi_j(x_j) = p_i/p_j$ that if p_i rises — other prices and income remaining unchanged — then if x_i fails to fall, all x_js ($j \neq i$) must rise, contradicting the budget equation. An only partially successful attempt along these lines was made by Walras in 1892; see the Jaffé translation [242, p. 466], and Stigler [229, pp. 320–1].

which is obtained by setting the numerator of the expression (2.8) equal to zero.[7]

To obtain the "law of demand" Pareto now substitutes (2.8) back in (2.7) for $i = j$, to obtain

$$\frac{\partial h_j(p,I)}{\partial p_j} = \frac{-p_j h_j(p,I) + \dfrac{\varphi_j[h_j(p,I)]}{p_j}\displaystyle\sum_{i\neq j}\frac{p_i^2}{\varphi_j'[h_i(p,I)]}}{\varphi_j'[h_j(p,I)]\displaystyle\sum_{i=1}^{n}\frac{p_i^2}{\varphi_i'[h_i(p,I)]}}, \tag{2.9}$$

and this is negative.

Pareto went on to relax the assumption of additive separability, and adopted Edgeworth's [67] assumption that the marginal utility of any commodity is a function of the quantities of all the commodities consumed. First of all he presented a brief discussion of the question of constancy of the marginal utility of income [154, Part 2, pp. 494–6]; however, he failed to obtain the characterization subsequently obtained by Samuelson [207], namely that the income elasticity of demand would have to be unitary and that the utility function would have to be of the form $U(x) = a\log\Phi(x) + b$, where $\Phi(x)$ is homogeneous of degree 1. Pareto proceeded, however, to investigate the "law of demand" under these more general conditions [154, Part 5, pp. 304–6]. His procedure may be interpreted as follows. Denote $U_j = \partial U/\partial x_j$ and $U_{jk} = \partial^2 U/\partial x_j \partial x_k$. As before, one has the budget identity

$$\sum_{k=1}^{n} p_k h_k(p,I) = I, \tag{2.10}$$

and the marginal utility of income is now defined by

$$\mu(p,I) = U_j[h(p,I)]/p_j \quad (j = 1,2,\ldots,n). \tag{2.11}$$

Differentiating (2.10) and (2.11) with respect to p_i one obtains the equations

$$\begin{cases} h_i + \displaystyle\sum_{k=1}^{n} p_k \frac{\partial h_k}{\partial p_i} = 0 \\ \displaystyle\sum_{k=1}^{n} U_{jk}\frac{\partial h_k}{\partial p_i} = \frac{\partial \mu}{\partial p_i}p_j + \mu\delta_{ij} \quad (j = 1,2,\ldots,n) \end{cases} \tag{2.12}$$

where δ_{ij} is the Kronecker delta.

[7] According to Wilson [251], Pareto made a slip at this point, failing (according to him) to note that constancy of the marginal utility of money, which Wilson identified with φ_1, would imply that $\varphi_1' = 0$, hence that the denominator of the expression on the right side of (2.8) was infinite; on this basis Wilson rejected Pareto's argument (presented in the *Manuel* [185, Appendix §58, pp. 585–6]) to the effect that constancy of the marginal utility of income implied that the numerator of (2.8) would have to vanish. However, Wilson seems to have (quite understandably) misinterpreted Pareto on this point, as explained in footnote 2 above. On this and related points see also the discussion in Georgescu-Roegen [100, p. 179].

The well-known modern procedure, first employed by Slutsky [223], is to solve the system (2.12) simultaneously in the form

$$\begin{bmatrix} 0 & p_1 & p_2 & \cdots & p_i & \cdots & p_n \\ p_1 & U_{11} & U_{12} & \cdots & U_{1i} & \cdots & U_{1n} \\ p_2 & U_{21} & U_{22} & \cdots & U_{2i} & \cdots & U_{2n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ p_i & U_{i1} & U_{i2} & \cdots & U_{ii} & \cdots & U_{in} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ p_n & U_{n1} & U_{n2} & \cdots & U_{ni} & \cdots & U_{nn} \end{bmatrix} \begin{bmatrix} -\partial\mu/\partial p_i \\ \partial h_1/\partial p_i \\ \partial h_2/\partial p_i \\ \vdots \\ \partial h_i/\partial p_i \\ \vdots \\ \partial h_n/\partial p_i \end{bmatrix} = \begin{bmatrix} -h_i \\ 0 \\ 0 \\ \vdots \\ \mu \\ \vdots \\ 0 \end{bmatrix} \tag{2.13}$$

yielding the expression

$$\frac{\partial h_j}{\partial p_i} = \frac{-h_i B_{0j} + \mu B_{ij}}{B}, \tag{2.14}$$

where B is the determinant of the bordered matrix of (2.13) and B_{ij} the cofactor of its ith row and jth column, $i, j = 0, 1, \ldots, n$. By a similar procedure — differentiating (2.10) and (2.11) with respect to I — one then obtains $\partial h_j/\partial I = B_{0j}/B$, yielding the symmetric Slutsky term

$$s_{ji} = \frac{\mu B_{ij}}{B} = \frac{\partial h_j}{\partial p_i} + \frac{h_i \partial h_j}{\partial I}.$$

Pareto's procedure was instead to solve the last n equations of (2.12) first, and then substitute the results in the first equation, yielding the much more awkward expression

$$\frac{\partial h_j}{\partial p_i} = \frac{N^j}{B}\left[h_i + \mu\left(\frac{N_i}{D} - \frac{D_{ij}B}{DN^j}\right)\right] = \frac{N^j}{B}h_i + \left(\frac{N^j N_i}{BD} - \frac{D_{ij}}{D}\right)\mu, \tag{2.15}$$

where D is the determinant of $[U_{ij}]$, D_{ij} the cofactor of its ith row and jth column, and

$$\begin{aligned} N_i &= \sum_{k=1}^{n} p_k D_{ik}, \quad N^j = \sum_{k=1}^{n} p_k D_{kj}, \\ B &= \sum_{j=1}^{n} p_j N_j = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j B_{ij} \end{aligned} \tag{2.16}$$

(cf. [154, Part 5, p. 306]). Pareto stopped at (2.15) and did not try to interpret the coefficients of h_i and μ; in fact he did not notice that symmetry of D, which implies $N_i = N^i$, would imply symmetry of the second term in the third expression in (2.15) (the substitution term). He therefore stopped just short of establishing the Slutsky equation.[8] He did succeed, however, in showing that

[8]Slutsky [223] later showed how (2.15) could be reduced to (2.14) by means of Jacobi's theorem on determinants.

in the special case in which $U_{ij} = 0$ for $i \neq j$ and $U_{ii} < 0$, (2.15) reduced to the expression he had previously obtained [154, Part 3, p. 124] and that the coefficients of μ and h_i were both negative for $i = j$; i.e. that the "law of demand" held for this case.

Pareto's investigations into the nature of demand functions were not restricted to the question of whether utility maximization implied that the demand for a commodity was a decreasing function of its price. He also, in his last major theoretical work [188, pp. 615–620], took up questions such as that of whether offer curves could be linear, or whether a demand function could have the constant-elasticity form $x_1 = Cp_1^{-n}$ (other prices and income being held constant) put forward by Marshall [145, Mathematical Appendix, Note IV of 1st ed., Note III of 2nd and subsequent editions]. While Pareto's treatment of the latter problem required subsequent correction by Wilson [252], who showed that the class of utility functions that could generate such a demand function was somewhat wider than Pareto thought, the point that should be stressed is that Pareto was apparently the first to even raise, let alone tackle, the question of whether certain functional forms frequently employed were compatible with the assumption of utility maximization on the part of a single consumer. This type of investigation set Pareto apart from his contemporaries and predecessors, who for the most part showed little interest in examining the empirical implications of utility theory. As Stigler has so well observed [229, p. 395]: "In this respect Pareto was the great and honorable exception. Despite much backsliding and digression, he displayed a powerful instinct to derive the refutable empirical implications of economic hypotheses ... Pareto — and he alone of the economists — constantly pressed in this direction."

2.2 Ordinal utility and the theory of related commodities

It was in 1898 that Pareto made the discovery that measurability of utility could be dispensed with in the explanation of consumer behavior in competitive markets (cf. [172]). This idea was further elaborated in a draft of a new Treatise [174] which never materialized as such; here, Pareto provided a beautiful exposition of the subject which was not to be matched again for its clarity until Hicks's exposition 39 years later [113]. Pleasure and pain were replaced by preference and "the fact of choice;" utility functions were replaced by "index functions," unique only up to monotone increasing transformations. This aspect of Pareto's work was manifestly influential, having provided the impetus for Allen's [7] contribution to the theory of exchange and for Hicks and Allen's [115] pioneering contribution to the theory of consumer behavior. It should nevertheless be mentioned that Pareto's ideas on this subject did not become generally known to the profession until after the publication of the *Manuel* [185] and the *Encyclopédie* article [188] a decade later, by which time they were already in the process of being independently rediscovered by Johnson [119].[9]

Also of unquestioned influence was Pareto's theory of interrelated demands elaborated in Chapter 4 and the Appendices of the *Manuale* [182, §§12–16, pp. 505–510] and the *Manuel* [185, §§47–49, pp. 575–578], with its celebrated discussion of complementarity and substitutability — which Pareto described as "dependence of the first and second kinds," respectively. The definition used by Pareto — $\partial^2 U/\partial x_i \partial x_j > 0$ for complementarity and < 0 for substitutability between commodities i and j — was not new, having originated with Auspitz and Lieben [17, §§36, 41; Appendix II, §2, p. 482], and having been adopted by Edgeworth [74, p. 21n]; nor did Pareto make any claims of novelty. Moreover, the influence it had on subsequent developments was due as much or more to what was left unsaid as to what was said — to the questions Pareto's treatment raised rather than resolved. For, Allen, who perhaps did more than anyone else to bring Paretian thought into the mainstream of English economics, soon noticed the incongruity between Pareto's position on ordinal utility and his adherence to a concept of related commodities that depended on the adoption of a particular cardinal utility index. Specifically, he pointed out [8, p. 171n], [115, p. 60n] that the signs of the second-order partial derivatives of a utility function were not invariant with respect to monotonic transformations. This led to the well-known alternative criteria for complementarity and substitutability put forward by Hicks and Allen [115] and Hicks [113], in which Johnson's [119] contributions had also played an important role.

The net result of these developments as far as Pareto's work was concerned has been, until recently at least, a rather negative assessment. Thus, Hicks [113, p. 43] maintained that "the Edgeworth-Pareto definition sins against Pareto's own principle of the immeasurability of utility." Stigler [229, p. 385] stated that "Pareto was inconsistent; he made extensive use of [the Auspitz-Lieben definition of complementarity] at the same time that he was rejecting the measurability of utility." And sharpest of all is Samuelson's recent assessment [213, p. 1280]: "It is puzzling to understand this inconsistency of using in his literary text the non-invariant cross derivative of a utility concept which he had already thrown away, both in his mathematical appendix and text. Undoubtedly, Pareto was confused."[10]

It appears, however, that most such assessments result from reading Pareto

[9] Priority for the recognition of the inessentiality of measurability of utility in consumer demand theory must nevertheless be ascribed to Antonelli [12].

[10] Again [213, p. 1256n]: "As is well known, Pareto was inconsistent in espousing $\left[\text{the criterion } \partial^2 U/\partial x_i \partial x_j \gtreqless 0\right]$ *after* he had given up *cardinal* utility in favor of better-or-worse, *ordinal* utility." On the other hand, despite such passages Samuelson's treatment is largely appreciative of Pareto's contribution, as when he says (p. 1277) : "We also see that the Pareto error, of continuing to talk about utility complementarity after he became an ordinalist, was not quite so bad an error after all." And his defense of Pareto in [213, p. 1281], in which he puts forward "another possibility other than utter confusion to rationalize Pareto's seeming inconsistencies," corresponds precisely to the interpretation I shall set out below (see also [51, p. 374n]) in support of my contention that Pareto, after all, was not in error; in fact, Professor Samuelson has authorized me to state that he now accepts this interpretation as being the more probable one.

through Hicksian glasses, rather than from taking him at his own word. The Hicksian position is well summarized by the following passage in *Value and Capital* [113, p. 18]: "Pareto's discovery only opens a door, which we can enter or not as we feel inclined. But from the technical point of view there are strong reasons for supposing that we ought to enter it. The quantitative concept of utility is not necessary to explain market phenomena. Therefore, on the principle of Occam's razor, it is better to do without it."

Pareto's position was quite different from that of Hicks. In the first place, he continued to believe that pleasure was in principle capable of measurement, even if such measurement was inessential to the explanation of economic equilibrium. He had previously studied the Bernoulli principle in the theory of behavior under risk [154, Part 4], where measurability is essential, as the later work of von Neumann and Morgenstern [147] amply confirmed. He had also, as we have seen, studied the Jevons-Walras-Marshall case of additively separable utility, and had shown that, when combined with the assumption of diminishing marginal utility, it had empirical implications — namely that demand for a commodity was an increasing function of income and a decreasing function of its price — not shared by all cases in the more general Edgeworth [67] class of utility functions. He had remarked [182, Appendix, §9, p. 501] that given any additively separable utility function, with the marginal utility of each commodity depending only on the amount of that commodity, there would always exist an equivalent utility index (a monotone increasing function of the given one) such that the marginal utility of each commodity depended on the quantities of all the commodities consumed; and that the converse was true *in some cases.* In those cases — provided it exhibited diminishing marginal utility for each commodity — he gave a privileged role to the additively separable utility index, which he regarded as the true measure of "ophelimity" (see especially [183]). In his recognition of the significance of the additive, or "cardinal", structure of such utility functions, Pareto's insight has also been supported by later developments, notably that of Debreu [66]. While we can agree with Hicks (as Pareto would have done himself) that, in the case in which an additively separable utility indicator exists, exhibiting diminishing marginal utility for each commodity, any monotone transformation of this indicator will do equally well in explaining the particular market phenomenon of equilibrium, nevertheless, the knowledge that preferences belong to this particular class — which can be represented by a utility function with a convenient additive structure, and which carries empirical implications — is, as Samuelson [207, Ch. 5] has stressed, extremely valuable in comparative statics and in making predictions — which is, or at least ought to be, what the subject is all about.

Thus, altering slightly a suggestive metaphor due to Georgescu-Roegen [99, p. 2n], just because the equilibrium of a table is determined by three of its legs, we are not required by any scientific principle to assume that actual tables have only three legs, especially if direct observation suggests that they may have four.

A second thing to keep in mind in contrasting Pareto's position with that of Hicks is that, in taking up the analysis of related commodities in terms of the signs of the partial derivatives $U_{ij} = \partial^2 U/\partial x_i \partial x_j$, Pareto was doing so *by* *way of departing from* the Jevons-Walras-Marshall assumption of independent commodities (additively separable utilities). Ironically, then, his analysis of complementarity and substitutability was a step away from the measurement structure inherent in additively separable utility. However, Pareto did not want to abandon all of the measurement structure in the process, because he recognized that it contained empirical implications. In particular — and this seems to have been overlooked by his critics — he supplemented his definition of complementarity by the requirement that for any set of mutually complementary (or independent) commodities $C_j, j \in J$ (J being some subset of the set of integers $\{1, 2, \ldots, n\}$), marginal utility should be strongly decreasing along any ray $\xi(t) = (\alpha_1 t, \alpha_2 t, \ldots, \alpha_n t)$ such that $\alpha_j > 0$ for $j \in J$ and $\alpha_j = 0$ otherwise, i.e., $d^2U(\xi(t))/dt^2 < 0$ for $t > 0$ [185, Ch. IV, §42, p. 270; Appendix, §48, p. 577]. Believing that this implied the negative definiteness of the Hessian matrix $[U_{ij}(x)], i, j \in J$, he specifically asserted that, for such utility functions, if all commodities were either independent or complementary to each other, i.e., if $U_{ij} \geqq 0$ for $i \neq j$, then the demand for each commodity would be a decreasing function of its price and an increasing function of income [182, Ch. IV, §49, pp. 260–1], [185, pp. 272–3]. Unfortunately his proof of negative definiteness of the Hessian was faulty (see §3.5 below), and he does not appear to have provided a proof of the proposition, even assuming negative definiteness to hold; however, a proof of the latter proposition can be readily supplied,[11] thus justifying Pareto's intuition that his cardinal criterion, when combined with a strong concavity assumption, does, after all, have meaningful empirical implications.

Pareto did provide a precise criterion for measurability of utility [182, Ch. IV, §32, pp. 252–3], [185, pp. 264–5] — one that was subsequently referred to and taken up by von Neumann and Morgenstern [147, pp. 18, 23]. This was based on the hypothesis that an individual could compare differences in pleasure and say whether the pleasure in moving from x^1 to x^2 is greater than the pleasure obtained in moving from x^2 to x^3. He noted in particular that if it were possible at the limit to find a bundle x^2 such that the preference for x^2 over x^1 was just equal to the preference x^3 over x^2, one would be entitled to say that the pleasure obtained in passing from x^1 to x^3 was double the pleasure obtained in passing from x^1 to x^2. He expressed strong reservations, however, about the possibility of an individual's being able to arrive at this much precision.[12] Nevertheless,

[11] For a proof see Chipman [50]. It seems to be an open question whether the negative definiteness does in fact follow from Pareto's assumptions — which might need to be supplemented by a quasi-concavity assumption.

[12] It is perhaps for this reason that Stigler concluded [229, p. 381]: "But Pareto believed the consumer could not rank utility differences." However, Pareto expressly stated that the consumer *could* rank at least *some* utility differences; what he was reluctant to assume was the kind of continuity or Archimedean axiom that would permit the construction of an interval scale. For discussion of these questions see Chipman [48] as well as the papers by Ragnar Frisch and Franz Alt translated in [51]. It should be added, however, that even if Pareto's reluctance to assume the continuity axiom is granted, it by no means follows that the construction of a cardinal utility scale is not possible; on the contrary, these threshold problems are very well known in mathematical psychology where methods have been developed, based on the concept of "just noticeable differences," to construct such scales precisely under the kinds of conditions specified by Pareto.

he certainly believed that the strict comparisons could often be made, and concluded as follows:

> Among the infinite number of systems of indices that are possible, only those should be retained that have the property that if more pleasure is experienced in passing from I to II than from II to III, the difference between the indices of I and II should be greater than that between the indices of II and III. In this manner the indices will always provide a better representation of ophelimity.

It was only after providing this explicit statement concerning the numerical representation of utility that Pareto proceeded with his theory of related commodities. And he provided an elaborate scheme [182, Ch. IV, §§66–68] for the hierarchy of commodities that foreshadowed in a remarkable way the modern developments along these lines carried out by Strotz [230] and others.

One interesting development of Pareto's theory deserves to be mentioned, particularly because it brings out in a special case the significance of the considerations mentioned above. Schultz [215] adopted Pareto's definitions of complementarity and substitutability, and sought to apply these concepts empirically. In doing so he added an additional assumption, namely constancy of the marginal utility of income. It is not hard to see from his development that, implicitly or otherwise, this meant for him constancy in the Marshallian sense, i.e., independence of prices. Schultz worked with the *inverse* demand functions, expressing real prices as functions of the quantities consumed, i.e.,

$$\frac{p_i}{I} = \frac{U_i(x)}{\displaystyle\sum_{k=1}^{n} U_k(x)x_k} \equiv \psi_i(x) \quad (i = 1, 2, \ldots, n). \tag{2.17}$$

Now it was shown by Samuelson [207, p. 84] that if the marginal utility of income is independent of prices, it must have the form $\mu(p, I) = m(I) = a/I$, where $a = m(1)$; accordingly,

$$\frac{p_i}{I} = \frac{U_i(x)}{Im(I)} = \frac{1}{a}U_i(x), \tag{2.18}$$

so that (2.17) has the simple form

$$\frac{p_i}{I} = \frac{1}{a}U_i(x) = \psi_i(x). \tag{2.19}$$

It follows that $U_{ij}(x)$ has the same sign as $\psi_{ij}(x)$, and Pareto's criterion can be applied to the empirically estimable inverse demand functions (2.17).

Schultz followed Marshall [145, 8th ed., p. 335] in justifying the assumption that the marginal utility of income was, in his words [215, p. 474], "sensibly constant," by saying that "the expenditure of the individual for any one commodity constitutes a small fraction of his income." Such a justification does not appear to have been warranted; but in any event Schultz's procedure is quite legitimate if it is assumed that preferences are homothetic and the utility index is chosen to be a linear function of the logarithm of a function which is homogeneous of degree 1. When preferences are homothetic, this is always possible; and being possible, it is also convenient. Such a log-homogeneous utility function may therefore be regarded as a natural cardinal utility representation of the consumer's preferences, satisfying the strong concavity conditions postulated by Pareto for the case of complementary commodities.[13]

2.3 Integrability

From the time of his earliest investigations into the theory of the consumer [154], Pareto questioned whether a total utility function could properly be assumed to exist, and maintained that such an assumption was unnecessary to explain demand behavior. His point of departure [154, Part 1, pp. 414–5] was the proposition that the individual's mind was capable of absorbing the idea of small variations in utility (marginal utility) but not large ones (total utility):

> None of us has a clear conception of the total utility of eating, drinking, being clothed, and having a house to protect us; we understand the advantages of these things only in terms of small variations: more or less. In other words, our mind is able to absorb only the notion of the final degree of utility.

Thus, the concept of marginal utility ("final degree of utility" in Jevons's [118] terminology, or "elementary ophelimity" in Pareto's subsequent terminology [163]) was regarded as the basic primitive concept. While this concept was later to be set aside, along with his abandonment of measurable utility — in his analysis of the equilibrium of the consumer, but not elsewhere — this did not essentially change his views regarding the local or myopic nature of the consumer's perception of the utility of consumption. The question naturally arose in Pareto's mind as to whether (a) the assumption that a total utility function exists which the consumer maximizes, or acts as if he is maximizing, is needed in the theory of the consumer, and (b) the existence of such a total utility function can be deduced from the partial differential equations characterizing the equilibrium conditions of the consumer in the market. The second of these is the integrability problem.

[13] In terms of the framework recently put forward by Samuelson [213], this would correspond to taking what he calls the "Weber-Fechnerized money metric" (p. 1263). I might suggest "Dupuit-Marshallized" as an alternative terminology for Schultz's case, since it is the form that is needed to justify consumer's surplus analysis. The development presented in the text suggests that Schultz's treatment, despite its blunder of specifying linear forms for the inverse demand functions, is deserving of more sympathetic interpretation than allowed by Samuelson [213, p. 1284]. Samuelson in his paper also made the valuable observation [213, pp. 1272, 1277] that one can employ an ordinally invariant criterion such as $U_{ij}/U_i - U_{kj}/U_k > 0$, which can be interpreted to mean that "commodity i is more complementary to commodity j than is commodity k." See also Georgescu-Roegen's seminal contribution [99], which establishes a formal identity between the Pareto and Hicks-Allen definitions of complementarity.

Given the original and path-breaking nature of Pareto's approach to the theory of the consumer, it is all the more disappointing that his own solutions to the problems he posed were far from satisfactory, and contained a number of technical defects and confusions. In particular, Pareto kept confusing two quite different conditions, both of which are, unfortunately, often referred to in the mathematical literature by the same term "integrability conditions," although the term more properly applies only to the first. These are:

(1) The conditions that must be satisfied by the marginal utility functions φ_i to ensure the existence of a non-constant solution U to the system of partial differential equations

$$\varphi_i(x)\frac{\partial U(x)}{\partial x_1} - \varphi_1(x)\frac{\partial U(x)}{\partial x_i} = 0 \quad (i = 2, 3, \ldots, n); \tag{2.20}$$

or equivalently, the conditions the φ_i's must satisfy in order to ensure the existence of a solution (or "integral") of the "total differential equation"

$$\varphi_1(x)dx_1 + \varphi_2(x)dx_2 + \cdots + \varphi_n(x)dx_n = 0, \tag{2.21}$$

by which is meant that there exists an integrating factor $\iota(x)$ (not identically zero) such that the product $\iota(x)\sum_{i=1}^n \varphi_i(x)$ is an exact differential, i.e., that there exists a non-constant function U such that

$$dU(x) = \iota(x)[\varphi_1(x)dx_1 + \varphi_2(x)dx_2 + \cdots + \varphi_n(x)dx_n], \tag{2.22}$$

so that $U_i(x) \equiv \partial U(x)/\partial x_i = \iota(x)\varphi_i(x)$.

(2) The conditions under which the total differential equation (2.21) is exact, i.e., is already the differential of some function U, so that $\iota(x) = 1$ in (2.22) is an integrating factor.

The confusion between these two concepts runs through all of Pareto's writings on the integrability problem, from his early 1892–3 treatise [154] right up to his last major theoretical work [188]. He sometimes indicated an awareness of the distinction, but more often than not he forgot. This curious gap in his otherwise excellent grasp of mathematics seems to be what is mainly responsible for the unsatisfactory nature of his treatment of the integrability problem.

A succession of slips made in his early treatise [154] — spotted and corrected by Stigler [229, pp. 379–80] in his penetrating survey — is indicative of Pareto's unsure grasp of this subject. The first was the assertion [154, Part I, p. 415] that if, in the case $n = 2$, $\varphi_1(x_1, x_2)$ and $\varphi_2(x_1, x_2)$ were not partial derivatives of one and the same function, a total utility function $U(x_1, x_2)$ would not exist; whereas in the case $n = 2$ it is well known that there always exists an integrating factor $\iota(x_1, x_2)$ such that $\partial(\iota\varphi_1)/\partial x_2 = \partial(\iota\varphi_2)/\partial x_1$ and hence a function U such that $\partial U/\partial x_i = \iota\varphi_i$ for $i = 1, 2$. This was presumably a minor lapse on Pareto's part, since in the final installment of the same treatise [154, Part 5, pp. 296–300] he remarked that the total differential equation (2.21) need not be integrable if $n > 2$ (p. 297), and is always integrable when $n = 2$ (p. 299); however, for the case $n = 3$ he furnished the wrong integrability conditions, namely

$$\frac{\partial\varphi_i}{\partial x_j} = \frac{\partial\varphi_j}{\partial x_i} \quad (i \neq j), \tag{2.23}$$

which are necessary and sufficient conditions for the differential equation (2.21) to be exact, and are thus sufficient but not necessary for integrability in sense (1) above. The correct integrability conditions, subsequently specified by Evans [82, p. 120], Allen [7, pp. 222–3n], and Hotelling [116, p. 592], and in an equivalent form by Hicks and Allen [115, p. 211n], are

(2.24)
$$\varphi_i\left(\frac{\partial\varphi_j}{\partial x_k}-\frac{\partial\varphi_k}{\partial x_j}\right)+\varphi_j\left(\frac{\partial\varphi_k}{\partial x_k}-\frac{\partial\varphi_i}{\partial x_k}\right)+\varphi_k\left(\frac{\partial\varphi_i}{\partial x_j}-\frac{\partial\varphi_j}{\partial x_j}\right)=0\ (i\neq j\neq k);$$

in fact, these are equivalent to the conditions that had already been stipulated by Antonelli [12].[14]

Pareto always associated lack of integrability with a state of affairs in which preferences depend on the order of consumption, i.e., in which the value of a line integral

$$\int_0^1\sum_{i=1}^n\varphi_i(\xi(t))d\xi_i(t) \tag{2.25}$$

along a path $\xi(t), 0 \leqq t \leqq 1$ connecting two points $\xi(0) = x^0$ and $\xi(1) = x^1$ depends on the particular path ξ chosen. Statements to this effect will be found in all his major works on utility theory, starting from his early treatise [154, Part 5, pp. 280–1], going on to his seminal work in which he abandoned measurable utility [174, Part 2, p. 521], again in his German Encyclopaedia article [178, pp. 1103–4], and culminating in his notorious article on "non-closed cycles" [181], the main argument of which was reproduced in the *Manuel* [185, Appendix, §§12–61, pp. 546–557]. Pareto's critics have rightly taken him to task for his literal interpretation of such integral paths as indicating the temporal order in which an individual consumes the commodities he has purchased, as if the consumer "eats his way" along the path (cf. Wilson [248, p. 468], Wicksell [245], Wold [255, Part 1, pp. 115–6], Stigler [229, pp. 380–1], Samuelson [210, pp. 361–2]). In fairness to Pareto it should be stated that he nearly always qualified his discussions on this subject by saying that they were of purely psychological interest and of no relevance to the problem of economic equilibrium, sometimes quite emphatically, as when he stated [188, §18, p. 414]: "It must be pointed out that these investigations remain foreign to the determination of equilibrium." What has not been so clearly recognized is that there is a much

[14] The equivalent form by Hicks and Allen was the one appropriate to the differential equation obtained by replacing the marginal utilities φ_i of (2.21) by the marginal rates of substitution $R_i(x) = \varphi_i(x)/\varphi_1(x)$, and reduces to

$$\frac{\partial R_j}{\partial x_k}-\frac{\partial R_k}{\partial x_j}+R_j\frac{\partial R_k}{\partial x_i}-R_k\frac{\partial R_j}{\partial x_i}=0\quad(i\neq j\neq k). \tag{2.24$'$}$$

It is in this form (for $i = 1$) that the conditions were stated by Antonelli; see p. 347, formula (21b) of the English translation of [12]. For discussion of the conditions see also Wold [255, Part 1, pp. 114–5] and Stigler [229, pp. 379n, 380n]. The conditions are well known (and were well known in Pareto's time), and can be found in standard mathematical textbooks, e.g., Wilson [247, p. 255]; what was new, and so long in coming, was recognition of their relevance to the particular economic problem.

more fundamental logical objection to Pareto's theory of open and closed cycles which remains, even if his interpretation of path-dependence of line integrals is overlooked.

The fact is that the necessary and sufficient conditions for a line integral (2.25) connecting x^0 and x^1 to be independent of the path ξ connecting those points are precisely the conditions (2.23) for the total differential equation (2.21) to be exact. The problem of path-independence arises just as much in the two-commodity case as in the general case, and is therefore irrelevant to the problem of integrability. The integrability problem is the problem of whether there exists an integrating factor ι such that the value of the line integral

$$\int_0^1 \sum_{i=1}^n \iota(\xi(t))\varphi_i(\xi(t))d\xi_i(t) \tag{2.26}$$

is independent of the path ξ joining x^0 and x^1. In the case $n = 2$ such an integrating factor always exists; but if one computes the line integral (2.25) neglecting first to multiply each term $\varphi_i(\xi(t))$ by $\iota(\xi(t))$, one will obtain path-dependence and "open cycles," even in the case $n = 2$. Unfortunately this is just what Pareto did.

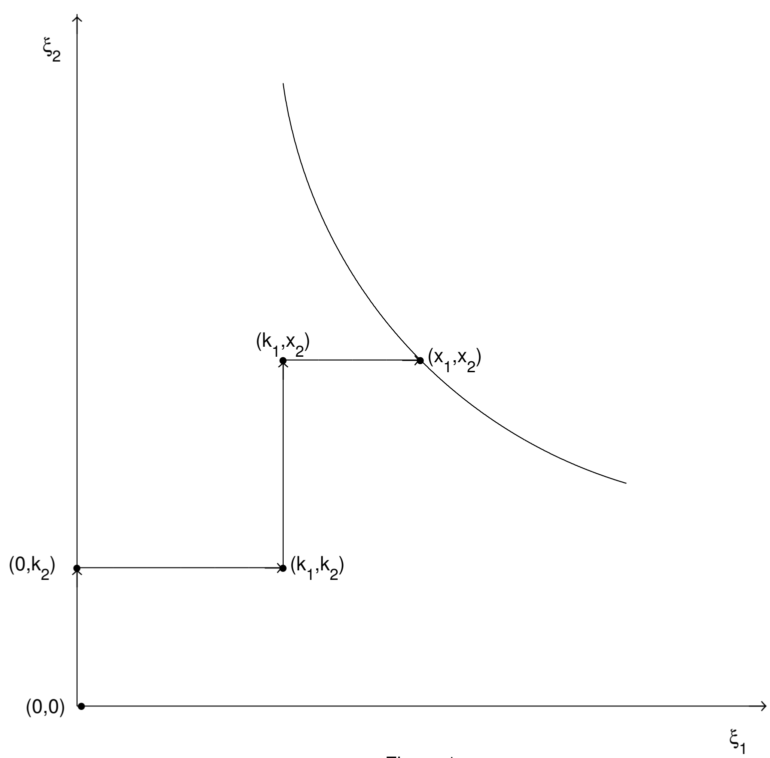

Figure 1

In his 1900 work [174], Pareto treated the problem of going from indifference directions to indifference maps and utility functions, but in a very intuitive and geometric manner. A similar treatment was contained in the Appendix of the *Manuale* [182], but with no recognition of the need to assume integrability conditions. Volterra [239] in his famous review of the *Manuale* raised this question, and Pareto's detailed response [183] to Volterra's suggestion is of interest to historians of economic thought mainly because of the basic misconception it reveals in Pareto's approach to the integrability problem. An explanation of the nature of this misconception may therefore be worth providing, if only to clear away the mystery that has puzzled readers of Pareto's works for so long.

The situation is best brought out by considering the case $n = 2$, where no integrability problem arises. Pareto considered polygonal paths $\xi(t) = (\xi_1(t), \xi_2(t))$ joining the origin (0,0) and some given point (x_1, x_2), of the form (where we define $s = x_1 + x_2$)

(2.27)
$$(\xi_1(t), \xi_2(t)) = \begin{cases} (0, st) & \text{for } 0 \leqq t \leqq k_2/s; \\ (st - k_2, k_2) & \text{for } k_2/s \leqq t \leqq (k_1 + k_2)/s; \\ (k_1, st - k_1) & \text{for } (k_1 + k_2)/s \leqq t \leqq (k_1 + x_2)/s; \\ (st - x_2, x_2) & \text{for } (k_1 + x_2)/s \leqq t \leqq 1, \end{cases}$$

where $0 < k_1 < x_1$ and $0 < k_2 < x_2$ (see Figure 1). He then considered an example of a total differential equation

$$\varphi_1(x_1, x_2)dx_1 + \varphi_2(x_1, x_2)dx_2 = 0 \tag{2.28}$$

which was not exact [183, p. 23, formula (11)]; in his words: "The integrability conditions are not satisfied; therefore the ophelimity depends on the order of consumption." (By "integrability" he could only have meant "exactness," since a page later he found an integrating factor and a solution for his example of (2.28).) He proceeded to try to prove that if the "elementary ophelimities" φ_i were assumed a priori to be independent of (k_1, k_2), and an individual was constrained by experiment to follow the above path (2.27), it would be possible to estimate the φ_is empirically (up to a proportionality factor) from observations on his market behavior.

His attempted proof goes as follows. First, he computes the line integral of (2.28) along the path (2.27) to obtain a function

$$\begin{aligned} U^k(x) &= \int_0^{k_2} \varphi_2(0, \xi_2)d\xi_2 + \int_0^{k_1} \varphi_1(\xi_1, k_2)d\xi_1 \\ &\quad + \int_{k_2}^{x_2} \varphi_2(k_1, \xi_2)d\xi_2 + \int_{k_1}^{x_1} \varphi_1(\xi_1, x_2)d\xi_1, \end{aligned} \tag{2.29}$$

which he then differentiates to get

$$\varphi_1(x_1, x_2)dx_1 + \Big[\varphi_2(k_1, x_2) + \int_{k_1}^{x_1} \varphi_{12}(\xi_1, x_2)d\xi_1\Big]dx_2 = 0. \tag{2.30}$$

He then asserts that the market behavior of an individual constrained to follow the path (2.27) would be given by an equation of the form $dx_1 + R^k(x_1, x_2)dx_2 = 0$, where $R^k(x_1, x_2)$ is the ratio of the coefficient of dx_2 in (2.30) to that of dx_1; and that by using the constraint that φ_1 must be independent of k_1, k_2, one could identify $\varphi_1(x_1, x_2)$ up to a proportionality factor.

Finally, Pareto performs another experiment in which the individual is now free to choose whatever path he pleases. The relevant differential equation is then (2.28), and by combining the empirical data on

$$R(x_1, x_2) \equiv \varphi_2(x_1, x_2)/\varphi_1(x_1, x_2)$$

obtained by observing his market behavior under these conditions with the estimate of $\varphi_1(x_1, x_2)$ (up to a proportionality factor) obtained by observing his market behavior in the constrained conditions, one obtains an estimate (up to a proportionality factor) of $\varphi_2(x_1, x_2)$.

There are a number of difficulties with this argument. In the first place, the consumer's marginal rate of substitution is postulated, to begin with (from (2.28)), to be given by $R(x_1, x_2) = \varphi_2(x_1, x_2)/\varphi_1(x_1, x_2)$; and this is simply inconsistent with (2.30) unless the coefficient of dx_2 is $\varphi_2(x_1, x_2)$, which it can be only if $\varphi_{12} = \varphi_{21}$, i.e., only if the equation (2.28) is exact.[15] Secondly, it is not clear why the coefficient of dx_2 in (2.30) should not *also* be assumed a priori to be independent of k_1 and k_2 (it is obviously independent of k_2); but it is independent of k_1 if and only if, once again, $\varphi_{12} = \varphi_{21}$ Thus, there seems to be no way to justify Pareto's procedure, and it must be considered to be a blunder.[16]

The exploration has been of interest, however, in shedding light on two matters. First of all, it is clear that Pareto was never aware of the necessity of conditions (2.24), and of what could be meant by their failure. This remained for Allen [7, 115], Georgescu-Roegen [98], and Samuelson [210] to investigate.[17] Secondly, Pareto is shown to be seeking experiments which will make it possible to *measure* ophelimity. Even though the attempt must be considered unsuccessful, the fact that he made it should help dispel the belief that he was inconsistent in adhering to a notion of cardinal utility in his theory of related demands, even

[15] That is, we require

$$\begin{aligned} 0 &= \varphi_2(x_1, x_2) - \varphi(k_1, x_2) - \int_{k_1}^{x_1} \varphi_{12}(\xi_1, x_2)d\xi_1 \\ &= \int_{k_1}^{x_1} \varphi_{21}(\xi_1, x_2)d\xi_1 - \int_{k_1}^{x_1} \varphi_{12}(\xi_1, x_2)d\xi_1, \end{aligned}$$

independently of the choice of k_1, k_2. Differentiatiing with respect to k_1 we obtain $\varphi_{21}(x_1, x_2) = \varphi_{21}(x_1, x_2)$.

[16] On a related question concerning Pareto's 1911 approach [188], see Georgescu-Roegen [97, p. 713].

[17] Georgescu-Roegen [98] was also apparently the first to point out that the integrability conditions (2.24) are still not sufficient for the existence of a utility function in a meaningful economic sense. This could be the case if, in the neighborhood of a singular point where all $\varphi_i(x) = 0$, the solution is a logarithmic spiral.

though he recognized that in general such a cardinal indicator could not be identified from market data.

Chapter 3

Welfare Economics and International Trade

3.1 Historical background

Although we think of Pareto today as an academic and even a highly abstract economist, in fact he was in 1891 a practical man who was extremely well informed about and deeply concerned with current economic developments in Italy. He had just been defeated in parliamentary elections in which he had run as a free trader in opposition to the protectionist and militaristic trends in the government.[1] Not being able to publish his analysis of these trends in Italy, he achieved notoriety by publishing it in France [150], along with another widely disseminated pamphlet providing statistical details [149]. In brief, Pareto showed that the tariff approved by the Italian parliament in 1887, and the simultaneous rupture of Italy's commercial treaty with France, had virtually wiped out Italy's wine and silk exports, and had led to a fall in aggregate exports and imports in the years 1888–90 following a steady rise from 1884 to 1887; whereas during these post-tariff years other European countries had continued to experience increases in both exports and imports. Pareto also provided statistics to show that the tariff had resulted in a fall in per capita wheat consumption in Italy, as well as a significant rise in emigration.

It is with this background that Pareto, who was at the time avidly absorbing Walras's *Elements* [242], set out to provide a deductive proof of the optimality of free trade. The background just described can serve to explain both the energy and originality that Pareto brought to the task, and the blind spots which prevented him from developing the theory to its logical conclusions.

[1] This was the period of Italy's military ventures in Eritrea.

3.2 Theory of tariffs

Having successfully uncovered a blunder by Cournot [52] in his treatment of the gains from international trade (cf. [151]) — a task which was subsequently to be repeated by Fisher [88] and Viner [238, pp. 586-9] without knowledge of Pareto's discussion — Pareto went on [153] to tackle the large work by Auspitz and Lieben [17]. These authors had provided an argument in favor of tariff protection by means of a geometric analysis employing supply-and-demand curves similar to Marshallian offer curves. Pareto took exception to their assumptions of constancy of the marginal utility of income and constancy of the $n-1$ remaining prices. However, he had sense enough to recognize that these were not fundamental objections and went on to meet their argument on their own terms, within the framework of a two-commodity model. While his analysis was interesting in its own right, the outcome was disappointing; he triumphantly concluded that the home country could gain only if the foreign country loses, hence: "This method of inflicting damage on one's neighbor is tantamount to reducing oneself to poverty in order to harm the shopkeepers one has been patronizing" [153, p. 231] — which is not a correct analogy at all. He added that the tariff revenues would be squandered by the government anyway — which is no doubt true, but what it meant to have recourse to this argument was that he had not succeeded in setting out what he had aimed to do, which was to prove that tariffs are necessarily harmful to the country imposing them.

3.3 Welfare economics

A few years later, Pareto made one of his fundamental contributions to welfare economics [157]. He started out [156] with an exposition of the Walrasian system, allowing — in response to objections raised by Edgeworth [68] — for variable factor supply. We may set this out as follows. Letting x_{ij} denote the amount of commodity j consumed by individual i, and z_{ik} the amount of the services of the kth factor of production supplied by individual i, the ith individual has a utility function

$$U_i(x_{i1}, x_{i2}, \ldots, x_{in}, -z_{i1}, -z_{i2}, \ldots, -z_{ir}) \quad (i = 1, 2, \ldots, m) \tag{3.1}$$

where n is the number of consumer goods and r the number of factors of production, m being the number of individuals. Denoting the total output of commodity j by y_j and the total input of factor k by ℓ_k, equality of supply and demand entails

$$\sum_{i=1}^{m} x_{ij} = y_j, \quad \sum_{i=1}^{m} z_{ik} = \ell_k \quad (j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, r). \tag{3.2}$$

Factor demand is determined by

$$\ell_k = \sum_{j=1}^{n} b_{jk} y_j \quad (k = 1, 2, \ldots, r), \tag{3.3}$$

where the b_{jk} are fixed production coefficients, and equality of prices and unit costs entails

$$p_j = \sum_{k=1}^{r} b_{jk} w_k \quad (j = 1, 2, \ldots, n), \tag{3.4}$$

where p_j is the price of commodity j and w_k the price of the services of factor k. Choosing commodity 1 as numéraire and setting $p_1 = 1$, the system is completed by specifying the equilibrium conditions for individuals as consumers of products and suppliers of productive services,[2] namely the conditions expressing equality of marginal rates of substitution and price ratios:

$$\begin{gathered} U_{i1}(x_i, -z_i) = \frac{1}{p_j} U_{ij}(x_i, -z_i) = \frac{1}{w_k} U_{i,n+k}(x_i, -z_i) \\ (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, r), \end{gathered} \tag{3.5}$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $z_i = (z_{i1}, z_{i2} \ldots, z_{ir})$, and where $U_{ij} = \partial U_i / \partial x_{ij}$ and $U_{i,n+k} = \partial U_i / \partial(-z_{ik})$; and the budget identities:

$$\sum_{j=1}^{n} p_j x_{ij} - \sum_{k=1}^{r} w_k z_{ik} = 0 \quad (i = 1, 2, \ldots, m). \tag{3.6}$$

As Pareto observed, equations (3.2)-(3.6) constitute $(m+2)(n+r)-1$ independent equations in the $(m+2)(n+r)-1$ unknowns $x_{ij}, z_{ik}, y_j, \ell_k, p_j, w_k$. Pareto also remarked [156, p. 145] that if any factor was offered in inelastic supply, its marginal utility would be zero and the corresponding equation would be deleted from (3.5).

Pareto's point of departure in analyzing this system of equations was the remark [156, p. 149] that: (i) "their solution satisfies the condition of providing maximum welfare. This is shown by the fact that equations (3.5) which express this very condition, form a part of these formulae"; and (ii) as a result of the equality between the number of equations and number of unknowns, "the problem is entirely determinate." With respect to (ii) we of course know now that this counting rule is not sufficient to establish the existence of equilibrium; of greater interest than (ii) for our purposes is, however, remark (i): what did Pareto mean by "providing maximum welfare"? Did he mean what we now describe as "Pareto optimal"? Pareto attributed the result to Walras; but a consideration of Walras's original statement sheds very little additional light on this question [242, p.121]: "The exchange of two commodities for each other in a market regulated by free competition is an operation by which all holders of either one or the other of the two commodities, or of both, can obtain the greatest possible satisfaction of their wants consistent with the condition that the two commodities be bought and sold at a common and identical rate of

[2] Actually, Pareto initially adopted Walras's assumption of additive separability of each U_i but subsequently adopted Edgeworth's more general formulation.

interchange."[3] On the face of it, this statement seems to express no more than a truism, namely that at the equilibrium prices, each trading individual maximizes his satisfaction subject to the budget constraint determined by those prices and his holdings. This indeed is how the proposition was subsequently to be interpreted by Scorza [218];[4] but Walras regarded it as fundamental, and as we shall see presently, Pareto did interpret it as meaning "Pareto optimality."

Pareto was dissatisfied with the assumption of fixed production coefficients in Walras's formulation, and sought to generalize the model by allowing them to be variable. He therefore introduced the functions [5]

$$f_j(b_{j1}, b_{j2}, \ldots, b_{jr}) = 1 \quad (j = 1, 2, \ldots, n). \tag{3.7}$$

Writing $v_{jk} = b_{jk}y_j$ for the amount of the kth factor employed in the jth industry, where $\sum_{j=1}^n v_{jk} = \sum_{i=1}^m z_{ik}$, this of course defines the production functions

$$y_j = f_j(v_{j1}, v_{j2}, \ldots, v_{jr}) \quad (j = 1, 2, \ldots, n), \tag{3.8}$$

which are homogeneous of degree 1.[6] Since (3.7) adds the nr new variables b_{jk} and only n new equations, Pareto added the condition that unit production costs, given by the expression on the right side of equation (3.4), be minimized with respect to the b_{jk} for each fixed set of factor prices w_k, yielding the $n(r-1)$ additional equations

$$w_1 \frac{\partial f_j/\partial b_{jk}}{\partial f_j/\partial b_{j1}} - w_k = 0 \quad (j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, r). \tag{3.9}$$

He then enunciated the following fundamental theorem [157, p. 55]: "that the production coefficients determined by free competition have values identical with those obtained by determining these coefficients by the condition that they produce maximum utility with minimum sacrifice." His proof is difficult to follow, but a few years later an elegant streamlined proof was presented in the *Cours* [163, I, $(385)^2$, p. 257; II, $(719)^2$, pp. 88–9 and $(721)^2$, pp. 92–4], which will now be briefly summarized.

Pareto formulated the problem as one which would be faced by a socialist state, with a Minister of Production whose job it was to determine the production coefficients, and a Minister of Justice whose job it was to allocate the outputs so produced among the individuals in such a way as to obtain maximum

[3] The passage occurs again on p. 99 of the 1926 edition. See p. 143 of the Jaffé translation as well as the discussion there on pp. 510-11.

[4] Wicksell [244] had already made the same criticism prior to Scorza [218].

[5] Pareto actually wrote them in the form $F_j(b_{j1}, b_{j2}, \ldots, b_{jr}) = 0$; (cf. [157, p. 50]). The difference of course is trivial since one can define $F_j = f_j - 1$.

[6] Pareto's contribution preceded by a few months that of Wicksteed [246]. As Jaffé [117]] recounts, Barone shortly thereafter submitted a review of Wicksteed's monograph to the *Economic Journal* which was turned down by its then editor, Edgeworth. This is one of the events that soured Pareto's relations with Edgeworth. Pareto subsequently [159] generalized his treatment to allow for variable returns to scale.

utility for each citizen [157, pp. 52, 65]. He took it for granted that prices would be used by the socialist state,[7] but replaced (3.6) by the condition (with $p_1 = 1$)

$$\sum_{j=1}^{n} p_j x_{ij} - \sum_{k=1}^{r} w_k z_{ik} = \lambda_i \quad (i = 1, 2, \ldots, m) \tag{3.10}$$

where the parameters λ_i were under state control; nevertheless, consumers would maximize their utility functions (3.1) subject to (3.10), so that the conditions (3.5) would still hold. On the other hand, he assumed that production would all be under government control, so that conditions (3.9) (and (3.4)[8]) would not necessarily hold; in fact, the problem is precisely to show that (3.4) and (3.9) follow from the conditions of social optimality. Thus, from (3.1), (3.5), and (3.10) one has, for any fixed set of prices $p = (p_1, p_2, \ldots, p_n)$ and $w = (w_1, w_2, \ldots, w_r)$,

$$dU_i = \mu_i d\lambda_i \quad (i = 1, 2, \ldots, m), \tag{3.11}$$

where

$$\mu_i = U_{i1} = U_{ij}/p_j = U_{i,n+k}/w_k \tag{3.12}$$

is the marginal utility of the ith person's income. Pareto then considers the aggregate net value variable

$$\Lambda = \sum_{i=1}^{m} \lambda_i \tag{3.13}$$

and observes that as long as $d\Lambda > 0$, it is possible by an appropriate reallocation of the x_{ij} and z_{ik} among the m individuals to make each $d\lambda_i > 0$; accordingly his criterion for a social optimum is $d\Lambda = 0$, or that Λ should be a maximum. From (3.2), (3.3), (3.10) and (3.13), one has

$$\Lambda = \sum_{j=1}^{m} y_j \left(p_j - \sum_{k=1}^{r} b_{jk} w_k \right). \tag{3.14}$$

When this is maximized with respect to the y_j and b_{jk}, subject to (3.7) — or equivalently, subject to $f_j(b_{j1}y_j, b_{j2}y_j, \ldots, b_{jr}y_j) = y_j, j = 1, 2, \ldots, n$ — one obtains precisely (3.4) and (3.9).[9]

---

[7] It should be remarked, however, that there is a noteworthy treatment in the final chapter of the *Cours* [163, II, §(1017)[1], pp. 366–8] in which Pareto takes up the role of non-desired capital goods in a socialist society (i.e., capital goods that do not directly enter utility functions) and shows that implicit accounting prices would have to be used for the purpose of achieving Pareto optimality.

[8] Actually, Pareto assumed (3.4) to hold and established that (3.9) followed by maximization of the function Λ (given by (3.13) below) with respect to the b_{jk} subject to (3.7); the treatment given here is therefore somewhat more general than Pareto's.

[9] As was stated above, the argument just summarized is that of the *Cours*. In his earlier 1894 treatment Pareto considered partial derivatives of x_{ij} and z_{ik} with respect to the b_{jk}s, which seems to imply the existence of some implicit allocation rule. The difficulty in trying to make this argument rigorous is in finding a precise way to define such a rule.

The above argument, it will be recognized, contains not only the concept of "Pareto optimality" but also the essence of the Compensation Principle. Lest there should be any doubt concerning the degree of Pareto's understanding of this principle, the following passage should dispel it [157, p. 60]:[10]

> If the $d\lambda_i$s are all positive, every individual undergoes a gain in utility and we shall say *elliptically* that social utility increases. In that case it will be advantageous to *all* members of society for b_{jk} to be increased by db_{jk}. Likewise if the $d\lambda_i$s are all negative, each individual will suffer a loss in utility, and we shall say *elliptically* that social utility decreases. In that case it will be advantageous not to increase b_{jk} but rather to decrease it by db_{jk}. And so on we will proceed in one way or another as long as all the $d\lambda_j$s have the same sign. But, when we reach the point at which some are positive and others negative, we shall not be able to proceed any further because we shall be favoring some individuals at the expense of others. When b_{jk} increases by db_{jk}, if some $d\lambda_i$s are positive and others negative, this means that the utility enjoyed by certain individuals increases and that of others decreases. It is not possible to offset these various utilities against each other since they are reckoned in different units.
>
> Expressions (3.10) on the other hand all represent quantities of commodity 1, the other commodities being thus expressed in the same unit. Letting $d\zeta$ be the sum of the positive $d\lambda_i$s and $d\sigma$ that of the negative $d\lambda_i$s, if $d\zeta - d\sigma$ is positive we can take as much of commodity 1 away from individuals for whom the $d\lambda_i$s are positive (the others being reckoned in terms of commodity 1) as is needed to bring the negative $d\lambda_i$s back to zero, and there will still be a residual. Hence, society as a whole will have made a gain. This gain can be distributed among all the members of society, or among only some; this is a question which I shall not now examine — it is enough to point out its existence. Hence it will be desirable to increase b_{jk} by db_{jk} and only later to examine how to distribute that residual. When should we stop increasing b_{jk}? Precisely when $d\zeta - d\sigma = 0$; because proceeding still further, so that $d\zeta - d\sigma$ becomes negative, it would no longer be possible to take so and so much of the commodity away from those for whom the $d\lambda_i$s are positive to compensate the individuals for whom the $d\lambda_i$s are negative. Society as a whole would therefore undergo a loss and no longer a gain.

It would be difficult to find in later writings a clearer statement of the case for productive efficiency than this; and Pareto's argument contains the essence of the New Welfare Economics launched by Hicks [112] forty-five years later. An interesting illustration of the principle presents the case perhaps even more sharply [157, p. 56]:

[10] The symbols and formulas in the above development (and similar passages to be quoted below) have been substituted for Pareto's own.

> The question has been raised whether entrepreneurs would not find it more to their advantage to employ more labor and less capital in order thereby to acquire consumers, on the grounds that the workers, having a larger sum total of wages to spend, would buy more commodities.
>
> My theorem answers this question in the negative. If it were desired to procure some advantages for workers, it would be better to give them a certain sum directly which would be taken from other citizens, and to leave the production coefficients as they are, since they ensure the greatest possible total of utility with the minimum disutility.

Unfortunately, Pareto slightly misinterpreted his own result in stating that the production coefficients would have the same *values* under conditions of "maximum social utility" as under free competition; this of course is not the case, since the equilibrium prices need not be the same in the two situations. What Pareto actually showed, and should have said, is that the production coefficients are the same *functions* of factor prices in the two situations. This bit of carelessness does not essentially detract from Pareto's argument in the illustration just cited; but it appears unfortunately to be largely what led him astray in applying his theorem to the theory of international trade.

3.4 Application to international trade

In a follow-up article [159] Pareto proceeded to apply his new results to the theory of international trade, and in particular to the problem of the welfare effects of an import duty; subsequently he provided a somewhat clearer exposition of his treatment in the *Cours* [163, II, §§862–873, pp. 222–7]. Briefly, his approach consisted in dealing with what he set forth as being an equivalent problem: that of assessing the additional requirements in terms of services of the factors of production (which by a suitable income distribution could be translated into an increased disutility for each member of society) that would result if the consumption of each commodity were to be maintained at its previous level while imports fell as a result of the tariff, it being assumed that domestic prices were kept equal to unit costs and trade kept balanced. His analysis appears to contain a fairly serious slip: he used domestic prices rather than exclusively foreign prices in setting down the balance-of-payments identity, as a result of which proper account was not taken of the tariff revenues. He came to the conclusion that there would be a welfare loss proportionate to the change in the *domestic* price ratio of the export good to the import good, rather than to the *foreign* ratio (i.e., the terms of trade). As a result, he was able to arrive at the statement [163, II, §873, p. 227]: "every protective measure produces a destruction of wealth. This is one of the surest and most important theorems to which economic science leads."

This result is certainly correct from the international point of view; that is, whatever a country could gain by a tariff it could gain in equal measure by a tribute exacted from foreign countries, at less cost to the rest of the world. However, Pareto did not put the matter this way. He distinguished between protective tariffs and other kinds, characterizing the former by the criterion that they had as their aim the reduction of imports. He allowed, as a theoretical possibility, export taxes on a product in which the exporting country held a monopoly position, and agreed with Edgeworth's analysis [69, Part 2] that if the foreign offer curve were inelastic a tariff would have the effect of raising imports and improving the country's terms of trade, and thus improving the country's welfare. However, this result need not hold unless the proceeds of the tariff are included among the welfare benefits; for example, if the government spends all the proceeds on the import good and this consumption is not included among the gains, there will actually be a welfare loss to the country's citizens even though total imports increase (cf. Lerner [131]). On the other hand we know from the Bickerdike-Edgeworth theory [26, 27, 77] that as long as the foreign offer curve is less than infinitely elastic and the home transformation surface is differentiable and strictly concave to the origin, there will be some sufficiently small but positive "incipient" tariff that will improve the home country's welfare if suitable income transfers are undertaken. This flatly contradicts Pareto's position. And, as we have seen, his position cannot be rescued by resorting to the hypothesis that the tariff proceeds are squandered by the government, because this would not square with his acceptance of Edgeworth's argument that tariffs are beneficial when they result in increased imports.

Despite his acceptance of Edgworth's argument as indicating a theoretical possibility, Pareto argued that the case did not apply to the analysis of protective import duties, since the types of commodities a country imports usually constitute a small proportion of the market for these products, and therefore a tariff is unlikely to have an appreciable effect on import prices. The trouble with this very plausible sounding argument is, of course, that it violates Lerner's symmetry theorem [131]: even if it is true that a tariff may have very little (absolute) effect on the world prices of a country's imports, if resources are internally mobile the rise in the internal price of imports will induce a movement of resources out of the country's export industries into the import-competing industries, resulting in a contraction of output in the export industries and thus — as long as the country is an important supplier of this product — in a rise in the world price of its export good and hence an improvement in its terms of trade.

It must be concluded that while Pareto developed extremely important concepts in welfare economics — in fact, just about all the basic concepts underlying modern welfare economics — his application of these concepts to international trade theory was less than satisfactory. Possibly it is a case of strong preconceptions blocking the logical course of analysis.[11] On the other hand, Pareto did

[11] In fact, he admitted as much in the Preface to the *Manuale* [182, p. viii]. Nevertheless, he never abandoned his view that protection involves a "destruction of wealth", but took the position that economic disadvantages might in certain practical cases be outweighed by non-economic advantages, and that one should therefore not jump to policy conclusions solely on the basis of one's economic analysis.

make three additional contributions to welfare economics, one a very interesting cost-benefit analysis of the railroad industry [152], and two further contributions which I shall now discuss, which are of a fundamental nature.[12]

3.5 The controversy with Scorza

We have already remarked with respect to equations (3.2)–(3.6) of the Walrasian system that Pareto had left things in a rather unsatisfactory state in accepting Walras's propositions that (i) competitive equilibrium entailed "maximum welfare" and (ii) the system was determinate owing to the equality of the number of equations and the number of unknowns. These two problems gave rise to a series of critical papers by a mathematician, Gaetano Scorza [217, 218, 219].

Scorza began by pointing out [217] that it would be possible for the system to have no solution (although no example was presented); he went on to give still greater emphasis to the possibility of multiple solutions, thereby questioning the logical validity of Walras's demonstration that a determinate solution would be obtained in the market by means of the tâtonnement process, and incidentally deploring the lack of an explicit dynamic formulation of this process. He also pointed out — without reference to Wicksell [244], who had already put forward the same argument — that since individuals would have different levels of utility in different equilibrium positions, the "maximum utility" they enjoyed in one of these positions could not all be absolute maxima — thereby leaving in doubt the meaning or logical validity of the claims made by Walras concerning the optimality of free competition. It is clear that the points raised by Scorza were all quite legitimate ones; those that had not already been raised before were all raised again by subsequent writers, and settled: by Wald [240, 241], Arrow and Debreu [16], and others, in connection with the existence of equilibrium; by Samuelson [208], Arrow, Block, and Hurwicz [15], and others, in connection with the explicit dynamic formulation of the tâtonnement process; and by Arrow [13], Debreu [64], and Koopmans [128], in connection with the proposition that every competitive equilibrium is a Pareto optimum, combined with the converse proposition that every Pareto optimum can be sustained by a competitive equilibrium — the so-called Fundamental Theorems of the New Welfare Economics.

Scorza followed up his critique with a review of an article by Cassel in which he expressed approval of Cassel's position [44, p. 431] with regard to Pareto's proposition [163, I, §(100)^1, p. 46n] that every participant in the market achieves maximum satisfaction, saying that "the reasoning according to which this conclusion is drawn from the presence of the equations of maximum satisfaction [i.e., (3.5) above] in the system which determines the equilibrium is nothing but a gross sophistry" [218, p. 194]. He elaborated this criticism in a further article [219] in which he went wide off the mark by suggesting that under free com-

[12] It is worth mentioning, in addition, a contribution [177] which does not fit in to any of the main categories under consideration, namely a formulation of the concept of dynamic equilibrium in terms of a system of differential equations.

petition each participant should — after calculating his excess demand at each set of prices so as to maximize utility — make a second calculation to select that (equilibrium?) price at which his utility would be maximized. In short, he failed to grasp the idea of the parametric role of market prices. This made him easy prey for Pareto who followed [180] with a scathing attack on Scorza, which he made the occasion for an attempt at a precise statement and proof of the proposition that, as we now describe it, every competitive equilibrium is a Pareto optimum (cf. Koopmans [128, p. 49]). Pareto proceeded as follows [180, pp. 410–12]. Let there be m traders with utility functions

$$U_i(x_{i1}, x_{i2}, \ldots, x_{in}) \quad (i = 1, 2, \ldots, m) \tag{3.15}$$

where x_{ij} is the amount of commodity j consumed by individual i, and a_{ij} is his initial endowment of commodity j. Defining excess demand by

$$q_{ij} = x_{ij} - a_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n), \tag{3.16}$$

any solution to the system (including that of free competition) is assumed to satisfy equality of demand and supply

$$\sum_{i=1}^{m} q_{ij} = 0 \quad (j = 1, 2, \ldots, n); \tag{3.17}$$

likewise, for any price constellation $p = (p_1, p_2, \ldots, p_n)$ the corresponding excess demands are assumed to satisfy the budget equalities

$$\sum_{j=1}^{n} p_j q_{ij} = 0 \quad (i = 1, 2, \ldots, m). \tag{3.18}$$

Finally, if p_j^* and q_{ij}^* are the prices and excess demands corresponding to a particular competitive equilibrium solution, where $p_1^* = 1$, this solution is assumed by definition to satisfy

$$U_{i1}^* = \frac{1}{p_j^*} U_{ij}^* \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n) \tag{3.19}$$

where $U_{ij}^* = U_{ij}(x_{i1}^*, x_{i2}^*, \ldots, x_{in}^*)$ and $x_{ij}^* = q_{ij}^* + a_{ij}$. Equations (3.17), (3.18), and (3.19) are those which together determine the competitive equilibrium, and the point at issue between Pareto and Scorza is whether the presence, in this system, of equations (3.19) (which correspond to (3.5) for the special case of fixed factor supplies) is what permits one to conclude that the competitive equilibrium is "Pareto optimal." In order to demonstrate this point Pareto examines the consequences of abandoning (3.19) while retaining not only the feasibility constraints (3.17) but also the budget equations (3.18). Now, Pareto expresses a small departure from the competitive equilibrium position in terms of the differentials

$$dU_i^* = \sum_{j=1}^{n} U_{ij}^* dq_{ij} \quad (i = 1, 2, \ldots, m), \tag{3.20}$$

and from (3.17) and (3.18) derives the equations

$$\sum_{i=1}^{n} dq_{ij} = 0 \quad (j = 1, 2, \ldots, n) \tag{3.21}$$

and

$$\sum_{j=1}^{n} p_j^* dq_{ij} + \sum_{j=1}^{n} q_{ij}^* dp_j = 0 \quad (i = 1, 2, \ldots, m). \tag{3.22}$$

Substituting (3.19) in (3.22) and making use of (3.20) he then obtains

$$\frac{1}{U_{i1}^*} dU_i^* + \sum_{j=1}^{n} q_{ij}^* dp_j = 0 \quad (i = 1, 2, \ldots, m). \tag{3.23}$$

Finally, summing equations (3.23) over individuals and taking account of (3.17), he obtains

$$\sum_{i=1}^{m} \frac{1}{U_{i1}^*} dU_i^* = 0. \tag{3.24}$$

He then points out that since $U_{i1}^* > 0$ by hypothesis, the dU_i^*s cannot all be positive or all be negative, and concludes [180, pp. 412–13]:

> It follows from all of this that the equilibrium position of maximum ophelimity defined by equations (3.19) is such that one cannot, in departing from it by infinitesimal variations in the quantities, increase (or decrease) all the ophelimities, conferring advantages (or disadvantages) on all the individuals, but that if some ophelimities increase others necessarily decrease; if some individuals are favored, others will be harmed.

On the face of it, this statement appears to contain a slip, in the view of the parenthetical phrases; obviously, in departing from a competitive equilibrium it is possible for all individuals to lose, although it is not possible for them all to gain. Similar statements, which appeared later in the *Manuale* [182, Ch. IV, §33, p. 337; Appendix, §45, pp. 543–4] and the *Manuel* [185, p. 354; Appendix, §89, pp. 617–18], led Wicksell [244] to puzzle about their meaning. As Wicksell perceived, however, the crucial word to be stressed in the above passage is the elusive term "infinitesimal"; for Pareto immediately followed the above passage by one containing an analogous statement for "finite" variations:

> For a finite displacement, this proposition remains true provided we remove the restriction that not all ophelimities can decrease, and state that the equilibrium position defined by the equations (3.19) of maximum ophelimity is such that one cannot, in departing from it, cause all the ophelimities to increase, thus conferring advantages on all individuals; but that at least same of the ophelimities must decrease, i.e., some of the individuals must become worse off.

Thus, it seems that Pareto was in his statements simply distinguishing between first-order and second-order conditions; presumably in the case of a maximum of a single function he would have said that for "infinitesimal variations" the value of the function would neither increase nor decrease, whereas for "finite variations" it would decrease. It remains only to attach an analogous and precise meaning to the corresponding statement for a *set* of functions; below I shall argue that this can be done in a perfectly rigorous way.

Pareto followed his "finite" statement by an attempted proof which will be analyzed in detail below. He also provided an illustration for the two-individual two-commodity case in which he took quadratic approximations to the utility functions. Continuing with this approach, he also displayed an example of an equilibrium in which one trader was a price taker, and the other a monopolist maximizing his utility subject to the knowledge of the price-taker's excess demand function, and showed that there existed a competitive equilibrium in which both participants were better off.

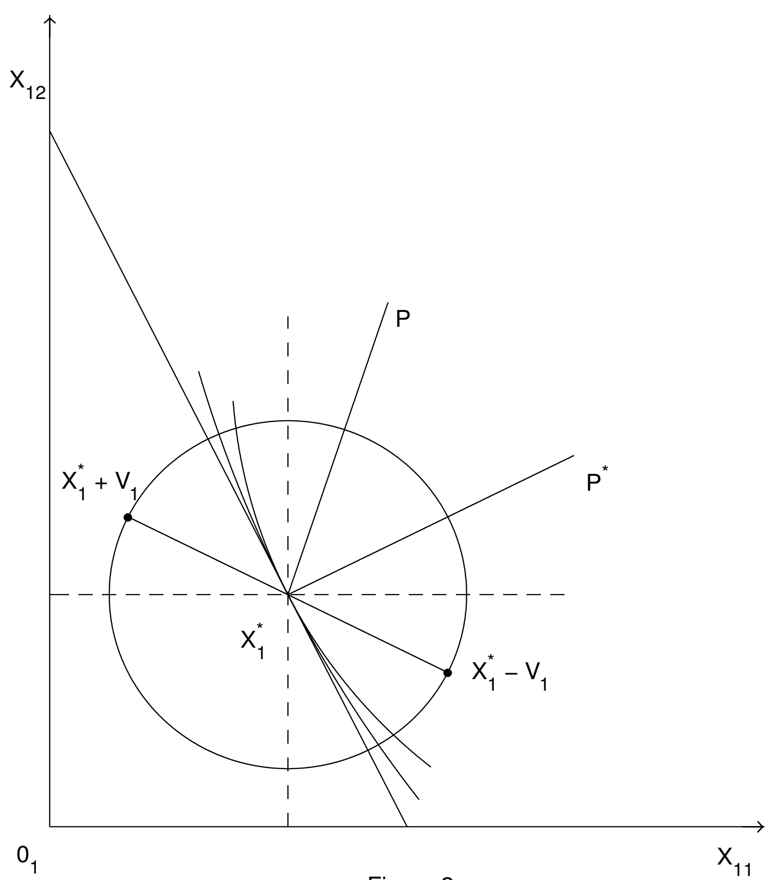

Figure 2

Pareto's general method of proof was illustrated by a diagrammatic argument which he presented for the two-individual two-commodity case [180, p. 421]. His procedure was to superimpose the two individuals' preference maps in such a way as to make their equilibrium consumption points coincide (see Figure 2 where, as in Pareto's figure, only the coordinate axes for individual 1 are displayed). Pareto then considered movements in individual 1's consumption $x_1 = (x_{11}, x_{12})$ along a line segment

$$x_1 = \xi_1(t; v_1) = x_1^* + tv_1 \quad (t \geqq 0) \tag{3.25a}$$

which was considered to be part of a budget line

$$p_1 x_{11} + p_2 x_{12} = p_1 x_{11}^* + p_2 x_{12}^*,$$

so that the new prices $p = (p_1, p_2)$ are orthogonal to the vector $v_1 = (v_{11}, v_{12})$, i.e., $(v_{11}, v_{12}) = \lambda(-p_2, p_1)$. Since Pareto regarded these prices as intrinsically positive, he limited himself to linear paths proceeding in the northwest and southeast directions, i.e., to vectors v_1 whose components were of opposite sign (e.g., $v_{11} < 0, v_{12} > 0$ in Figure 2). So that individual 2 should satisfy the corresponding budget constraint

$$p_1 x_{21} + p_2 x_{22} = p_1 x_{21}^* + p_2 x_{22}^*$$

at the same prices, and in such a way as to maintain equality of total consumption and total initial endowment, Pareto assumed that individual 1's movements (3.25a) would be balanced by equal and opposite movements on the part of individual 2, given by

$$x_2 = \xi_2(t; v_1) = x_2^* - tv_1. \tag{3.25b}$$

Pareto noted that along such a path given by (3.25a) and (3.25b), for sufficiently small t one individual would gain and the other lose — a correct proposition provided $p \neq p^*$. He went on to note that this property was characteristic of all competitive equilibrium points; since his terminology "maximum ophelimity" has sometimes misled readers into thinking that Pareto was unaware of the non-uniqueness of Pareto-optimal points, the following passage is worth quoting to dispel any doubts [180, p. 422]:

> The equilibrium point in the case under consideration therefore enjoys the stated property; but it is not the only one to do so. Also enjoying this property are all the points where two indifference lines (one appertaining to one individual, the other to the other individual) are tangential. We thus have a locus which could be called the locus of maxima of ophelimities; and if the initial quantities possessed by the individuals are given, and if it is determined that exchange is to take place at constant prices, free competition leads precisely to equilibrium at a point on this locus. Other different systems of organization, provided they are appropriately chosen, could also lead to equilibrium at a point on this locus.

In the *Manuale* [182, Ch. VI, §35, p. 338], [185, p. 355], Pareto repeated the same geometric argument as given above, but this time with the innovation of the back-to-back diagram (Figure 3) — now known as the "Edgeworth box"![13] Once again he noted that by moving along a straight line (3.25) from the equilibrium point $t = 0$, one individual would necessarily benefit and the other lose; and he added: "It is therefore not possible in departing from [the equilibrium point] to benefit or harm both individuals simultaneously." He later recognized, however, that if the path (3.25) was chosen so as to coincide with the original budget line through x_1^* at prices p^*, i.e, so that $v_1 = \lambda(-p_2^*, p_1^*)$, both individuals would lose; in his words [185, Appendix, §122, p. 652]: "If the movement takes place in the direction in which all the $\delta U_1, \delta U_2, \ldots$ are zero, all the ophelimities will decrease, This is what takes place when one follows the path along which the equilibrium of the consumers is established."

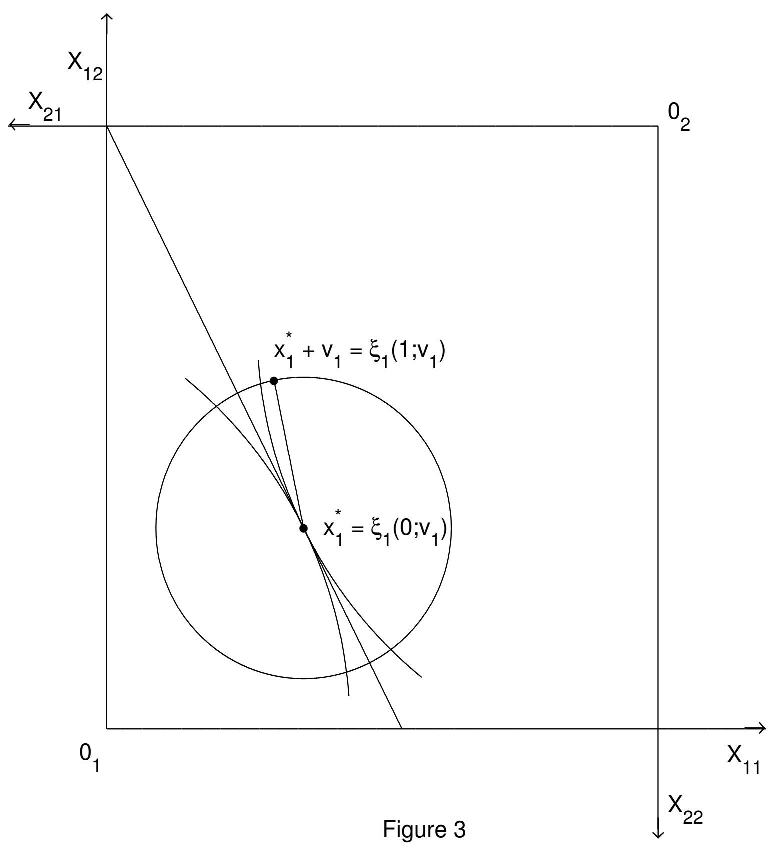

Figure 3

[13] The diagram may be said to have been implicit in Edgeworth [67, p. 28], however, but without either the axes or the indifference curves.

The proofs of the Pareto optimality of competitive equilibrium that have been offered since Pareto's time have been quite different in nature from Pareto's attempted proof. Lange [129] adopted the device of maximizing one person's utility subject to the utilities of the remaining $m-1$ individuals being held constant; using classical calculus techniques he was able to show that the problem was equivalent to that of maximizing a linear combination of individual utility functions, and obtained as a necessary condition for an interior solution the equality among individuals of their marginal rates of substitution between any two commodities.[14] Arrow (13) objected to Lange's approach on the ground that it did not take account of non-negativity constraints and corner solutions; and up until very recently the standard treatments (Koopmans [128, pp. 46–53], Debreu [65, pp. 68–71, 94–96]) have been based on the methods introduced by Arrow [13] and Debreu [64] in which the differential calculus is eschewed and reliance is instead placed on the theory of convex sets, and on the equivalence of utility maximization subject to a budget constraint and expenditure minimization subject to a utility constraint. Very recently, however, under the impetus of the work of Smale [224], there has been a renewed interest in calculus methods made more powerful by advances in global analysis and differential topology, and in the light of the new point of view it is worth reexamining Pareto's original approach to see whether his proof can be recast in a correct and rigorous manner, and to see to what extent it constitutes a legitimate and complete proof of the proposition that a competitive equilibrium is a Pareto optimum. It will be convenient to start by presenting the following argument which is adapted from a more general treatment due to Hans Weinberger [243].

For simplicity only (since the new methods allow one to handle inequality constraints) suppose we consider an initial equilibrium in which $x^*_{ij} > 0$ for each individual, i, and each commodity, j, and all prices are positive. Consider an $(m-1) \times n$ matrix V whose ith row is $v_i = (v_{i1}, v_{i2}, \ldots, v_{in})$ and which is normalized so that $\sum_{i=1}^{m-1} \sum_{j=1}^{n} v_{ij}^2 = 1$; for any such V and for sufficiently small $t > 0$ we may define the path $\xi(t; V)$ by

$$(3.26) \qquad \begin{cases} \xi_i(t;V) &= x^*_i + tv_i \qquad (i = 1, 2, \ldots, m-1), \\ \xi_m(t;V) &= \sum_{i=1}^{m} a_i - \sum_{i=1}^{m-1} \xi_i(t;V) = x^*_m - t\sum_{i=1}^{m-1} v_i, \end{cases}$$

where $a_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ is the vector of the ith individual's initial endowments. By defining the vector $v_m = (v_{m1}, v_{m2}, \ldots, v_{mn})$ by

$$(3.27) \qquad v_m = -\sum_{i=1}^{m-1} v_i,$$

[14] In one respect Lange's method is similar to Pareto's. The weights λ_i in Lange's linear combination of utility functions, which he chose (arbitrarily, it seems) to be equal to 1, could instead be chosen to be precisely the weights $1/U^*_{i1}$ in Pareto's formula (3.24). The latter is the more suitable and natural choice, since Pareto's weights are proportional to the reciprocals of the individuals' marginal utilities of income, and these reciprocals are precisely the "prices" corresponding to the slope of the utility possibility surface at the equilibrium point; cf. Samuelson [209], [211, p. 11].

we may write the mth function of (3.26) in the same form as the first $m-1$. Note that if V has rank $\leqq n-1$ (which will be the case if $m \leqq n$) there will exist a vector $p \neq 0$ satisfying $Vp = 0$; this could be identified with Pareto's disequilibrium price vector p, although it should be noted that it could have negative components. Note further that if $m = n$ and V has full rank $m-1$, p is defined uniquely up to a proportionality factor; this special assumption could thus be used to justify Pareto's procedure.

Now, along a path ξ_i defined by (3.26), for $t \geqq 0$, the ith individual's utility is defined by

$$u_i(t;V) = U_i(\xi_i(t;V)) \quad (i = 1, 2, \ldots, m). \tag{3.28}$$

Denoting $\dot{u}_i = \partial u_i/\partial t$ and $\ddot{u}_i = \partial^2 u_i/\partial t^2$, we may expand $\dot{u}_i$ in a Taylor's series around $t = 0$ to one term, to obtain

$$\dot{u}_i(t;V) = \dot{u}_i(0;V) + t\ddot{u}_i(0;V) + o(t), \tag{3.29}$$

where $o(t)$ is a function such that $o(t)/t \to 0$ as $t \to 0$. Upon applying the chain rule to (3.28) and using (3.26) and (3.27), (3.29) becomes

$$\dot{u}_i(t;V) = \sum_{j=1}^{n} U_{ij}^* v_{ij} + t\sum_{j=1}^{n}\sum_{k=1}^{n} U_{i,jk}^* v_{ij} v_{ik} + o(t), \tag{3.30}$$

where U_{ij}^* and $U_{i,jk}^*$ denote respectively the partial derivatives $\partial U_i/\partial x_{ij}$ and $\partial^2 U_i/\partial x_{ij}\partial x_{ik}$ evaluated at x_i^*. Note that the first term on the right in (3.30) corresponds to that of Pareto's formula (3.20).

To obtain the analogue of Pareto's formula (3.24) we note from (3.19) and (3.27) that

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{U_{ij}^*}{U_{i1}^*} v_{ij} = \sum_{j=1}^{n} p_j^* \sum_{i=1}^{m} v_{ij} = 0. \tag{3.31}$$

Thus, from (3.30) and (3.31) we obtain the required formula

$$\sum_{i=1}^{m} \frac{1}{U_{i1}^*} \dot{u}_i(t;V) = t\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{1}{U_{i1}^*} v_{ij} v_{ik} + o(t). \tag{3.32}$$

The argument can now proceed as follows. Suppose first that

$$\sum_{j=1}^{n} U_{ij}^* v_{ij} < 0$$

for some individual i; then from (3.30) it follows that, for sufficiently small $\varepsilon > 0$, $\dot{u}_i(t;V) < 0$ for $0 < t < \varepsilon$, i.e., the ith individual must become worse off over a sufficiently small distance along such a path. This corresponds to the kind of path considered by Pareto in Figures 2 and 3 above. Now suppose, on the contrary, that

$$\sum_{j=1}^{n} U_{ij}^{*} v_{ij} \geqq 0 \quad \text{for all } i = 1, 2, \ldots, m.$$

From (3.31), a positively weighted sum of these m terms must vanish; hence each term must do so, i.e.,

$$\sum_{j=1}^{n} U_{ij}^{*} v_{ij} = 0 \quad \text{for } i = 1, 2, \ldots, m. \tag{3.33}$$

Geometrically, (3.33) states that v_i must be orthogonal to p^* for each i, i.e., it must belong to the ith individual's budget hyperplane. Now, U_i may be said to be *strongly quasi-concave* in a neighborhood of x_i^* if, for all x_i in that neighborhood, the quadratic form

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 U_i}{\partial x_{ij} \partial x_{ik}}(x_i) v_{ij} v_{ik} \tag{3.34}$$

is negative definite subject to

$$\sum_{j=1}^{n} \frac{\partial U_i}{\partial x_{ij}}(x_i) v_{ij} = 0. \tag{3.35}$$

Assuming this strong quasi-concavity property to hold for each i (it would actually be enough to assume it for a single i), it follows that the weighted sum of quadratic forms in (3.32) is negative; and since the remainder term is $o(t)$, it becomes negligible in comparison with the negative term for sufficiently small t. Thus, the expression (3.32) is negative for sufficiently small $t > 0$. Since the coefficients $1/U_{i1}^*$ are positive, at least one person must therefore become worse off as one starts to move along such a path.[15]

We may now compare this argument with Pareto's. One difference that should immediately be noted is that Pareto's procedure of representing movements away from equilibrium as movements along parallel budget hyperplanes determined by non-equilibrium prices is not only unnecessary but unduly limiting when $m > n$; for, by requiring the v_is all to be orthogonal to a single

[15] It is clear from the above argument that it is enough to assume that the quadratic form (3.34) is negative *at* x_i^* whenever (3.35) holds at x_i^* for the given v_i, for $i = 1, 2, \ldots, m$ (i.e., when (3.33) holds). When the sufficient condition is stated in this weak fashion, however, the $t > 0$ for which at least one individual is worse off may have to be very small. The question then arises as to whether there is a positive greatest lower bound to such maximal t for all V. This can be shown to be the case, by a compactness argument; details are provided in Weinberger [243]. It follows that the given equilibrium is a "strict local Pareto optimum" in Smale's [224] sense, namely that there is a neighborhood of the given equilibrium $(x_1^*, x_2^*, \ldots, x_m^*)$ such that for every other allocation in that neighborhood satisfying the constraints (3.17), at least one individual is worse off.

price vector p, the movements (3.26) are constrained to a subset of possible V matrices so that, on the face of it at least, it might be possible for all individuals to benefit along one of the paths that is necessarily excluded by this procedure when $m > n$ (or along one of the paths excluded by his requirement of positive prices). Pareto also supplemented the paths (3.26) by similar linear paths for the prices, whereas in the above argument (as well as in Pareto's own diagrammatic analysis) the direction of movement of the paths followed by the quantities consumed was initially perturbed (in relation to the original budget hyperplanes) but then fixed once and for all.[16] If one examines Pareto's proof closely, however, one will see that no essential use is made of his assumed price changes, and that the logic of his argument is essentially the same as that contained in the argument presented above, but with one important difference: his failure to fully exploit the quasi-concavity of the utility functions, and his reliance instead on concavity.

In essence, Pareto's argument proceeded as follows. At the equilibrium point $x_i^* = \xi(0; V), i = 1, 2, \ldots, m$, formula (3.32) reduces to

$$\sum_{i=1}^{m} \frac{1}{U_{i1}^*} \dot{u}_i(0; V) = 0, \tag{3.36}$$

which we can interpret as corresponding to Pareto's formula (3.24). Pareto's verbal statement of the "infinitesimal" version of his proposition can be justified as being simply a description of the possible directions of movement $\dot{u}_i(0; V)$ permitted by (3.36); that is, (3.36) implies that the $\dot{u}_i(0; V)$ cannot all be positive and, likewise, cannot all be negative. Pareto's "infinitesimal" proposition states this and nothing else. This leaves open, then, only two remaining possibilities: (1) $\dot{u}_i(0; V) = 0$ for $i = l, 2, \ldots, m$; (2) some $\dot{u}_i(0; V)$s are positive and others are negative. In his 1902 paper [180, p. 412], Pareto excluded the first possibility on the fallacious grounds that, since it implied that the direction of movement had to be along the original budget hyperplanes, it would contradict the hypothesis that the paths departed from equilibrium. In the *Manuel* [185, Appendix, §122, p. 652], however, in the passage quoted above, he correctly noted — quite possibly as a result of Scorza's criticism on this point [220, p. 55] — that such paths had to be and could be accounted for. Also in his 1902 paper Pareto implicitly assumed that each of the utility functions U_i was additively separable [180, p. 414]; only later, in the *Manuale* [182, Appendix, §49, p. 551n] was this assumption both made explicit and relaxed.

[16] That is, Pareto assumed the prices p_i to move along the paths $\pi_i(t) = p_i^* + \sigma_i t$, whereas he would have done better, if he was going to use prices out of equilibrium at all, to adopt the paths $\pi(t) = p_i$; and in fact his argument does not depend on his price equations at all. Scorza [220] was later to object that assuming the σ_is and $v_{ij}s$ to be constants, Pareto's price equations when combined with the quantity equations (3.26) together violated the constraints (3.17) and (3.18). In his later formulations in the *Manuale* [182, Appendix, §48, p. 549] and *Manuel* [185, Appendix, §121, p. 651] Pareto pointed out that "it should be carefully noted that the σ_is are not constants" but are assumed to vary in such a way as to satisfy the constraints. Undoubtedly this remark was in response to Scorza's objection, and we may interpret the "minor blunder" that Pareto later ascribed to Scorza as being Scorza's failure to note that the σ_is were variable. The slip was surely Pareto's however, rather than Scorza's.

Accordingly, the 1902 argument proceeded as follows. Each utility function was assumed to be additively separable and to exhibit diminishing marginal utility for each commodity, i.e.,

$$U_{i,jj} < 0; \quad U_{i,jk} = 0 \text{ for } j \neq k. \tag{3.37}$$

Pareto now implicitly considered ([180, p. 414] — and did so explicitly in the *Manuale* [182, Appendix, §48, p. 550]) — the second derivative of u_i:

$$\ddot{u}_i(t; V) = \sum_{j=1}^{n} \sum_{k=1}^{n} U_{i,jk}(\xi(t; V)) v_{ij} v_{ik}. \tag{3.38}$$

When (3.37) holds this obviously reduces to $\sum_{j=1}^{n} U_{i,jj}(\xi(t;V))v_{ij}^2 < 0$. Assuming case (2) to hold, Pareto then simply observed that since each of the $\dot{u}_i(t; V)$ was decreasing, those that were negative at $t = 0$ had to remain negative for $t > 0$, and those that were positive could either remain positive or become negative; in any event, at least one individual would necessarily remain worse off. Later, in the passage previously cited in the *Manuel* [185, p. 652], Pareto realized that the same argument could be used in case (1), that is, if each $\dot{u}_i(t; V) = 0$ at $t = 0$, they would all have to become negative for $t > 0$, i.e., all individuals would necessarily become worse off.

In the *Manuale* [182, Appendix, §§48-9, pp. 549–552] (see also the *Manuel* [185, Appendix, §§121–4, pp. 650–654]), Pareto relaxed the assumption of additively separable utilities, and noted that his "finite" proposition followed if the expressions (3.38) were all negative for $t > 0$. It remained then simply to find sufficient conditions for this to hold; and he specified the conditions that the principal minors of the Hessian matrix $[U_{i,jk}]$ should be alternately negative and positive.

A proof of virtually complete generality was now just within Pareto's reach. In place of strong quasi-concavity, all he had to assume was the slightly stronger condition that each individual's preferences were "strongly concavifiable," i.e., that for each i there exist *some* strongly concave utility function U_i (i.e., one whose Hessian had strictly oscillating principal minors) capable of representing his preferences. Of course, the same preferences could be represented by non-concave utility functions, and in terms of such utility functions no conclusion could be drawn one way or the other by means of such an argument. In terms of the argument presented above, strong quasi-concavity can be used when (3.33) holds, and when (3.33) does not hold one need not consider second derivatives at all, since by (3.30) they are of negligible importance in comparison with the first-order terms. (The geometric meaning of this is obvious in terms of Figures 2 and 3; the first case, when (3.33) holds, corresponding to movements along the budget lines, and the second case to other movements.) Nevertheless, it would have made for a perfectly correct argument to employ (3.32) for both cases — that in which (3.33) did not hold as well as that in which it did — provided preferences could be and were represented by strongly concave utility functions.

Although such an argument was just within his reach, it did not occur to

Pareto to avail himself of it. Instead, he had recourse to his theory of interrelated commodities.

As was noted in §2.2 above, Pareto defined the n commodities to be complementary (in his words, to have "a dependence of the first kind") for the ith consumer if

$$U_{i,jj} < 0, \quad \text{and} \quad U_{i,jk} > 0 \quad \text{for } j \neq k, \tag{3.39}$$

and if, moreover, along any ray from the origin,

$$\xi_i^0(t; v_i) = (v_{i1}t, v_{i2}t, \ldots, v_{in}t), \; v_{ij} > 0, t > 0, \tag{3.40}$$

marginal utility was decreasing, i.e.,

$$\frac{d^2}{dt^2} U_i(\xi_i^0(t; v_i)) = \sum_{j=1}^{n} \sum_{k=1}^{n} U_{i,jk}(\xi_i^0(t; v_i) v_{ij} v_{ik} < 0. \tag{3.41}$$

The intuitive idea behind (3.41) was that when n commodities are mutually complementary, then when they are combined in fixed proportions they may be treated as a single commodity which exhibits diminishing marginal utility (*Manuale* [182, Ch. IV, §42, p. 258], [185, p. 270]). Pareto believed that (3.39) and (3.41) implied that (3.38) was negative definite; unfortunately his proof was faulty ([182, Appendix, §49, p. 552], [185, Appendix, §142, pp. 653–4]), since it overlooked the dependence of the terms $U_{i,jk}(\xi_i^0(t; v_i))$ in (3.41) on the v_{ij}s, and thus, in effect, assumed what had to be proved.[17]

Thus it happened that Pareto associated the negative definiteness of (3.38) with circumstances in which commodities were complementary or independent, and failure of negative definiteness with circumstances in which commodities were substitutes. And, equally erroneously, he considered the negative definiteness of (3.38) to be indispensable for the proof of the Pareto optimality of competitive equilibrium. This accounts for the surprising qualifying sentence at the end of the following passage in the *Manuale* [182, Ch. VI, §37, p. 339], [185, p. 356]:

> Consequently, if one departs from the equilibrium position by a finite distance along a straight line, the ophelimities enjoyed by the two individuals may vary in such a way that one increases while the other decreases, or both decrease; but they cannot both increase. This is true, however, only for commodities whose ophelimities are independent, or in the case in which the commodities have a dependence of the first kind.

We may conclude with respect to Pareto's attempted demonstration of the "Pareto optimality" of competitive equilibrium that while neither his statement

[17] He reasoned that if $\sum_{j=1}^{n} \sum_{k=1}^{n} U_{i,jk} v_{ij} v_{ik} < 0$ for all positive v_{ij}, v_{ik}, then it would certainly hold if v_{ij} and v_{ik} had opposite sign, since $U_{i,jk} > 0$. What this argument overlooked was that $U_{i,jk}$ was a function of the v_{ij}s and v_{ik}s, and defined only for positive values of these arguments.

nor his proof of the proposition was fully satisfactory, it is nevertheless clear that in the course of meeting Scorza's objections he had made a major reformulation of Walras's doctrine that could no longer be accused of being tautological.[18] And the profession had to wait for another fifty years to pass before a fully satisfactory proof was developed by Arrow [13] and Debreu [64]. In the meantime it could only be questioned whether Pareto's proposition was logically correct.

In a final riposte [220, p. 48], Scorza did just that. He accepted the correctness of Pareto's proposition regarding infinitesimal variations, but rejected his proof for the case of finite variations on the ground that Pareto's price variations $\pi_i(t) = p_i^* + \sigma_i t$ violated the feasibility and budget constraints (see footnote 16 above). He went on, however, to provide his own proof of Pareto's proposition for infinitesimal and "very small" — but not "large" — variations. His method anticipated Lange's [129], in that he formulated the problem in terms of maximizing the sum of the individuals' utility functions subject to the constraints (following Pareto, he included the budget constraints (3.18) in addition to the feasibility constraints (3.17)) employing the method of Lagrangean multipliers. He even went on to obtain second-order conditions, so as to confirm Pareto's "finite" proposition *locally*; but so as not to give Pareto any more credit than was possible, he remarked that the proposition was not new but was known to Launhardt and Edgeworth! However, having conceded this much, he went on to criticize Pareto for not taking into account the case of multiple competitive equilibria; he assumed — and in the absence of an explicit assumption of (global) quasi-concavity of utility functions, and assuming the conditions (3.19) to correspond only to a *local* maximum for each individual, this could well be true — that it would be possible to find such a case in which individuals were all better off in one equilibrium than in another, and thus concluded that Pareto's proposition was locally, but not globally, true [220, pp. 60–61]. In his last sally [181] Pareto accused Scorza of having committed a minor and a major blunder, refusing to say what they were. The minor one was presumably the slip mentioned in footnote 16 above — which was really Pareto's and not Scorza's. The major one could well have referred to Scorza's assumption concerning multiple equilibria; however, in the *Manuale* [182, Appendix, §23, p. 515] and *Manuel* [185, Appendix, §35, pp. 564–5], Pareto indicated (but without mentioning Scorza by name) that Scorza's major error was his failure (but in his earlier article [219]) to comprehend that competitive behavior was characterized by the treatment of prices as parameters by individual agents. Thus, both men ended their acrimonious dispute at cross purposes, without meeting the point. Such was the

[18] The proposition is hinted at in Edgeworth [67, p. 23]: "Or, again, we may consider that motion is possible so long as, one party not losing, the other gains. The point of equilibrium, therefore, may be described as a relative maximum, the point at which, e.g., Π being constant, P is a maximum" (where Π and P represent the utility functions of two individuals). A similar although slightly less explicit statement will be found in the second edition of Marshall's *Principles*, with which Pareto was familiar [145, p. 506; 8th ed., p. 471]: "It is true then that a position of equilibrium of demand and supply is a position of maximum satisfaction in this limited sense, that the aggregate satisfaction of the two parties concerned increases until that position is reached . . . ". Pareto's achievement was to turn these casual remarks into careful, precise statements, to recognize that they required proof, and to provide a formal proof.

origin of the concept of Pareto optimality and of the proposition that it was a property of competitive equilibrium!

3.6 Social welfare

Ten years later it is interesting to find Pareto changing roles. In a remarkable paper [189] he introduced the concept of a differential form

$$dW = \sum_{i=1}^{m} \delta_i dU_i, \tag{3.42}$$

which we may interpret as the differential of a social welfare function.[19] He contrasted this with the expression (3.24), which he now represented

$$dF = \sum_{i=1}^{m} \frac{1}{U_{i1}} dU_i, \tag{3.43}$$

(although without providing a complete justification)[20] as an exact differential remarking that the aggregation of heterogeneous utilities defined by (3.43) was justified by the fact that the terms in the sum were all expressed in units of commodity 1 (see also the *Manuel* [185, Appendix, §127, p. 655]). He went on

[19] It could be argued that Pareto did not quite arrive at the concept of a social welfare function since the expression on the right in (3.42) might not be integrable. However, Pareto did write it as an exact differential, implying (at least implicitly) that the expression was not only integrable but exact. Since there is every reason to assume that he regarded the δ_is as constants, no question of integrability arises. Of course, in view of Pareto's poor understanding of the integrability problem (see §2.3 above), as well as the ambiguity of the differential expression (3.43) below (see footnote 20), we cannot be sure that Pareto perceived the matter so clearly himself, and it is certainly just as well that he was not challenged by Volterra or someone else to investigate the integrability of (3.42)! What we can say, however, is that in the use to which he put (3.42), he treated it just as if it was the differential of a social welfare function.

[20] Pareto's notation was a bit ambiguous and did not make entirely clear whether one should interpret (3.43) to mean

$$dF(x_1, x_2, \ldots, x_m) = \sum_{i=1}^{m} \frac{1}{U_{i1}(x_i)} dU_i(x_i) \tag{3.43$'$}$$

or

$$dF(x_1, x_2, \ldots, x_m; x_1^*, x_2^*, \ldots, x_m^*) = \sum_{i=1}^{m} \frac{1}{U_{i1}(x_i^*)} dU_i(x_i). \tag{3.43$''$}$$

However, the expression (3.43') cannot be an exact differential unless each U_{i1} is constant, yielding the Auspitz-Lieben-Wilson-Samuelson utility function of footnote 2 (§2.1 above); since the choice of numéraire is arbitrary this means that each U_i would have to be linear, and the indifference surfaces of each individual would have to be parallel to the budget hyperplane. This is absurd. Thus, only the interpretation (3.43") makes sense, and this defines a family of indifference surfaces, one for each equilibrium point $(x_1^*, x_2^*, \ldots, x_m^*)$; these are precisely (when $x_i = x_i^*$ and $n = 2$) the so-called "Scitovsky indifference curves" (cf. Samuelson [211], Koopmans [128, pp. 51–2], and the remark made in footnote 14 above).

to point out that although interpersonal comparisons of utility were not needed to explain market phenomena — which he identified as being in the province of economics — they were now needed to help explain non-market phenomena such as collective (governmental) decisions — which he placed in the domain of sociology. It is impossible to read this extraordinary paper without sensing its relevance to problems of our own day, when Pareto distinguishes societies that give greater relative weights to the utilities of humanitarians and criminals than to those of the criminals' victims, from societies that do the opposite. It was clear which way Pareto's preferences ran; in 1913 Pareto was definitely a "law and order" man. Nevertheless he did not propose (3.42) as a device for forming bases for policy prescriptions, as Bergson [23] and Samuelson [208] were to do with the social welfare function after him (and quite independently of Pareto); he had become too wise — or some might think, too cynical — to believe that policy exhortations on the part of economists could have any effect. As we might express it today, for Pareto the social welfare function was an expression of the revealed preferences of a centrally organized society, as exhibited by its observed behavior; it was an instrument not for recommending or denouncing, but for describing and classifying.

This paper, together with the further elaboration in the *Treatise on General Sociology* [191, §§2111–2139], also provide some interesting light on Pareto's earlier dispute with Scorza. Pareto, now taking the side of Cassel and Scorza (but without mentioning their names), quotes approvingly a passage from a textbook by Pierre Bovin criticizing Walras for reasoning in a circle in claiming that the presence of equations (3.19) in the system of equations characterizing competitive equilibrium proves that each participant achieves maximum utility. And he suggests [191] that his own optimality criterion — which he now sharply disassociates from that of Walras — should be described as yielding "maximum utility *for* society" as opposed to "*of* society"; or better still [189, 191], that positions satisfying his criterion should be described simply as "points of type **P**" — in the hope that this would induce people to consult the definition rather than the etymology of the terms. He added [189, p. 338]: "I . . . pray the Lord that people will not start looking for the etymology of that letter, as they did in the case of the word *ophelimity*"; on the other hand, by his choice of that particular letter we can safely assume that he was aware of the nature of his own contribution to general equilibrium theory in developing the subject beyond the point at which it had been left by Walras.

Chapter 4

Population and Income Distribution

4.1 Infant mortality

Pareto first introduced what we now know as the Pareto distribution in the course of what must be one of the earliest pieces of econometric research to have been carried out in the area of "human capital" [155]. Availing himself of mortality tables and survival curves supplied by Bodio [28], he found that for the first 20 years of life a good fit was provided by the formula[1]

$$N(x) = \frac{a}{(b+x)^{\alpha}} \quad (0 \leqq x \leqq 20), \tag{4.1}$$

where $N(x)$ denotes the number of children surviving to age x out of an initial number $N(0)$. He estimated the parameters a, b, and α for a number of countries, including in particular Switzerland (where $b = .023$ and $\alpha = .054$) and Bavaria (where $b = .004$ and $\alpha = .065$). Thus, Bavaria displayed a markedly higher infant mortality rate than Switzerland.[2] Pareto then posed the following

[1] Apparently, neither Pareto nor Bodio was familiar with the Gompertz [110] and Makeham [138, 139] mortality laws, so the idea of fitting a curve to mortality data while original was not new. However, the Gompertz and Makeham laws, while they continue to be very useful in actuarial work, fit well only for ages $x \geqq 20$. Pareto's law fits well for $0 \leqq x \leqq 20$ but could not be extrapolated any further. In a letter dated March 26, 1891, to Pantaleoni [195, I, Letter 11, p. 27] Pareto revealed that he had carried his researches on mortality distributions quite a bit further than his published work indicates, having fitted the following formula to Italian data during the years 1881–3 on the number of survivors at age x:

$$N(x) = 1.548(x+2)^{-2} + 0.613 - 0.00206x - 0.0000412x^2 - 0.130\exp\{-0.0046(x-82)^2\}.$$

He asserted that this provided a good fit between the ages 0 and 88, the first term preponderating in the early years, the last term in the late years, and the remaining terms in the intermediate years.

[2] Defining the instantaneous death rate, or "force of mortality" as it is known in the actuarial literature (cf. Brillinger [41]), by

$$V(x) = -N'(x)/N(x)$$

question: Is it true, as had been alleged, that a country with a low infant mortality has an economic advantage over one with a high infant mortality, in that fewer children need be raised in order to produce the same number of adults?

To answer this question Pareto availed himself of the results of a study by Engel [81] indicating that cumulative expenditures per child as a function of age could be represented by the quadratic equation

$$s(x) = s(0) + px + \tfrac{1}{2}qx^2 \quad (0 \leqq x \leqq 20), \tag{4.2}$$

where $s(0) = 100, p = 105$, and $q = 10$. Cumulative expenditures up to time n on an initial group of $N(0)$ children would then be

$$S(n) = N(0)s(0) + \int_0^n N(x)ds(x) = 100N(0) + \int_0^n \frac{a}{(b+x)^\alpha}(105 + 10x)dx. \tag{4.3}$$

Integrating (4.3) by parts and dividing by the number of survivors at time n, $a/(b+n)^\alpha$ Pareto then obtained a complicated expression for the cost of production of a human being aged n. Inserting his estimates for a, b, and α, he found the cost of production of an adult aged 20 to be 4418 marks for Switzerland and 4485 marks for Bavaria — a very slight difference. The explanation is quite straightforward [155, pp. 454–5]: "in countries with a low infant mortality many children are saved in the first years who die before reaching adulthood. . . From the economic point of view it is an advantage for Bavaria that the [high] mortality during the first few years of life should have reduced mortality between the ages of 15 and 20."

Pareto extended his calculations to several other European countries and found the same stability in the cost of an adult. He then applied these results to obtain an estimate of the cost to Italy of the emigration during the years 1887–1893 [153, I, §§253-4, pp. 151–2] — which emigration he had already attributed to Italy's protectionist policies [149].

4.2 The law of demand

Pareto next turned his attention to the empirical "law of demand" [158]. His starting point was a relationship that had been found by Gregory King [127] in seventeenth century England between the wheat harvest and the price of wheat. He sought to reconcile this result with economic theory by: (i) postulating

(written $\mu_x = -\dot{\ell}_x/\ell_x$ in actuarial notation), we find that with the Pareto law (4.1),

$$V(x) - \frac{\alpha}{b+x}.$$

Given Pareto's values for the parameters this is uniformly lower for Switzerland than for Bavaria. The Gompertz Law and the Makeham First Law are instead respectively

$$V(x) = Be^{\lambda x}, \quad V(x) = A + Be^{\lambda x} \ (A, B, \lambda > 0).$$

identical individual demand functions for wheat, which he took to be of the form

$$x_1 = h_1(p_1, I) = a\frac{(b_1 I/p_1)^c}{1 + (b_1 I/p_1)^c}, \tag{4.4}$$

where x_1 and p_1 are the quantity and price of wheat and I is the individual's income; and (ii) postulating a density function for the distribution of income, which he took to be of the form

$$f(I) = \frac{C}{I^{1+\alpha}} \quad (0 < m \leqq I < \infty;\ \alpha > 1) \tag{4.5}$$

where m is the minimum (subsistence) income. For this to be a density function the constant C must satisfy

$$C = \alpha m^\alpha. \tag{4.6}$$

Given (4.4) and (4.5) the aggregate demand for wheat is determined by

$$X_1 = \int_m^\infty x_1 f(I) dI = aC \int_m^\infty \frac{(b_1 I/p_1)^c}{1 + (b_1 I/p_1)^c} I^{-1-\alpha} dI. \tag{4.7}$$

This expresses X_1 as a complicated function of p_1. By means of elaborate computations Pareto obtained an approximating expression for this function. He then estimated α to be 1.6 in the seventeenth century (on the basis of estimates of 1.52 and 1.38 in 1843 and 1881 and backward extrapolation), and then found that by choosing $c = .46$ and $(b_1 m)^c = .2$ the resulting function closely reproduced the relationship found by Gregory King.

4.3 Income distribution and taxation

When introducing his formula for the income curve [158], Pareto drew a number of interesting conclusions. For convenience let us now revert to Pareto's notation and denote income by x rather than I, so that the density function given by (4.5) and (4.6) becomes[3].

$$f(x) = \frac{\alpha m^\alpha}{x^{\alpha+1}} \quad (0 < m \leqq x < \infty;\ \alpha > 1).^4 \tag{4.8}$$

[3] The dependence of the function f on m and α is not indicated explicitly. However, I shall in §4.4 employ the notations $f(x; m, \alpha)$, $G(x; m, \alpha)$, etc., when discussing comparisons of distributions for different values of the parameters m and α

[4] The assumption $\alpha > 1$ is required for the existence of a finite mean (4.15) of the distribution. Pareto [162] briefly considered the case of a truncated distribution defined over the finite domain $m \leqq x \leqq M$ for $\alpha \leqq 1$; however, he expressed grave doubts (in Pareto [164]) as to the possibility of implementation of $\alpha < 1$ by a socialist regime, given his description of α as an "inequality" coefficient (see §4.4 below). The case $m \leqq x \leqq M$ and $0 < \alpha \leqq 1$ was analyzed in great detail by Bortkiewicz [30, pp. 241–251].

Let $N(x)$ be the number of individuals with incomes exceeding x; then the cumulative distribution function, defining the proportion p of the population earning incomes less than or equal to x, is given by

$$p = F(x) = \int_m^x f(\xi)d\xi = 1 - \frac{N(x)}{N(m)} = 1 - \left(\frac{m}{x}\right)^\alpha, \tag{4.9}$$

hence

$$N(x) = N(m)\int_m^\infty f(\xi)d\xi - N(m)\left(\frac{m}{x}\right)^\alpha = \frac{A}{x^\alpha}, \tag{4.10}$$

where

$$A = N(m)m^\alpha. \tag{4.11}$$

The proportion $q = 1 - p$ of the total population with incomes exceeding x is then given by

$$q = G(x) \equiv 1 - F(x) = \frac{N(x)}{N(m)} = \int_x^\infty f(\xi)d\xi = \left(\frac{m}{x}\right)^\alpha. \tag{4.12}$$

Let $R(x)$ be the sum total of incomes exceeding x; then

$$R(x) = \int_x^\infty \xi\left(-\frac{dN(\xi)}{d\xi}\right)d\xi = N(m)\int_x^\infty \xi f(\xi)d\xi = \frac{A\alpha}{\alpha - 1}\frac{1}{x^{\alpha-1}}. \tag{4.13}$$

The proportion r of total income which exceeds x, i.e., the relative share of those earning more than x, is then

$$r = \Psi(x) \equiv \frac{R(x)}{R(m)} = \frac{1}{\mu}\int_x^\infty \xi f(\xi)d\xi = \left(\frac{m}{x}\right)^{\alpha-1}, \tag{4.14}$$

where

$$\mu = \int_m^\infty \xi f(\xi)d\xi = \frac{R(m)}{N(m)} = \frac{\alpha}{\alpha - 1}m \tag{4.15}$$

is the mean of the distribution, equal to per capita income.

To find the income x^* which divides total income into two equal parts,[5] so that the total of incomes exceeding x^* is equal to the total of incomes less than x^*, we set $\Psi(x^*) = \frac{1}{2}$ to get $x^* = 2^{1/(\alpha-1)}m$. Then, as Pareto observed, when $\alpha = 1.5$ we obtain $x^* = 4m$, so that "the income which divides the total social income into equal parts is equal to four times the minimum income," and consequently "a tax of one per cent on incomes up to four times the lowest will yield as much as a tax, also of one per cent, on incomes exceeding four times the lowest" [158, p. 62]. Likewise (setting $\Psi(x^*) = \frac{1}{3}$) "the amount paid by taxpayers with incomes higher than nine times the minimum will be only

[5] Not to be confused, of course, with the median, which divides the *population* into two equal parts, and is equal to $2^{1/\alpha}m$; nor with the mean income $\mu = [\alpha/(\alpha - 1)]m$.

half that paid by those with incomes lower than nine times the minimum," and so on. Thus: "even with taxes at an equal percentage of incomes, the rich contribute far less to public expenditures than the poor, whereas they benefit much more from them. For whom, if not for the vain rich, are funds expended on armaments and the like?"

Pareto went on to use his income distribution formula to illustrate the relative ineffectiveness of purely redistributive measures as a means of improving the lot of the poor: "If, for incomes exceeding 49 times the minimum, the excess were to be distributed equally among those with incomes below that limit, each of the latter would receive about 1/171 of the income 49 times the minimum." Setting $x_1/m = 49$ and $\alpha = 1.5$ we obtain

$$\frac{\int_x^\infty (x - x_1) f(x) dx}{1 - G(x_1)} = \frac{x_1}{\alpha - 1} \frac{1}{(x_1/m)^\alpha - 1} = \frac{x_1}{171}. \tag{4.16}$$

In a subsequent paper [162] Pareto fitted his formula $N(x) = Ax^{-\alpha}$ to data from different countries and epochs and came to the conclusion not only that the fit was good but also that — unlike the case of mortality curves — the estimated values of α were remarkably close to one another, varying between 1.45 and 1.72. He went on to consider the yields from proportional and progressive taxes, the latter assumed to be determined by the formula

$$t(x) = \tau(1 - m/x)^n \quad (n > 0), \tag{4.17}$$

where $t(x) \cdot x$ is the tax paid on an income of x, and τ is the asymptotic proportional rate of the progressive tax. Denoting by p the constant rate of the proportional tax, Pareto found the asymptotic rate τ, as a function of α and n, which would give rise to the same total yield as the proportional tax at rate p, given by the formula [162, p. 383]:

$$p = \tau \frac{\Gamma(n+1)\Gamma(\alpha)}{\Gamma(n+\alpha)} \tag{4.18}$$

where Γ is the gamma function. Thus, for example, for $\alpha = 1.5$ and $n = 1$ this gives $p/\tau = .667$, so that the asymptotic rate τ must be fifty percent higher than the proportional rate p.

4.4 The measurement of inequality of incomes

. In his early work [158, p. 61] Pareto remarked, without further explanation, that α could be taken as an index of the inequality of incomes. This criterion, which has been the subject of considerable controversy, was reiterated in the *Cours* [163, II, §§964-5, pp. 318–322], and provided with a detailed justification which is worth quoting:

> But what is the true meaning of the expression *less inequality of incomes*, or of those which are employed in about the same sense, *less inequality of wealth* and *less inequality in living conditions?*

> If it was a question of complete inequality of incomes, wealth, or living conditions, no ambiguity would be possible. But one can approach such a state in two essentially different ways: either by the rich becoming poorer or by the poor becoming richer.
>
> These are two different phenomena, and if it were a matter of positive science, where facts count for everything and words for nothing, one would not hesitate to designate things that are so different by different words. But political economy is still, very often, no more than a kind of literature. Great importance is attached to words at the expense of facts. People therefore engage in debates as to which of the two phenomena should have reserved to it the denomination *less inequality of wealth*.

Pareto settled on the second definition:

> A reduction in this inequality will therefore be defined by the fact that the number of poor people decreases relatively to the number of rich people, or equivalently, relatively to the total number of members of society. This is the sense that seems to have prevailed, and is therefore the one I shall adopt. Except in cases of absolute necessity, such as the need to avoid the muddle that the confusion between ophelimity and utility led to, it is better to resign oneself to not running too much afoul of current prejudices.
>
> In general, when the number of persons with incomes less than x decreases[6] relatively to the number of persons with incomes greater than x, I shall say that the inequality of incomes decreases. But the reader is duly warned that by these terms I mean simply to designate this thing and nothing else.

Despite his strong warning, subsequent authors have criticized this definition as being "erroneous," instead of simply accepting it as a definition. Others, e.g. Sorel [226], while acknowledging Pareto's right to define any term as he pleases, nevertheless found it misleading and inappropriate; while some authors, e.g. Benini [20, 21] and Bresciani-Turroni [33, 34]; actually adopted an exactly opposite definition in which the terms were reversed. Subsequently, in the *Manuel*

[6] The text of the *Cours* reads "increases" instead of "decreases" [163, II, §964, p. 320]. In his review article, Bortkiewicz [29, p. 120] interpreted this as a slip. Sorel [226, p. 586] drew attention to the anomaly, pointing out that he had drawn it to Pareto's attention directly and had received the following reply: "The term *less inequality in living conditions* does not seem to capture what is sought to me any more than it does to you. I would gladly have substituted another term, but I have already been reproached for my luckless 'ophelimity', and I wanted to avoid providing my critics with more verbal tangles. At least I have made a point of explaining carefully what it was that was designated by this term." In his formal reply [169] Pareto expressed himself in substantially the same terms (see below), but did not correct the passage. However, nine years later, in the *Manuale* [182, VII, §24, p. 371n] he corrected "increases" to "decreases," calling it a "misprint." In the *Manuel* [185, p. 390n] he added: "This is a misprint which I spotted immediately after the publication of the *Cours*." Bortkiewicz later described it as a "slip of the pen" [30, p. 221n]. Whether it was the pen or the type that slipped, the original passage succeeded in leading Gibrat astray [103, p. 83].

[185, VII, §24, p. 390], Pareto, in response to such reactions, changed his terminology to "inequality of *relative* incomes. "

Pareto's mathematical definition of *equality* of incomes is given by the index $G(x)$ of (4.12), which we may express as

$$G(x; m, \alpha) = \frac{N(x; m, \alpha)}{N(m; m, \alpha)} = \int_x^\infty f(\xi; m, \alpha) d\xi = \left(\frac{m}{x}\right)^\alpha \tag{4.19}$$

to indicate the dependence on the parameters explicitly. In general, in terms of any cumulative distribution function $F_\nu(x)$ in a certain class indexed by ν, Pareto's index of *inequality* would be simply

$$1 - G_\nu(x) = F_\nu(x), \tag{4.20}$$

and this would provide only a partial ordering of distributions, since we could presumably say that F_1 is more unequal than F_2 only if $F_1(x) \geqq F_2(x)$ for *all* x, with equality holding only for a finite number of values of x. In Pareto's case (4.8) however, it is clear from (4.19), as Pareto showed [163, II, §(965)^1, p. 321], that $\partial G/\partial m > 0$ and that $\partial G/\partial \alpha < 0$ for $x > m$. Thus, inequality of incomes varies directly with α and inversely with m.

After noting that $G(x; m, \alpha)$ varies inversely with α Pareto unfortunately altered his definition and proceeded to identify α as the parameter representing inequality.[7] In terms of this terminology he laid down the following proposition [163, II, §965, p. 320]: "(1) A rise in the minimum income, and (2) a decline in the inequality of incomes, cannot be produced either singly or together unless total income increases faster than population." I.e., it is impossible for m to rise with α constant, or for α to fall with m constant, or for m to rise and α to fall together, unless per capita income μ rises. This is immediate from formula (4.15) above (cf. [163, II, §(965)^1, p. 321, formula (3)]). For Pareto this proposition was fundamental: it meant that to reduce pauperism and inequality, the only basic remedy was to increase production relative to population; and as he had already argued, purely redistributive measures could accomplish little.

Because of the importance of the proposition, it is not surprising that use of the parameter α as a measure of inequality should have been the subject of further scrutiny. It was in fact rejected as such by Benini [21], Lorenz [134], Bresciani-Turroni [33, 34, 38, 39, 40], Gini [106, 107], Dalton [61], and Pigou [201, p. 25n].

Sorel [226, p. 599] objected that in stating his proposition Pareto had brushed aside on grounds of lack of realism the possibility that m and α might move in

[7] This was undoubtedly a slip on his part. See the discussion of this point in Bortkiewicz [30, pp. 221–2], D'Addario [56, pp. 181–4], and Bresciani-Turroni [40, pp. 109–12]. See also the discussion by Tarascio [231, p. 528] who argues that "what has been almost completely overlooked is the fact that Pareto never used α *directly* as a measure of income inequality, although such a tendency is generally attributed to him," and that when the income distributrion has the form (4.10), the formula (4.19) "will be influenced by α" (p. 529). But this is the source of the slip; (4.19) is influenced by m as well as by α, whereas in the celebrated passage to be quoted presently in the text, as well as in Pareto's demonstration of it [163, II, p. 322n], the term "inequality" is applied only to α and not to m.

the same direction (see footnote 17, below); and that when restricted to cases in which m and α move in opposite directions the proposition loses much of its interest. A similar observation was made even more pointedly by Bresciani-Turroni [40, p. 112], who remarked that since with m constant a decline in α can take place only as a consequence of a rise in per capita income μ, this was "merely a consequence of the peculiar definition of inequality given by Pareto." This is made immediately evident by expressing Pareto's coefficient in the form $\alpha = 1/(1-m/\mu)$. In fact, as Bortkiewicz pointed out [30, p. 222], even in terms of Pareto's original definition (4.20), when applied to any family of distribution functions $F_\nu(x)$ with finite means μ_ν and the same minimum income m (i.e., the same domain $x \geqq m$), the fulfillment of the inequality $F_1(x) < F_2(x)$ for all $x \geqq m$ m necessarily implies that $\mu_1 > \mu_2$.

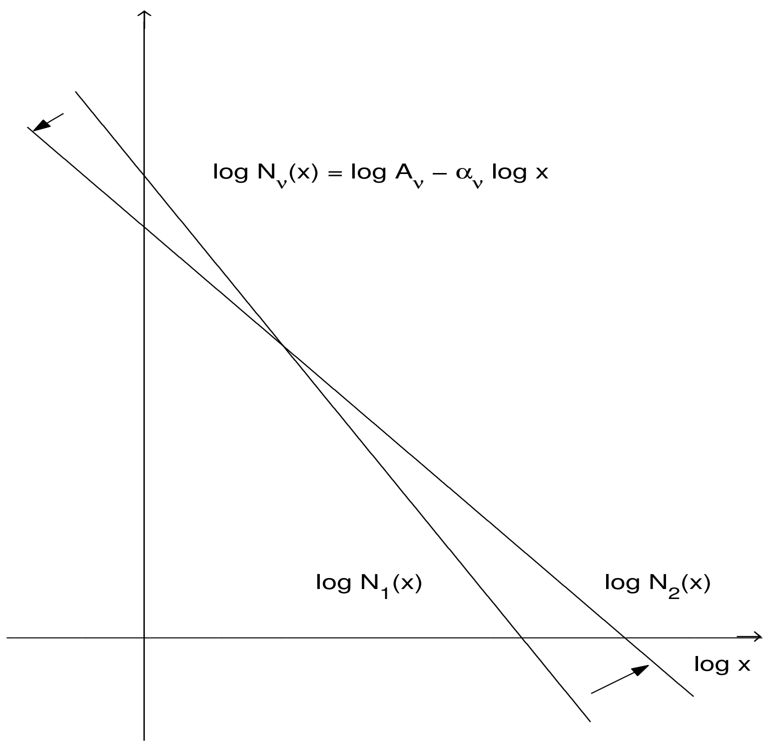

Figure 4

Benini [20, p. 194], in commenting on the values for α of 1.42 and 2.88 obtained by Pareto [167] for salaries of professionals and for property incomes respectively, took the position — but without taking direct issue with Pareto — that "the larger α, the more rapidly do property owners or taxpayers decline in numbers as incomes increase. Thus, starting from the same initial position and moving towards the highest income level, the ranks of those who owe their income to their own labor quickly thin out as far as to disappear, whereas the ranks of the privileged who live off the yield of capital are still considerable." Sorel [226, p. 593] read into this statement the implication that for Benini a decline in α was a detriment, rather than a benefit as implied by Pareto. Later, Benini was still more explicit [21, p. 227]: "The larger is α, the more rapidly do income recipients diminish at every step of the income scale, that is, the less unequal is the distribution of wealth.[8] In fact, if α were very large it would be enough to move up a few steps in the income scale to fail to find any more income recipients; which means precisely that the various strata of the population differ little in their economic conditions." Subsequently, Bresciani [33, p. 117] pointed to the contrast between this interpretation and Pareto's, and adopted Benini's. In a subsequent article [34, pp. 20–29] Bresciani subjected the concept of inequality of incomes to a searching examination. He objected with respect to Pareto's original definition (4.20) that Pareto had failed to take account of the case in which low incomes rise less rapidly than high incomes, i.e., $F_1(x) < F_2(x)$ for incomes x below a certain level, and $F_1(x) > F_2(x)$ for incomes above this level. This would in particular include the case in which actually "the rich get richer and the poor get poorer." However, within the Pareto family this can only mean that α decreases (see Figure 4). One would then have to reverse Pareto's conclusion [163, II, §965, pp. 324–5] that inequality of wealth has been declining with the development of capitalism.

Pursuing the question further, Bresciani-Turroni [35, 36] introduced, following an idea suggested by Bortkiewicz [29, p. 121], [9] the relative mean deviation

$$\theta = \frac{1}{2\mu}\int_m^\infty |x-\mu| f(x)dx \tag{4.21}$$

as a measure of inequality, for any density function with finite mean μ and defined over the interval $[m, \infty)$. It was subsequently shown by Bortkiewicz [30, p. 224] (and was implicit in Bresciani-Turroni, [35, p. 807, formula (7)]) that in

[8] Commenting on this statement, which was reproduced in Benini's text [22, p. 187], Dalton remarked [61, p. 359n]: "Professor Benini inverts Professor Pareto's measure, but apparently without realizing that he has done so." However, the contrast was immediately noticed by both Sorel and Bresciani, and it can hardly be believed that Benini was unaware of it himself; presumably he was unwilling to provoke any conflict with Pareto.

[9] Bortkiewicz suggested use of either the mean deviation (the numerator of (4.21)) or the standard deviation. For Pareto distributions with $\alpha < 2$, however, the standard deviation does not exist.

the case of Pareto distributions this reduces to[10]

$$\theta = \frac{(\alpha - 1)^{\alpha-1}}{\alpha^\alpha}. \tag{4.22}$$

which varies *inversely* with α. Later, Bresciani-Turroni [38, 39] suggested two more indices: one, as an indicator of *equality*, the proportion of the population earning more than the mean income, which for Pareto distributions becomes

$$\frac{N(\mu)}{N(m)} = \Big(\frac{\alpha - 1}{\alpha}\Big)^{\alpha}; \tag{4.23}$$

and another, as an indicator of *inequality*, the proportion of total income earned by people earning more than the mean income, which for Pareto distributions is

$$\frac{R(\mu)}{R(m)} = \Big(\frac{\alpha - 1}{\alpha}\Big)^{\alpha-1}. \tag{4.24}$$

As Bresciani-Turroni pointed out, the equality index (4.23) varies directly with Pareto's inequality index α, and the inequality index (4.24) varies inversely with α.[11]

Lorenz [134] put forward as a requirement for a measure of the concentration of wealth that it should be invariant with respect to multiplication of all incomes by a scale factor. Defining the relative share, s, of those earning an income less than or equal to x as

$$s = \Phi(x) = \frac{1}{\mu}\int_m^x \xi f(\xi) d\xi = 1 - \Psi(x) \tag{4.25}$$

(this is the "incomplete first moment distribution" — cf. Kendall & Stuart [125, p. 48]), the famous "Lorenz curve", expressing the relative income, s, of those earning x or less, as a function of the proportion, p, of the population earning x or less, is defined by (see Figure 5)

$$s = L(p) = \Phi[F^{-1}(p)] \quad (0 \leqq p, s \leqq 1) \tag{4.26}$$

[10] This is easily seen by computing $\theta_+ = \mu^{-1}\int_\mu^\infty (x-\mu)f(x)dx$ and $\theta_- = \mu^{-1}\int_m^\mu (\mu - x)f(x)dx$, both of which are equal to (4.22). The coefficient θ_+ was used by D'Addario [56, p. 184] as a measure of inequality.

[11] Proof: Integrate the inequality

$$\frac{1}{\alpha} < \frac{1}{t} < \frac{1}{\alpha - 1}$$

over the interval $\alpha - 1 \leqq t \leqq \alpha$, to obtain the inequality

$$\frac{1}{\alpha} < \log\alpha - \log(\alpha - 1) < \frac{1}{\alpha - 1}$$

Then differentiation of the logarithms of (4.23) and (4.24) with respect to α gives the results asserted.

where p and F are defined by (4.9). L is monotone increasing since both F and Φ are, hence since

$$q = 1 - p,\ r = 1 - s \quad (0 \leqq q,\ r \leqq 1) \tag{4.27}$$

the same Lorenz curve may be described by the function

$$q = L^*(r) \equiv 1 - L^{-1}(1-r) = G[\Phi^{-1}(1-r)] = G[\Psi^{-1}(r)]. \tag{4.28}$$

According to Lorenz, as between two distributions F_1 and F_2, F_1 displays greater concentration of wealth (i.e., greater inequality) than F_2 if $L_1(p) < L_2(p)$ for $0 < p < 1$, the greatest degree of equality occurring when $L(p) = p$ for all p, and the greatest degree of inequality when $L(p) = 0$ for $0 \leqq p < 1$ and $L(1) = 1$.

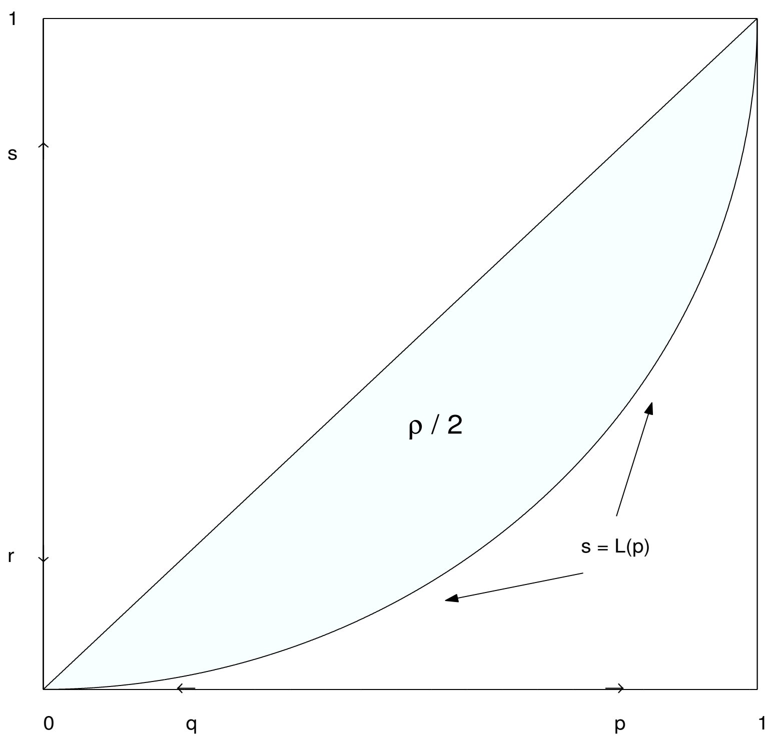

Figure 5

As is the case with Pareto's measure $G(x)$, Lorenz's measure provides only a partial ordering of distributions in general. However, as applied to the Pareto distribution we see immediately from (4.12), (4.14), and (4.28) that

$$q = L^*(r) = r^{\delta} \quad \text{where } \delta = \frac{\alpha}{\alpha - 1}. \tag{4.29}$$

As between two Pareto distributions, therefore, the one with the higher α, hence the lower δ, has a Lorenz curve uniformly *closer* to the diagonal. Thus, for

Lorenz, a higher α indicates less rather than greater inequality of incomes. This was pointed out by Dalton [61], who used the fact to argue that α should be interpreted as a measure of relative *equality* rather than inequality when the Pareto distribution holds.

Gini [105, p. 72], [106, p. 19] introduced his "index of concentration and dependence" (not to be confused with his concentration *ratio* — see below) defined by

$$\delta(x) = \frac{\log G(x)}{\log \Psi(x)} = \frac{\log[N(x)/N(m)]}{\log[R(x)/R(m)]}. \tag{4.30}$$

This is simply the variable exponent satisfying (4.29) in place of δ; in the case of Pareto distributions it is clearly constant and equal to

$$\delta(x) = \frac{\alpha}{\alpha - 1} = \delta. \tag{4.31}$$

In general (4.30) furnishes only a partial ordering, but Gini's *concentration ratio* [107] (also called the *Gini coefficient*)[12]

$$\rho = \frac{1}{2\mu} \int_m^\infty \int_m^\infty |x - y| f(x) f(y) dx dy \tag{4.32}$$

has the property indicated by Gini [108] (for a proof, see Kendall & Stuart [125, p. 49]) that it is equal to the ratio of the area between the Lorenz curve and the diagonal, to the area below the diagonal, i.e. (see Figure 5),

$$\rho = 1 - 2\int_0^1 L(p)dp = 1 - 2\int_0^1 L^*(r)dr. \tag{4.33}$$

In the case of Pareto distributions this is computed at once to be

$$\rho = \frac{\delta - 1}{\delta + 1} = \frac{1}{2\alpha - 1}, \tag{4.34}$$

as is well known (cf. Dalton [61, p. 360], Aitchison & Brown [2, p. 101]). Thus, like the Bresciani ratio θ, the Gini ratio ρ — which is an index of *inequality* — is inversely related to Pareto's index of *inequality* α.[13]

Still another inequality index has been suggested by Frechet [94, p. 106], namely the ratio of the mean to the median, which in the case of Pareto distributions becomes (see footnote 5 above)

$$\frac{\mu}{2^{1/\alpha} m} = \frac{\alpha}{\alpha - 1} 2^{-1/\alpha}. \tag{4.35}$$

[12] The numerator of (4.32) is Gini's *mean difference* [107]. It can actually be traced back as far as F. R. Helmert in 1876. For a history and interesting discussion of the concept see David [62].

[13] The relation between Gini's concentration ratio and Bresciani-Turroni's relative mean deviation was the subject of intensive study on the part of Pietra [199, 200], and was also analyzed in detail by Bortkiewicz [30].

Again, this varies inversely with Pareto's α. Allais [4, II, p. 36] has proposed as an inequality index the ratio, to the median, of the average of incomes above the median; in the case of Pareto distributions this coincides with the Gini index $\delta = \alpha/(\alpha - 1)$. Thus it happens that every single alternative index that has been proposed gives precisely the reverse of Pareto's definition of inequality!

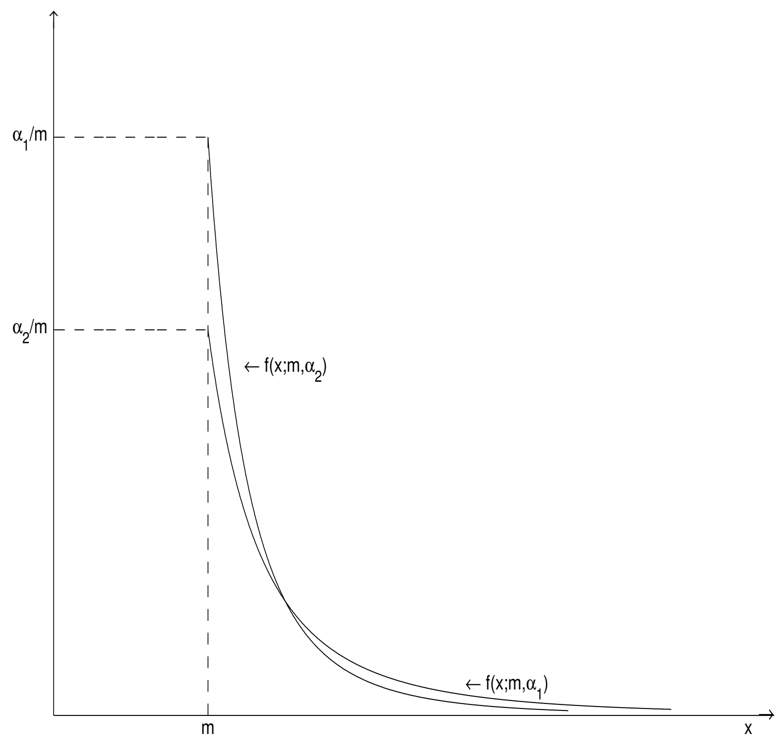

Figure 6

The paradox is easily explained. Since these indices are designed to measure *relative* inequality in income distribution, it is perhaps not surprising that in the case of the Pareto family of distributions they should be independent of the minimum income m. But then, within that family, for given m, what does it mean for the distribution of income to be "equal" in this sense? It can only mean that all incomes are concentrated towards the minimum income. The density function corresponding to higher α is therefore both higher and steeper at $x = m$ (since $f(m; m, \alpha) = \alpha/m$ and $\partial f(m; m, \alpha)/\partial\alpha = -(2\alpha+1)$; see Figure 6). It is a case, therefore, of incomes becoming more equal by reason of the rich becoming poorer rather than the poor becoming richer — a case which Pareto had explicitly set aside as not corresponding to what he considered to be the usual meaning of "equality of incomes."

How should the matter be settled? Pigou [201, p. 25n] provided the criterion

$$W(m,\alpha) = \int_m^\infty U(x) f(x; m, \alpha) dx \tag{4.36}$$

where $U(x)$ is a utility function, expressed as a function of individual income and assumed identical for all individuals, and assumed to be increasing and strictly concave.[14] Pigou did not apply this welfare criterion to the Pareto distribution as such, however, and instead argued that by expanding $U(x)$ in a Taylor's series around mean income one could express social welfare by

$$\int_m^\infty U(x) f(x) dx = U(\mu) + \frac{1}{2!} U''(\mu) \int_m^\infty (x-\mu)^2 f(x) dx + \ldots; \tag{4.37}$$

then, neglecting terms beyond that of the second order, it followed from concavity that for income distributions with the same mean the distribution with smaller variance would be socially preferred. He used this to justify the reciprocal of Pareto's α as an inequality criterion. It was correctly pointed out by Dalton [61], however, that higher-order terms could not in general be neglected. Indeed, in the case of the Pareto distribution, with $1 < \alpha < 2$, no moments higher than the first exist, so the right side of (4.37) is not convergent.

Substituting (4.8) in (4.36) it can be shown by a direct computation that[15]

$$\begin{aligned} \frac{\partial W(m,\alpha)}{\partial m} &= \alpha m^{\alpha-1} \int_m^\infty U'(x) x^{-\alpha} dx > 0; \\ \frac{\partial W(m,\alpha)}{\partial \alpha} &= -m^\alpha \int_m^\infty U'(x) x^{-\alpha} \log(x/m) dx < 0. \end{aligned} \tag{4.38}$$

That is, welfare can be increased either by an increase in the minimum income m, or by a decrease in Pareto's measure of inequality α, or both. In terms of the criterion (4.36), Pareto's intuition is completely vindicated.

Nevertheless, this is not the whole story. From (4.15) we have, for each value of the Pareto coefficient greater than 1, a one-to-one relationship between minimum income m and mean income μ. The Pareto distribution can, therefore,

[14] Such a criterion has, of course, the theoretically objectionable feature of employing income rather than a vector of physical amounts of commodities consumed as argument of the utility function, thus partaking of the "shibboleth" quality described by Samuelson [208, 211]. It has the further objectionable feature of indexing (identifying) individuals by their income levels only, so that individuals with the same income are required to have the same tastes — a feature that is shared by Pareto's own aggregation procedure for demand functions (4.7). To surmount this latter difficulty one could parametrize the differences in tastes by, say, a single parameter, and replace (4.36) by the corresponding double integral. This is the procedure followed by Farrell [83] in aggregating demand functions. Alternatively one could follow the much deeper approach of Pessoa de Araujo [198] and Sondermann [225] of introducing a measure of the distance between peoples' preferences.

[15] The computation makes use of the fact that $0 \leqq U'(x) \leqq U'(m)$ from monotonicity and concavity, and hence

$$\lim_{x\to\infty} U(x) x^{-\alpha} = \lim_{x\to\infty} U(x) x^{-\alpha} \log x = 0.$$

just as well be characterized by the two parameters μ, α as by the two parameters m, α. In comparing two distributions *with the same mean*, would the one having the higher Pareto coefficient α (hence the lower Gini index δ, and lower Gini ratio ρ) now be preferable? This was the contention of Gini [106, pp. 49–50], [107, pp. 72, 79] and Dalton [61], and we shall see that it is indeed correct.

Writing the density function (4.8) as a function of μ and α we have

$$f^*(x;\mu,\alpha) = \alpha^{1-\alpha}(\alpha-1)^{\alpha}\mu^{\alpha}x^{-\alpha-1}, \tag{4.39}$$

and the corresponding welfare function may be written

$$W^*(\mu,\alpha) = \alpha^{1-\alpha}(\alpha-1)^{\alpha}\mu^{\alpha}\int_{\frac{\alpha-1}{\alpha}\mu}^{\infty} U(x)x^{-\alpha-1}dx. \tag{4.40}$$

It is found, as to be expected, that welfare is an increasing function of μ:

$$\frac{\partial W^*(\mu,\alpha)}{\partial\mu} = \frac{\alpha}{\mu}\int_{\frac{\alpha-1}{\alpha}\mu}^{\infty} U'(x)x^{-\alpha}dx > 0. \tag{4.41}$$

After lengthy computations it is found also that (cf. Chipman [49])

$$\frac{\partial W^*(\mu,\alpha)}{\partial\alpha} = -\frac{1}{\alpha-1}\int_{\frac{\alpha-1}{\alpha}\mu}^{\infty} U''(x)x^{1-\alpha}\Big[\log\Big(\frac{x}{\mu}\Big) - \log\Big(\frac{\alpha-1}{\alpha}\Big)\Big]dx > 0. \tag{4.42}$$

The above conclusion might seem to be at variance with a result of Samuelson's [212], but it is actually not. Samuelson dealt with the four-parameter family of stable distributions as defined by Levy [132], with location parameter μ, scale parameter ν, skewness parameter η, and kurtosis parameter α, the right tails of which are well approximated by Pareto's distribution with parameter α $(1 < \alpha < 2)$ and mean μ (cf. Mandelbrot [140, p. 87], Feller [86, p. 576].[16] What Samuelson showed was that, in terms of the criterion (4.36) applied to the Lévy densities $f(x)$ with $1 < \alpha < 2$, welfare is a decreasing function of ν for fixed μ, η, and α. This still leaves open the question of the sign of $\partial W/\partial\alpha$ in the general case. Samuelson conjectured [212, p. 250n] that it could go either way, which would no doubt be correct if the density were symmetric (i.e., $\eta = 0$). However, if it is sufficiently skewed to the right (i.e., if η is positive and large), the approximation by the Pareto distribution may be expected to be good and the general presumption (4.42) may be expected to hold.

The question of whether Pareto's α coefficient should be considered as a measure of "equality" or "inequality" is thus resolved, in the case of Pareto distributions, if what is desired is a dispersion (scale) parameter such that social welfare is inversely related with this parameter (if it is described as an "inequality" parameter) for given values of a location parameter. If the location parameter is minimum income m, then Pareto's α as a measure of "inequality" serves this purpose. However, if the location parameter is mean income μ, then it is $1/\alpha$ that would serve this purpose, or any other parameter that is inversely

[16] Specifically, $G(x) = 1 - F(x) = O(x^{-\alpha})$ as $x \to \infty$, hence $f(x) = F'(x) = O(x^{\alpha-1})$.

related to α, such as the Gini index $\delta = \alpha/(\alpha-1)$, the Gini ratio $\rho = 1/(2\alpha-1)$, or the relative mean deviation θ of (4.22). If we compare two Pareto distributions with the same minimum income m, the one with the higher α will have a Lorenz curve uniformly closer to the diagonal, yet it will necessarily have lower mean income μ and give rise to lower welfare.[17]

Dalton, who took issue with Pareto's conclusions, furnished the criterion that [61, p. 349] "if a given income is to be distributed among a given number of persons, it is evident that economic welfare will be a maximum, when all incomes are equal. It follows that the inequality of any given distribution may conveniently be defined as the ratio of the total economic welfare attainable under an equal distribution to the total economic welfare attainable under the given distribution." Thus Dalton would characterize the Pareto distribution in terms of the parameters μ and ρ. Dalton therefore concluded that (p. 359) "the Pareto measure should be inverted, so that, the greater α, the smaller inequality. But such an inversion will explode Professor Pareto's alleged economic harmonies, and it will follow, according to his law, that increased production per head will always mean increased inequality!" In terms of Dalton's terminology this conclusion is certainly correct, as long as the minimum income rises less rapidly than income per head (since $1/\alpha = 1-m/\mu$); but in these circumstances it is also true that Dalton's "increased inequality" gives rise to increased welfare! The dispute is therefore one about words.

But words are important, as Pareto recognized very well, and it may be admitted that Pareto's choice of words in this case was unfortunate, particularly when occurring in a passage such as the following one [163, II, §965, p. 325] "It is therefore not true that under existing circumstances the inequality of wealth is on the rise, and all the deductions that have been drawn from this erroneous proposition come to nothing." Sorel [226, p. 587] objected that in the absence of a neutral term such a statement "opens the doors to the sophistries of the apologists of capitalism." Pareto, in reply [169], conceded that Sorel was right and that "inequality" had not been a happy expression for describing the α coefficient: "It is better to avoid this ambiguous expression. If I have used it, it was simply to avoid further idle discussions such as those to which the new word *ophelimity* gave rise." His use of an emotionally charged word turns out to have been a greater mistake, however, then it would have been to coin a new word. A neutral expression for α that could not give rise to any ambiguity or disagreement, would surely be the "Pareto coefficient."

The controversies just recounted all had to do with the question of whether Pareto's index was a satisfactory *ordinal* indicator of inequality in the distribution of income. However, Pareto's well-known assertion that α differed little

---

[17] Before introducing his definition of "inequality," Pareto [163, II, §964, p. 318, Fig. 52] had actually used the term in the Gini sense of high dispersion, and had discussed two hypothetical income distributions, one with a low minimum income and low dispersion, and the other with a higher minimum income and high dispersion. He remarked with respect to the former (using "inequality" in the sense of "high dispersion") that it displayed "very little inequality of income but a very intense degree of pauperism." He added, however, that "we do not find examples of this in the real world."

among widely separated countries and epochs gave rise to the additional question of whether, even in its inverted form, it could be regarded as a suitable *cardinal* indicator. The first to raise this issue was Sorel [226, p. 595] who objected that "individual appraisals of the degree of inequality of wealth vary very rapidly when α undergoes only very slight changes," and it was the intensity of these individual sentiments concerning inequality which should properly be measured by an inequality index. He went on to propose [226, p. 601] as a measure of equality the degree to which the bottom incomes kept pace with average incomes, i.e., the ratio m/μ, which equals $(\alpha - 1)/\alpha$ in the case of Pareto distributions. The corresponding inequality index is therefore precisely the "concentration index" δ introduced some years later by Gini [105, 106] , without reference to Sorel. Gini himself [106, p. 49] noted that δ was more sensitive to differences in income distribution than α, and used this observation to strike at Pareto's dictum that income distribution showed great stability among different countries and epochs, and to demonstrate the superiority of his own index δ. This led to a good deal of fruitless controversy which D'Addario characterized [53, p. 280] as "a monotonous contest between α and δ." It follows of course from the very nature of the transformation $1/\alpha + 1/\delta = 1$ that $|d\delta/d\alpha| > 1$ for $1 < \alpha < 2$, but the choice of the best cardinal index must be decided on the basis of the considerations Sorel was interested in rather than on the kinds of mathematical considerations appealed to by Gini. For example, one might wish to appeal directly to a welfare function such as (4.36) to decide on the most appropriate index. In his reply to Sorel, Pareto [169] said that he did not disagree with Sorel's position in any respect, but was simply not interested in studying the distribution of utilities — only the distribution of incomes. This was a characteristic bit of obstinacy on his part, for such an attitude is hard to reconcile with the social significance which he attached to the stability of income distribution over time and among different social systems.

Pareto's treatment also came under a different type of criticism on the part of both Dalton and Lorenz. Lorenz objected [134, p. 216]: "imagine a community in which the wealth is nearly equally distributed, and then assume that the richest individual becomes a multi-millionaire, with no change in the wealth of the remainder. Professor Pareto's curve would tell us nothing about this change." What Lorenz should have said is that Pareto's curve would fit very poorly in the second of these situations; and what Pareto would surely have retorted is that the types of hypothetical distributions Lorenz described are never observed.[18] Dalton also objected [61, p. 360] that "when distribution may depart widely from Pareto's law, the measure has, of course, no significance at all." But as general considerations show (cf. Chipman [49]), and as Newbery [148] has proved in a particular case (that of the Gini coefficient), this is a statement that can be made with respect to all inequality measures that have been proposed, including the relative mean deviation favored by Dalton himself. Unless one is willing to impose severe restrictions on the types of income distribution

[18] Pareto also took pains to repeat many times (e.g. [163, II, §965, p. 321]), that his analysis was designed to deal with general and average phenomena and not with accidental ones of which he gave an example very similar to that of Lorenz.

under comparison, or alternatively on the form of the utility function U, one cannot hope to find a system for ranking income distributions in terms of the values of a small number of parameters.[19]

4.5 The controversy with Edgeworth and its aftermath

Pareto's work on income distribution came to Edgeworth's attention after the latter had already published a paper [70] discussing Pearson's "generalized probability curve" [196, p. 357]

$$y = y_0(1 + x/\alpha)^{\gamma\alpha} e^{-\gamma x} \tag{4.43}$$

which Pearson had used as a density function to describe skewed data such as the distribution of house rents. Pareto had already introduced [162, p. 379] a generalized income curve — his "third approximation"

$$N(x) = \frac{A}{(x+a)^\alpha} e^{-\beta x} \quad (0 < m \leqq x < \infty) \tag{4.44}$$

which specializes to his "second approximation" when $\beta = 0$ and to his "first approximation" (4.10) when $a = \beta = 0$. Edgeworth [71, p. 534] remarked upon the similarity between (4.44), and (4.43) but added that "the identity is only apparent, since N denotes in Professor Pareto's formula all the incomes *at or above* a certain x, while the corresponding symbol [y] in Professor Pearson's formula denotes only the valuations *at* a certain x," i.e., (4.44) is a right-cumulative distribution function (after division by $N(m)$) rather than a density function. Edgeworth added two criticisms: (i) that, as he had remarked in [70] in criticism of Pearson's formula: "It appears to me to confirm the opinion which I expressed last September, that a close fit of a curve to given statistics is not, *per se* and apart from *a priori* reasons, a proof that the curve in question is *the* form proper to the matter in hand;" and (ii) that in the special case $a = \beta = 0$, "according to the formula given above there ought to be an infinite number of *null* incomes, and an indefinitely large number of incomes in the neighborhood of zero." He added [71, p. 538] that "there is not the same objection to an infinite *ordinate* that there is to the infinite *integral* involved in Professor Pareto's first approximation," i.e., $N(x)$ in (4.10) must be finite at its lower limit if N is to define a distribution function $F(x) = 1 - N(x)/N(m)$, but this does not require $dN(x)/dx$ to be finite at its lower limit. Finally, Edgeworth summed

[19] The proposition is formally equivalent to the corresponding proposition concerning expected utility functions in portfolio analysis. What Newbery showed was that if the welfare function was assumed to be increasing and strictly quasi-concave, it cannot be expressed as a function of mean and Gini ratio if the income distributions are unrestricted as to type. The result no longer follows if the hypothesis is relaxed to allow for either "Rawlsian" (L-shaped) or Benthamic (straight-line) welfare contours; see Sheshinski [222], Kats [123], Rothschild & Stiglitz [204]. For additional references see Chipman [49], to which may be added Aigner & Reins [1], Sen [221].

up the situation by saying that since both Pearson's and Pareto's curves appeared to have a good fit, and since neither had a strong a priori theoretical basis (though he allowed that Pearson's had some), the deciding criterion for choosing between them was that of authority: "Where opinions on a matter of this sort differ, the presumption is certainly in favor of the author who has made the greatest advance in the science of Probabilities since the era of Poisson," namely Karl Pearson.

It is hardly surprising that these remarks should have given rise to what is perhaps the most ill-tempered and acrimonious dispute in the history of economic thought.

Pareto [165], seeing in them a thinly veiled accusation of plagiarism, reacted with biting sarcasm: "It must have displeased Mr. Edgeworth to see me poach on territory which is apparently reserved for Professor Pearson, just as political economy is reserved for Professor Marshall ..." His answer to (i) was that nobody denied that a rational law carried more weight than an empirical law, but that the latter had to precede the former rather than vice versa. Comparing himself by implication to Kepler, he suggested that some day a Newton would come along and provide a rational basis for his empirical law; he went on to suggest that a theory of "social heterogeneity" such as that proposed by Ammon [10] and others would provide such a basis; and in fact the final chapter of the *Cours* [163, Vol. II, Book III, Ch. II, pp. 347–369], as well as much of the work of his later years developing a theory of the "circulation of the élites" [176, 179, 182], was devoted precisely to the search for such a "rational law." As for (ii), Pareto pointed out that his income curve $N(x)$ was defined only for $x \geqq m > 0$, and thus there was no question of it diverging at its lower limit; he reminded Edgeworth also that he was concerned not with partial incomes such as house rents, nor with the source of a person's income, but with an individual's *total income*, which he had defined as his actual consumption — "whether it has been inherited, earned, stolen, or given to him by charity" [165, p. 439]. Pareto also feigned — so it seems — misunderstanding of what Edgeworth was driving at in saying that it is legitimate for a density function to have an infinite ordinate but not an infinite integral, by retorting that any integral is itself the derivative of its primitive integral. Had Pareto defined his function $N(x)$ for all $x > 0$ rather than for $x \geqq m > 0$, Edgeworth's objection would of course have been valid. But with characteristic sarcasm Pareto pointed out that "those who have a total income equal to zero will be angels" and formula (4.10) "does not hold for angels" [165, p. 444].

Edgeworth devoted the first part of his rejoinder [73] to explaining himself on his point concerning infinite ordinates and integrals — a point which Pareto obviously understood perfectly well — rather than to meeting Pareto's point regarding the domain of definition of the function $N(x)$. He remarked also that in his opinion it would be better for the food of the poor "not to be calculated as part of *their* income ... but as deducted from the income attributed to the rich."[20] Finally, he returned to the general question which was his chief concern,

[20] We learn also from Pareto's letters to Edgeworth (there is no trace of Edgeworth's letters

as to whether there could be found a probabilistic basis for the law of income distribution, and in particular whether this law could be deduced from the normal (Gaussian) law of error. In his "Final Reply" [168], Pareto reiterated with heightened irritation that his income curve was defined only for "total incomes" as he had defined them, and over their observable range; and that he did not understand why he should be reproached for studying this limited problem rather than engaging in loftier speculations.

The dispute actually continued even after Pareto's death in 1923 and only shortly before Edgeworth's death in 1926. Once again [171, p. 376] Pareto repeated that his formula was valid only for incomes approaching subsistence level, and that despite his having repeated this many times, Edgeworth "insisted on judging [it] by the results it gives for zero incomes." Edgeworth quickly retorted [75, p. 675] that "Professor Pareto, indeed, the inventor of a beautiful and useful representation of the frequency of incomes of different sizes, seems averse even to entertain the idea of a generalized probability curve," and: "My incidental allusion to the eminent statistician's income curve ... could hardly have provoked his implacable retorts ... if he had realized that my subject was 'General laws which govern the grouping of members of species' ... and my thesis, 'that a close fit of a curve to given statistics is not, *per se* and apart from *a priori* reasons, a proof that the curve in question is *the* form proper to the matter in hand.'" In fact, Pareto [165, p. 442] had quoted this passage and accepted the point in the sense that his was an empirical rather than a rational law. And in the addendum ("Additions") added in proof at the end of the *Cours* [163, II, pp. 416–19], he made a fascinating contribution — which has been analyzed by Vinci [237] — to the subject of inferring the law of distribution of human aptitudes "from that of the distribution of incomes; according to D'Addario [60, p. 100], whose belief seems well founded, this was a direct outgrowth of Pareto's controversy with Edgeworth.

Petty though the controversy appears to have been on the surface, it helped give birth to Edgeworth's very original "method of translation" [75, 76], whereby the density function $f(x)$ of incomes is determined through a one-to-one function $x = T(u)$, or "translation," by the density function $g(u)$ of aptitudes u. Accordingly,

$$f(x) = \frac{1}{|T'(T^{-1}(x))|} g(T^{-1}(x)). \tag{4.45}$$

Edgeworth devised a method for empirically estimating the form of $T(u)$ given

to Pareto) that Edgeworth considered it to be "the better opinion" that income should be defined as earnings rather than consumption (cf. [195, III, p. 406]). Such remarks, and the generally haughty tone of Edgeworth's comments combined with his servile attitude towards Pearson (and Marshall), so enfuriated Pareto as to provoke him to write to Pantaleoni to say: "He's Irish, very likely Catholic and Jesuit. In any case he's worthy of it" [195, I, pp. 474–5], and "I keep repeating to you that I don't want to have anything more to do with Edgeworth, I don't care a rap about that Jesuit" [195, II, p. 40]. In his final shaft Edgeworth commented on Pareto's "somewhat acrimonious explanation ... which is of interest as throwing light not only on the character of the curve, but also on that of its discoverer [80, p. 712], and showed himself still unwilling to accept Pareto's definition of income and his specification of the domain of definition of his distribution function.

the empirical form of $f(x)$ and the hypothesis that $g(u)$ was the normal density function. A special case of Edgeworth's procedure had already been worked out by McAlister [136] for the case in which

$$T(u) = \omega + \kappa e^{\lambda u} \quad (\kappa, \lambda > 0), \tag{4.46}$$

on the basis of Galton's suggestion [96, p. 367] that such a transformation should be used to take account of the circumstance that "a capital employed in a business increases in proportion to its size." The resulting density function, further analyzed by Kapteyn [122] (who independently formulated the general method of translation), is of course that of the now famous lognormal distribution [3][21]

$$f(x) = \frac{1}{\lambda(x-\omega)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\frac{1}{\lambda} \log(x-\omega) - \frac{1}{\lambda} \log \kappa - \theta \right]^2 \right\}, \tag{4.47}$$

defined for $x > \omega$, where θ and σ are the mean and standard deviation of the normal random variable representing aptitudes.

In the course of perfecting his method of translation, Edgeworth returned to the subject of his dispute with Pareto [76, p. 547]:

> There is much to be said for the procedure which Professor Pareto has adopted with brilliant success: finding a simple formula which fits the descending right hand branch of the given group most accurately. On the other hand, it is natural to wish to treat the whole group of incomes, the lower with the higher, as what Professor Pearson calls "homogeneous material." But for *this* purpose it would be better, I think, to define income as what a man wins in *economic* ways, not as Pareto has defined it, quite properly for *his* purpose. ... If we consider the total group of incomes, it clearly does not conform to the [normal] probability curve: Pareto has shown this convincingly. And yet, after all, there may be something in the hypothesis which he combats — that of Dr. Otto Ammon. Income does not vary according to the normal law of error, but it may be *dependent* on some attribute which does so vary, to wit, *ability*, as Dr. Ammon has suggested. Thus the frequency of *incomes* might form a *translated* probability-curve ...."[22]

Why incomes should be defined as earnings rather than disposable income, Edgeworth never explained. Presumably because of a relationship that could be assumed to hold between ability and marginal productivity. However, it is

[21] Priority for discovery of the lognormal distribution is still often attributed to Gibrat [102, 103] — in fact, it is still sometimes called the "Gibrat distribution" — despite the fact, as was brought out in detail in D'Addario's controversy with Gibrat [54, 55], [104], that Gibrat not only added nothing to the theory that had not already been fully developed by Dalton, McAlister, Edgeworth, and Kapteyn, but in fact committed a serious blunder — the omission of the term $1/[\lambda(x-\omega)]$ in (4.47) — which largely vitiated his empirical results.

[22] Edgeworth considered the distribution of divorces at different dates after marriage, and found that $T(u)$ was well approximated by a cubic.

doubtful whether Pareto would have accepted such an explanation, since he was too good a Walrasian not to remember that the value of a factor's marginal product is the outcome of a general equilibrium process, as is the determination of the ownership of assets; he also took a broader view of social phenomena than Edgeworth, and would not have been satisfied with a theory of income distribution that could not account for the acquisition of income by transfer as well as by earnings. Specifically, he would not under the heading of "ability" have excluded ability to appropriate the earnings of others by means of sharp practices, strikes, intimidation, or the use of government as an instrument of spoliation [163, II, §1026, p. 371], [182, Chs. II, VII].

After temporarily abandoning his and Kapteyn's approach in favor of Pearson's in the wake of a dispute between Kapteyn and Pearson,[23] Edgeworth returned to it later and suggested that "Pareto's celebrated income curve is perhaps to be explained on this principle" [79, p. 573] ; but no details were provided. In the meantime, however, a way had been shown by Cantelli [42], working independently of Edgeworth.[24] Cantelli showed that with $T(u)$ given by (4.46) (he actually worked with the inverse function $T^{-1}(x) = c\log(x+a)+b$, where $a = -\omega$, $c = 1/\lambda$, and $b = -(\log\kappa)/\lambda$), if the generating density is of the negative exponential type

$$g(u) = \rho e^{-\rho u} \quad (\rho > \lambda, 0 \leqq u < \infty), \tag{4.48}$$

then the resulting income density is precisely the Pareto density (second approximation) given by

$$f(x) = \alpha(m+a)^{\alpha}(x+a)^{-1-\alpha} \quad (0 < m \leqq x < \infty), \tag{4.49}$$

where $a = -\omega$, $m = \kappa + \omega$, and $\alpha = \rho/\lambda$. Cantelli's principal contribution, however, was to provide a theoretical justification for (4.48). Because of its importance it is worth while presenting Cantelli's argument here, as this can be done very briefly.

Let the income scale be subdivided into a large number r of income brackets, and let x_i be a representative income in the ith bracket. Let it be assumed that each person that falls in the ith bracket receives precisely the income x_i. The problem may then be formulated as a classical occupancy problem (cf. Feller [85, pp. 38–9]); let n_i be the number of individuals falling in the ith income bracket, where

$$\sum_{i=1}^{r} n_i = n. \tag{4.50}$$

[23] For a discussion of this dispute see Aitchison & Brown [3, pp. 20-22]. Pearson's main objection was that u could not in general be directly measured. His own alternative solution was the "method of separation," i.e., identification of subpopulations each following the normal law, which Edgeworth [78, p. 402n] accepted in 1911 in preference to the method of translation "suggested by the present writer [75], and independently by Professor J. C. Kapteyn [122]."

[24] In his last work Edgeworth [80, p. 713] cited Vinci's article [237], which analyzes Pareto's "Additions" and briefly mentions Cantelli's work, but he was apparently not familiar either with Cantelli's paper [42] or with Vinci's exposition and extension of it [236].

Assuming that all r^n placements are equally probable, the probability of obtaining the given "occupancy numbers" $n_1, n_2, \ldots, n_r$ is

$$\frac{n!}{n_1!n_2!\ldots n_r!}r^{-n}. \tag{4.51}$$

Cantelli now introduces an additional constraint, namely

$$\sum_{i=1}^{r} n_i U(x_i) = w, \tag{4.52}$$

where w is a certain parameter and U a certain function (namely T^{-1}). Finally, Cantelli postulates that the observed distribution of occupancy numbers is the most probable one given the above constraints, i.e., that it is such as to maximize (4.51) subject to (4.50) and (4.52) for any given w. For large n_is the solution to this constrained maximization problem is approximately[25]

$$n_i = ke^{-\rho U(x_i)}. \tag{4.53}$$

Using a limiting argument, as r increases (4.53) yields the density function (4.48). Thus, under the scheme (4.51), (4.48) is the most probable density function g satisfying $\int ug(u)du =$ constant.

Cantelli interpreted U as a utility function, appealing to Bernoulli [24] for justification of its logarithmic form. The constraint (4.52) then consists in holding the level of welfare constant when the social welfare function is of Benthamic type. Vinci [236] objected to this interpretation, and substituted the interpretation that $U(x)$ consisted of the obstacles encountered by the recipient of an income x.[26] To D'Addario [57, p. 55], neither interpretation was satisfactory, and he substituted the interpretation that $U(x)$ was the capacity to acquire an

[25] The proof goes as follows (cf. Castelnuovo [46], pp. 276–7 of the first edition, or Vol. II, pp. 107–9 of the second edition). For large n_i, the terms may be approximated by Stirling's formula (see, e.g., Feller [85, p. 52]), whence, taking account of the constraint (4.50) we have

$$\log \prod_{i=1}^{r} n_i! \sim r \log \sqrt{2\pi} - n + \sum_{i=1}^{r} (n + \tfrac{1}{2}) \log n_i,$$

where the sign $\sim$ indicates that the difference between the two sides approaches zero as the n_i approach infinity. The third term on the right may then be minimized subject to (4.50) and (4.52). This leads to

$$\log n_i = -1 - \nu - \rho U(x_i) - 1/n_i$$

where ν and ρ are Lagrangean multipliers. Neglecting the last term which becomes negligible for large n_i, integration of the resulting expression leads to the formula (4.53) of the text.

Castelnuovo supplements this argument, similar to that of maximum likelihood estimation, by one to the effect that there is a small confidence interval, i.e., a high probability that the true density will be close to (4.53); cf. [46], pp. 285–9 of the first edition or Vol. II, pp. 119–124 of the second edition. A modern treatment of this theory is contained in Khinchin [126].

[26] According to Vinci's interpretation one might think of the constraint (4.52) as defining a linear utility possibility frontier (in utility space) in Samuelson's sense [209], as opposed to a linear Benthamic welfare contour which would correspond to Cantelli's interpretation. In the space of incomes, the constraint (4.52) defines a hyper-surface that is convex to the origin; this would suggest that Cantelli's interpretation is more plausible than Vinci's.

income of x. This of course identifies $U(x)$ with $T^{-1}(x)$. D'Addario's interpretation certainly seems to be the most solidly based. D'Addario later showed [58] that if to the constancy of the mean of the aptitude variable was added the constancy of its variance, then Cantelli's approach would yield the normal density for g, hence the lognormal density for f — always on the assumption that T is of exponential form.[27] D'Addario [60] went on to provide a general synthesis of Edgeworth's method of translation, finding generator functions g and translations T corresponding to the income-distribution functions suggested by March [144] , Vinci [236], Amoroso [11], and Davis [63, pp.408-418].[28]

The soundness of the exponential specification (4.46) is well based on Kapteyn's amplification of Galton's remark quoted above: "let the price of an article in which both A and B have invested their capital rise or fall. Then it will be evident that ... the effect of this cause will not be independent of the capital, but proportional to it" [122, p. 13], and again, "it seems plausible enough to admit that most of the principal causes of deviation in wealth, that is, the main causes of gain and loss of capital, have an effect roughly proportional to the capital possessed" [122, p. 43]. This is what Gibrat [103] later described as the "law of proportionate effect."

Basing himself on Edgeworth's method of translation, Fréchet [93] accepted the law of proportionate effect, i.e., the exponential translation (4.46), but replaced the normal density $g(u) = \pi^{-1/2}e^{-u^2}$ by the Laplace density $g(u) = \frac{1}{2}e^{-|u|}, -\infty < u < \infty$. He then showed that the resulting income density $f(x)$ was continuously differentiable and unimodal, and coincided with the Pareto density (second approximation) for all values of x higher than the median income. In support of the Laplace density, however, he offered only empirical evidence obtained by Wilson [249] and himself [92], combined with skepticism concerning the applicability of the normal law to these situations. On the other hand he provided an a priori argument to the effect that $f(x)$ would have to be unimodal, since $N(x)$ would necessarily be constant for incomes below subsistence level, hence dN/dx would tend to zero with small as well as with large incomes and thus $|dN/dx|$ would have to have a maximum. This argument

[27] An alternative rationalization of the lognormal distribution was proposed by Van der Wijk [234], who, like Cantelli, interpreted T^{-1} as a utility function. He appealed to the analogy with Fechner's law (which had been discussed by both Galton and Kapteyn) and to Bernoulli's "moral worth" for justification of the logarithmic form of U and for the interpretation of $U(x)$ as "psychic income." Like Sorel [226], Van der Wijk was interested in the distribution of utility and psychic income, and he postulated that this distribution should follow the normal law. However, no indication was provided as to what social mechanisms might bring this about; in particular he does not appear to have attempted, as might seem a reasonable thing to do, to establish a link between psychic income and aptitudes.

[28] D'Addario [59] also studied a probabilistic model leading to a new income distribution formula of his own. Cantelli [43] applied his model to the distribution of labor income interpreting (4.46) as a formula for the remuneration of labor, one part being independent of the amount of work done and the other part varying exponentially with it, in harmony with its interpretation by Edgeworth and D'Addario. Castellani [45] extended Cantelli's method so as to find the appropriate constraints leading to Pearson's curves. Finally, a study by Rhodes [203] may be mentioned which investigates relationships between distributions of incomes and of talents.

assumes that there are incomes close to subsistence.

Formally, Cantelli's approach is appealing because it provides an a priori basis for exactly Pareto's law. But Pareto's law does not fit the data exactly at the lower extremities. In what is perhaps an oblique criticism of Cantelli's approach, Fréchet remarked in a subsequent paper [95, p. 16]: "a theory has a better chance of being soundly based if it does not assign itself a result chosen in advance, and contents itself with expressing the diverse conditions of the problem as well as possible. We shall reject it if the approximation obtained is bad. In the contrary case, it will be natural to accord it greater confidence than a theory which has been provided with blinkers to lead it to a determined formula." Cantelli's constraint (4.52) was inspired by the analogous constraint in physical statistics, which corresponds to the law of conservation of energy; the fact that three different economic interpretations have been given to it proves, if proof was needed, that there is no equally compelling principle in economics. Moreover, plausible as is the hypothesis that formula (4.51) — which is the Maxwell-Boltzmann distribution in physics — is true, it has had to be rejected in physics in favor of less plausible hypotheses whose implications are closer to the facts. As Feller [85, p. 51] remarks: "We have here an instructive example of the impossibility of selecting or justifying probability models by *a priori* arguments." Which is just what Pareto kept insisting on by saying [163, II, Additions, p. 418]: "It is Kepler's observations which were the basis of Newton's theory, not Newton's theory which proved the facts observed by Kepler."

The search for a rational basis of Pareto's law has continued up to the present day. A notable advance was made by Champernowne [47] who was the first to set up an explicitly dynamic model of social mobility leading to Pareto's law. Hints of such a model had already appeared in Sorel's study [226] (to be discussed in §4.6 below), which contained both the concept of a transition probability of moving from one income bracket to an adjacent one, and the idea of subdividing the income brackets in geometric progression. An outline of a possible model along these lines was also sketched by Winkler [254, p. 386], who described it as a "sifting process" (*Siebungsprozess*). However, it was not until Champernowne's study that a precise model of this kind was presented.

Champernowne subdivides the income scale into discrete brackets arranged in geometric progression:

$$x_i = x_0 e^{\lambda_i}; \quad \lambda > 0, \ i = 0, 1, 2, \ldots. \tag{4.54}$$

By itself this assumption is not limiting, but it becomes so in conjunction with Champernowne's assumption that the probability of moving up or down one bracket (other than into the bottom bracket) is independent of the income bracket one is in. I shall consider only a special case of Champernowne's model in which it is impossible to move up or down at any step by more than one bracket. Champernowne also assumes that it is impossible to move below the bottom (zeroth) bracket — possibly an unfortunate assumption, since it rules out the real possibility of starvation as a selective force. Let the probability of moving up one bracket be p_1, that of staying in the same (other than the bottom) bracket p_0, and that of moving down one bracket (other than from the bottom bracket) p_{-1}, where $p_{-1} + p_0 + p_1 = 1$. The probability of moving down from the bottom bracket being zero, the probability of remaining in this (zeroth) bracket is then $p_{-1} + p_0$.

Let $n_i(t)$ be the number (strictly speaking, the expected number) of individuals in the ith bracket at time t; this is then determined by the equations

$$\begin{aligned} n_i(t+1) &= p_1 n_{i-1}(t) + p_0 n_i(t) + p_{-1} n_{i+1}(t) \quad (i \geqq 1) \\ n_0(t+1) &= (p_{-1} + p_0) n_0(t) + p_{-1} n_1(t), \end{aligned} \tag{4.55}$$

the last equation following from the rest by reason of the constraint

$$\sum_{i=0}^{\infty} n_i(t) = n = N(x_0). \tag{4.56}$$

An equilibrium solution to the first set of equations of (4.55) exists if and only if a solution exists to

$$p_{-1} n_{i+1} - (p_{-1} + p_1) n_i + p_1 n_{i-1} = 0 \quad (i \geqq 1), \tag{4.57}$$

itself a homogeneous linear difference equation in n_i with constant coefficients. The only possible solutions are linear combinations of particular solutions of the form $n_i = \nu^i$, which when substituted in (4.57) yield

$$(\nu - 1)(p_{-1}\nu - p_1)\nu^{k-1} = 0.$$

The solution $\nu = 1$ is ruled out since it would violate the constraint (4.56). The solution $\nu = p_1/p_{-1}$ would likewise have to be ruled out *unless* $p_1 < p_{-1}$, which is Champernowne' s condition for the process to be "non-dissipative", i.e., convergent to an equilibrium solution; in this case it is necessary and sufficient for the existence of an equilibrium solution. That is: *a necessary and sufficient condition for the existence of a limiting distribution is that the probability of moving up one bracket be less than the probability of moving down one bracket.* This seems to be fully in the spirit of Pareto's theory of the "circulation of the élites" [176], [179] (see §4.7 below) to the effect that there is a preponderant degenerative tendency on the part of classes above the subsistence level, and to this extent it portrays the selective nature of the bottom bracket. Assuming $0 < p_1 < p_0$ we then obtain a unique solution

$$n_i = c\left(\frac{p_1}{p_{-1}}\right)^i = n\left(1 - \frac{p_1}{p_{-1}}\right)\left(\frac{p_1}{p_{-1}}\right)^i, \tag{4.58}$$

the evaluation of the constant c in the third term following upon substitution of the second term in the constraint (4.56).

The number of individuals with incomes greater than or equal to x_i is now, from (4.58),

$$N(x_i) = \sum_{j=1}^{\infty} n_j = n\left(\frac{p_1}{p_{-1}}\right)^i = N(x_0)e^{-\rho i} \tag{4.59}$$

where by definition

$$\rho = -\log(p_1/p_{-1}) > 0. \tag{4.60}$$

Defining $\alpha = \rho/\lambda$ and combining (4.59) and (4.54) we immediately obtain the Pareto form

$$N(x_i) = N(x_0)x_0^{\alpha}x_i^{-\alpha} \tag{4.61}$$

in accordance with (4.10).

One additional assumption — not mentioned by Champernowne — needs to be added, however, in order to ensure the finiteness of total income. This is that $\alpha = \rho/\lambda > 1$ or, from (4.60) and (4.54),

$$\frac{p_1}{p_{-1}} = e^{-\rho} < e^{-\lambda} = \frac{x_{i-1}}{x_i} \quad (i = 1, 2, \ldots). \tag{4.62}$$

Thus, the steeper is the gradation of incomes as defined by λ, the greater must be the probability of falling relative to that of rising in the income scale.

Oversimplified as such a model may be, it does provide the kind of explanation for Pareto's law that, in principle at least, could be tested against separate data, say on social mobility; that is, it makes it possible to extend the realm of applicability of the law and confirm it with a broader variety of facts. Surely this is all that can be meant by saying that a "rational law" carries more weight than an "empirical law." Even if such a wider verification were to be obtained, the question would still remain — important in deciding whether it would be possible to establish a new social equilibrium with, say, greater equality of incomes — whether the empirical regularity could be attributed to human nature and not to the accidental nature (if indeed it is accidental) of particular social institutions. Not only that; there is nothing in Champernowne's model — nor, for that matter, in Cantelli's model considered above — to set an upper limit on α. If the Benini-Bresciani interpretation is accepted, perfect equality would be approached as $\alpha \to \infty$, and there is nothing in these formal models to rule this out.

Mandelbrot [141] has studied generalizations of Champernowne's process that lead to limiting distributions which are "Paretian" in his sense, i.e., whose right tails are closely approximated by the (strict) Pareto distribution. However, one of Mandelbrot's chief contributions [140] has been to suggest a model that can explain Pareto's empirical finding that $1 < \alpha < 2$ for aggregate incomes. This is the hypothesis that total incomes are an aggregate of a large number of independently distributed partial incomes with finite means but infinite variance; this leads to Lévy's [132] stable distributions with parameter α in the interval $1 < \alpha < 2$, which have right tails which are closely approximated by the Pareto distribution, i.e., $1 - F(x) \sim Cx^{-\alpha}$ for large x, hence are "Paretian" in Mandelbrot's sense. Empirical studies (e.g. Lydall [135]) indicate that α is not limited to this interval for partial incomes, but can even exceed 3; and for these, Mandelbrot [142], [143] has made the extremely interesting observation that Fréchet's [91] distribution defined by $F(x) = e^{-x^{-\alpha}}$ for

$x > 0$ (and $F(x) = 0$ for $x \leqq 0$) might apply, since it also has the asymptotic Paretian property and is the limiting distribution of the maximum of a large number of independently distributed random variables.[29] The fact that total but not partial incomes are found empirically to satisfy $1 < \alpha < 2$ makes Mandelbrot's hypothesis quite compelling. Also in its favor is the fact that it allows a more realistic representation of income distributions at the lower end of the scale, without sacrificing — as is the case with the lognormal distribution — the empirically well-established Paretian nature of the right tail of the distribution.

Wold and Whittle [256] have developed a model that yields the Pareto distribution for wealth, and which explicitly introduces demographic factors. Assuming a constant force of mortality, γ, and a constant rate of growth of assets, β, they obtain (in the simplest version of their model) the relationship $\alpha = \gamma/\beta$, which they justify as being plausible by taking as an illustration the values 1.6 = .04/.025. They do not provide a justification for the assumed constancy of γ, which is perhaps to be explained on the basis of aggregation over different age groups. One of the desirable features of the model, which they stress, is that it is capable of being tested against independent estimates of α, β, and γ. On the other hand, the question immediately arises as to whether the relationship $\alpha = \gamma/\beta$, even if it should hold up in a particular country at a particular time, shows any empirical stability across countries and over different periods. This is the kind of test that would be needed in order to be able to ascribe any sort of explanatory power to the model.

Lydall [135, pp. 128–9] has introduced an interesting model of quite a different kind which assumes, "firstly, that on the average managers in a given grade supervise a constant number of people in the grade below them, . . . and, secondly, that the salary of managers in a given grade is a constant proportion of the aggregate salaries of the people whom they directly supervise." He shows that these hypotheses lead to the Pareto law. What is particularly interesting about Lydall's study is his interpretation of the predictive power of his model; he finds (p. 130) that countries where incomes tend to follow a lognormal rather than Pareto distribution (such as those of eastern Europe) tend to be countries where managers are paid on the basis of "ability" rather than responsibility. It is not necessary to accept this particular inference in order to appreciate that this is just the kind of insight one hopes to obtain in looking for a rational basis for Pareto's law.

[29] The basis for Fréchet's distribution is that it is one of the three possible types of distribution of the maximum of a series of independently and identically (except for scale and translation factor) distributed random variables, the other two being $F(x) = e^{-(-x)^{-\alpha}}$ for $x \leqq 0$ and $F(x) = 1$ for $x > 0$ (which obviously does not apply to incomes, although the corresponding distribution obtained by replacing $F(x)$ by $1 - F(-x)$, for the minimum of the series conceivably could), and $F(x) = e^{-e^{-\alpha x}}$ (which, when $F(x)$ is replaced by $1 - F(-x)$ and maxima are replaced by minima, forms the basis for the Gompertz distribution — cf. Brillinger [41]). These results are due to Fréchet [91], Fisher & Tippett [89], and Gnedenko [109] (see also Feller [86, p. 278], Kendall & Stuart [125, pp. 330–4]). Mandelbrot [142] has provided an economic interpretation to justify the first of the Fréchet distributions, in terms of the seeking of maximum income, which is related to models proposed by Roy [205] and Tinbergen [232]. The empirical justification is the Paretian property $1 - e^{-x^{-\alpha}} \sim Cx^{-\alpha}$.

4.6 Characterizations of Pareto's law

In his penetrating discussion of Pareto's law, which won Pareto's [169] high praise, Sorel [226] sought to obtain an interpretation of the meaning of Pareto's formula. He drew the analogy, as did Edgeworth [75] also only a short time later, between Pareto's law and the laws of mortality of Gompertz [110] and Makeham [138], [139]. According to Gompertz's law, the instantaneous death rate, or "force of mortality," increases exponentially with age, and can thus be interpreted as a measure of vulnerability to death. Makeham's first law, which fits the data still better, adds a constant term representing chance events independent of age. As Sorel remarked, neither of these laws does justice to the complex biological phenomena underlying the causes of death; they can be considered as explanations only on a very formal plane. And yet, they do without doubt have a ring of plausibility about them which enhances one's confidence in the mortality distributions which they imply.[30] Pursuing his objective of finding an analogous interpretation of Pareto's law, Sorel began by appealing to Fechner's [84] law as well as to Bernoulli's [24] formula for evaluating income, as a way to justify subdivision of the income scale in geometric progression, so that the typical incomes in the various income brackets could be represented by $x_i = x_0 e^{\lambda i}$, $i = 0, 1, 2, \ldots$. In accordance with the logarithmic evaluation, or utility function, $u_i = U(x_i) = \log x_i$, the widths of these income brackets would be perceived by individuals to be the same, i.e., as corresponding to equal differences in utility $u_{i+1} - u_i = \lambda$. The number of people with utility higher than u is given by

$$P(u) = N(e^u) = N(x), \tag{4.63}$$

so the number of people in the ith bracket is

$$-\Delta P(u_i) = P(u_i) - P(u_{i+1}) = P(u_i) - P(u_i + \lambda).$$

Now, Sorel considers the ratio between the number of people in the $(i+1)$th bracket and the number of people in the ith bracket as an indication of the prospects for advancement from the ith to the $(i+1)$th bracket. One minus this ratio is then what Sorel calls the *difficulty of rising.* Dividing this quantity by the width of the brackets, $\Delta u_i = \lambda$, we obtain

$$\frac{-\Delta P(u_i) + \Delta P(u_{i+1})}{-\Delta P(u_i)\Delta u_i} = \frac{P(u_i) - 2P(u_i + \lambda) + P(u_i + 2\lambda)}{[P(u_i) - P(u_i + \lambda)]\lambda}.$$

Dropping the i subscript in the expression on the right and letting the width of the brackets approach zero, we obtain by l'Hospital's rule

$$\lim_{\lambda \to 0} \frac{P(u) - 2P(u + \lambda) + P(u + 2\lambda)}{[P(u) - P(u + \lambda)]\lambda} = -\frac{P''(u)}{P'(u)}.$$

[30] See footnote 2 of Chapter 4, above. There is more to these laws than just plausibility, however; they can be shown to result from the hypothesis that a mechanism with many components fails with the first failure of one of these components — cf. Brillinger [41] and the discussion in the preceding footnote.

Sorel's characterization reads [226, p. 588]: *"the difficulty of rising is constant and measured by α."* That this follows from Pareto's law is immediately verified. But the converse is also true; that is, the condition

$$D(u) \equiv -\frac{P''(u)}{P'(u)} = \alpha \tag{4.64}$$

implies that $N(x)$ satisfies Pareto's law. For, as a differential equation in P' it has the solution $P'(u) = -ce^{-\alpha u}$, which when integrated gives (putting $c = A\alpha$ and taking account of the fact that $P(u) \to 0$, as $u \to \infty$)

$$P(u) = Ae^{-\alpha u}. \tag{4.65}$$

Substituting (4.65) in (4.63) we obtain the Pareto law $N(x) = Ax^{-\alpha}$.

Sorel also noted that the "difficulty of rising" function $D(u)$ of (4.64) could also be interpreted as the "ease of falling" in the income scale. In fact, as is clear from the discussion in §4.5 of Champernowne's model [47], it can be related directly to the ratio of the probability of falling to the probability of rising in the income scale.[31] The interpretation is a very satisfying one, since with a fixed minimum income m, a greater difficulty of rising will lead to a smaller expected income, μ, hence a lower α. And in terms of the Benini-Bresciani interpretation of α, if per capita income μ is fixed, then the greater is the difficulty of rising and ease of falling, the less does it matter whether one rises or falls.

The next important contribution to the characterization of Pareto's law was that of Hagstroem [111], who presented three separate and equivalent characterizations which I shall take up in reverse order. Being an actuary, it was natural for Hagstroem to consider the analogue of the "force of mortality" function measuring vulnerability to death. In the case of the Pareto distribution this function has the form

$$V(x) = -\frac{N'(x)}{N(x)} = \frac{\alpha}{x}. \tag{4.66}$$

Thus, unlike Gompertz's law whereby the vulnerability increases geometrically with age as $V(x) = Bq^x (q > 1)$, viewed as a force of mortality function (4.66) diminishes with age. In fact, as was noted in §4.1 Pareto employed very nearly the form (4.66) for *infant* mortality. It is immediate that (4.66), which may be written in the elasticity form[32]

$$\varepsilon(x) = -\frac{xN'(x)}{N(x)} = -\frac{d\log N(x)}{d\log x} = \alpha, \tag{4.67}$$

[31] Specifically, the "difficulty of rising" is, from (4.60), given by

$$\alpha = -[\log(p_1/p_{-1})]/\lambda$$

.

[32] Giaccardi [101] studied income distributions implied by assigning various values to the parameters when the function (4.67) is of the form $\varepsilon(x) = A + Bx + Cx^{\gamma}$, and found the appropriate characterizations for the Amoroso distribution [11] as well as for the lognormal distribution ($C = 0$).

yields the Pareto law $\log N(x) = \log A - \alpha \log x$. Hagstroem interprets (4.66) as follows [111, p. 83]: "the difficulty of acquiring a large income may be compared with that of reaching an old age. The difficulty is measured by $V(x)$ and evidently has to do with the popular thesis that it is the first million that counts." An alternative and illuminating interpretation of the characterization (4.66) has been suggested by Mackey [137] who inverts the expression (4.67) — which is the elasticity with respect to income x of the number of people with incomes higher than x — to obtain the function

$$I(x) = -\frac{N(x)}{xN'(x)} \tag{4.68}$$

which he interprets as a measure of the *incentive* of moving up the income scale, since it measures the percentage increase in a person's income resulting from, and expressed relative to, the percentage of those people above him that he passes in the income scale. In terms of this interpretation, the incentive to rise is independent of one's position in the income scale if and only if the income distribution has the Pareto form.

The second of Hagstroem's characterizations is of considerably greater interest. In fact, it had already been anticipated by Bowley [32], and a closely related (and equivalent) formula had been obtained by Bresciani-Turroni [35, 36] earlier still. And since the publication of Hagstroem's paper, the characterization has been independently rediscovered no less than three times, by Fréchet [90], Van der Wijk [233], and Mackey [137]. What Bresciani-Turroni showed ([35, p. 800]; see also [40, p. 13l]) was that the Pareto law implied, for any income levels x and y in the observable range,

$$R(x) - R(y) = \delta[xN(x) - yN(y)], \tag{4.69}$$

where the function R is defined by (4.13) and δ is given by (4.29). This is a useful formula for computing the amount of income in any given bracket. Four years later Bowley [31, p. 264] observed that under the Pareto law, "the average income above £x is £$[\alpha/(\alpha-1)]x$"; the following year he described Pareto's formula by saying [32, p. 106]: "Its simplest expression is that the average of all incomes above any amount £x varies directly as x." Denoting this average by $M(x) = R(x)/N(x)$, the characterization states that the function

$$E(x) = \frac{M(x)}{x} = \frac{R(x)}{xN(x)} = \delta \tag{4.70}$$

is equal to a constant $\delta > 1$ for all x. An explicit proof that (4.70) implies that x follows the Pareto law was apparently first provided by Hagstroem [111, p. 82], who obtained (4.70) as a consequence of his first characterization to be described below. The proof is obtained most simply by differentiating

$$R(x) = \delta x N(x) \tag{4.71}$$

to obtain

$$xN'(x) = R'(x) = \delta[N(x) + xN'(x)] \tag{4.72}$$

(the first equality following directly from the definition (4.13) of the function R). This immediately yields (4.67), where $\alpha = \delta/(\delta - 1)$. Obviously, differentiation of Bresciani-Turroni's formula (4.69) with respect to x, for any given y, yields the same result.

Fréchet [90, p. 548] noticed that the constancy of $M(x)/x$ followed from strict adherence of x to the Pareto law, and remarked that "if it were practically constant it would furnish a coefficient of great usefulness, which could be considered as measuring the inequality in the distribution of incomes," and "which would be very close to unity if incomes were nearly all equal." He went on to assert that if the function reciprocal to (4.70) satisfied the equation

$$\frac{xN(x)}{R(x)} = \varphi(\log N(x)) \tag{4.73}$$

for some function φ, then R could be obtained as a function of N, yielding an explicit relation between N and x. For the proof he referred the reader to another paper which, however, never appeared.[33] The solution was finally presented by him thirteen years later [93, pp. 34–5], in the form of the inverse function

$$x(N) = \frac{\varphi(\log N)}{N} e^{\int \varphi(\log N) d \log N}. \tag{4.74}$$

Choosing φ to be the constant function equal to $1/\delta$, the Pareto form follows immediately. Fréchet put forward this approach as a way of generalizing Pareto's law, and originally suggested [90] choosing φ to be a polynomial or, alternatively, either the Laplace or normal density function. This idea was apparently abandoned, but later Fréchet showed [94, pp. 104–5], [95, p. 30] that φ could be chosen so as to yield the lognormal distribution, for which case he found that $\lim_{x\to\infty} M(x)/x = 1$.

Van der Wijk [233, p. 574] came upon formula (4.70) by way of Gini 's formula for income distribution [105, p. 72], [106, p. 49]. Gini had estimated the parameter δ of (4.29) by fitting the equation

$$\log N(x) = \delta \log R(x) - \log K \tag{4.75}$$

to empirical data. What this procedure amounts to is assuming the function $\delta(x)$ of (4.30) to be a constant. Gini argued vigorously in favor of his formula in preference to Pareto's, and even stated explicitly [106, pp. 48–9] that Pareto's law implied (4.75) *but not conversely.* He thus considered his law $N = R^\delta/K$ to

[33] Fréchet's characterization was discussed by Roy [206, pp. 113-5] who, however, was led astray by some misprints in Fréchet's paper, and by Gibrat [103, pp. 44–6), who attempted to reconstruct the then missing argument with only partial success. The characterization has since been stressed in the writings of Allais [4, II, Appendix IIA, pp. 34–7], [5, p. 406], [6, p. 79], who refers to Gibrat [103] and Fréchet [93]. Fréchet in his early article [90] suggested using an inequality index of the form $\int E(x)f(x)dx$, which would of course reduce to the Gini index δ in the case of Pareto distributions.

be more general and of wider validity.[34] Gini's assertion remained unchallenged until D'Addario [56, pp. 189–90], twenty-four years later, proved that on the contrary (4.75) implies Pareto's law. In fact, the elements of the proof were already contained in the literature, since what Van der Wijk showed was that Gini's law implies (4.70), and Hagstroem had already shown that (4.70) implies Pareto's law. To obtain Van der Wijk's result one need merely differentiate (4.75) to obtain, with (4.13) ,

$$\frac{N'(x)}{N(x)} = \delta \frac{R'(x)}{R(x)} = \delta \frac{xN'(x)}{R(x)}. \tag{4.76}$$

Cancelling $N'(x)$ from the left and right expressions one obtains precisely (4.70).

Van der Wijk described (4.70) as the "average-law," and correctly asserted that the Pareto law, Gini law, and "average-law" were mutually equivalent [233, pp. 573–4], contenting himself to supply only the "easiest" implication. He also gave (4.70) an interpretation in terms of a Bernoulli-Fechner utility function, as expressing the constancy of the difference $\log M(x) - \log x$, where $\log x$ was described as "psychic income." In his words [233, p. 581]: "within very broad income limits every social group feels equally poor."[35]

Mackey [137] has described the function (4.70) suggestively as a measure of "how envious people with income x should be of their economic superiors." Allais [5, p. 406], [6, p. 79], like Fréchet [90], has interpreted (4.70) as an inequality index, and in a further elaboration which calls to mind Sorel's [226] as well as Van der Wijk's [233] interpretations, he says [6, p. 79]: "The constancy of the ratio $M(x)/x$ for all values of x in a society at a given time appears to mean simply that the sociological equilibrium that is established is such that at each income level the sentiment of inequality is independent of this income."

Hagstroem's principal characterization is an interesting application of the previous one, and one which also has important practical implications. It is stated as follows [111, p. 78]: if, as a consequence of a monetary inflation, every citizen's income is multiplied by the same constant $\lambda > 0$, the law of income distribution will be such that the average income of those whose incomes exceed x is unchanged. This statement is to be interpreted as holding no matter what values of $\lambda > 0$ and x (in the domain of the distribution) one chooses. Defining

[34] Gini also stressed [106, p. 49], in support of his argument for the superiority of his formula to Pareto's, that it gave a better fit to the data. This could well be true; but the fact was misinterpreted by Gini. That is, it might indeed be preferable to estimate α by first estimating δ by Gini's formula and then substituting in the formula $\alpha = \delta/(\delta - 1)$, rather than directly by Pareto's formula; however, this is simply a problem in statistical estimation, and has nothing to do with the relative merits of Gini's and Pareto's formulae — which would be impossible since they are equivalent. Fréchet [95] also expressed a preference for the Gini method of estimation, but without reference to Gini.

[35] Van der Wijk, who was generous in acknowledging others' priority, mentioned [233, pp. 574–5] that he discovered Bresciani-Turroni's earlier formula (4.69) after he had already arrived at (4.70). However, Van der Wijk made a slip in stating that (4.69) reduces to (4.70) when $y = 0$; rather, it reduces to (4.70) when $y \to \infty$, but to establish this one needs first to show that (4.69) implies that x satisfies Pareto's law. Later [235, pp. 51–2], Van der Wijk came across the Fréchet [90] and Bowley [32] references, and acknowledged their priority.

$N_\lambda(x)$ as the number of persons whose incomes exceed x after the inflation, we have

$$N_\lambda(x) = N(x/\lambda).$$

Each such person will have a money income λ times as high as before, so that defining $R_\lambda(x)$ as the sum total of money incomes accruing after the inflation to those with money incomes in excess of x, we have

$$R_\lambda(x) = \lambda R(x/\lambda).$$

The average money income, after the inflation, of those earning more than x will then be

$$M_\lambda(x) \equiv \frac{R_\lambda(x)}{N_\lambda(x)} = \frac{\lambda R(x/\lambda)}{N(x/\lambda)} = \lambda M(x/\lambda). \tag{4.77}$$

Now, Hagstroem's hypothesis states that

$$M_\lambda(x) = M(x) \tag{4.78}$$

for all x and λ. Combining (4.77) and (4.78) we then have

$$M(x) = \lambda M(x/\lambda). \tag{4.79}$$

Introducing the change of variable $y = x/\lambda$, this becomes

$$\frac{M(x)}{x} = \frac{M(y)}{y},$$

and this is to hold for all x and y in the domain of the distribution. This of course means that $M(x)/x$ is constant and (4.70) holds. This in turn by the previous arguments (all in Hagstroem's paper) implies that (4.66) holds, which implies Pareto's law.

The result brings out a great advantage of the Pareto distribution in applications: in making comparisons of inequality in income distributions over time or between countries, there is no need to correct for inflation or to find appropriate exchange rates for converting currencies to a common unit. Indeed, Pareto made no such attempt in his early empirical studies [163, II, §959, p. 312];[36] on the other hand it is significant that in his last empirical study of income distribution [175], in which he employed his "second approximation" in making a comparison of income distributions in Paris in the years 1292 and

[36] In his commentary on Sorel's [226] review of his work, Pareto stated [169, p. 260]: "*A quoi s'applique la théorie?* A la répartition des revenues exprimés en numéraire." It might be thought that Pareto was using the term *numéraire* here in Walras's sense, meaning some commodity that plays the role of money. However, up to that time Pareto in fact used only money income in his empirical calculations, so that the above passage must be translated as: "*To what does the theory apply?* To the distribution of incomes expressed in money." This interpretation is further reinforced by the fact that Pareto goes on in the above passage to contrast this proposition with the one Sorel was interested in: "The distribution of what I have called *ophelimity* is not considered."

1896, he was very careful to reduce all units to wheat equivalents. The above characterizations of course do not apply to the second and third approximations (see (4.44) above).

Another characterization that has great relevance to applications is the following one due to Bhattacharya [25]. Bhattacharya posed the followng problem: Since the most frequent sources of data on incomes is income-tax statistics, one will ordinarily not have data on all incomes in the interval $[m, \infty)$, but only on incomes in an interval $[y, \infty)$, where $y > m$. An investigator is therefore forced to limit himself to studying the truncated distribution over the latter interval. The question then arises: when is the form of the Lorenz curve invariant with respect to the choice of the point of truncation y?

Now, from (4.12), (4.14), and the definition of the Lorenz curve (4.28), it follows that this Lorenz curve may be expressed parametrically as the function L^* satisfying the following identity in x:

$$\frac{N(x)}{N(m)} = L^* \left(\frac{R(x)}{R(m)} \right). \tag{4.80}$$

If the distributions are replaced by truncated distributions with point of truncation y, then the corresponding Lorenz curves have the expression (4.80) with m replaced by y, for $x \geqq y$. Bhattacharya's requirement is that the functional form should be independent of the choice of y, that is, that one and the same L^* should satisfy the equation

$$\frac{N(x)}{N(y)} = L^* \left(\frac{R(x)}{R(y)} \right) \tag{4.81}$$

identically in both x and y, for $x > y$. The solution, as expected, is that this requirement is sufficient as well as necessary for x to follow the Pareto law (first approximation).

The result may readily be proved by replacing (4.81) by the equivalent identity

$$\log N(x) - \log N(y) = \psi(\log R(x) - \log R(y)), \tag{4.82}$$

and differentiating both sides of (4.82) with respect to x and y. This yields

$$\frac{R(x)}{N(x)} \frac{N'(x)}{R'(x)} = \psi'(\log R(x) - \log R(y)) = \frac{R(y)}{N(y)} \frac{N'(y)}{R'(y)}$$

whence (differentiating (4.13)) the function

$$\frac{R(x)}{N(x)} \frac{N'(x)}{R'(x)} = \frac{R(X)}{xN(x)}$$

is a constant, i.e., (4.70) holds.

Thus, (4.81) holds identically in y and $x > y$ if and only if it is of the form

$$\frac{N(x)}{N(y)} = \left(\frac{R(x)}{R(y)} \right)^\delta \tag{4.83}$$

and in particular (4.80) must be of this form for $y = m$. This is, of course, Gini's form, and the result is of interest in view of Gini's contention [106, p. 48] that Pareto's formula held only for incomes above a certain limit whereas his own formula held equally well for such truncated distributions and the entire distribution all the way to $m = 0$. The contention is, of course, groundless and incorrect.

The final characterization I now come to is due to Mackey [137]. He assumes that individuals have already been arranged in order of merit. Let u be a suitable measure of merit (talent, or aptitude), and let $g(u)$ as in §4.5 be the density function of individual aptitudes or abilities. Denote by

$$B(u) = \int_u^\infty g(t)dt \tag{4.84}$$

the proportion of the population that is higher (better) than u on the aptitude scale. Let $x = T(u)$ (as in §4.5) denote the income of a person with aptitude u, and assume that T is strictly increasing. Mackey's hypothesis is that given any two individuals at positions u and v on the talent scale, their relative incomes should depend only on their relative position on the scale, i.e.,

$$\frac{T(u)}{T(v)} = \varphi\left(\frac{B(u)}{B(v)}\right) \tag{4.85}$$

for some function φ. It is a remarkable fact that property (4.85) is also characteristic of the Pareto distribution.

To establish the result we may first observe that from the strict monotonicity of T we have $B(u) = G(T(u))$, where G is defined by (4.12) Then, (4.85) is equivalent to the condition

$$\frac{x}{y} = \varphi\left(\frac{G(x)}{G(y)}\right) = \varphi\left(\frac{N(x)}{N(y)}\right). \tag{4.86}$$

This is in turn equivalent to the condition

$$\log x - \log y = \psi(\log N(x) - \log N(y)). \tag{4.87}$$

Differentiating (4.87) with respect to x and y (which is permissible since the left side is certainly differentiable, hence the right side must be) we see that ψ' must be negative (hence φ' as well), and that the equations

$$\frac{N(x)}{xN'(x)} = \psi'(\log N(x) - \log N(y)) = \frac{N(y)}{yN'(y)}$$

must hold for all x and y, implying the constancy of the function (4.68).

4.7 Invariance of the law of income distribution

Pareto always insisted that his formula for the distribution of incomes was only a first approximation; that the fit was good only for the right tail of the distribution (the only part for which he had solid empirical evidence), and even then not perfect, i.e., the logarithms of the observed points, which he fitted by the straight line

$$\log N(x) = \log A - \alpha \log x, \tag{4.88}$$

actually followed a slightly convex curve. Nothing in Pareto's work leaves one to believe that he would not have welcomed the kinds of generalizations made possible by Fréchet's [93] extended Pareto distribution discussed in §4.5 above, or by the family of asymptotically Paretian distributions, i.e., distributions F such that $1 - F(x) \sim Cx^{-\alpha}$, introduced by Mandelbrot [140], [141], [142], [143]. At the same time Pareto was impressed by the fact that, according to his own evidence, there was remarkable stability in the slope of the observed distribution, over time and cutting across social systems as diverse as European states in the 19th century and Peru at the time of the Spanish conquest. This led him to the following statement [163, II, §960, p. 312]:

> These results are very remarkable. It is absolutely impossible to assume that they are due solely to chance. There must certainly be a *cause* which produces a tendency for incomes to be distributed along a certain curve. The form of this curve seems to depend only slightly on different economic conditions of the countries considered, since the effects are about the same for countries in which economic conditions are as diverse as those of England, Ireland, Germany, Italian cities, and even Peru.
>
> True, since we are dealing only with empirical laws, we cannot be too prudent. In any case, the consequences we shall draw from this law will at least always be valid for peoples for whom we have seen that they are confirmed.

Cautious as this statement surely is, it did not prevent either his detractors or some of his followers from pressing the point much more strongly. The question inevitably rose as to whether it meant that it would be impossible for a socialist state to bring about complete equality of incomes. With respect to this question Pareto pointed out (i) that even if a socialist state should succeed in doing this, the resulting gain to the poor would still be insignificant [163, II, §967, p. 328], and (ii) that the most effective way for a socialist state to accomplish this program would be to carry out the redistribution directly, while organizing production according to the principles of the competitive price mechanism. This does not answer the question of whether a socialist state could in fact succeed in accomplishing purely redistributive egalitarian measures — ineffective as they might be. Here, Pareto fully accepted Edgeworth's position that one would need a "rational law" of income distribution rather than a purely "empirical law" in order to attach enough weight to it that it could be extrapolated to new systems of social organization. The following passage well expresses his position [170, p. 501]:

> The laws of the distribution of wealth evidently depend on the nature of man and on the economic organization of society. We

> might derive these laws by deductive reasoning, taking as a starting point the data of the nature of man and of the economic organization of society. Will this work some time be completed? I cannot say; but at present it is certain that we lack sufficient data for undertaking it. At present the phenomena must be considered synthetically, and every endeavor must be made to discover if the distribution of wealth presents any uniformity at all. Fortunately, the figures representing the distribution of wealth group themselves according to a very simple law, which I have been able to deduce from unquestioned statistical data. The law being empirical, it may not always remain true, especially not for all mankind. At present, however, the statistics which we have present no exceptions to the law; it may therefore provisionally be accepted as universal. But exceptions may be found, and I should not be greatly surprised if some day a well-authenticated exception were discovered.

Pareto did in fact go a considerable distance in developing such a "rational law," although only in qualitative outlines. First presented in the *Cours*, [163, II, §§1002–1012, 1024–1031, pp. 356–363, 371–5], it gradually evolved into his elaborate theory of the "circulation of the élites" [176], [179, I, pp. 6–15, 34–38; II, pp. 130–169], [182, Ch. II, VII]. This was an essentially dynamic theory according to which natural selection operated in the lower classes to produce a relative preponderence of talented and aggressive individuals, while at the same time the absence of such selective forces among the more well-to-do would produce a relative preponderance of degenerate (in the technical sense) and regressive individuals. Depending on the respective birth and death rates among the various strata, there would be a stronger or weaker pressure for the vigorous elements to move up and the decadent elements to move down. If the system allowed for social mobility the process would be smooth and continuous; if not, it would be discontinuous and accompanied by disorder and revolution [182, Ch. VII, §21]. But one way or another, the process would continue. Pareto never formulated this as a stochastic process, but the formalization by Champernowne [47] discussed in §4.5 above may be regarded as a first approximation to such a process and in many ways an admirable one.

In Pareto's description of the theory of the circulation of the élites, an important part is played by sentiment. Here, his ideas are strongly stimulated by those of Spencer [227], [228]. Pareto held that human beings, in the desire for self respect, have a deep-felt need to rationalize their actions and make them seem meritorious. Thus, the rising elements who are in the process of ousting the old élites feel they must justify their actions on the basis of some moral principle such as the common good. And these sentiments are not knowingly hypocritical; on the contrary, they would not take hold unless they were genuinely and sincerely felt, by the leaders as well as the followers.

In terms of this theory, socialism, humanitarianism, egalitarianism, etc., are emerging religions which serve to rationalize the power plays on the part of the talented groups emerging from the lower classes. They are unlike older religions in being more abstract, since the advance of science and increased sophistication have made the idea of personal deities and physical miracles less plausible than in former times; but they are religions nonetheless, just as dogmatic and self-righteous as the ones they supplant, and just as blind to certain facts. Government plays an important role in the theory, being, among other things, an instrument of spoliation of other classes on the part of the emerging élites; and "what limits spoliation is rarely the resistance of the despoiled; rather it is the losses inflicted upon the entire country which redound in part on the spoliaters" [163, II, §1049, p. 384]. And "the demand for equality is nothing but a disguised manner of demanding a privilege" [191, §1222]; more specifically still [191, §1227]:

> The sentiment that is very inappropriately named equality is fresh, strong, alert precisely because it is not, in fact, a sentiment of equality and is not related to any abstraction, as a few naive "intellectuals" still believe; but because it is related to the direct interests of individuals who are bent on escaping certain inequalities not in their favor, and setting up new inequalities that will be in their favor, this latter being their chief concern.

Thus, for Pareto, the socialist movement is just one more example in history, of which there are many others which have made similar promises, of an ideological vestment covering the activities of a rising élite. And this élite in its turn will be superseded by others. Complete equality, even if attained momentarily, would be an unstable equilibrium which could not last [163, II, §1009, p. 360]; soon the aggressive elements would take over from the regressive ones and the former distribution of income would be restored. In sum [163, II, §1012, p. 363]:

> The inequality in the distribution of income seems therefore to depend much more on the very nature of men than on the economic organization of society. Profound modifications in this organization could well have only slight influence in modifying the law of distribution of income.

Much has been made of the fact that in later years Pareto arrived at a considerably more cautious assessment than the one given in the above passage from the *Cours*, seeming even to reverse himself in the following passage in the *Manuale* [182, Ch. VII, §23]:

> The data available to determine the shape of the [income curve] refer mainly to the 19th century and to civilized nations; consequently, the conclusions drawn from the data cannot be extended beyond those limits. It remains only as a more or less probable induction that in other times and among other peoples it has a shape somewhat similar to that found today.
>
> Likewise, we cannot assert that this shape would not change if the social constitution were to change radically; if, for instance, collectivism were to take the place of private property.

Pigou [202, pp. 652–5] interpreted this to be a complete capitulation from Pareto's previous position, and triumphantly concluded: "This means that, even if the statistical basis of the 'law' were much securer than it is, the law would but rarely enable us to assert that any contemplated change *must* leave the form of income distribution unaltered" — as if Pareto had ever asserted such necessity — and that "in view of the weakness of its statistical basis [i.e., the limitation to 19th century civilized nations], it can *never* enable us to do this." Pigou therefore attached no credence at all to the empirical uniformity Pareto discovered. However, in calling Pareto in as a witness against himself he overlooked the fact that the position Pareto held in 1906 was one he had come to hold with respect to all "laws," empirical, rational, and even logical [191, §69, pp. 4–5]:

> We look for the uniformities presented by facts, and those uniformities we may even call laws; but the facts are not subject to the laws: the laws are subject to the facts. Laws imply no *necessity*. They are hypotheses serving to epitomize a more or less extensive number of facts and so serving only until superseded by better ones.
>
> . . . Every proposition that we state, not excluding propositions in pure logic, must be understood as qualified by the restriction *within the limits of the time and experience known to us.*

Thus, as to whether a socialist regime could succeed in making radical alterations in income distribution in the direction of complete equality, Pareto in the end maintained the same kind of reserve as he did with respect to all scientific laws, and even propositions in logic. It could hardly be inferred, therefore, that as one who believed that science "is just a quest for uniformities, and that is the end of it" [191, §89n] he could so easily dismiss the empirical uniformities he had found, particularly when they were in agreement with his inductions concerning the heterogeneity of human beings, the historical processes of class circulation, and the real meaning to be attached to slogans calling for greater equality. In the last analysis, then, Pareto based his cautiously held belief with respect to the invariance of the distribution of income on a large number of facts — not just on the empirical constancy he observed, although this was paramount, but also on a detailed historical and qualitative study of heterogeneity in the distribution of abilities and the process of social mobility and historical evolution. Whether or not his assessment turns out to be right, it seems fair to say that those egalitarians who reject his view have so far failed to come up with equally convincing empirical evidence on the other side.

Chapter 5

Time Series Analysis and Methods of Interpolation

In early life Pareto was a social reformer and an ardent free trader. As he grew older he became more of a fatalist. Instead of trying to reform the world he tried to describe it and explain it. Having a very strong empirical orientation, he recognized that if liberal and laissez-faire economic policies made little headway despite their apparent a priori correctness, there must be a reason for this that needs to be explained. This does not mean that he became a determinist in the narrow sense of believing in the complete predictability of economic and social facts — although he did, as we have seen in the case of income distribution, hold a strong belief based on his study of the facts in the existence of uniformities among economic and social aggregates. As a social scientist his approach differed in fundamental ways from that of Karl Marx. In the first place he had no illusions that economic and social laws would be repealed after the Revolution, and he felt extremely skeptical about the scope and efficacy of policy measures in changing these basic laws, whether by reform or by revolution. Such a belief is just as consistent with a belief that the real world is basically random as it is with the belief that it is predictable and deterministic, and Pareto's views seem to have been somewhere between these two extremes. Secondly, by way of contrast to Marx, Pareto had learned his lessons about general interdependence very well from Walras, and believed that such interdependence existed between social and economic phenomena; he rejected the belief in a one-way causal relationship as being as naive as beliefs of early classical economists concerning the "cause" of value.

Pareto's belief in the mutual interrelationship of economic and social phenomena was fundamental to his approach to applied economics. In effect, there could be no such thing as applied economics, but only applied social science [191; §2022]:

> A number of economists today are aware that the results of their science are more or less at variance with concrete fact, and are alive

> to the necessity of perfecting it. They go wrong, rather, in their choice of means to that end. They try obstinately to get from their science alone the materials they know are needed for a closer approximation to fact; whereas they should resort to other sciences and go into them thoroughly — not just incidentally — for their bearing on the given economic problem.

He was to follow his own advice in the empirical studies that absorbed his interest in his later years.

Quite early in his career (cf. [161]) Pareto began displaying an interest in quantitative analysis of economic and social time series and the development of techniques for seasonal adjustment. He indicated a preference for Cauchy's interpolation formula

$$b = \frac{\sum_{t=1}^{n}\{(y_t - \bar{y}) \cdot \operatorname{sgn}(x_t - \bar{x})\}}{\sum_{t=1}^{n} |x_t - \bar{x}|} \quad \left(\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n}, \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}\right)$$

for the slope of a line $y = a + bx$ fitted to a series of observations, mainly on account of its computational simplicity compared (then) to that of the method of least squares, and it was this method (applied to formula (4.88)) that he used to estimate the coefficient α in fitting his income distribution curve. Pareto also developed elaborate tables for use in fitting polynomial functions of time by least squares (cf. [173]).

Among Pareto's chief interests was the search for empirical relationships between economic and social phenomena. In 1896 he fitted some functions relating the marriage rate in England to the level of exports (the latter being taken as an index of economic prosperity) and found a positive association (cf. [171]). Subsequently [184, 186] he developed an elaborate method of curve-fitting based on both Cauchy's method and the method of least squares, which he then applied [190] to comparisons of time series on trade statistics in various European countries with data on emigration and theater ticket sales, finding a positive association. He then went on to observe that periods of rapid economic progress were much less disturbed from a socio-political point of view than periods of economic depression, and imperialism was listed as one of the phenomena that tend to be associated with a sustained period of prosperity. A subsequent study [192] investigated the relation between the exchange rate on the lira and the Italian money supply, reserve ratio, and balance of trade, as well as the relation between tax revenues and various indices of economic prosperity. This was followed by a further study of exchange rate fluctuations [193].

Pareto's approach to interpolation of time series data will come as a surprise to many readers who think of him only as a theoretician [171, p. 376]:

> Any interpolation formula is, in my opinion, all the better the less it is influenced by theoretical concepts and a priori ideas. Of course, I do not mean by this that the statistician should renounce the use of a judicious critical spirit and seek, at random, relations between facts which obviously cannot have any. A theoretical preparation is

> always necessary. I only mean that when we consult the facts we must, as far as possible, abstain from dictating to them the reply we expect from them.

This strongly-expressed belief, which set him apart from Walras, was one from which he never swayed. His method of posing the estimation problem [184], [186] is also of great interest. He specifically rejected the notion that one should specify a model one takes as true, and seek to obtain an optimal estimate of its parameters; rather, he took the position that one should recognize explicitly that the actual model fitted is a very crude approximation to what one might regard as a true model, and that the estimation procedure should take account of this type of specification error along with errors of observation. His approach therefore anticipates in an interesting way some of the modern developments in econometric methodology which stress errors of aggregation and specification.

It may also come as a surprise to those who think of Pareto as a general equilibrium economist to learn of his views of economic dynamics [163, II, §926, p. 278]: "An economic crisis must not be considered as something accidental which interrupts the normal state of affairs. Rather, cyclical movement is to be regarded as the normal state ..." Further (p. 297): "It is not at all certain that rhythmic movement is not one of the conditions of economic progress. On the contrary it seems very likely that such movement is just one manifestation of the vitality of the economic organism." One cannot fail here to notice the influence of Juglar [l20], nor the influence which both Juglar and Pareto must have had on Schumpeter [216].

In his analysis of business cycles Pareto placed much emphasis on the mutual interrelationship between economic and social phenomena. Thus he expressed the belief [190] that imperialism, while being an outgrowth of prosperity, if it results in victorious wars also "increases and stimulates the general vigor of the population" [190, p. 511], and this in turn feeds the prosperity.

One of the great insights of Pareto's later period, in the analysis of "non-logical actions" [187, 191], was that one must not take people's own explanations for their actions at face value. People tend to act on the basis of a number of motives, some of them explainable by the heritage of the past and some of them by self-interest and the interest of the group; and for whatever they do they will find a rationalization that makes the action seem altruistic and noble. In his words [182, Ch. II, §108, p. 132]: "Men are moved by sentiment and interest, but they like to imagine that they are moved by reason; hence, they seek — and always find — a theory which, *a posteriori*, gives a veneer of logic to their actions." For Pareto such rationalization is not sinister hypocracy, but largely unconscious and sincerely felt. But there is another equally important aspect of Pareto's theory of non-logical actions: While rationalizations should not be taken seriously as explanations for human actions, this does not mean that they can be ignored; on the contrary, they are extremely important facts which, when their role is understood, can help explain human behavior. Pareto's message is that it is the job of the social scientist not only to see through such explanations that people provide, but also to use them as an objective fact to help explain the true basis of their actions.

While economists have by now absorbed, rediscovered, and gone beyond Pareto's early ideas — many of which he himself abandoned in later life — they have hardly begun to think in terms of the still bolder ideas he introduced in his final years.

Chapter 6

REFERENCES

Bibliography

[1] Aigner, D. J., and A. J. Heins, "A Social Welfare View of the Measurement of Income Inequality," *Review of Income and Wealth*, 13 (March 1967), 12–25.

[2] Aitchison, J., and J. A. C. Brown, "On Criteria for Descriptions of Income Distributions," *Metroeconomica*, 6 (December 1954), 88–107.

[3] Aitchison, J., and J. A. C. Brown, *The Lognormal Distribution*. Cambridge: at the University Press, 1957.

[4] Allais, Maurice, *Le Tiers monde au carrefour, centralisation autoritaire ou planification concurrentielle*, in two volumes. Brussels: Editions des Cahiers africains, and Paris: Créations de Presse, 1962.

[5] Allais, Maurice, "Pareto, Vilfredo. Contributions to Economics," *International Encyclopedia of the Social Sciences*, The Macmillan Company & The Free Press, 1968, Vol II, pp. 399–411.

[6] Allais, Maurice, "Inégalité et civilisations," in *Mélanges en l'honneur de Raymond Aron*, Paris: Editions Calmann-Lévy, 1971, pp. 71–95. English translation: "Inequality and Civilizations," *Social Science Quarterly*, 54 (December 1973), 508–524.

[7] Allen, R. G. D., "The Foundations of a Mathematical Theory of Exchange," *Economica*, 12 (May 1932), 197–226.

[8] Allen, R. G. D., "A Comparison Between Different Definitions of Complementary and Competitive Goods," *Econometrica*, 2 (April 1934), 168–175.

[9] Allen, R. G. D., "Nachfragefunktionen für Güter mit korreliertem Nutzen," *Zeitschrift für Nationalökonomie*, 5 (September 1934), 486–506.

[10] Ammon, Otto, *Die Gesellschaftsordnung und ihre natürlichen Grundlagen*. Jena: Verlag von Gustav Fischer, 1865. 2nd edition, 1896. French translation: *L'Ordre social et ses bases naturelles*. Paris: Albert Fontemoing, Editeur, 1900. English translation of excerpts from the second edition: "Some Social Applications of the Doctrine of Probability," *Journal of Political Economy*, 7 (March 1899), 204–237.

[11] Amoroso, Luigi, "Ricerche intorno alla curva dei redditi," *Annali di matematica pura ed applicata* [4], 2 (1924–25), 123–159.

[12] Antonelli, G. B., *Sulla teoria matematica della economia politica.* Pisa: nella Tipografia del Folchetto, 1886. English translation: "On the Mathematical Theory of Political Economy," Chapter 16 in [51], pp. 332–364. Abstract in [19], pp. 33–39.

[13] Arrow, Kenneth J., "An Extension of the Basic Theorems of Classical Welfare Economics," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (edited by J. Neyman), Berkeley and Los Angeles: University of California Press, 1951, pp. 507–532.

[14] Arrow, Kenneth J., *Social Choice and Individual Values*, 2nd edition. New York: John Wiley & Sons, Inc., 1963.

[15] Arrow, Kenneth J., H. D. Block, and Leonid Hurwicz, "On the Stability of the Competitive Equilibrium, II,"*Econometrica*, 27 (January 1959), 82–109.

[16] Arrow, Kenneth J., and Gerard Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, 22 (July 1954), 265–290.

[17] Auspitz, Rudolf, and Richard Lieben, *Untersuchungen über die Theorie des Preises*. Leipzig: Duncker & Humblot, 1889.

[18] Barone, Enrico, "Il ministro della produzione nello stato collettivista," *Giornale degli Economisti* [2], 37 (August, October 1908), 267–293, 391–414. Abridged English translation: "The Ministry of Production in the Collectivist State," in *Collectivist Economic Planning* (edited by F. A. Hayek), London: Routledge & Kegan Paul, Ltd., 1935, pp. 245–290.

[19] Baumol, William J., and Stephen M. Goldfeld (editors), *Precursors in Mathematical Economics: An Anthology*. London: The London School of Economics and Political Science, 1968.

[20] Benini, Rodolfo, "Di alcune curve descritte da fenomeni economici aventi relazione colla curva del reddito o con quella del patrimonio," *Giornale degli Economisti* [2], 14 (March 1897), 177–214.

[21] Benini, Rodolfo, "I diagrammi a scala logaritmica," *Giornale degli Economisti* [2], 30 (March 1905), 222–231.

[22] Benini, Rodolfo, *Principii di Statistica Metodologica*. Turin: Unione Tipografico-Editrice Torinese, 1906.

[23] Bergson (Burk), Abram, "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, 52 (February 1938), 310–334.

[24] Bernoulli, Daniel, "Specimen theoriae novae de mensura sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5 (1738), 175–192. English translation: "Exposition of a New Theory of Risk Evaluation," in [19], pp. 15–26.

[25] Bhattacharya, N., "A Property of the Pareto Distribution," *Sankhyā* [B], 25 (December 1963), 195–196.

[26] Bickerdike, C. F., "The Theory of Incipient Taxes," *Economic Journal*, 16 (December 1906), 529–535.

[27] Bickerdike, C. F., Review of *Protective and Preferential Import Duties* by A. C. Pigou, *Economic Journal*, 17 (March 1951), 98–102.

[28] Bodio, Luigi, "Quelques renseignements sur les conditions hygiéniques et sanitaires de l'Italie," *Bulletin de l'Institut International de Statistique*, 2 (1887), 1re Livraison, 264–287.

[29] Bortkiewicz, Ladislaus von, "Die Grenznutzentheorie als Grundlage einer ultra-liberalen Wirtschaftspolitik," [*Schmollers*] *Jahrbuch für Gesetzgebung, Verwaltung und Volkwirtschaft im Deutschen Reich*, 22 (Heft. 4, 1898), 89–128 [1177–1216].

[30] Bortkiewicz, Ladislaus von, "Die Disparitätsmasse der Einkommensstatistik," *Bulletin de l'Institut International de Statistique*, 25 (1931), 3e Livraison, 189–298. "Observations," by Corrado Gini, F. Savorgnan, G. Pietra, and L. von Bortkiewicz, and "Note" by Corrado Gini, G. Pietra, 299–320 M.

[31] Bowley, A. L., "The British Super-Tax and the Distribution of Income," *Quarterly Journal of Economics*, 28 (February 1914), 255–268.

[32] Bowley, A. L., *The Nature and Purpose of the Measurement of Social Phenomena*. London: P. S. King & Son, Ltd., 1915.

[33] Bresciani, Costantino, "Dell'influenza delle condizioni economiche sulla forma della curva dei redditi," *Giornale degli Economisti* [2], 31 (August 1905), 115–138.

[34] Bresciani, Costantino, "Sull'interpretazione e comparazione di seriazioni di redditi o di patrimoni," *Giornale degli Economisti* [2], 34 (January 1907), 13–47.

[35] Bresciani-Turroni, Costantino, "Di un indice misuratore della disugualianza nella distribuzione della richezza," in *Studi in onore di Biagio Brugi nel XXX anno del suo insegnamento*, Palermo: Stab. Tip. Ditta L. Gaipa — Editore, 1910, pp. 793–812.

[36] Bresciani-Turroni, Costantino, "Über einen quantitativen summarischen Ausdruck für die Ungleichmässigkeit der Einkommensverteilung," *Deutsches Statistisches Zentralblatt*, Vol. 3 No. 6, (1911) columns 173–176.

[37] Bresciani-Turroni, Costantino, “Kritische Betrachtung über einige Methoden der Einkommenstatistik,” *Statistische Monatschrift*, 22 (1915), 1–30.

[38] Bresciani-Turroni, Costantino, “On Pareto’s Law,” *Journal of the Royal Statistical Society*, 100 (Part III, 1937), 421–432.

[39] Bresciani-Turroni, Costantino, “On Some Methods of Measuring the Inequality of Incomes,” *Majallat al-qānūn wa-al-iqtisād* (*Revue Al Qanoun Wal Iqtisad*), Cairo, 8 (December 1938), 371–403.

[40] Bresciani-Turroni, Costantino, “Annual Survey of Statistical Data: Pareto’s Law and the Index of Inequality of Incomes,” *Econometrica*, 7 (April 1939), 107–133.

[41] Brillinger, David R., “A Justification of Some Common Laws of Mortality,” *Society of Actuaries, Transactions*, 13 (December 1961), 116–119.

[42] Cantelli, Francesco Paolo, “Sulla deduzione delle leggi di frequenza da considerazioni di probabilità,” *Metron*, 1 (April 1921), 83–91.

[43] Cantelli, Francesco Paolo, “Sulla legge di distribuzione dei redditi,” *Giornale degli Economisti e Rivista di Statistica* [4], 69 (November 1929), 850–852.

[44] Cassel, Gustav, “Grundriss einer elementaren Preislehre,” *Zeitschrift für die gesamte Staatswissenschaft*, 55 (No. 3, 1899), 395–458.

[45] Castellani, Maria, “On Multinomial Distributions with Limited Freedom: A Stochastic Genesis of Pareto’s and Pearson’s Curves,” *Annals of Mathematical Statistics*, 21 (June 1950), 289–293.

[46] Castelnuovo, Guido, *Calcolo delle probabilità*. Milan-Rome-Naples: Società editrice Dante Alighieri di Albrighi, Segati & C., 1919. 2nd edition (in two volumes), Bologna: Nicola Zanichelli, Editore, 1926, 1928.

[47] Champernowne, D. G., “A Model of Income Distribution,” *Economic Journal*, 63 (June 1953), 318–351.

[48] Chipman, John S., “Introduction to Part II,” in [51], pp. 321–331.

[49] Chipman, John S., “The Welfare Ranking of Pareto Distributions,” *Journal of Economic Theory*, 9 (November 1974), 275–282.

[50] Chipman, John S., “An Empirical Implication of Auspitz-Lieben-Edgeworth-Pareto Complementarity,” *Journal of Economic Theory*, 14 (February 1977), 228–282.

[51] Chipman, John S., Leonid Hurwicz, Marcel K. Richter, and Hugo F. Sonnenschein, *Preferences, Utility, and Demand*. New York: Harcourt Brace Jovanovich Inc., 1971.

[52] Cournot, Augustin, *Recherches sur les principes mathématiques de la théorie des richesses*, Paris: L. Hachette, 1838. English trans1ation: *Researches into the Mathematical Principles of the Theory of Wealth*. New York: The Macmillan Co., 1897.

[53] D'Addario, Raffaele, "Un'indagine sulla dinamica distributiva," *Rivista italiana di Statistica*, 2 (July–September 1930), 269–304.

[54] D'Addario, Raffaele, "La curva dei redditi," *Rivista italiana di Statistica*, 3 (April–September 1931), 140–161.

[55] D'Addario, Raffaele, "Intorno ad una curva di ripartizione," *Rivista italiana di Statistica, Economia e Finanza*, 4 (December 1932), 739–778.

[56] D'Addario, Raffaele, "Intorno alla validità dei due teoremi paretiani sulla dinamica distributiva," *Atti dell'Istituto nazionale delle assicurazioni*, 6 (1934), 179–198.

[57] D'Addario, Raffaele, "Sulla curva dei redditi di Amoroso," *Annali dell' Istituto di Statistica*, Bari, 10 (1936), 1–57.

[58] D'Addario, Raffaele, "La curva dei redditi: condizioni per le quali l'equazione di McAlister e la più probabile," *Atti del secondo congresso dell' Unione Matematica Italiana*, Rome: Edizioni Cremonese, 1942, p. 715.

[59] D'Addario, Raffaele, "Sulla curva dei salari dedotta da considerazioni di probabilititià," in *Studi in memoria di Guglielmo Masci* (in two volumes), Milan: A. Giuffre, 1943 , Vol. I, pp. 81–91.

[60] D'Addario, Raffaele, "Ricerche sulla curva dei redditi," *Giornale degli Economisti e Annali de Economia*, N.S., 8 (January–February 1949), 91–114.

[61] Dalton, Hugh, "The Measurement of the Inequality of Incomes," *Economic Journal*, 30 (September 1920), 348–361. Reprinted as Appendix to Hugh Dalton, *Some Aspects of the Inequality of Incomes in Modern Communities*, London: George Routledge & Sons, and New York: E. P. Dutton & Co., 1925.

[62] David, H. A., "Gini's Mean Difference Rediscovered," *Biometrika*, 55 (November 1968), 573–575.

[63] Davis, Harold T., *The Analysis of Economic Time Series*. Bloomington, Indiana: The Principia Press Inc., 1941.

[64] Debreu, Gerard, "Valuation Equilibrium and Pareto Optimum," *Proceedings of the National Academy of Sciences*, 40 (July 1954), 588–592.

[65] Debreu, Gerard, *Theory of Value*. New York: John Wiley & Sons, 1959.

[66] Debreu, Gerard, "Topological Methods in Cardinal Utility Theory," in *Mathematical Methods in the Social Sciences, 1959* (edited by Kenneth J. Arrow, Samuel Karlin, and Patrick Suppes), Stanford, California: Stanford University Press, 1960, pp. 16–26.

[67] Edgeworth, F. Y., *Mathematical Psychics*. London: C. Kegan Paul and Company, 1881.

[68] Edgeworth, F. Y., "La théorie mathématique de l'offre et de la demande et le coût de production," *Revue d'Economie politique*, 5 (January 1891), 10–28.

[69] Edgeworth, F. Y., "The Theory of International Values, I, II, III," *Economic Journal*, 4 (March, September, December 1894), 35–50, 424–443, 606–638.

[70] Edgeworth, F. Y., "On Some Recent Contributions to the Theory of Statistics," *Journal of the Royal Statistical Society*, 58 (September 1895), 506–515.

[71] Edgeworth, F. Y., "Supplementary Notes on Statistics," *Journal of the Royal Statistical Society*, 59 (September 1896), 529–539.

[72] Edgeworth, F. Y., Review notice concerning Vilfredo Pareto's "Courbe des Revenus," *Economic Journal*, 6 (December 1896), 666.

[73] Edgeworth, F. Y., "La curva delle entrate e la curva di probabilità," *Giornale degli Economisti* [2], 14 (March 1897), 215–218.

[74] Edgeworth, F. Y., "La teoria pura del monopolio," *Giornale degli Economisti* [2], 15 (July, October, November 1897), 13–31, 307–320, 405–414. English translation: "The Pure Theory of Monopoly," [19], pp. 201–227.

[75] Edgeworth, F. Y., "On the Representation of Statistics by Mathematical Formulae," Part I, *Journal of the Royal Statistical Society*, 61 (December 1898), 670–700.

[76] Edgeworth, F. Y., "On the Representation of Statistics by Mathematical Formulae," Part IV, *Journal of the Royal Statistical Society*, 62 (September 1899), 534–555.

[77] Edgeworth, F. Y., "Appreciations of Mathematical Theories, III," *Economic Journal*, 18 (September, December 1908), 392–403, 541–556.

[78] Edgeworth, F. Y., "Probability," *Encyclopaedia Britannica*, 11th edition, Vol. 22, Cambridge: at the University Press, 1911, 376–403.

[79] Edgeworth, F. Y., "Untried Methods of Representing Frequency," *Journal of the Royal Statistical Society*, 87 (July 1924), 571–594.

[80] Edgeworth, F. Y., "Pareto, Vilfredo (1848–1923)" and "Pareto's Law," in *Palgrave's Dictionary of Political Economy* (edited by Henry Higgs), London: Macmillan and Co., Limited, 1926, Vol. III, pp. 710–711, 712–713.

[81] Engel, E., "La consommation comme mesure du bien-être des individus, des familles et des nations," *Bulletin de l'Institut International de Statistique*, 2, (1887), 1re Livraison, 50–75.

[82] Evans, Griffith C., *Mathematical Introduction to Economics*. New York: McGraw-Hill Book Company, Inc., 1930.

[83] Farrell, M. J., "Some Aggregation Problems in Demand Analysis," *Review of Economic Studies*, 21 (No. 3, 1953–54), 193–203.

[84] Fechner, Gustav Theodor, *Elemente der Psychophysik*, in two volumes. Leipzig. Breitkopf und Härtel, 1860. English translation of Volume I: *Elements of Psychophysics*. New York: Holt, Rinehart and Winston, Inc., 1966.

[85] Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edition. New York: John Wiley & Sons, Inc., 1968.

[86] Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edition. New York: John Wiley & Sons, Inc., 1971.

[87] Fisher, Irving, "Mathematical Investigations in the Theory of Value and Prices," *Transactions of the Connecticut Academy of Arts and Sciences*, 9 (July 1892), 1–124.

[88] Fisher, Irving, "Cournot and Mathematical Economics," *Quarterly Journal of Economics*, 12 (January 1898), 119–138, 238–244.

[89] Fisher, R. A., and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Proceedings of the Cambridge Philosophical Society*, 24 (1928), 180–190.

[90] Fréchet, Maurice, "Une nouvelle représentation analytique de la répartition des revenus," *Bulletin de l'Institut International de Statistique*, 22 (1926), 3e Livraison, 547–549.

[91] Fréchet, Maurice, "Sur la loi de probabilité de l'écart maximum," *Annales de la Société polonaise de mathématique* (*Rocznik polskiego tow. matematycznego*), Krakow, 5 (1927), 93–116.

[92] Fréchet, Maurice, "Sur l'hypothèse de l'additivité des erreurs partielles," *Bulletin des Sciences mathematiques* [2], 52 (May 1928), 203–216.

[93] Fréchet, Maurice, "Sur les formules de répartition des revenus," *Revue de l'Institut International de Statistique*, 7 (No. 1, 1939), 32–38.

[94] Fréchet, Maurice, "La formule de Pareto comme instrument statistique: Discussion," *Journal de la Société de Statistique de Paris*, 82 (April 1942), 100–106.

[95] Fréchet, Maurice, "Nouveaux essais d'explication de la répartition des revenus," *Revue de l'Institut International de Statistique*, 13 (No. 1, 1945), 16–32.

[96] Galton, Francis, "The Geometric Mean in Vital and Social Statistics," *Proceedings of the Royal Society of London*, 29 (November 1879), 365–367.

[97] Georgescu-Roegen, Nicholas, "Note on a Proposition of Pareto," *Quarterly Journal of Economics*, 49 (August 1935), 706–714.

[98] Georgescu-Roegen, Nicholas, "The Pure Theory of Consumer's Behavior," *Quarterly Journal of Economics*, 50 (August 1936), 545–593.

[99] Georgescu-Roegen, Nicholas, "A Diagrammatic Analysis of Complementarity," *Southern Economic Journal*, 19 (July 1952), 1–20.

[100] Georgescu-Roegen, Nicholas, "Revisiting Marshall's Constancy of Marginal Utility of Money," *Southern Economic Journal*, 35 (October 1968), 176–181.

[101] Giacardi, Fernando, "Sulla 'curva dei redditi'," *Giornale degli Economisti e Annali di Economia*, N.S., 8 (January–February 1949), 115–121.

[102] Gibrat, R., "Une loi des répartitions économiques: l'effet proportionnel," *Bulletin de la statistique générale de la France*, 19 (July–September 1930), 469–513.

[103] Gibrat, R., *Les inégalités économiques*. Paris: Librairie du Recueil Sirey, 1931. English translation of Chs. V–VII: *International Economic Papers*, No. 7, London: Macmillan and Company Limited, 1957, pp. 53–70.

[104] Gibrat, R., "Une loi des répartitions économiques," *Rivista italiana di Statistica*, 3 (October–December 1931), 332–334.

[105] Gini, Corrado, "Il diverso accrescimento delle classi sociali e la concentrazione della richezza," *Giornale degli Economisti* [2], 38 (January 1909), 27–83.

[106] Gini, Corrado, "Indici di concentrazione e di dipendenza," *Biblioteca dell'Economista*, Series 5, Vol. 50, Turin: Unione Tipografico-Editrice Torinese, 1910, 1920, pp. 1–151.

[107] Gini, Corrado, *Variabilità e Mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche*. Bologna: Tipografia di Paolo Cuppini, 1912.

[108] Gini, Corrado, "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 73 (Part 2, 1913–14), 1203–1248.

[109] Gnedenko, B., "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, 44 (July 1943), 423–453.

[110] Gompertz, Benjamin, "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies," *Philosophical Transactions of the Royal Society of London*, 115 (Part II, 1825), 513–585.

[111] Hagstroem, K.-G., "La loi de Pareto et la réassurance," *Skandinavisk Aktuarietidskrift*, 8 (Häft 1, 1925), 65–88.

[112] Hicks, J. R., "The Foundations of Welfare Economics," *Economic Journal*, 49 (December 1939), 696–712.

[113] Hicks, J. R., *Value and Capital*. Oxford: at the Clarendon Press, 1939.

[114] Hicks, J. R. "The Rehabilitation of Consumers' Surplus," *Review of Economic Studies*, 8 (February 1941), 108–116.

[115] Hicks, J. R., and R. G. D. Allen, "A Reconsideration of the Theory of Value, I, II," *Economica*, N.S., 1 (February, May 1934), 52–75, 196–219.

[116] Hotelling, Harold, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions," *Journal of Political Economy*, 40 (October 1932), 577–616.

[117] Jaffé, William, "New Light on an Old Quarrel: Barone's Unpublished Review of Wicksteed's Essay on the Coordination of the Laws of Distribution and Related Documents," *Cahiers Vilfredo Pareto*, 3 (1964), 61–102.

[118] Jevons, W. Stanley, *The Theory of Political Economy*. London and New York: Macmillan and Co., 1871. 2nd edition, 1879. 3rd edition, 1888.

[119] Johnson, W. E., "The Pure Theory of Utility Curves," *Economic Journal*, 23 (December 1913), 483–513. Reprinted in [19], pp. 97–124.

[120] Juglar, Clément, *Des crises commerciales et de leur retour périodique en France, en Angleterre et aux Etats-Unis*. Paris: Guillaumin et Cie., 1862. 2nd edition, 1889.

[121] Kaldor, Nicholas, "Welfare Propositions in Economics and Interpersonal Comparisons of Utility," *Economic Journal*, 49 (September 1939), 549–552.

[122] Kapteyn, J. C., *Skew Frequency Curves in Biology and Statistics*. Groningen: P. Noordhoff, Publisher, 1903.

[123] Kats, Amoz, “On the Social Welfare Function and the Parameters of the Income Distribution,” *Journal of Economic Theory*, 5 (December 1972), 377–382.

[124] Katzner, Donald W., “A Note on the Constancy of the Marginal Utility of Income,” *International Economic Review*, 8 (February 1967), 128–130.

[125] Kendall, Maurice G., and Alan Stuart, *The Advanced Theory of Statistics*, Vol. I. London: Charles Griffen & Company, Limited, 1958.

[126] Khinchin, A. I., *Analytical Foundations of Physical Statistics*. Delhi: Hindustan Publishing Corpn. (India), 1961.

[127] King, Gregory, *Natural and Political Observations upon the State and Condition of England*. London, 1696.

[128] Koopmans, Tjal1ing C., *Three Essays on the State of Economic Science*. New York: McGraw-Hil1 Book Company, Inc., 1957.

[129] Lange, Oskar, “The Foundations of Welfare Economics,” *Econometrica*, 10 (July–October 1942), 215–228.

[130] Lerner, Abba P., “The Concept of Monopoly and the Measurement of Monopoly Power,” *Review of Economic Studies*, 1 (June 1934), 157–175.

[131] Lerner, Abba P. , “The Symmetry between Import and Export Taxes,” *Economica*, N.S., 3 (August 1936), 306–313.

[132] Lévy, Paul, *Théorie de l’addition des variables aléatoires*. Paris: Gauthiers-Villars, 1937.

[133] Little, I. M. D., *A Critique of Welfare Economics*. Oxford: at the Clarendon Press, 1950.

[134] Lorenz, M. O. , “Methods of Measuring the Concentration of Wealth,” *Quarterly Publications of the American Statistical Association*, N.S., 9 (June 1905), 209–219.

[135] Lydall, Harold, *The Structure of Earnings*. Oxford: at the Clarendon Press, 1968.

[136] McAlister, Donald, “The Law of the Geometric Mean,” *Proceedings of the Royal Society of London*, 29 (November 1879), 367–376.

[137] Mackey, George W., Letter to J. S. Chipman, July 3, 1975. Appendix to this paper.

[138] Makeham, W. M. , “On the Law of Mortality, and the Construction of Annuity Tables,” *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8 (January 1860), 301–310.

[139] Makeham, W. M., "On the Further Development of Gompertz's Law," *Journal of the Institute of Actuaries*, 28 (October 1889, January 1890, April 1890), 152–159, 185–192, 316–332.

[140] Mandelbrot, Benoit, "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, 1 (May 1960), 79–106. "Errata," *ibid.*, 2 (May 1961), 238. "Errata and Deletions," *ibid.*, 4 (January 1963), 114–115.

[141] Mandelbrot, Benoit, "Stable Paretian Random Functions and the Multiplicative Variation of Income," *Econometrica*, 29 (October 1961), 517–543.

[142] Mandelbrot, Berloit, "Paretian Distributions and Income Maximization," *Quarterly Journal of Economics*, 76 (February 1962), 57–85.

[143] Mandelbrot, Benoit, "New Methods in Statistical Economics," *Journal of Political Economy*, 71 (October 1963), 421–440.

[144] March, Lucien, "Quelques exemples de distribution de salaires," *Journal de la Société de Statistique de Paris*, 39 (June, July 1898), 193–206, 241–248.

[145] Marshall, Alfred, *Principles of Economics*. London: Macmillan and Co., Ltd., 1890. 2nd ed., 1891. 5th ed., 1907. 8th ed., 1920.

[146] Menger, Karl, "Das Unsicherheitsmoment in der Wertlehre," *Zeitschrift für Nationalökonomie*, 5 (September 1934), 459–485. English translation: "The Role of Uncertainty in Economics," in *Essays in Mathematical Economics in Honor of Oskar Morgenstern* (edited by Martin Shubik). Princeton, N. J.: Princeton University Press, 1967, pp. 211–231.

[147] Neumann, John von, and Oskar Morgenstern, *Theory of Games and Economic Behavior*, 2nd edition. Princeton: Princeton University Press, 1947.

[148] Newbery, David, "A Theorem on the Measurement of Inequality," *Journal of Economic Theory*, 2 (September 1970), 264–266.

[149] Pareto, Vilfredo, "Le protectionisme en Italie, ses résultats," Paris: Publications de l'Union pour la franchise des matières premières, 1891. Amplified Italian version: "Il protezionismo in Italia. Suoi effeti," *Rassegna di scienze sociale e politiche* (September 15, 1891), 68–83.

[150] Pareto, Vilfredo, "L'Italie économique," *Revue des Deux Mondes* (October 15, 1891), 904–944.

[151] Pareto, Vilfredo, "Di un errore del Coumot nel trattare l'economia politica colla matematica," *Giornali degli Economisti* [2], 4 (January 1892), 1–14.

[152] Pareto, Vi1fredo, "Dell'utile che procurano al paese le ferrovie," *Giornali degli Economisti* [2], 4 (February 1892), 154–162.

[153] Pareto, Vilfredo, "La teoria dei prezzi dei Signori Auspitz e Lieben e le osservazioni del professore Walras," *Giornale degli Economisti* [2], 4 (March 1892), 201–239. "Ancora della *Theorie des Preises* dei Signori Auspitz e Lieben," *ibid.*, 5 (August 1892), 168.

[154] Pareto, Vilfredo, "Considerazioni sui principi fondamentali dell' economia politica pura," *Giornale degli Economisti* [2], 4 (May 1892), 389–420; 4 (June 1892), 485–512; 5 (August 1892), 119–157; 6 (January 1893), 1–37; 7 (October 1893), 279–321.

[155] Pareto, Vilfredo, "La mortalità infantile ed il costo dell'uomo adulto," *Giornale degli Economisti* [2], 7 (November 1893), 451–456.

[156] Pareto, Vilfredo, "Teoria matematica dei cambi forestieri," *Giornale degli Economisti* [2], 8 (February 1894), 142–173.

[157] Pareto, Vilfredo, "Il massimo di utilità dato dalla libera concorrenza," *Giornale degli Economisti* [2], 9 (July 1894), 48–66.

[158] Pareto, Vilfredo, "La legge della domanda," *Giornale degli Economisti* [2], 10 (January 1895), 59–68.

[159] Pareto, Vilfredo, "Teoria matematica del commercio internazionale," *Giornale degli Economisti* [2], 10 (April 1895), 476–498.

[160] Pareto, Vilfredo, "Theorie mathématique des changes étrangers," *Association Française pour l'Avancement des Sciences*, Compte Rendu de la 23e session, Caen, 1894, Seconde partie, Notes et Mémoires. Paris: au Secrétariat de l'Association, 1895, pp. 1180–1191.

[161] Pareto, Vilfredo, "Il modo di figurare i fenomeni economici," *Giornale degli Economisti* [2], 12 (January 1896), 75–87.

[162] Pareto, Vilfredo, "La courbe de la répartition de la richesse," Université de Lausanne, recueil publié par la Faculté de Droit à l'occasion de l'Exposition nationale suisse, Genève 1896. Lausanne: Ch. Viret-Genton Impr., 1896, pp. 373–387.

[163] Pareto, Vilfredo, *Cours d'Economie politique*. Vol. I, Lausanne: F. Rouge, éditeur, 1896. Vol. II, Lausanne: F. Rouge, libraire-editeur, 1897.

[164] Pareto, Vilfredo, "La courbe des revenus," *Le Monde Economique*, 6 (July 25, 1896), 99–100.

[165] Pareto, Vilfredo, "La curva delle entrate e le osservazioni del prof. Edgeworth," *Giornale degli Economisti* [2], 13 (November 1896), 439–448.

[166] Pareto, Vilfredo, "Les revenus des contribuables," *Gazette de Lausanne*, December 9, 1896, and December 21, 1896.

[167] Pareto, Vilfredo, "Aggiunta allo studio sulla curva delle entrate," *Giornale degli Economisti* [2], 14 (January 1897), 15–26.

[168] Pareto, Vilfredo, "Ultima risposta al prof. Edgeworth," *Giornale degli Economisti* [2], 14 (March 1897), 219–220.

[169] Pareto, Vilfredo, "La répartition des revenus," *Le Monde Economique*, 7 (August 28, 1897), 259–261.

[170] Pareto, Vilfredo, "The New Theories of Economics," *Journal of Political Economy*, 5 (September 1897), 485–502.

[171] Pareto, Vilfredo, "Quelques exemples d'application des méthodes d'interpolation à la statistique," *Journal de la Société de Statistique de Paris*, 38 (November 1897), 367–379.

[172] Pareto, Vilfredo, "Comment se pose le problème de l'économie pure," Mémoire présenté en décembre 1898, à la Société Stella. Lausanne, 1898 (privately published), 12 pp.

[173] Pareto, Vilfredo, "Tables pour faciliter l'application de la méthode des moindres carrés," *Zeitschrift für Schweizerisch Statistik*, 35 (I. Band, 1899), 121–150.

[174] Pareto, Vilfredo, "Sunto di alcuni capitoli di un nuovo trattato di economia pura del prof. Pareto," *Giornale degli Economisti* [2], 20 (March, June 1900), 216–235, 511–549.

[175] Pareto, Vilfredo, "Distribution de la richesse d'après le rôle de la taille, à Paris, en 1292," *Journal de la Société de Statistique de Paris*, 41 (July 1900), 224–225.

[176] Pareto, Vilfredo, "Un'applicazione di teorie sociologiche," *Rivista italiana di Sociologia*, 4 (July 1900), 401–456. English translation: *The Rise and Fall of the Elites. An Application of Theoretical Sociology* (edited by Hans L. Zetterberg). Totowa, New Jersey: The Bedminster Press, 1968.

[177] Pareto, Vilfredo, "Le equazione dell'equilibrio dinamico," *Giornale degli Economisti* [2], 23 (September 1901), 253–259.

[178] Pareto, Vilfredo, "Anwendungen der Mathematik auf Nationalökonomie," *Encyklopedie der Mathematischen Wissenschaften mit einschluss ihrer Andwendungen*, 1 (September 1902), 1094–1120. Italian translation: "Applicazioni della matematica all'economia politica," *Giornale degli Economisti* [2], 33 (November 1906), 424–453.

[179] Pareto, Vilfredo, *Les systèmes socialistes*, in two volumes. Paris: V. Giard & E. Brière, 1902, 1903.

[180] Pareto, Vilfredo, "Di un nuovo errore nello interpretare le teorie dell'economia matematica," *Giornale degli Economisti* [2], 25 (November 1902), 401–433.

[181] Pareto, Vilfredo, "A proposito del massimo di ofelimità," *Giornale degli Economisti* [2], 26 (February 1903), 177–178.

[182] Pareto, Vilfredo, *Manuale di economia, con una introduzione alla scienza sociale*. Milan: Società Editrice Libraria, 1906.

[183] Pareto, Vilfredo, "L'ofelimità nei cicli non chunisi," *Giornale degli Economisti* [2], 33 (July 1906), 15–30. English translation: "Ophelimity in Non-Closed Cycles," Chapter 18 in [51], pp. 370–385.

[184] Pareto, Vilfredo, "L'interpolazione per la ricerca delle leggi economiche," *Giornale degli Economisti* [2], 34 (May 1907), 366–385; 36 (June 1908), 423–453.

[185] Pareto, Vilfredo, *Manuel d'Economie politique*. Paris: V. Giard et E. Brière, 1909.

[186] Pareto, Vilfredo, "Nouvelle méthode d'interpolation pour les phénomènes donnés par l'experience," *VI^{e} Congrès International de Psychologie*, tenu à Genève du 2 au 7 août 1909, *Rapports et Comptes Rendus*. Geneva: Librairie Kundig, 1910, pp. 552–559.

[187] Pareto, Vilfredo, "Le azione non logiche," *Rivista italiana di Sociologia*, 14 (May–August 1910), 305–364.

[188] Pareto, Vilfredo, "L'économie mathématique," *Encyclopédie des Sciences Mathématiques*, Tome I, vol. 4, fascicule 4 (August 1911), 591–640. English translation: "Mathematical Economics," in [19], pp. 52–95.

[189] Pareto, Vilfredo, "Il massimo di utilità per una collettività in Sociologia," *Giornale degli Economisti e Rivista di Statistica* [3], 46 (April 1913), 337–341.

[190] Pareto, Vilfredo, "Alcuni relazioni fra lo stato sociale e le variazioni della prosperità economica," *Rivista italiana di Sociologia*, 17 (September–December 1913), 501–518.

[191] Pareto, Vilfredo, *Trattato di Sociologia generale*, in two volumes. Florence: Barbera, 1916. English translation: *The Mind and Society*, in four volumes. New York: Harcourt, Brace and Company, 1935.

[192] Pareto, Vilfredo, "Forme di fenomeni economiche e previsione," *Rivista di scienze bancaria*, 5 (August–September 1917), 73–84.

[193] Pareto, Vilfredo, "L'aggio e il cambio," in *In Onore di Tullio Martello, Scritti Varii.* Bari: Gius. Laterza & Figli, 1917, pp. 223–242.

[194] Pareto, Vilfredo, "Economia sperimentale," *Giornale degli Economisti e Rivista di Statistica* [3], 57 (July 1918), 1–18.

[195] Pareto, Vilfredo, *Lettere a Maffeo Pantaleoni* (edited by Gabriele de Rosa), in three volumes. Rome: Banca Nazionale del Lavoro, 1960.

[196] Pearson, Karl, "Contributions to the Mathematical Theory of Evolution. — II. Skew Variation in Homogeneous Material," *Transactions of the Royal Society of London* [A], 186 (Part I, 1895), 343–415.

[197] Pen, Jan, *Income Distribution* (translated from the Dutch by Trevor S. Preston). New York and Washington: Praeger Publishers, 1971.

[198] Pessoa De Araujo, Aloisio, "On Aggregation of Demand Correspondences," mimeographed, University of California, Berkeley, 1974. Published as "Notes on the Smoothing of Aggregate Demand" by Aloisio Araujo and Andreu Mas-Colell, *Journal of Mathematical Economics*, 5 (September 1978), 113–127.

[199] Pietra, Gaetano, "Delle relazioni tra gli indici di variabilità," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 74 (Part 2, 1914–15), 775–792, 793–804.

[200] Pietra, Gaetano, "Intorno alIa discordanza fra gli indici di variabilità e di concentrazione," *Bulletin de l'Institut International de Statistique*, 28 (1935), 2e Livraison, 171–191.

[201] Pigou, A. C., *Wealth and Welfare*. London: Macmillan and Co., Ltd., 1912.

[202] Pigou, A. C., *The Economics of Welfare*, 4th edition. London: Macmillan and Co., Ltd., 1932.

[203] Rhodes, E. C., "The Pareto Distribution of Incomes," *Economica*, N.S., 11 (February 1944), 1–11.

[204] Rothschild, Michael, and Joseph E. Stiglitz, "Some Further Results on the Measurement of Inequality," *Journal of Economic Theory*, 6 (April 1973), 188–204.

[205] Roy, A. D. , "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, N.S., 3 (June 1951), 135–146.

[206] Roy, René, "La demande dans ses rapports avec la répartition des revenues," *Metron*, 8 (February 1930), 101–153.

[207] Samuelson, Paul A. , "Constancy of the Marginal Utility of Income," in *Studies in Mathematical Economics and Econometrics. In Memory of Henry Schultz* (edited by Oscar Lange, Francis McIntyre, and Theodore O. Yntema), Chicago: The University of Chicago Press, 1942, pp. 75–91.

[208] Samuelson, Paul A., *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press, 1947.

[209] Samuelson, Paul A. , "Evaluation of Real National Income," *Oxford Economic Papers*, N.S., 1 (January 1950), 1–29.

[210] Samuelson, Paul A., "The Problem of Integrability in Utility Theory," *Economica*, N.S., 17 (November 1950), 355–385.

[211] Samuelson, Paul A. , "Social Indifference Curves," *Quarterly Journal of Economics*, 70 (February 1956), 1–22.

[212] Samuelson, Paul A., "A Fallacy in the Interpretation of Pareto's Law of Alleged Constancy of Income Distribution," *Rivista Internazionale di Scienze economiche e commerciali*, 12 (April 1965), 246–250.

[213] Samuelson, Paul A., "Complementarity — An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory," *Journal of Economic Literature*, 12 (December 1974), 1255–1289.

[214] Schneider, Erich, "Vilfredo Pareto. The Economist in the Light of his Letters to Maffeo Pantaleoni," *Banca Nazionale del Lavoro Quarterly Review*, 14 (September 1961), 247–295.

[215] Schultz, Henry, "Interrelations of Demand," *Journal of Political Economy*, 41 (August 1933), 468–512. Reproduced in revised form as Chapter XVIII of Henry Schultz, *The Theory of Measurement of Demand*. Chicago: The University of Chicago Press, 1938.

[216] Schumpeter, Joseph A., *Business Cycles*, in two volumes. New York: McGraw-Hill Book Co. Inc., 1939.

[217] Scorza, Gaetano, "Osservazioni sulla teoria del baratto secondo il Prof. Walras," *Giornale degli Economisti* [2], 24 (April 1902), 282–302.

[218] Scorza, Gaetano, Review of G. Cassel, "Grundriss einer elementaren Preislehre," *Giornale degli Economisti* [2], 25 (August 1902), 191–194.

[219] Scorza, Gaetano, "Osservazioni su alcuni teorie di economia pura," *Giornale degli Economisti* [2], 25 (December 1902), 503–516.

[220] Scorza, Gaetano, "A proposito del massimo di ofelimità dato dalla libera concorrenza," *Giornale degli Economisti* [2], 26 (January 1903), 41–62.

[221] Sen, Amartya, *On Economic Inequality*. Oxford: Clarendon Press, 1973.

[222] Sheshinski, Eytan, "Relation between a Social Welfare Function and the Gini Index of Income Inequality," *Journal of Economic Theory*, 4 (February 1972), 98–100.

[223] Slutsky, Eugenio, "Sulla teoria del bilancio del consumatore," *Giornale degli Economisti e Rivista di Statistica* [3], 51 (July 1915), 1–26. English translation: "On the Theory of the Budget of the Consumer," in *Readings in Price Theory* (edited by G. J. Stigler and K. J. Boulding). Homewood, Ill.: Richard D. Irwin, Inc., 1952, pp. 27–56.

[224] Smale, S., "Global Analysis and Economics, III. Pareto Optima and Price Equilibria," *Journal of Mathematical Economics*, 1 (August 1974), 107–117.

[225] Sondermann, Dieter, "Smoothing Demand by Aggregation," *Journal of Mathematical Economics*, 2 (June–September 1975), 201–223.

[226] Sorel, G., "La loi des revenus," *Le Devenir Social*, 3 (July 1897), 577–607.

[227] Spencer, Herbert, "The Militant Type of Society," *Contemporary Review*, 40 (September 1881), 337–360.

[228] Spencer, Herbert, "The Industrial Type of Society," *Contemporary Review*, 40 (October 1881), 507–533.

[229] Stigler, George J., "The Development of Utility Theory," *Journal of Political Economy*, 58 (August, October 1950), 307–327, 373–396.

[230] Strotz, Robert H. , "The Empirical Implications of a Utility Tree," *Econometrica*, 25 (April 1957), 269–280.

[231] Tarascio, Vincent J., "The Pareto Law of Income Distribution," *Social Science Quarterly*, 54 (December 1973), 525–533.

[232] Tinbergen, Jan, "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77 (1956), 155–175.

[233] Van der Wijk, J., "De Gemiddeldenwet," *De Economist*, 77 (1928), 565–590.

[234] Van der Wijk, J., "Psychisch inkomen en psychisch vermogen," *De Economist*, 80 (1931), 112–140, 232–245, 305–341.

[235] Van der Wijk, J., *Inkomens- en vermogensverdeling*. Haarlem: De Erven F. Bohn N. V., 1939.

[236] Vinci, Felice, "Nuovi contributi allo studio della distribuzione dei redditi," *Giornale degli Economisti e Rivista di Statistica* [3], 61 (November 1921), 365–369.

[237] Vinci, Felice, "Calcolo delle probabilità e distribuzione dei redditi nel pensiero di Pareto," *Giornale degli Economisti e Rivista di Statistica* [4], 64 (January–February 1924), 127–129.

[238] Viner, Jacob, *Studies in the Theory of International Trade*. New York: Harper & Brothers, Publishers, 1937.

[239] Volterra, Vito, "L'economia matematica ed il nuovo manuale del prof. Pareto," *Giornale degli Economisti* [2], 32 (April 1906), 296–301. English translation: "Mathematical Economics and Professor Pareto's New Manual," Chapter 17 in [51], pp. 365–369.

[240] Wald, Abraham, "Über die eindeutige positive Lösbarkeit der neuen Produktionsgleichungen," *Ergebnisse eines mathematischen Kolloquiums*, Heft 6 (1933–1934), 12–20. English translation: "On the Unique Non-Negative Solvability of the New Production Equations," in [19], pp. 281–288.

[241] Wald, Abraham, "Über die Produktionsgleichungen der ökonomischen Wertlehre," *Ergebnisse eines mathematischen Kolloquiums*, Heft 7 (1934–1935), 1–6. English translation: "On the Production Equations of Economic Value Theory," in [19], pp. 289–293.

[242] Walras, Léon, *Eléments d'économie politique pure, ou Théorie de la richesse sociale*, 2nd edition. Lausanne: F. Rouge, Editeur, 1889. 5th edition (édition définitive), 1926. English translation (by William Jaffé) of the 5th edition: *Elements of Pure Economics or The Theory of Social Wealth.* London: George Allen and Unwin Ltd., 1954.

[243] Weinberger, H. F., "Conditions for a Local Pareto Optimum," Department of Mathematics, University of Minnesota, May 1975.

[244] Wicksell, Knut, *Über Wert, Kapital und Rente*. Jena: Verlag von Gustav Fischer, 1893. English translation: *Value, Capital and Rent*. London: George Allen and Unwin Ltd., 1954.

[245] Wicksell, Knut, "Vilfredo Paretos Manuel d'économie politique," *Zeitschrift für Volkswirtschaft, Sozialpolitik und Verwaltung*, 22 (1913), 132–151. English translation: "Vilfredo Pareto's *Manuel d'économie politique*," in Knut Wicksel1, *Selected Papers on Economic Theory*. London: George Allen and Unwin Ltd., 1958, pp. 159–175.

[246] Wicksteed, Philip H., *The Co-ordination of the Laws of Distribution*. London: Macmillan & Co., 1894.

[247] Wilson, Edwin B., *Advanced Calculus*. New York: Ginn and Company, 1911.

[248] Wilson, Edwin B., Review of Vilfredo Pareto's *Manuel d'Economie politique*, *Bulletin of the American Mathematical Society*, 18 (June 1912), 462–474.

[249] Wilson, Edwin B., "First and Second Laws of Error," *Journal of the American Statistical Association*, 18 (September 1923), 841–851.

[250] Wilson, Edwin B., " Generalization of Pareto's Demand Theorem," *Quarterly Journal of Economics*, 49 (August 1935), 715–717.

[251] Wilson, Edwin B., "Pareto versus Marshall," *Quarterly Journal of Economics*, 53 (August 1939), 645–650.

[252] Wilson, Edwin B., "Pareto on Marshall's Demand Curve," *Quarterly Journal of Economics*, 58 (November 1943), 141–145.

[253] Wilson, Edwin B., "Notes on Utility Theory and Demand Equations," *Quarterly Journal of Economics*, 60 (May 1946), 453–460.

[254] Winkler, Wilhelm, "Einkommen," *Handwörterbuch der Staatswissenschaften*, 4th edition, Vol. III, Jena: Verlag von Gustav Fischer, 1926, pp. 367–400.

[255] Wold, Herman, "A Synthesis of Pure Demand Analysis," *Skandinavisk Aktuarietidskrift*, 26 (1943), 85–118, 220–263; 27 (1944), 69–120.

[256] Wold, H. O. A., and P. Whittle, "A Model Explaining the Pareto Distribution of Wealth," *Econometrica*, 25 (October 1957), 591–595.

Chapter 7

APPENDIX

APPENDIX
Some Characterizations of the Pareto Distribution
By George W. Mackey[1]
Harvard University

I made my observation about the Pareto distribution six or seven years ago and found it new to many economists. However last summer in the Toronto airport I discovered a Pelican book entitled "Income Distribution" by a Dutchman named Jan Pen. On page 238 he makes my observation and attributes it to one J. van der Wijk who published it in Dutch in 1939. Thus I was anticipated by about thirty years. My only consolation is that Pen refers to it as "a great step forward" — a rather considerable overstatement in my opinion.

The observation is a rather trivial one but illuminating all the same. Here it is. For each income level y let M_y denote the mean of all incomes x with $x \geq y$. Then M_y/y is a number independent of the units employed and in a sense measures how envious people with income y should be of their economic superiors. One naturally wonders how M_y/y varies with y and my observation is that it is a constant if and only if the distribution has the Pareto form. The proof which I leave to you is a simple exercise in elementary differential equations. The constant value of M_y/y is related to the exponent in Pareto's law by a linear fractional transformation. In Pareto's data M_y/y was around three for the countries of Western Europe. In the late nineteen sixties it was a bit less than 2 and in the U.S. at least had been more or less constant since the forties. Of course in giving M_y/y a value I am assuming that the Pareto law holds which in fact it does more or less for incomes greater than or equal to twice the mean.

It is interesting to compute how things would differ if the Pareto law in the U.S. held right on down to a minimum income which would necessarily be one half the mean income (if $M_y/y \equiv 2$). I made this computation and discovered

[1]Excerpts from a letter to J. S. Chipman from Professor George W. Mackey, Department of Mathematics, Harvard University, July 3, 1975.

that the bottom 20% would get twice what they do now, the *second* 20% (from the top) would get a lot less and the other three quintiles would get more or less what they do now. The second 20% are too near the middle to get much more than the mean in any reasonable distribution scheme — but they do.

The Pareto distribution strikes me as an admirable one for several reasons.
(1) It is very easy to compute with. Once one knows the constant value f of M_y/y and the mean income $\bar{M}$ the entire distribution is known and it is easy to devise exact formulas for such things as the median, the income of the upper α% for all α etc., etc. Try it. All sorts of interesting and informative investigations can be carried out.
(2) Of all income distributions which do not give the same income to everyone the Pareto distributions are the "most equal" in that everyone has the same reason to be envious of those above him. Looking upward everyone sees the same thing. Moreover by making the constant value of M_y/y sufficiently close to one we can approach absolute equality as closely as we please. Thus the real issue is "What value of M_y/y is "best."

A few years later I discovered another and more "local" characterization of the Pareto distribution. Let N_y be the number of people with incomes $\geq y$. Then $(N_y/y)/\dfrac{dN_y}{dy}$ measures *incentive* in terms of the fractional increase in salary achieved by passing any given fraction of those above you. To say that incentive is the same at all levels is again equivalent to saying that the distribution is Pareto.

Still another characterization is the following. Let us say that we have solved the problem of defining "merit" and arrange everyone in order of merit. Suppose we decide that the ratio of the incomes of any two individuals shall depend only on the ratios of their positions in the merit order. We measure position by the percentage above them, but do not specify the functional relationship. It can be proved that the Pareto distribution will result.