CHAPTER ONE

Introduction

Certum est quia impossibile est. TERTULLIAN, 155/160 A.D.—after 220 A.D.

This book is devoted to the parametric statistical distributions of economic size phenomena of various types—a subject that has been explored in both statistical and economic literature for over 100 years since the publication of V. Pareto's famous breakthrough volume Cours d'économie politique *in 1897. To the best of our knowledge, this is the first collection that systematically investigates various parametric models—a more respectful term for distributions—dealing with income, wealth, and related notions. Our aim is marshaling and knitting together the immense body of information scattered in diverse sources in at least eight languages. We present empirical studies from all continents, spanning a period of more than 100 years.*

We realize that a useful book on this subject matter should be interesting, a task that appears to be, in T. S. Eliot's words, "not one of the least difficult." We have tried to avoid reducing our exposition to a box of disconnected facts or to an information storage or retrieval system. We also tried to avoid easy armchair research that involves computerized records and heavy reliance on the Web.

Unfortunately, the introduction by its very nature is always somewhat fragmentary since it surveys, in our case rather extensively, the content of the volume. After reading this introduction, the reader could decide whether continuing further study of the book is worthwhile for his or her purposes. It is our hope that the decision will be positive. To provide a better panorama, we have included in the Appendix brief biographies of the leading players.

1.1 OUR AIMS

The modeling of economic size distributions originated over 100 years ago with the work of Vilfredo Pareto on the distribution of income. He apparently was the first to

Statistical Size Distributions in Economics and Actuarial Sciences, By Christian Kleiber and Samuel Kotz.
ISBN 0-471-15064-9 © 2003 John Wiley & Sons, Inc.

observe that, for many populations, a plot of the logarithm of the number of incomes N_x above a level x against the logarithm of x yields points close to a straight line of slope $-\alpha$ for some $\alpha > 0$. This suggests a distribution with a survival function proportional to $x^{-\alpha}$, nowadays known as the Pareto distribution.

"Economic size distributions" comprise the distributions of personal incomes of various types, the distribution of wealth, and the distribution of firm sizes. We also include work on the distribution of actuarial losses for which similar models have been in use at least since Scandinavian actuaries (Meidell, 1912; Hagstrœm, 1925) observed that—initially in life insurance—the sum insured is likely to be proportional to the incomes of the policy holders, although subsequently there appears to have been hardly any coordination between the two areas. Since the lion's share of the available literature comprises work on the distribution of income, we shall often speak of income distributions, although most results apply with minor modifications to the other size variables mentioned above.

Zipf (1949) in his monograph *Human Behavior and the Principle of Least Effort* and Simon (1955) in his article "On a class of skew distribution functions" suggest that Pareto-type distributions are appropriate to model such different variables as city sizes, geological site sizes, the number of scientific publications by a certain author, and also the word frequencies in a given text. Since the early 1990s, there has been an explosion of work on economic size phenomena in the physics literature, leading to an emerging new field called *econophysics* (e.g., Takayasu, 2002). In addition, computer scientists are nowadays studying file size distributions in the World Wide Web (e.g., Crovella, Taqqu, and Bestavros, 1998), but these works are not covered in this volume. We also exclude discrete Pareto-type distributions such as the Yule distribution that have been utilized in connection with the size distribution of firms by Simon and his co-authors (see Ijiri and Simon, 1977).

Regarding the distribution of income, the twentieth century witnessed unprecedented attempts by powerful nations such as Russia (in 1917) and China (in 1949) and almost all Eastern European countries (around the same time) to carry out far-reaching economic reforms and establish economic regimes that will reduce drastically income inequality and result in something approaching the single-point distribution of income when everyone is paid the same wages.

The most radical example is, of course, the blueprint for the economy of the Peoples' Republic of China (PRC) proclaimed by Mao Tse Tung on October 1, 1949 (his delivery of this plan was witnessed by one of the authors of this book in his youth). Mao's daring and possibly utopian promise of total economic equality for close to 1 billion Chinese and a guaranteed "iron bowl" of rice for every citizen totally receded into the background over the next 30 years due partially to blunders, unfavorable weather conditions, fanaticism, and cruelty, but perhaps mainly because of—as claimed by Pareto in the 1890s— the inability to change human nature and to suppress the natural instinct for economic betterment each human seems to possess.

It is remarkable that only eight years after Mao's death, the following appeared in the Declaration of the Central Committee of the Communist Party of China in October 1984:

> There has long been a misunderstanding about the distribution of consumer goods under socialism as if it meant egalitarianism.
>
> If some members of society got higher wages through their labor, resulting in a wide gap in income it was considered polarization and a deviation from socialism. This egalitarian thinking is utterly incompatible with scientific Marxist views of socialism.

In modern terminology, this translates to "wealth creation seems to be more important than wealth redistribution." Even the rigid Stalinist regime in North Korea began flirting with capitalism after May 2002, triggering income inequality.

Much milder attempts at socialism (practiced, e.g., in Scandinavian countries in the early years of the second part of the twentieth century) to reduce inequality by government regulators, especially by substantial taxation on the rich, were a colossal failure, as we are witnessing now in the early years of the twenty-first century. Almost the entire world is fully entrenched in a capitalistic market economy that appears to lead to a mathematical expression of the income distribution close to the one discovered by Pareto, with possibly different α and insignificant modifications. In fact, in our opinion the bulk of this book is devoted to an analysis of Pareto-type distributions, some of them in a heavily disguised form, leading sometimes to unrecognizable mathematical expressions.

It was therefore encouraging for us to read a recent book review of Champernowne and Cowell's *Economic Inequality and Income Distribution* (1998) written by Thomas Piketty, in which Piketty defends the "old-fashionedness" of the authors in their frequent reference to Pareto coefficients and claims that due to the tremendous advances in computer calculations "at the age of SAS and STATA," young economists have never heard of "Pareto coefficients" and tend to assume "that serious research started in the 1980s or 1990s." We will attempt to provide the background on and hopefully a proper perspective of the area of parametric income distributions throughout its 100-year-plus history.

It should be admitted that research on income distribution was somewhat dormant during the period from 1910–1970 in Western countries, although periodically publications—mainly of a polemical nature—have appeared in basic statistical and economic journals (see the bibliography). (An exception is Italy where, possibly due to the influence of Pareto and Gini, the distribution of income has always been a favorite research topic among prominent Italian economists and statisticians.) This changed during the last 15 years with the rising inequality in Western economies over the 1980s and a surge in inequality in the transition economies of Eastern Europe in the 1990s. Both called for an explanation and prompted novel empirical research. Indeed, as indicated on a recent Web page of the Distributional Analysis Research Programme (DARP) at the London School of Economics (http://darp.lse.ac.uk/),

> the study of income distribution is enjoying an extraordinary renaissance: interest in the history of the eighties, the recent development of theoretical models of economic growth that persistent wealth inequalities, and the contemporary policy focus upon the concept of social exclusion are evidence of new found concern with distributional issues.

Readers are referred to the recently published 1,000-page *Handbook of Income Distribution* edited by Atkinson and Bourguignon (2000) for a comprehensive discussion of the economic aspects of income distributions. We shall concentrate on statistical issues here.

On the statistical side, methods can broadly be classified as parametric and nonparametric. The availability of ever more powerful computer resources during recent decades gave rise to various nonparametric methods of density estimation, the most popular being probably the kernel density approach. Their main attraction is that they do not impose any distributional assumptions; however, with small data sets—not uncommon in actuarial science—they might result in imprecise estimates. These inaccuracies may be reduced by applying parametric models.

A recent comment by Cowell (2000, p. 145) seems to capture lucidly and succinctly the controversy existing between the proponents and opponents to the parametric approach in the analysis of size distributions:

> The use of the parametric approach to distributional analysis runs counter to the general trend towards the pursuit of non-parametric methods, although [it] is extensively applied in the statistical literature. Perhaps it is because some versions of the parametric approach have had bad press: Pareto's seminal works led to some fanciful interpretations of "laws" of income distribution (Davis, 1941), perhaps it is because the non-parametric method seems to be more general in its approach.
>
> Nevertheless, a parametric approach can be particularly useful for estimation of indices or other statistics in cases where information is sparse [such as given in the form of grouped data, *our addition*].... Furthermore, some standard functional forms claim attention, not only for their suitability in modelling some features of many empirical income distributions but also because of their role as equilibrium distributions in economic processes.

We are not concerned with economic/empirical issues in this book that involve the choice of a type of data such as labor of nonlabor earnings, incomes before or after taxes, individual or household incomes. These are, of course, of great importance in empirical economic studies. Nor are we dealing with the equally important aspects of data quality; we refer interested readers to van Praag, Hagenaars, and van Eck (1983); Lillard, Smith, and Welch (1986); or Angle (1994) in this connection. This problem is becoming more prominent as more data become available and new techniques to cope with the incompleteness of data such as "top-coding" and outliers are receiving significant attention. In the latter part of the twentieth century, the works of Victoria-Feser and Ronchetti (1994, 1997), Cowell and Victoria-Feser (1996), and more recently Victoria-Feser (2000), provided a number of new tools for the application of parametric models of income distributions, among them robust estimators and related diagnostic tools. They can protect the researcher against model deviations such as gross errors in the data or grouping effects and therefore allow for more reliable estimation of, for example, income distributions and inequality indices. (The latter task has occupied numerous researchers for over half a century.)

1.2 TYPES OF ECONOMIC SIZE DISTRIBUTIONS

In this short section we shall enumerate for completeness the types of size distributions studied in this book. Readers who are interested solely in statistical aspects may wish to skip this section. Those inclined toward broader economic-statistical issues may wish to supplement our brief exposition by referring to numerous books and sources, such as Atkinson and Harrison (1978), Champernowne and Cowell (1998), Ijiri and Simon (1977), Sen (1997), or Wolff (1987) and books on actuarial economics and statistics.

Distributions of Income and Wealth

As Okun (1975, p. 65) put it, "income and wealth are the two box scores in the record book of people's economic positions." It is undoubtedly true that the size distribution of income is of vital interest to all (market) economies with respect to social and economic policy-making. In economic and social statistics, the size distribution of income is the basis of concentration and Lorenz curves and thus at the heart of the measurement of inequality and more general social welfare evaluations. From here, it takes only a few steps to grasp its importance for further economic issues such as the development of adequate taxation schemes or the evaluation of effectiveness of tax reforms. Income distribution also affects market demand and its elasticity, and consequently the behavior of firms and a fortiori market equilibrium. It is often mentioned that income distribution is an important factor in determining the amount of saving in a society; it is also a factor influencing the productive effort made by various groups in the society.

Distributions of Firm Sizes

Knowledge of the size distribution of firms is important to economists studying industrial organization, to government regulators, as well as to courts. For example, courts use firm and industry measures of market share in a variety of antitrust cases. Under the merger guidelines of the U.S. Department of Justice and the Federal Trade Commission, whether mergers are challenged depends on the relative sizes of the firms involved and the degree of concentration in the industry. In recent years, for example, the Department of Justice challenged mergers in railroads, banks, soft drink, and airline industries using data on concentration and relative firm size.

As of 2002 tremendous upheavals in corporate institutions that involve great firms are taking place throughout the world especially in the United States and Germany. This will no doubt result in drastic changes in the near future in the size distribution of firms, and the recent frequent mergers and occasional breakdowns of firms may even require a new methodology. We will not address these aspects, but it is safe to predict new theoretical and empirical research along these lines.

Distributions of Actuarial Losses

Coincidentally, the unprecedented forest fires that recently occurred in the western United States (especially in Colorado and Arizona) may challenge the conventional wisdom that "fire liabilities are rare." The model of the total amount of losses in a given period presented below may undergo substantial changes: In particular, the existing probability distributions of an individual loss amount $F(x)$ will no doubt be reexamined and reevaluated.

In actuarial sciences, the total amount of losses in a given period is usually modeled as a risk process characterized by two (independent) random variables: the number of losses and the amount of individual losses. If

- $p_n(t)$ is the probability of exactly n losses in the observed period $[0, t]$,
- $F(x)$ is the probability that, given a loss, its amount is $\leq x$,
- $F^{*n}(x)$ is the nth convolution of the c.d.f. of loss amount $F(x)$,

then the probability that the total loss in a period of length t is $\leq x$ can be expressed as the compound distribution

$$G(x, t) = \sum_{n=0}^{\infty} p_n(t) F^{*n}(x).$$

Although the total loss distribution $G(x, t)$ is of great importance for insurers in their task of determining appropriate premiums or reinsurance policies, it is the probability distribution of an *individual loss amount*, $F(x)$, that is relevant when a property owner has to decide whether to purchase insurance or when an insurer designs deductible schedules. Here we are solely concerned with the distributions of individual losses.

1.3 BRIEF HISTORY OF THE MODELS FOR STUDYING ECONOMIC SIZE DISTRIBUTIONS

A statistical study of personal income distributions originated with Pareto's formulation of "laws" of income distribution in his famous *Cours d'économie politique* (1897) that is discussed in detail in this book and in Arnold's (1983) book *Pareto Distributions*.

Pareto was well aware of the imperfections of statistical data, insufficient reliability of the sources, and lack of veracity of income tax statements. Nonetheless, he boldly analyzed the data using his extensive engineering and mathematical training and succeeded in showing that there is a relation between N_x—the number of taxpayers with personal income greater or equal to x—and the value of the income x given by a downward sloping line

$$\log N_x = \log A - \alpha \log x \tag{1.1}$$

or equivalently,

$$N_x = \frac{A}{x^{\alpha}}, \quad A > 0,\ \alpha > 0,\ x > x_0, \tag{1.2}$$

x_0 being the minimum income (Pareto, 1895). Economists and economic statisticians (e.g., Brambilla, 1960; Dagum, 1977) often refer to α (or rather $-\alpha$) as the elasticity of the survival function with respect to income x

$$\frac{d\log\{1-F(x)\}}{d\log x} = -\alpha.$$

Thus, α is the elasticity of a reduction in the number of income-receiving units when moving to a higher income class. The graph with coordinates $(\log x,\ \log N_x)$ is often referred to as the *Pareto diagram*. An exact straight line in this display defines the Pareto distribution.

Pareto (1896, 1897a) also suggested the second and third approximation equations

$$N_x = \frac{A}{(x+x_0)^{\alpha}}, \quad A > 0, \alpha > 0, x_0 + x > 0, \tag{1.3}$$

and

$$N_x = \frac{A}{(x+x_0)^{\alpha}} e^{-\beta x}, \quad A > 0, \alpha > 0, x > x_0, \beta > 0. \tag{1.4}$$

Interestingly enough, equation (1.2) provided the most adequate fit for the income distribution in the African nation of Botswana, a republic in South Central Africa, in 1974 (Arnold, 1985).

The fact that empirically the values of parameter α remain "stable" if not constant (see Table 1.1—based on the fitting of his equations for widely diverse economies such as semifeudal Prussia, Victorian England, capitalist but highly diversified Italian cities circa 1887, and the Communist-like regime of the Jesuits in Peru during Spanish rule (1556–1821)—caused Pareto to conclude that human nature, that is, humankind's varying capabilities, is the main cause of income inequality, rather than the organization of the economy and society. If we were to examine a community of thieves, Pareto wrote (1897a, p. 371), we might well find an income distribution similar to that which experience has shown is generally obtained. In this case, the determinant of the distribution of income "earners" would be their *aptitude for theft*. What presumably determines the distribution in a community in which the production of wealth is the only way to gain an income is the aptitude for work and saving, steadiness and good conduct. This prevents necessity or desirability of legislative redistribution of income. Pareto asserted (1897a, p. 360),

> This curve gives an equilibrium position and if one diverts society from this position automatic forces develop which lead it back there.

In the subsequent version of his *Cours*, Pareto slightly modified his position by asserting that "we cannot state that the shape of the income curve would not change

Table 1.1 Pareto's Estimates of α

Country	Date	α
England	1843	1.50
	1879–1880	1.35
Prussia	1852	1.89
	1876	1.72
	1881	1.73
	1886	1.68
	1890	1.60
	1894	1.60
Saxony	1880	1.58
	1886	1.51
Florence	1887	1.41
Perugia (city)	1889	1.69
Perugia (countryside)	1889	1.37
Ancona, Arezzo, Parma, Pisa (total)	1889	1.32
Italian cities (total)	1889	1.45
Basle	1887	1.24
Paris (rents)	1887	1.57
Augsburg	1471	1.43
	1498	1.47
	1512	1.26
	1526	1.13
Peru	ca. 1800	1.79

Source: Pareto, 1897a, Tome II, p. 312.

if the social constitution were to radically change; were, for example, collectivism to replace private property" (p. 376). He also admitted that "during the course of the 19th century there are cases when the curve (of income) has slightly changed form, the type of curve remaining the same, but the constants changing." [See, e.g., Bresciani Turroni (1905) for empirical evidence using German data from the nineteenth century.]

However, Pareto still maintained that "statistics tells us that the curve varies very little in time and space: different peoples, and at different times, give very similar curves. There is therefore a notable stability in the figure of this curve."

The first fact discovered by Ammon (1895, 1898) and Pareto at the end of the nineteenth century was that "the distribution of income is highly skewed." It was a somewhat uneasy discovery since several decades earlier the leading statistician Quetelet and the father of biometrics Galton emphasized that many human characteristics including mental abilities were normally distributed.

Numerous attempts have been made in the last 100 years to explain this paradox.

Firstly it was soon discovered that the original Pareto function describes only a portion of the reported income distribution. It was originally recognized by Pareto but apparently this point was later underemphasized.

Pareto's work has been developed by a number of Italian economists and statisticians. Statisticians concentrated on the meaning and significance of the parameter α and suggested alternative indices. Most notable is the work of Gini (1909a,b) who introduced a measure of inequality commonly denoted as δ. [See also Gini's (1936) Cowles Commission paper: *On the Measurement of Concentration with Special Reference to Income and Wealth*.] This quantity describes to which power one must raise the fraction of total income composed of incomes *above a given level* to obtain the fraction of all income earners composed of high-income earners.

If we let $x_1, x_2, \ldots, x_n$ indicate incomes of progressively increasing amounts and r the number of income earners, out of the totality of n income earners, with incomes of x_{n-r+1} and up, the distribution of incomes satisfies the following simple equation:

$$\left(\frac{x_{n-r+1} + x_{n-r+2} + \cdots + x_n}{x_1 + x_2 + \cdots + x_n}\right)^{\delta} = \frac{r}{n}. \tag{1.5}$$

If the incomes are equally distributed, then $\delta = 1$. Also, δ varies with changes in the selected limit (x_{n-r+1}) chosen and increases as the concentration of incomes increases. Nevertheless, despite its variation with the selected limit, in applications to the incomes in many countries, the δ index does not vary substantially.

Analytically, for a Pareto type I distribution (1.2)

$$\delta = \frac{\alpha}{\alpha - 1}, \tag{1.6}$$

however, repeated testing on empirical income data shows that calculated δ often appreciably differs from the theoretical values derived (for a known α) from this equation.

As early as 1905 Benini in his paper "I diagramma a scala logarithmica," and 1906 in his *Principii de Statistica Metodologica*, noted that many economic phenomena such as savings accounts and the division of bequests when graphed on a double logarithmic scale generate a parabolic curve

$$\log N_x = \log A - \alpha \log x + \beta(\log x)^2, \tag{1.7}$$

which provides a good fit to the distributions of legacies in Italy (1901–1902), France (1902), and England (1901–1902). This equation, however, contains two constants that may render comparisons between countries somewhat dubious. Benini thus finally proposes the "quadratic relation"

$$\log N_x = \log A + \beta(\log x)^2. \tag{1.8}$$

Mortara (1917) concurred with Benini's conclusions that the graph with the coordinates ($\log x$, $\log N_x$) is more likely to be an upward convex curve and suggested an equation of the type

$$\log N_x = a_0 + a_1 \log x + a_2(\log x)^2 + a_3(\log x)^3 + \cdots.$$

In his study of the income distribution in Saxony in 1908, he included the first four terms, whereas in a much later publication (1949) he used only the first three terms for the distribution of the total revenue in Brazil in the years 1945–1946. Bresciani Turroni (1914) used the same function in his investigation of the distribution of wealth in Prussia in 1905.

Observing the fragmentary form of the part of the curve representing lower incomes (which presumably must slope sharply upward), Vinci (1921, pp. 230–231) suggests that the complete income curve should be a Pearson's type V distribution with density

$$f(x) = Ce^{-b/x}x^{-p-1}, \quad x > 0, \tag{1.9}$$

or more generally,

$$f(x) = Ce^{-b/(x-x_0)}(x - x_0)^{-p-1}, \quad x > x_0, \tag{1.10}$$

where $b, p > 0$, x_0 denotes as above the minimum income, and C is the normalizing constant.

Cantelli (1921, 1929) provided a probabilistic derivation of "Pareto's second approximation" (1.3), and similarly D'Addario (1934, 1939) carried out a detailed investigation of this distribution that (together with the initial first approximation) has the following property: The average income $\zeta(x)$ of earners above a certain level x is an increasing linear function of the variable x. However, this is not a characterization of the Pareto distribution(s). D'Addario proposed an ingenious *average excess value* method that involves indirect determination of the graph of the function $f(x)$ by means of $\zeta(x)$ utilizing the formula

$$f(x) = \frac{\alpha\zeta'(x)}{x - \zeta(x)} \exp\left\{\int_x^\infty \frac{\zeta(z)}{z - \zeta(z)}\, dz\right\}.$$

This approach requires selecting the average $\zeta(x)$ and its parameters based on the empirical data. The method was later refined by D'Addario (1969) and rechecked by Guerrieri (1969–1970) for the lognormal and Pearson's distributions of type III and V.

For a complete income curve, Amoroso (1924–1925) provided the density function

$$f(x) = Ce^{-b(x-x_0)^{1/s}}(x - x_0)^{(p-s)/s}, \quad x > x_0, \tag{1.11}$$

x_0 being the minimum income, C, b, $p > 0$, and s a nonzero constant such that $p + s > 0$ and fit it to Prussian data. This distribution is well known in the English language statistical literature as the generalized gamma distribution introduced by Stacy in 1962 in the *Annals of Mathematical Statistics*—which is an indication of lack of coordination between the European Continental and Anglo-American statistical literature as late as the sixties of the twentieth century. The cases $s = 1$ and $s = -1$ correspond to Pearson's type III and type V distributions, respectively.

Rhodes (1944), in a neglected work, succeeded in showing that the Pareto distribution can be derived from comparatively simple hypotheses. These involve constancy of the coefficient of variation and constancy of the type of distribution of income of those in the same "talent" group, and require that, on average, the consequent income increases with the possession of more talents.

D'Addario—like many other investigators of income distributions—was concerned with the multitude of disconnected forms proposed by various researchers. He attempted to obtain a general, relatively simply structured formula that would incorporate numerous special forms. In his seminal contribution *La Trasformate Euleriane*, he showed how transforming variables in several expressions for the density of the income distribution lead to the general equation

$$f(x) = \frac{1}{\Gamma(p)} e^{-w(x)} [w(x)]^{p-1} |w'(x)| \qquad (1.12)$$

[here $\Gamma(p)$ is the gamma function]. Given a density $g(z)$, transforming the variable $x = u(z)$ and obtaining its inverse $z = w(x)$, we calculate the density of the transformed variable, $f(x)$, say, by the formula

$$f(x) = g[w(x)]|w'(x)|.$$

Here, if we use D'Addario's terminology, $g(z)$ is the *generating* function, $z = w(x)$ the *transforming* function, and $f(x)$ the *transformed* function. If the generating function is the gamma distribution

$$g(z) = \frac{1}{\Gamma(p)} e^{-z} z^{p-1}, \quad z \geq 0,$$

then the *Eulerian transform* is given by (1.12). This approach was earlier suggested by Edgeworth (1898), Kapteyn (1903) in his *Skew Frequency Curves in Biology and Statistics*, and van Uven (1917) in his *Logarithmic Frequency Distributions*, but D'Addario applied it skillfully to income distributions. More details are provided in Section 2.4.

In 1931 Gibrat, a French engineer and economist, developed a widely used lognormal model for the size distributions of income and of firms based on Kapteyn's (1903) idea of the proportional effect (by adding increments of income to an initial income distribution in proportion to the level already achieved). Champernowne (1952, 1953) refined Gibrat's approach and developed formulas that often fit better than Gibrat's lognormal distribution. However, when applied to U.S. income data of 1947 that incorporate low-income recipients, his results are not totally satisfactory. Even his four-parameter model gives unacceptable, gross errors. Somewhat earlier Kalecki (1945) modified Gibrat's approach by assuming that the increments of the income are proportional to the excess in ability of given members of the distribution over the lowest (or median) member. (A thoughtful observation by Tinbergen, made as early as 1956, prompts to distinguish between two underlying causes for income distribution. One is dealing here simultaneously with the distribution of abilities to earn income as well as with a distribution of preferences for income.)

A somewhat neglected (in the English literature) contribution is the so-called *van der Wijk's law* (1939). Here it is assumed that the average income above a limit x, $\sum_{x_i>x} x_i/N_x$, is proportional to the selected income level x, leading to the "law"

$$\frac{\sum_{x_i>x} x_i}{N_x} = \eta x, \tag{1.13}$$

where η is a constant of proportionality. For instance, if $\eta = 2$, then the average income of people with at least \$20,000 must be in the vicinity of \$40,000 and so on. Bresciani Turroni proposed a similar relationship in 1910, but it was not widely noticed in the subsequent literature.

Van der Wijk in his rather obscure volume *Inkomens- en Vermogensverdeling* (1939) also provided an interpretation of Gibrat's equation by involving the concept of *psychic income*. This was in accordance with the original discovery of the lognormal distribution inspired by the Weber–Fechner law in psychology (Fechner, 1860), quite unrelated to income distributions.

Pareto's contribution stimulated further research in the specification of new models to fit the whole range of income. One of the earliest may be traced to the French statistician Lucien March who as early as 1898 proposed using the gamma distribution and fitted it to the distribution of wages in France, Germany, and the United States. March claimed that the suggestion of employing the gamma distribution was due to the work of German social anthropologist Otto Ammon (1842–1916) in his book *Die Gesellschaftsordnung und ihre natürlichen Grundlagen* (1896 [second edition]), but we were unable to find this reference in any one of the three editions of Ammon's text. Some 75 years later Salem and Mount (1974) fit the gamma distribution to U.S. income data (presumably unaware of March's priority).

Champernowne (1952) specified versions of the log-logistic distribution with two, three, and four parameters. Fisk (1961a,b) studied the two-parameter version in detail.

Mandelbrot (1960, p. 79) observed that

> over a certain range of values of income, its distribution is not markedly influenced either by the socio-economic structure of the community under study, or by the definition chosen for "income." That is, these two elements may at most influence the values taken by certain parameters of an apparently universal distribution law.

and proposed nonnormal stable distributions as appropriate models for the size distribution of incomes.

Metcalf (1969) used a three-parameter lognormal distribution. Thurow (1970) and McDonald and Ransom (1979a) dealt with the beta type I distribution.

Dagum in 1977 devised two categories of properties for a p.d.f. to be specified as a model of income or wealth distribution: The first category includes essential properties, the second category important (but not necessary) properties. The essential properties are

- Model foundations
- Convergence to the Pareto law
- Existence of only a small number of finite moments
- Economic significance of the parameters
- Model flexibility to fit both unimodal and zeromodal distributions

(It seems to us that property 3 is implied by property 2.) Among the important properties are

- Good fit of the whole range of income
- Good fit of distributions with null and negative incomes
- Good fit of the whole income range of distributions starting from an unknown positive origin
- Derivation of an explicit mathematical form of the Lorenz curve from the specified model of income distributions and conversely

Dagum attributed special importance to the concept of income elasticity

$$\eta(x, F) = \frac{x}{F(x)} \frac{dF(x)}{dx} = \frac{d \log F(x)}{d \log x}$$

of a distribution function as a criterion for an income distribution.

He noted that the observed income elasticity of a c.d.f. behaves as a nonlinear and decreasing function of F. To represent this characteristic of the income elasticity, Dagum specified (in the simplest case) the differential equation

$$\eta(x, F) = ap\{1 - [F(x)]^{1/p}\}, \quad x \geq 0,$$

subject to $p > 0$ and $ap > 0$, which leads to the Dagum type I distribution

$$F(x) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, \quad x > 0,$$

where $a, b, p > 0$.

It was noted by Dagum (1980c, 1983) [see also Dagum (1990a, 1996)] that it is appropriate to classify the income distributions based on three generating systems:

- Pearson system
- D'Addario's system
- Generalized logistic (or Burr logistic) system

Only Champernowne's model does not belong to any of the three systems.

The pioneering work of McDonald (1984) and Venter (1983) led to the generalized beta (or transformed beta) distribution given by

$$f(x) = \frac{ax^{ap-1}}{b^{ap}B(p,q)[1+(x/b)^a]^{p+q}}, \quad x > 0. \tag{1.14}$$

It is also known in the statistical literature as the generalized F (see, e.g., Kalbfleisch and Prentice, 1980) and was rediscovered in a slightly different parameterization by Majumder and Chakravarty (1990) a few years later. This family includes numerous models used as income and size distributions, in particular the Singh and Maddala (1976) model, the Dagum type I model (Dagum, 1977), the Fisk model, and evidently the beta distribution of the second kind. In actuarial science the Singh–Maddala and Dagum models are usually referred to as the Burr and inverse Burr distributions, respectively, since they are members of the Burr (1942) system of distributions.

We also mention the natural generalization of the Pareto distribution proposed by Stoppa in 1990b,c. It is given by

$$F(x) = \left[1 - \left(\frac{x}{x_0}\right)^{-\alpha}\right]^{\varphi}, \quad 0 < x_0 \le x. \tag{1.15}$$

This book is devoted to a detailed study of the distributions surveyed in this section and their interrelations. The literature is immense and omissions are unavoidable although we tried to utilize all the references collected during a six-month extensive search. Due to the rather sporadic developments in that area, only some isolated multivariate distributions are included.

1.4 STOCHASTIC PROCESS MODELS FOR SIZE DISTRIBUTIONS

Interestingly enough, income and wealth distributions of various types can be obtained as steady-state solutions of stochastic processes.

The first example is Gibrat's (1931) model leading to the lognormal distribution. He views income dynamics as a multiplicative random process in which the product of a large number of individual random variables tends to the lognormal distribution. This multiplicative central limit theorem leads to a simple Markov model of the "law of proportionate effect." Let X_t denote the income in period t. It is generated by a first-order Markov process, depending only on X_{t-1} and a stochastic influence

$$X_t = R_t X_{t-1}.$$

Here $\{R_t\}$ is a sequence of independent and identically distributed random variables that are independent of X_{t-1} as well. X_0 is the income in the initial period. Substituting backward, we see that

$$X_t = X_0 \cdot R_0 \cdot R_1 \cdot R_2 \cdot \ldots R_{t-1},$$

and as t increases, the distribution of X_t tends to a lognormal distribution provided $\text{var}(\log R_t) < \infty$.

In the Gibrat model we assume the independence of R_t, which may not be realistic. Moreover, the variance of $\log X_t$ is an increasing function of t and this often contradicts the data. Kalecki (1945), in a paper already mentioned, modified the model by introducing a negative correlation between X_{t-1} and R_t that *prevents* $\text{var}(\log X_t)$ from growing. Economically, it means that the probability that income will rise by a given percentage is lower for the rich than for the poor. (The modification is an example of an ingenious but possibly ad hoc assumption.)

Champernowne (1953) demonstrated that under certain assumptions the stationary income distribution will approximate the Pareto distribution irrespectively of the initial distribution. He also viewed income determination as a Markov process (income for the current period depends only on one's income for the last period and random influence). He subdivided the income into a finite number of classes and defined p_{ij} as the probability of being in class j at time $t+1$ given that one was in class i at time t. The income intervals defining each class are assumed (1) to form a *geometric* (not arithmetic) progression. The limits of class j are higher than those of class $j-1$ by a certain *percentage* rather than a certain absolute amount of income and the transitional probabilities p_{ij} depend only on the differences $j-i$. (2) Income cannot *move up* more than one interval nor down more than n intervals in any one period; (3) there is a lowest interval beneath which no income can fall, and (4) the average number of intervals shifted in a period is negative in each income bracket. Under these assumptions, Champernowne proved that the distribution eventually behaves like the Pareto law.

The assumptions of the Champernowne model can be relaxed by allowing for groups of people (classified by age, occupation, etc.) and permitting movement from one group to another. However, constancy of the transition matrix is essential; otherwise, no stationary distribution will emerge from the Markov process. Moreover, probabilities of advancing or declining ought to be independent of the amount of income. Many would doubt the existence of a society whose institutional framework is so static, noting that such phenomena as "inherited privilege," and cycles of poverty or prosperity are part and parcel of all viable societies.

To complicate the matter with the applicability of Champernowne's model, it was shown by Aitchison and Brown (1954) that if the transition probabilities p_{ij} depend on j/i (rather than $j - i$, as is the case in Champernowne's model) and further that the income brackets form an arithmetic (rather than geometric) progression, then the limiting distribution is lognormal rather than Pareto. In our opinion the dependence on $j - i$ may seem to be more natural, but it is a matter of subjective opinion.

It should also be noted that the Champernowne and Gibrat models and some others require long durations of time until the approach to stationarity is obtained. This point has been emphasized by Shorrocks (1975).

Rutherford (1955) incorporated birth–death considerations into a Markov model. His assumptions were as follows:

- The supply of new entrants grows at a constant rate.
- These people enter the labor force with a lognormal distribution of income.
- The number of survivors in each cohort declines exponentially with age.

Under these assumptions, the data eventually approximate the Gram–Charlier type A distribution, which often provides a better fit than the lognormal. In Rutherford's model the overall variance remains *constant* over time.

Mandelbrot (1961) constructed a Markov model that approximates the Pareto distribution similarly to Champernowne's model, but does not require the strict law of proportionate effect (a random walk in logarithms).

Wold and Whittle (1957) offered a rather general continuous-time model that also generates the Pareto distribution: It is applied to stocks of wealth that grow at a compound interest rate during the lifetime of a wealth-holder and are then divided among his heirs. Deaths occur randomly with a known mortality rate per unit time. Applying the model to wealth above a certain minimum (this is necessary because the Pareto distribution only applies above some positive minimum wealth), Wold and Whittle derived the Pareto law and expressed the exponent α as a function of (1) the number of heirs per person, (2) the growth rate of wealth, and (3) the mortality rate of the wealth owners.

The most complicated model known to us seems to be due to Sargan (1957). It is a continuous-time Markov process: The ways in which transitions occur are explicitly spelled out. His approach is quite general; it accommodates

- Setting of new households and dissolving of old ones
- Gifts between households
- Savings and capital gains
- Inheritance and death

It is its generality that makes it unwieldy and unintelligible.

As an alternative to the use of ergodic Markov processes, one can also explain wealth or income distributions by means of branching processes. Steindl (1972), building on the model of Wold and Whittle (1957) mentioned above, showed in this way that the distribution of wealth can be regarded as a certain transformation of an age distribution. Shorrocks (1975) explained wealth accumulations using the theory of queues. He criticized previously developed stochastic models for concentrating on equilibrium distributions and proposed a model in which the transition probabilities or parameters of the distribution are allowed to change over time.

These models were often criticized by applied economists who favor models based on human capital and the concept of economic man (Mincer, 1958; Becker, 1962, 1964). Some of them scorn size distribution of income and refer to them as antitheories. Their criticism often goes like this:

> Allowing a stochastic mechanism to be the sole determinant of the income distribution is TO GIVE UP BEFORE YOU START. The deterministic part of a model (in econometrics) is "what we think we know," the disturbance term is "what we don't know." The probabilistic approach allocates 100% variance in income to the latter.

In our opinion this type of argument shows a lack of understanding of the concept of stochastic model and by extension of the probabilistic-statistical approach.