# Toward an Economic Theory of Income Distribution

Alan S. Blinder

# Toward an Economic Theory of Income Distribution

This book was printed on Decision 94
and bound in Columbia Millbank Vellum MBV-4376
by The Colonial Press Inc.
in the United States of America.



Library of Congress Cataloging in Publication Data

Blinder, Alan S.
Toward an economic theory of income distribution.

(M.I.T. monographs in economics, 11)
Bibliography: p.
1. Income—Mathematical models. 2. Income—United States—Mathematical models. I. Title. II. Series.
HB601.B53 339.2'01 74-5417
ISBN 0-262-02114-5

# Preface

The research reported here originated as a doctoral dissertation written at the Massachusetts Institute of Technology during the 1970–1971 academic year. It has continued to evolve, in sporadic bursts of activity separated by lengthy periods of inaction, during the past three years at Princeton University.

While an author's evaluation of his own work should be treated with a healthy dose of skepticism, if not utterly disregarded, it seems to me that the intervening three years have not made this work as obsolete as I would have imagined in 1971. In those days it appeared, at least to a graduate student single-mindedly immersed in the study of income distribution, that the profession was on the verge of a burgeoning of interest in inequality, that the economic "pie" had at last grown large enough so that more attention could be paid to its division and less to its size.

The events of the past three years have belied these lofty expectations, especially in so far as theoretical work is concerned. While more research has probably been published in this field during the last three years than in the preceding three, it is hard to argue that the increase has exceeded the growth rate of economic literature in general. And the university that offers a course on income distribution is still the exception rather than the rule. This is unfortunate, since I believe that the field is a fertile one for the application of rigorous economic analysis. And I still think Ricardo was right when he remarked that no problem in economics is so important as the determination of the distribution of income. It is surely no false modesty to state that this book raises many more questions than it answers. My hope is that other economists, finding the questions interesting and the techniques worth pursuing, will join in the development of a coherent theory of income distribution.

My debts are many. Peter Diamond and Robert Solow were my principal advisers while I was working on the dissertation, and their perceptive evaluations of my early work assisted me in formulating many of the ideas expressed here and, most importantly, enabled me to avoid numerous pitfalls. Robert Hall joined my committee when the thesis was nearly finished, and offered astute comments on a complete draft. I also benefited from conversations with Christian von Weizsäcker, Robert Merton, and several others at MIT that year.

Since then many individuals have read bits and pieces of the manuscript as the thesis was laboriously being turned into a book. My former colleague at Princeton, Daniel Hamermesh, must be singled out for having had the perseverance to read and offer valuable comments upon the entire manuscript in its penultimate form. Ray Fair, A. B. Atkinson, and Gregory Chow also offered suggestions which materially improved the content of the final version.

Research assistance was provided, at various times during the 1972–1973 and 1973–1974 academic years by Dennis Warner, Barry Schwartz, Donald Coes, and Edward Meyer. I thank them all. Betty Kaminski did her usual fine job of typing the manuscript.

One ought never to forget his benefactors. As a graduate student, my research was supported by a National Defense Education Act Fellowship to MIT, and for the past two years the National Science Foundation has provided generous financial assistance under Grant GS-36027.

Finally, and most importantly, my sincerest "thank you" is due my wife, Madeline, who has been unfairly subjected to double jeopardy by this monograph—first as a dissertation which stubbornly refused to be banished from our household, and then as a book which preoccupied me far too long. She has had the patience to put up with me as I scratched out the manuscript, and even contributed to its completion by proofreading and correcting my faltering grammar. I am indeed grateful.

For myself, I am content to claim full credit for all remaining errors and omissions.

A. S. B.
Princeton, New Jersey
January 1974

# Contents

# 1
# Desiderata for an Economic Theory of Size Distribution

The body of economic analysis rather desperately needs a reliable theory of the distribution of incomes. Whether or not this approach is ultimately deemed to be satisfactory, it should demonstrate that such a theory need not be a patchwork of Pareto distributions, ability vectors, and ad hoc probability mechanisms, but can rely on the basic economic principles that have so often proven their worth elsewhere.

Gary S. Becker

The title of this study is used advisedly. Much of what has been offered in the literature as "economic" models of the size distribution of income and wealth hardly merits the name. That is, while often elegant and ingenious, these models have not been integrated into the mainstream of modern economic theory. This is both inexplicable and unfortunate, since there is a considerable body of economic theory which can be brought to bear on the subject. I hope, within the following chapters, to demonstrate that this is so, and to point the way toward a theory of income distribution which is part of the corpus of neoclassical economic thought. Of course, I take only a few small steps in this direction; hence the use of the word "toward" in the title. The most interesting contributions to the economic theory of size distribution are yet to come.

Most of the work in economics that goes by the name "income distribution theory" has focused on the distribution of income among *factors of production*, rather than the distribution among *individuals*. This orientation dates back at least to Ricardo and Marx and may have been appropriate to the capitalism of the day. While the behavior of distributive shares may still pose interesting intellectual problems in positive economics, its normative significance for inequality as a social problem is nowadays rather limited. But a comparably rigorous theory of size distribution has not been developed. Jan Tinbergen's remark of some years ago [1956, p. 244] seems equally appropriate today: "The fairly satisfactory state of affairs with respect to the statistical description of income distribution contrasts with an unsatisfactory state in the area of economic interpretation."

The remainder of this chapter outlines the requirements for a complete and exact microeconomic theory of the size distribution of income and wealth. I am by no means prepared to meet all of these requirements here. However, by synthesizing some established pieces of economic theory and filling in a few gaps, it is possible to develop a rigorous, though simplified, model of income distribution under capitalism. This is the program for the book. Chapters 2 and 3 provide the microeconomic building blocks, and Chapters 4 to 6 exploit these results to see what economic theory has to say about the size distribution of income in the United States.

## 1.1 Intragenerational and Intergenerational Models of Size Distribution

There are two separate aspects of distribution theory which are best distinguished at the outset. An *intragenerational* model is designed to answer the question, Why is the income distribution what it is today? Its principal components are models of the savings, consumption, investment, training, and labor supply behavior of individual consumer units. It takes as given the wealth, technology, and abilities inherited from previous generations. An *intergenerational* model is designed to answer the question, What factors determine the evolution of the income distribution over time? It focuses on decisions to bequeath wealth, both human (through education) and nonhuman (through inheritance), to one's heirs. The inheritance of genetic ability, though not subject to human choice (yet!), also plays a role here.

The two models complement each other in a straightforward way: each provides the "initial conditions" for the other. For example, a fully developed intragenerational model would have to generate the distribution of bequests since the latter is an integral part of savings behavior and wealth accumulation. Appending to this some model—and none has been suggested to date—of parental decisions to educate their offspring would close the loop between the income distribution in one generation and the income distribution among its successors.

This study is confined to the intragenerational model though it will have something to say about bequest behavior.[1] To the extent that they have contained any behavioral aspects at all, most previous efforts have also been confined to intragenerational aspects. The reader unfamiliar with the size distribution literature may be startled to learn that there are income distribution models devoid of behavioral content. But in fact, such models—generally based on some sort of stochastic process—are among the best known distribution theories.

## 1.2 Models of the Size Distribution: A Survey[2]

### 1.2.1 Stochastic Process Models

The fact that income distributions appear quite stable over time has suggested to several authors that the distribution might be the steady-state solution of some stochastic process. Robert Gibrat [1957] seems to have originated this line of thought when he noted that the *product* of a large number of independent random variables tends toward the lognormal distribution,[3] which has the positive skewness displayed by the data, rather than toward the symmetric normal distribution, which is the limit of the sum of additive errors. This multiplicative central limit theorem leads naturally to the following simple Markov model, which Gibrat dubbed "the law of proportional effect." Let income in period $t$ be denoted by $Y_t$. Assume that $Y_t$ is generated by a first-order Markov process, so that it depends only on $Y_{t-1}$ and stochastic influences. Specifically,

$$Y_t = R_{t-1} Y_{t-1},$$

where $\{R_t\}$ is a sequence of serially independent random variables which are independent of $Y_t$. If $Y_0$ is income in the initial period, it follows immediately that

$$Y_t = Y_0 \cdot R_0 \cdot R_1 \cdot R_2 \cdot \ldots \cdot R_{t-1}.$$

The multiplicative central limit theorem implies that as $t$ gets large, the distribution of $Y_t$ tends toward the lognormal.

1. For the beginnings of a crude intergenerational model, the reader is referred to Blinder [1973b]. Other relevant references are Stiglitz [1969], Atkinson [1971], Ishikawa [forthcoming], and Pryor [1969].

2. Other surveys of the theoretical literature have been offered by Bjerke [1961], Lydall [1968], and Mincer [1970].

3. If $y = \log x$, and $y$ is normal, then $x$ is said to have the lognormal distribution.

Other than the serial independence of the $R_t$, one troubling feature of this model is its implication that the variance of $\log(Y_t)$ is steadily increasing, a prediction which is belied by the data. Michal Kalecki [1945] has modified the simple Gibrat model by introducing a negative correlation between $Y_t$ and $R_t$ which is just sufficient to prevent the log variance of $Y_t$ from growing. Economically, this means that the probability that income will rise by a given percentage is lower for the rich than for the poor. It is far from obvious that this is true. In a way, Kalecki's contribution is a microcosm of the entire stochastic process approach: it is highly ingenious, but equally *ad hoc*.

Other than the lognormal, the analytical distribution which is used most frequently to fit the data is the Pareto distribution

$$N(Y)=AY^{-\alpha},$$

where $N(Y)$ is the fraction of the population having income greater than $Y$ and $A$ and $\alpha$ are constants. Over fifty years elapsed between Pareto's remarkable empirical discovery that the upper tails of almost all income distributions followed this law and D.G.Champernowne's [1953] elegant demonstration that, under suitable assumptions, the stationary income distribution must approximate the Pareto irrespective of the initial distribution. Like Gibrat, Champernowne views the income determination process as a Markov process, so that one's income for this period depends only on one's income for the last period and random influences. But, unlike Gibrat, he subdivides income into a finite number of classes and defines transitional probabilities $p_{ij}$ as the probability of being in class $j$ at time $t+1$, given that one was in class $i$ at time $t$. The crucial assumptions of Champernowne's analysis concern the definition of the income classes and the specification of the transitional probabilities. The income intervals defining each class are assumed to form a geometric progression rather than the conventional arithmetic progression. That is, the limits of class $k$ are higher than the limits of class $k-1$ by a certain *percentage* rather than by a certain absolute amount of income. Most crucial to his result is the assumption that the transitional probabilities $p_{ij}$ depend only on the differences $j-i$. Under these and certain other assumptions,[4] Champernowne proves that the distribution eventually behaves like the Pareto law.

4. Among the other assumptions are (1) incomes cannot move up more than one interval, nor down more than $n$ intervals, in any one year; (2) there is a lowest interval, beneath which no income can fall; (3) the average number of intervals shifted in a year is negative in every income bracket.

Champernowne's result can be generalized in several directions. For example, a Pareto distribution can be derived in a model where people fall into groups (say, by age or occupation), and where stochastic movements from one group to another are allowed. But, as he recognizes, several assumptions cannot be dispensed with. Of course, no Markov process yields a stationary distribution unless the matrix of transitional probabilities is constant forever. This is obvious enough; but it is hard to imagine a society whose institutional framework is so static as this. Secondly, his assumption that the probabilities of advancing or declining are independent of the size of income is crucial. Many people who believe in "inherited privilege" or the "cycle of poverty" will not find this a congenial notion. Finally, J. Aitchison and J. A. C. Brown [1954] have shown that a minor alteration in one of his assumptions—specifically making the $p_{ij}$ depend on the *ratio* $j/i$ rather than the *difference* $j-i$—makes the model generate the lognormal distribution rather than the Pareto. It is difficult to argue that either assumption is more plausible than the other.

Another difficulty with Champernowne's model, as with Gibrat's, is that stochastic processes like these may take a very long time to approach their stationary states. If initial conditions are to be unimportant, this requires that an "income" be passed on at death from one person to the next, so that we are not dealing with the incomes of finite-lived individuals but rather with the incomes of infinite families. R. S. G. Rutherford [1955] has explicitly incorporated birth-and-death considerations into a Markov model. Under the assumptions that (1) the supply of new entrants grows at a constant rate, (2) these people enter the labor force with a lognormal distribution of income, and (3) the number of survivors in each cohort declines exponentially with age, he deduces that incomes will eventually approximate the Gram-Charlier Type A distribution, which, he claims, fits the data better than the lognormal. Aside from being a step in the direction of greater realism, the advantage of Rutherford's model is that it offers an alternative to Kalecki's method for insuring that the log variance of income does not grow over time. In Rutherford's model, unlike Kalecki's, the shocks are independent of income, so that the variance of $\log(Y_t)$ grows over time *within each age cohort*; but the cohort with the largest variance dies each year, and a new cohort with a small variance is born. Thus Rutherford is able to show that the overall variance of $\log(Y_t)$ is constant over time.

Benoit Mandelbrot [1961], perhaps the chief proponent of the Pareto distribution, has shown that the income distribution must eventually approximate the Pareto in a Markov model very similar to Champernowne's, but one which does not require the strict law of proportionate effect (that is, that the random shocks be additive in the logs). He has also stressed several desirable statistical properties of what he calls "weak Pareto laws," that is, frequency distributions that are asymptotic to the Pareto. First of all, consider the overall distribution of income as a weighted average of many components, for example, incomes in different occupations or incomes from different sources. Suppose further that the distributions of these components all follow some probability law. If the overall income distribution also follows this probability law, Mandelbrot calls it a "stable distribution." It turns out that the only stable distributions are the normal—which is known not to fit income distribution data—and the family of weak Pareto laws [Mandelbrot, 1960]. The second convenient property of the Pareto family is as follows. If one considers the limit distribution of the sum of a large number of independent and identically distributed random variables, one arrives at the normal distribution only by further assuming that the *largest* of the components is negligible in size. If, as Mandelbrot believes is more common in economic applications, the largest component is not negligible, then the limit distribution follows a weak Pareto law [Mandelbrot, 1961].

A final stochastic model that generates the Pareto distribution was offered by H. O. A. Wold and P. Whittle [1957]. Their model is meant to apply to stocks of wealth, which grow at a compound interest rate during the lifetime of a wealth-holder and then are divided among the heirs at death. They assume that deaths occur randomly with a known mortality rate per unit time. Applying this model only to wealth above a certain minimum,[5] they derive the Pareto law and express the exponent $\alpha$ as a function of the number of heirs per person, the growth rate of wealth, and the mortality rate.

The probabilistic school of thought culminates in a brilliant but almost unknown paper by J. D. Sargan [1957]. Sargan's model can be thought of as a continuous Markov process, where the ways in which transitions occur are explicitly spelled out. The great virtue of the model is its generality: it can accommodate almost any probability distributions for (1) setting up of new households and dissolving of old ones, (2) gifts from one household to another, (3) savings and capital gains, (4) inheritances. This list incorporates, I believe, most

5. This is necessary because the Pareto distribution only applies above some positive minimum wealth.

of the reasons economists would give for changes in household wealth. Unfortunately, the very generality of the model makes it unwieldy (not to mention unintelligible), and Sargan has to settle for analysis of a special case. In this instance, the stochastic process eventually leads to a distribution which is approximately lognormal.

What do all these stochastic models contribute to an economist's understanding of income distribution? In my opinion, not very much. Assuming a stochastic mechanism, no matter how complex, to be the sole determinant of income inequality is to give up before one starts. It is antithetical to the mainstream of economic theory which seeks to explain complex phenomena as the end result of deliberate choices by decision-makers. Borrowing terms from the econometric literature, one may think of the deterministic part of any model as "what we (think we) know" and the stochastic disturbance as the measure of our ignorance. The probabilistic approach to distribution theory appears to allocate the entire variance in income to the latter. One would hope that economics could do better than that.

An important first step in this direction was taken in a paper by Milton Friedman [1953]. I classify Friedman's model with the stochastic theories since the income distribution that it generates is a drawing from a random process. But, unlike the other stochastic models, individual choices by persons differing in risk aversion help determine the shape of the distribution. Roughly speaking, Friedman views every person as having a certain income and an opportunity to participate in a lottery if he so desires. Each person consults his utility function, and the less risk averse enter the lottery while the more risk averse do not. The resulting income distribution is an amalgam of three distributions, each one of which could be symmetrical: (1) nonparticipants; (2) lottery losers, whose distribution has a slightly lower mean; and (3) lottery winners, whose distribution has a much higher mean. If the lottery has only a few winners of very large prizes, the resulting overall distribution is positively skewed with an elongated upper tail.

Certainly the papers by Friedman, Sargan, and Wold-Whittle make it clear that, if the stochastic process theories are to play any role in a model of size distribution, they are most appropriately used to analyze the accumulation of risky capital. Here random elements are likely to predominate, although there are still economic considerations in choosing an optimal portfolio. It may well be no accident that the upper tails of almost all income distributions, where returns to capital dominate and earnings play a minor role, exhibit a striking resemblance to the Pareto distribution. Models like those of Champernowne, Mandelbrot, and Wold-Whittle may well hold the key to this phenomenon.

The model of income distribution to be presented in the following chapters is exact and nonstochastic. A more complete and realistic model would allow for random elements, perhaps along the lines suggested by these models.

### 1.2.2 Ability-Earnings Models

Most of income consists of earnings, and stochastic models appear to have little to say about this type of income. Of course, this does not mean that the laws of probability theory are not useful in this context. If earnings depend on ability (however measured), and the distribution of ability follows some known frequency distribution, it may be possible to deduce the functional form of the income distribution from the distribution of abilities. A second school of thought, which seeks to exploit this simple idea, has arisen.

Theorizing of this sort appears to have been started by Otto Ammon's [1899] early observation that incomes follow a skewed distribution while abilities are apparently normally distributed.[6] Ammon attributed this discrepancy to quirks in the income tax data which he used and to altruism, which prevented those with unusually low ability from having such low incomes.

This explanation was deemed unsatisfactory by many economists. Most notably, A. C. Pigou, in his monumental *The Economics of Welfare* [1924], pointed out two reasons why the income distribution might be skewed despite the normal distribution of abilities. First, part of income is attributable to inherited wealth, including the opportunities for increasing one's earnings that large inheritances typically bring. And it was well known, even then, that inheritances follow a highly skewed distribution. Secondly, Pigou suggested that the overall distribution of earnings might be skewed because it is an amalgam of the distributions within "noncompeting" subgroups of the population. His example was "brain-workers" versus "hand-workers," and he suggested that the across-group competition was minimal. Pigou speculated that the distributions among brain-workers and hand-workers might each be normal, and yet the overall distribution could be skewed. Some years later, Hans Staehle [1943] offered some evidence from U.S. and German data to support

6. A convenient summary of the early literature, beginning wth Ammon, can be found in Staehle [1943], Part I.

Pigou's conjecture, and Herman Miller's [1955a, 1955b] thorough examinations of U.S. Census data have established that income distributions for relatively homogeneous subgroups of the population tend to be much more symmetric than the overall distribution.

Staehle offered yet another explanation of skewness, which ought to have been obvious to economists long before. Ability is, presumably, a proximate determinant of the *wage rate* (potential earnings) rather than of earnings. If individual supplies of work effort respond positively to higher wages, hours of work and wage rates are positively correlated, so that their product—earnings—is positively skewed even if both wages and hours are symmetric. Alternatively, if the variation in hours worked is mostly due to involuntary unemployment and if employers lay off their least skilled workers more freely than their skilled workers, positive correlation between wages and hours again arises.

The notion that products of normally distributed variates are generally positively skewed has stimulated several models of earnings distributions, beginning with C. H. Boissevain [1939]. He simply observed that if earnings depend in a multiplicative way on various factors ("skills"), then the distribution of earnings is skewed even if all the factors are uncorrelated and normally distributed. This result may be generalized and somewhat sharpened by noting that the multiplicative central limit theorem, which Gibrat applied to random shocks over time, can also be applied cross sectionally. That is, the lognormal distribution tends to result if earnings are a product of a large number of independent factors, even if those factors are not normally distributed.

A. D. Roy has applied these ideas to income distribution in a series of papers. In his simplest model [Roy, 1950a], he asserts that earnings are proportional to output produced, and that output is the product of speed, accuracy, and hours of work. Assuming each of these three factors to be normally distributed, though correlated, he appeals to some results of J. B. S. Haldane [1942] to show that the earnings distribution is approximately lognormal if the coefficients of variation of the three factors are about equal. Roy's case is strengthened once it is observed that the coefficients of variation of speed, accuracy, and hours worked are not in fact equal. For then his model places more people in the upper tail of the distribution than the lognormal would predict, and Harold Lydall [1968] has documented the fact that actual income distributions have fatter tails than the lognormal. In a later paper, Roy [1951] takes an important first step toward making his model less mechanistic by allowing each individual to choose the job in which he earns the highest income. He argues that the resulting income distribution still resembles the lognormal.[7]

A somewhat different ability-earnings model, using precisely that same mathematical result, was offered by Thomas Mayer [1960] some ten years after Roy. He argues for the empirical validity of the notion that earnings depend on the product of the probability of completing a task successfully (which he calls "ability") and the scale of the activity (which he calls "responsibility"). As I have just noted, if ability and responsibility are normally and independently distributed with equal coefficients of variation, this leads precisely to a lognormal earnings distribution. Of course, Mayer believes that these two determinants of earnings are positively correlated; but this still yields an "almost lognormal" distribution. Significantly, Mayer's paper may be the first example of an economist questioning the underlying assumption that abilities are normally distributed. Lydall [1968] has shown that this belief is based on perilously little evidence.

All of the ability-earnings models cited so far seek to explain how a skewed income distribution might arise from a normal distribution of abilities. A closely related set of models employs somewhat different, and often *ad hoc*, assumptions about individual talents. E. C. Rhodes [1944] suggested the following model to explain the Pareto distribution. Suppose people fall into a finite number of homogeneous classes defined by the number of talents they possess. Suppose further that the number of people with $k$ talents declines with $k$ in a geometric progression, and that the mean income rises with $k$ in a different geometric progression. Finally, suppose that the coefficients of variation are equal in each group, though the within-group distributions are not necessarily normal. Rhodes shows that these hypotheses imply a weak Pareto distribution, although he realizes that his assumptions about the wage structure and the distribution of talents come close to assuming the conclusion.

Lydall [1959] has constructed a similar model, though with a much stronger economic motivation, which he means to apply to the upper tail of the earnings distribution. He argues that in hierarchical organizations a person's earnings depend largely on the number of people he supervises. Let there be $k$ distinct grades in the bureaucracy in question, with one person in the highest grade (the

7. A slight change in Roy's assumptions for this model yields a Pareto distribution instead. See below.