OXFORD

Measuring Inequality

Frank A. Cowell

THIRD EDITION

LSE PERSPECTIVES IN ECONOMIC ANALYSIS

Measuring Inequality

LSE Perspectives in Economic Analysis

Series editors

Timothy Besley and Frank A. Cowell

The LSE Perspectives in Economic Analysis series provides concise and original insight into a wide range of topics in economics. Each book is accessibly written but scholarly to appeal to advanced students of economics, and academics and professionals wishing to expand their knowledge outside their own particular field.

Books in the series

A Primer in Social Choice Theory, Revised Edition by Wulf Gaertner

Strategy and Dynamics in Contests by Kai A. Konrad

Measuring Inequality by Frank A. Cowell

Forthcoming books

Econometric Analysis of Panel Data by Vassilis Hajivassiliou

Measuring Inequality

Frank A. Cowell

Third Edition

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States
by Oxford University Press Inc., New York

First edition 1977
Second edition 1995
Third edition first published 2011

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
2010940330

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Books Group, Bodmin and King's Lynn

ISBN 978–0–19–959403–0 (Hbk.)
978–0–19–959404–7 (Pbk.)

1 3 5 7 9 10 8 6 4 2

This book is dedicated to the memory of my parents.

This page intentionally left blank

Preface

> 'It is not the business of the botanist to eradicate the weeds. Enough for him if he can tell us just how fast they grow.'
>
> C. Northcote Parkinson (1958), *Parkinson's Law*

The maligned botanist has a good deal to be said for him in the company of rival gardeners, each propagating his own idea about the extent and the growth of thorns and thistles in the herbaceous border, and each with a patent weedkiller. I hope that this book will perform a similar role in the social scientist's toolshed. It does not deal with theories of the development of income distribution, of the generation of inequality, or of other social weeds, nor does it supply any social herbicides. However, it does give a guide to some of the theoretical and practical problems involved in an analysis of the extent of inequality thus permitting an evaluation of the diverse approaches hitherto adopted. In avoiding patent remedies for particular unwanted growths, one finds useful analogies in various related fields—for example, some techniques for measuring economic inequality have important counterparts in sociological and political studies. Thus, although I have written this as an economist, I would like to think that students in these related disciplines will be interested in this material.

This book is deliberately limited in what it tries to do as far as expounding theory, examining empirical evidence, or reviewing the burgeoning literature is concerned. For this reason, a set of notes for each chapter is provided on pages 178 ff. The idea is that if you have not already been put off the subject by the text, then you can follow up technical and esoteric points in these notes, and also find a guide to further reading.

A satisfactory discussion of the techniques of inequality measurement inevitably involves the use of some mathematics. However, I hope that people who are allergic to symbols will nevertheless read on. If you are allergic, you may need to toil a little more heavily round the diagrams that are used fairly extensively in Chapters 2 and 3. In fact the most sophisticated piece of notation which it is essential that all should understand in order to read the main body of the text is the expression

$$\sum_{i=1}^{n} x_i,$$

representing the sum of n numbers indexed by the subscript i, thus: $x_1 + x_2 + x_3 + \ldots + x_n$. Also it is helpful if the reader understands differentiation, though this is not strictly essential. Those who are happy with mathematical notation may wish to refer directly to Appendix A in which formal definitions are listed, and where proofs of some of the assertions in the text are given. Appendix A also serves as a glossary of symbols used for inequality measures and other expressions.

Associated with this book there is a website with links to data sources, downloadable spreadsheets of constructed datasets, and examples and presentation files showing the step-by-step developments of some arguments and techniques. Although you should be able to read the text without having to use the website, I am firmly of the opinion that many of the issues in inequality measurement can only be properly understood through experience with practical examples. There are quite a few numerical examples included in the text and several more within the questions and problems at the end of each chapter: you may well find that the easiest course is to pick up the data for these straight from the website rather than doing them by hand or keying the numbers into a computer yourself. This is described further in the Appendix A (page 177), but to get going with the data you only go to the welcome page of the website.

Acknowledgements

This book is in fact the third edition of a project that started a long time ago. So I have many years' worth of intellectual debt that I would like to break up into three tranches:

Acknowledgements from the First Edition

I would like to thank Professor M. Bronfenbrenner for the use of the table on page 98. The number of colleagues and students who wilfully submitted themselves to reading drafts of this book was most gratifying. So I am very thankful for the comments of Tony Atkinson, Barbara Barker, John Bridge, David Collard, Shirley Dex, Les Fishman, Peter Hart, Kiyoshi Kuga, H. F. Lydall, M. D. McGrath, Neville Norman, and Richard Ross; without them there would have been lots more mistakes. You, the reader, owe a special debt to Mike Harrison, John Proops, and Mike Pullen who persistently made me make the text more intelligible. Finally, I am extremely grateful for the skill and patience of Sylvia Beech, Stephanie Cooper, and Judy Gill, each of whom has had a hand in producing the text; 'so careful of the type she seems', as Tennyson once put it.

Acknowledgements from the Second Edition

In preparing the second edition I received a lot of useful advice and help, particularly from past and present colleagues in STICERD. Special thanks go to Tony Atkinson, Karen Gardiner, John Hills, Stephen Jenkins, Peter Lambert, John Micklewright, and Richard Vaughan for their comments on the redrafted chapters. Z. M. Kmietowicz kindly gave permission for the use of his recent work in Question 8 on page 152. Christian Schlüter helped greatly with updating the literature notes and references. Also warm appreciation to Elisabeth Backer and Jumana Saleheen without whose unfailing assistance the revision would have been completed in half the time.

Acknowledgements for the Present Edition

I am very grateful for extended discussions with and support from Guillermo Cruces and for detailed comments from Kristof Bosmans, Udo Ebert, Marc Fleurbaey, Wulf Gaertner, Stephen Jenkins, and Dirk Van de gaer. For much-needed help in updating the bibliography and data sources my thanks go to Yinfei Dong, Elena Pisano, Alex Teytelboym, and Zhijun Zhang.

STICERD, LSE

Contents

List of Figures

This page intentionally left blank

List of Tables

This page intentionally left blank

1

First Principles

> 'It is better to ask some of the questions than to know all of the answers.'
>
> James Thurber (1945), *The Scotty Who Knew Too Much*

'Inequality' is in itself an awkward word, as well as one used in connection with a number of awkward social and economic problems. The difficulty is that the word can trigger quite a number of different ideas in the mind of a reader or listener, depending on his training and prejudice.

'Inequality' obviously suggests a departure from some idea of equality. This may be nothing more than an unemotive mathematical statement, in which case 'equality' just represents the fact that two or more given quantities are the same size, and 'inequality' merely relates to differences in these quantities. On the other hand, the term 'equality' evidently has compelling social overtones as a standard which it is presumably feasible for society to attain. The meaning to be attached to this is not self-explanatory. Some years ago Professors Rein and Miller revealingly interpreted this standard of equality in nine separate ways

- *One-hundred-percentism*: in other words, complete horizontal equity—'equal treatment of equals'.
- *The social minimum*: here one aims to ensure that no one falls below some minimum standard of well-being.
- *Equalization of lifetime income profiles*: this focuses on inequality of future income prospects, rather than on the people's current position.
- *Mobility*: that is, a desire to narrow the differentials and to reduce the barriers between occupational groups.
- *Economic inclusion*: the objective is to reduce or eliminate the feeling of exclusion from society caused by differences in incomes or some other endowment.

- *Income shares*: society aims to increase the share of national income (or some other 'cake') enjoyed by a relatively disadvantaged group—such as the lowest tenth of income recipients.
- *Lowering the ceiling*: attention is directed towards limiting the share of the cake enjoyed by a relatively advantaged section of the population.
- *Avoidance of income and wealth crystallization*: this just means eliminating the disproportionate advantages (or disadvantages) in education, political power, social acceptability and so on that may be entailed by an advantage (or disadvantage) in the income or wealth scale.
- *International yardsticks*: a nation takes as its goal that it should be no more unequal than another 'comparable' nation.

Their list is probably not exhaustive and it may include items which you do not feel properly belong on the agenda of inequality measurement; but it serves to illustrate the diversity of views about the nature of the subject—let alone its political, moral or economic significance—which may be present in a reasoned discussion of equality and inequality. Clearly, each of these criteria of 'equality' would influence in its own particular way the manner in which we might define and measure inequality. Each of these potentially raises particular issues of social justice that should concern an interested observer. And if I were to try to explore just these nine suggestions with the fullness that they deserve, I should easily make this book much longer than I wish.

In order to avoid this mishap let us drastically reduce the problem by trying to set out what the essential ingredients of a Principle of Inequality Measurement should be. We shall find that these basic elements underlie a study of equality and inequality along almost any of the nine lines suggested in the brief list given above.

The ingredients are easily stated. For each ingredient it is possible to use materials of high quality—with conceptual and empirical nuances finely graded. However, in order to make rapid progress, I have introduced some cheap substitutes which I have indicated in each case in the following list:

- Specification of an individual social unit such as a single person, the nuclear family or the extended family. I shall refer casually to 'persons'.
- Description of a particular attribute (or attributes) such as income, wealth, land-ownership or voting strength. I shall use the term 'income' as a loose coverall expression.
- A method of representation or aggregation of the allocation of 'income' among the 'persons' in a given population.

The list is simple and brief, but it will take virtually the whole book to deal with these fundamental ingredients, even in rudimentary terms.

1.1 A preview of the book

The final item on the list of ingredients will command much of our attention. As a quick glance ahead will reveal we shall spend quite some time looking at intuitive and formal methods of aggregation in Chapters 2 and 3. In Chapter 2 we encounter several standard measurement tools that are often used and sometimes abused. This will be a chapter of 'ready-mades' where we take as given the standard equipment in the literature without particular regard to its origin or the principles on which it is based. By contrast the economic analysis of Chapter 3 introduces specific distributional principles on which to base comparisons of inequality. This step, incorporating explicit criteria of social justice, is done in three main ways: social welfare analysis, the concept of distance between income distributions, and an introduction to the axiomatic approach to inequality measurement. On the basis of these principles we can appraise the tailor-made devices of Chapter 3 as well as the off-the-peg items from Chapter 2. Impatient readers who want a quick summary of most of the things one might want to know about the properties of inequality measures could try turning to page 74 for an instant answer.

Chapter 4 approaches the problem of representing and aggregating information about the income distribution from a quite different direction. It introduces the idea of modelling the income distribution rather than just taking the raw bits and pieces of information and applying inequality measures or other presentational devices to them. In particular we deal with two very useful functional forms of income distribution that are frequently encountered in the literature.

In my view the ground covered by Chapter 5 is essential for an adequate understanding of the subject matter of this book. The practical issues which are discussed there put meaning into the theoretical constructs with which you will have become acquainted in Chapters 2 to 4. This is where you will find discussion of the practical importance of the choice of income definition (ingredient 1) and of income receiver (ingredient 2); of the problems of using equivalence scales to make comparisons between heterogeneous income units and of the problems of zero values when using certain definitions of income. In Chapter 5 also we shall look at how to deal with patchy data, and how to assess the importance of inequality changes empirically.

The back end of the book contains two further items that you may find helpful. Appendix A has been used mainly to tidy away some of the more cumbersome formulas which would otherwise have cluttered the text; you may want to dip into it to check up on the precise mathematical definitions and results that are described verbally or graphically in the main text. Appendix B (Notes on Sources and Literature) has been used mainly to cover literature references which would otherwise have also cluttered the text; if you want to follow up the principal articles on a specific topic, or to track down the reference containing detailed proof of some of the key results, this is where you should turn first; it also gives you the background to the data examples found throughout the book.

Finally, a word or two about this chapter. The remainder of the chapter deals with some of the issues of principle concerning all three ingredients on the list; it provides some forward pointers to other parts of the book where theoretical niceties or empirical implementation is dealt with more fully; it also touches on some of the deeper philosophical issues that underpin an interest in the subject of measuring inequality. It is to theoretical questions about the second of the three ingredients of inequality measurement that we shall turn first.

1.2 Inequality of what?

Let us consider some of the problems of the definition of a personal attribute, such as income, that is suitable for inequality measurement. This attribute can be interpreted in a wide sense if an overall indicator of social inequality is required, or in a narrow sense if one is concerned only with inequality in the distribution of some specific attribute or talent. Let us deal first with the special questions raised by the former interpretation.

If you want to take inequality in a global sense, then it is evident that you will need a comprehensive concept of 'income'—an index that will serve to represent generally a person's well-being in society. There are a number of personal economic characteristics which spring to mind as candidates for such an index—for example, wealth, lifetime income, weekly or monthly income. Will any of these do as an all-purpose attribute?

While we might not go as far as Anatole France in describing wealth as a 'sacred thing', it has an obvious attraction for us (as students of inequality). For wealth represents a person's total immediate command over resources. Hence, for each man or woman we have an aggregate which includes the money in the bank, the value of holdings of stocks and bonds, the value of the house and the car, his ox, his ass, and everything that he has. There are two difficulties with this. First, how are these disparate possessions to be valued and aggregated in money terms? It is not clear that prices ruling in the market (where such markets exist) appropriately reflect the relative economic power inherent in these various assets. Second, there are other, less tangible assets which ought perhaps to be included in this notional command over resources, but which a conventional valuation procedure would omit.

One major example of this is a person's occupational pension rights: having a job that entitles me to a pension upon my eventual retirement is certainly valuable, but how valuable? Such rights may not be susceptible to being cashed in like other assets so that their true worth is tricky to assess.

A second important example of such an asset is the presumed prerogative of higher future incomes accruing to those possessing greater education or training. Surely the value of these income rights should be included in the calculation of a person's wealth just as is the value of other income-yielding assets such as stocks or bonds? To do this we need an aggregate of earnings over the entire life span. Such an aggregate—'lifetime income'—in conjunction with other forms of wealth appears to yield the index of personal well-being that we seek, in that it includes in a comprehensive fashion the entire set of economic opportunities enjoyed by a person. The drawbacks, however, are manifest. Since lifetime summation of actual income receipts can only be performed once the income recipient is deceased (which limits its operational usefulness), such a summation must be carried out on anticipated future incomes. Following this course we are led into the difficulty of forecasting these income prospects and of placing on them a valuation that appropriately allows for their uncertainty. Although I do not wish to assert that the complex theoretical problems associated with such lifetime aggregates are insuperable, it is expedient to turn, with an eye on Chapter 5 and practical matters, to income itself.

Income—defined as the increase in a person's command over resources during a given time period—may seem restricted in comparison with the all-embracing nature of wealth or lifetime income. It has the obvious disadvantages that it relates only to an arbitrary time unit (such as one year) and thus that it excludes the effect of past accumulations except in so far as these are deployed in income-yielding assets. However, there are two principal offsetting merits:

- if income includes unearned income, capital gains, and 'income in kind' as well as earnings, then it can be claimed as a fairly comprehensive index of a person's well-being at a given moment;
- information on personal income is generally more widely available and more readily interpretable than for wealth or lifetime income.

Furthermore, note that none of the three concepts that have been discussed completely covers the command over resources for all goods and services in society. Measures of personal wealth or income exclude 'social wage' elements such as the benefits received from communally enjoyed items like municipal parks, public libraries, the police, and ballistic missile systems, the interpersonal distribution of which services may only be conjectured.

In view of the difficulty inherent in finding a global index of 'well-offness', we may prefer to consider the narrow definition of the thing called 'income'. Depending on the problem in hand, it can make sense to look at inequality in the endowment of some other personal attribute, such as consumption of a particular good, life expectancy, land ownership, etc. This may be applied also to publicly owned assets or publicly consumed commodities if we direct attention not to *interpersonal* distribution but to *intercommunity* distribution—for example, the inequality in the distribution of *per capita* energy consumption in different countries. The problems concerning 'income' that I now discuss apply with equal force to the wider interpretation considered in the earlier paragraphs.

It is evident from the foregoing that two key characteristics of the 'income' index are that it be *measurable* and that it be *comparable* among different persons. That these two characteristics are mutually independent can be demonstrated by two contrived examples. First, to show that an index might be measurable but not comparable, take the case where well-being is measured by consumption per head within families, the family rather than the individual being taken as the basic social unit. Suppose that consumption by each family in the population is known but that the number of persons is not. Then for each family, welfare is measurable up to an arbitrary change in scale, in this sense: for family A doubling its income makes it twice as well-off, trebling it makes it three times as well-off; the same holds for family B; but A's welfare scale and B's welfare scale cannot be compared unless we know the numbers in each family. Second, to show that an index may be interpersonally comparable, but not measurable in the conventional sense, take the case where 'access to public services' is used as an indicator of welfare. Consider two public services, gas and electricity supply—households may be connected to one or to both or to neither of them—and the following scale (in descending order of amenity) is generally recognized:

- access to both gas and electricity;
- access to electricity only;
- access to gas only;
- access to neither.

We can compare households' amenities—A and B are as well-off as each other if they are both connected only to electricity—but it makes no sense to say that A is *twice* as well-off if it is connected to gas as well as electricity.

It is possible to make some progress in the study of inequality without measurability of the welfare index, and sometimes even without full comparability. For most of the time, however, I shall make both these assumptions, which may be unwarranted. For this implies that when I write the word 'income', I assume that it is so defined that adjustment has already been made for non-comparability on account of differing needs, and that fundamental differences in tastes (with regard to relative valuation of leisure and monetary income, for example) may be ruled out of consideration. We shall reconsider the problems of non-comparability in Chapter 5.

The final point in connection with the 'income' index that I shall mention can be described as the 'constant amount of cake'. We shall usually talk of inequality freely as though there is some fixed total of goodies to be shared among the population. This is definitionally true for certain quantities, such as the distribution of acres of land (except perhaps in the Netherlands). However, this is evidently questionable when talking about income as conventionally defined in economics. If an arbitrary change is envisaged in the distribution of income among persons, we may reasonably expect that the size of the cake to be divided—national income—might change as a result. Or, if we try to compare inequality in a particular country's income distribution at two points in time, it is quite likely that total income will have changed during the interim. Moreover, if the size of the cake changes, either autonomously or as a result of some redistributive action, this change in itself may modify our view of the amount of inequality that there is in society.

Having raised this important issue of the relationship between interpersonal distribution and the production of economic goods, I shall temporarily evade it by assuming that a given whole is to be shared as a number of equal or unequal parts. For some descriptions of inequality this assumption is irrelevant. However, since the size of the cake as well as its distribution is very important in social welfare theory, we shall consider the relationship between inequality and total income in Chapter 3 (particularly page 48), and examine the practical implications of a growing—or dwindling—cake in Chapter 5 (see page 143.)

1.3 Inequality measurement, justice, and poverty

So what is meant by an inequality measure? In order to introduce this device which serves as the third 'ingredient' mentioned previously, let us try a simple definition which roughly summarizes the common usage of the term:

- a scalar numerical representation of the interpersonal differences in income within a given population.

Now let us take this bland statement apart.

Scalar Inequality

The use of the word 'scalar' implies that all the different features of inequality are compressed into a single number—or a single point on a scale. Appealing arguments can be produced against the contraction of information involved in this aggregation procedure. Should we don this one-dimensional strait-jacket when surely our brains are well-developed enough to cope with more than one number at a time? There are three points in reply here.

First, if we want a multi-number representation of inequality, we can easily arrange this by using a variety of indices each capturing a different characteristic of the social state, and each possessing attractive properties as a yardstick of inequality in its own right. We shall see some practical examples (in Chapters 3 and 5) where we do exactly that.

Second, however, we often want to answer a question like 'has inequality increased or decreased?' with a straight 'yes' or 'no'. But if we make the concept of inequality multi-dimensional we greatly increase the possibility of coming up with ambiguous answers. For example, suppose we represent inequality by two numbers, each describing a different aspect of inequality of the same 'income' attribute. We may depict this as a point such as B in Fig. 1.1, which reveals that there is an amount I_1 of type-1 inequality, and I_2 of type-2 inequality. Obviously all points like C represent states of society that are more unequal than B, and points such as A represent less unequal states. But it is much harder to compare B and D or to compare B and E. If we attempt to resolve this difficulty, we will find that we are effectively using a single-number representation of inequality after all.

Third, multi-number representations of income distributions may well have their place alongside a standard scalar inequality measure. As we shall

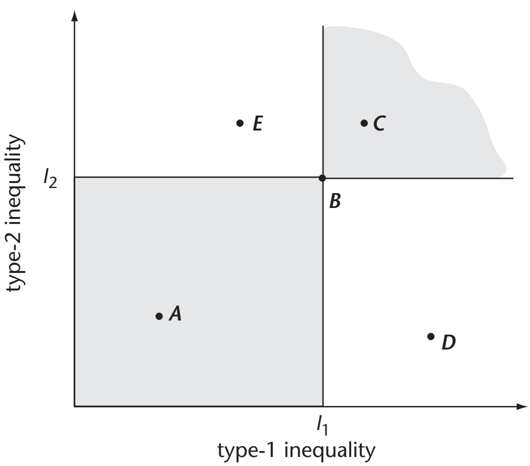

FIG. 1.1. Two types of inequality

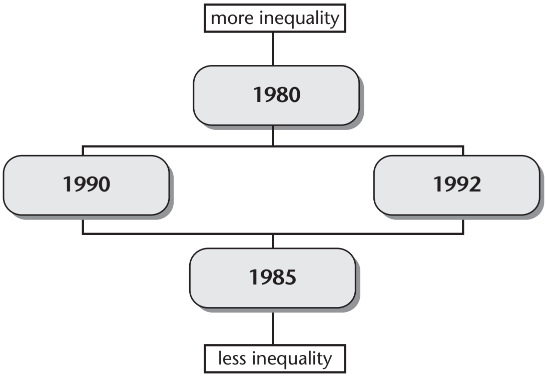

FIG. 1.2. An inequality ranking

see in later chapters, even if a single agreed number scale (I_1 or I_2) is unavailable, or even if a collection of such scales (I_1 and I_2) cannot be found, we might be able to agree on an inequality ranking. This is a situation where—although you may not be able to order or to sort the income distributions uniquely (most equal at the bottom, most unequal at the top)—you nevertheless find that you can arrange them in a pattern that enables you to get a fairly useful picture of what is going on. To get the idea, have a look at Fig. 1.2. We might find that over a period of time the complex changes in the relevant income distribution can be represented schematically as in the league table illustrated there: you can say that inequality went down from 1980 to 1985, and went up from 1985 to either 1990 or 1992; but you cannot say whether inequality went up or down in the early nineties. Although this method of looking at inequality is not decisive in terms of every possible comparison of distributions, it could still provide valuable information.

Numerical Representation

What interpretation should be placed on the phrase 'numerical representation' in the definition of an inequality measure? The answer to this depends on whether we are interested in just the ordering properties of an inequality measure or in the actual size of the index and of changes in the index.

To see this, look at the following example. Imagine four different social states A, B, C, D, and four rival inequality measures I_1, I_2, I_3, I_4. The first column in Table 1.1 gives the values of the first measure, I_1, realized in each of the four situations. Are any of the other candidates equivalent to I_1? Notice that I_3 has a strong claim in this regard. Not only does it rank A, B, C, D in the same order, it also shows that the percentage change in inequality in going from one state to another is the same as if we use the I_1 scale. If this is true for all social states, we will call I_1 and I_3 *cardinally*

Table 1.1. Four inequality scales

	I_1	I_2	I_3	I_4
A	.10	.13	.24	.12
B	.25	.26	.60	.16
C	.30	.34	.72	.20
D	.40	.10	.96	.22

equivalent. More formally, I_1 and I_3 are cardinally equivalent if one scale can be obtained from the other, multiplying by a positive constant and adding or subtracting another constant. In the above case, we multiply I_1 by 2.4 and add on zero to get I_3. Now consider I_4: it ranks the four states A to D in the same order as I_1, but it does not give the same percentage differences (compare the gaps between A and B and between B and C). So I_1 and I_4 are certainly not cardinally equivalent. However, if it is true that I_1 and I_4 always rank any set of social states in the same order, we will say that the two scales are *ordinally equivalent*.[1] Obviously cardinal equivalence entails ordinal equivalence, but not vice versa. Finally we note that I_2 is not ordinally equivalent to the others, although for all we know it may be a perfectly sensible inequality measure.

Now let A be the year 1970, let B be 1960, and D be 1950. Given the question, 'Was inequality less in 1970 than it was in 1960?', I_1 produces the same answer as any other ordinally equivalent measure (such as I_3 or I_4): 'numerical representation' simply means a ranking. But, given the question, 'Did inequality fall more in the 1960s than it did in the 1950s?', I_1 only yields the same answer as other cardinally equivalent measures (I_3 alone): here inequality needs to have the same kind of 'numerical representation' as temperature on a thermometer.

Income Differences

Should any and every 'income difference' be reflected in a measure of inequality? The commonsense answer is 'No', for two basic reasons: need and merit. The first reason is the more obvious: large families and the sick need more resources than the single, healthy person to support a particular economic standard. Hence in a 'just' allocation, we would expect those with such greater needs to have a higher income than other people; such income differences would thus be based on a principle of justice, and should

[1] A mathematical note: I_1 and I_4 are ordinally equivalent if one may be written as a monotonically increasing function of the other, say $I_1 = f(I_4)$, where $dI_1/dI_4 > 0$. An example of such a function is $\log(I)$. I_1 and I_3 are cardinally equivalent if f takes the following special form: $I_1 = a + bI_3$, where b is a positive number.

not be treated as inequalities. To cope with this difficulty one may adjust the income concept such that allowance is made for diversity of need, as mentioned in the last section; this is something which needs to be done with some care—as we will find in Chapter 5 (see the discussion on page 110).

The case for ignoring differences on account of merit depends on the interpretation attached to 'equality'. One obviously rough-and-ready description of a just allocation requires equal incomes for all irrespective of personal differences other than need. However, one may argue strongly that in a just allocation higher incomes should be received by doctors, heroes, inventors, Stakhanovites, and other deserving persons. Unfortunately, in practice it is more difficult to make adjustments similar to those suggested in the case of need and, more generally, even distinguishing between income differences that do represent genuine inequalities and those that do not poses a serious problem.

Given Population

The last point about the definition of an inequality measure concerns the phrase 'given population' and needs to be clarified in two ways. First, when examining the population over say a number of years, what shall we do about the effect on measured inequality of persons who either enter or leave the population, or whose status changes in some other relevant way? The usual assumption is that as long as the overall structure of income differences stays the same (regardless of whether different personnel are now receiving those incomes), measured inequality remains unaltered. Hence the phenomenon of social mobility within or in-and-out of the population eludes the conventional method of measuring inequality, although some might argue that it is connected with inequality of *opportunity*.[2] Secondly, one is not exclusively concerned with inequality in the population as a whole. It is useful to be able to decompose this 'laterally' into inequality within constituent groups, differentiated regionally or demographically perhaps, and inequality between these constituent groups. Indeed, once one acknowledges basic heterogeneities within the population, such as age or sex, awkward problems of aggregation may arise, although we shall ignore them. It may also be useful to decompose inequality 'vertically' so that one looks at inequality within a subgroup of the rich or of the poor, for example. Hence the specification of the given population is by no means a trivial prerequisite to the application of inequality measurement.

[2] Check Question 6 at the end of the chapter to see if you concur with this view.

Although the definition has made it clear that an inequality measure calls for a numerical scale, I have not suggested how this scale should be calibrated. Specific proposals for this will occupy Chapters 2 and 3, but a couple of basic points may be made here.

You may have noticed just now that the notion of justice was slipped in while income differences were being considered. In most applications of inequality analysis social justice really ought to be centre stage. That more just societies should register lower numbers on the inequality scale evidently accords with an intuitive appreciation of the term ‘inequality’. But, on what basis should principles of distributional justice and concern for inequality be based? Economic philosophers have offered a variety of answers. This concern could be no more than the concern about the everyday risks of life: just as individuals are upset by the financial consequences of having their car stolen or missing their plane, so too they would care about the hypothetical risk of drawing a losing ticket in a lottery of life chances; this lottery could be represented by the income distribution in the UK, the USA, or wherever; nice utilitarian calculations on the balance of small-scale gains and losses become utilitarian calculations about life chances; aversion to risk translates into aversion to inequality. Or the concern could be based upon the altruistic feelings of each human towards his fellows that motivates charitable action. Or again it could be that there is a social imperative toward concern for the least advantaged—and perhaps concern about the inordinately rich—that transcends the personal twinges of altruism and envy. It could be simple concern about the possibility of social unrest. It is possible to construct a coherent justice-based theory of inequality measurement on each of these notions, although that takes us beyond the remit of this book.

However, if we can clearly specify what a just distribution is, such a state provides the zero from which we start our inequality measure. But even a well-defined principle of distributive justice is not sufficient to enable one to mark off an inequality scale unambiguously when considering diverse unequal social states. Each of the apparently contradictory scales I_1 and I_2 considered in Fig. 1.1 and Table 1.1 might be solidly founded on the same principle of justice, unless such a principle were extremely narrowly defined.

The other general point is that we might suppose there is a close link between an indicator of the extent of poverty and the calibration of a measure of economic inequality. This is not necessarily so, because two rather different problems are generally involved. In the case of the measurement of poverty, one is concerned primarily with that segment of the population falling below some specified ‘poverty line’; to obtain the poverty measure one may perform a simple head count of this segment, or calculate the gap between the average income of the poor and the average income of the general population, or carry out some other computation on poor people's incomes in relation to each other and to the rest of the population. Now, in the case of inequality one generally wishes to capture the effects of income differences over a much wider range. Hence it is perfectly possible for the measured extent of poverty to be declining over time, while at the same time and in the same society measured inequality increases due to changes in income differences within the non-poor segment of the population, or because of migrations between the two groups. (If you are in doubt about this you might like to have a look at Question 5 on page 14.) Poverty will make a few guest appearances in the course of this book, but on the whole our discussion of inequality has to take a slightly different track from the measurement of poverty.

1.4 Inequality and the social structure

Finally we return to the subject of the first ingredient, namely the basic social units used in studying inequality—or the elementary particles of which we imagine society to be constituted. The definition of the social unit, whether it be a single person, a nuclear family, or an extended family depends intrinsically upon the social context, and upon the interpretation of inequality that we impose. Although it may seem natural to adopt an individualistic approach, some other 'collective' unit may be more appropriate.

When economic inequality is our particular concern, the theory of the development of the distribution of income or wealth may itself influence the choice of the basic social unit. To illustrate this, consider the classical view of an economic system, the population being subdivided into distinct classes of workers, capitalists, and landowners. Each class is characterized by a particular function in the economic order and by an associated type of income—wages, profits, and rents. If, further, each is regarded as internally fairly homogeneous, then it makes sense to pursue the analysis of inequality in class terms rather than in terms of individual units.

However, so simple a model is unsuited to describing inequality in a significantly heterogeneous society, despite the potential usefulness of class analysis for other social problems. A superficial survey of the world around us reveals rich and poor workers, failed and successful capitalists, and several people whose rôles and incomes do not fit into neat slots. Hence the focus of attention in this book is principally upon individuals rather than types, whether the analysis is interpreted in terms of economic inequality or some other sense.

Thus reduced to its essentials it might appear that we are dealing with a purely formal problem, which sounds rather dull. This is not so. Although the subject matter of this book is largely technique, the techniques involved are essential for coping with the analysis of many social and economic problems in a systematic fashion; and these problems are far from dull or uninteresting.

1.5 Questions

1. In Syldavia the economists find that (annual) household consumption c is related to (annual) income y by the formula

   $$c = \alpha + \beta y,$$

   where $\alpha > 0$ and $0 < \beta < 1$. Because of this, they argue, inequality of consumption must be less than inequality of income. Provide an intuitive argument for this.
2. Ruritanian society consists of three groups of people: Artists, Bureaucrats and Chocolatiers. Each Artist has high income (15,000 Ruritanian Marks) with a 50 per cent probability, and low income (5000 RM) with 50 per cent probability. Each Bureaucrat starts working life on a salary of 5000 RM and then benefits from an annual increment of 250 RM over the 40 years of his (perfectly safe) career. Chocolatiers get a straight annual wage of 10,000 RM. Discuss the extent of inequality in Ruritania according to annual income and lifetime income concepts.
3. In Borduria the government statistical service uses an inequality index that in principle can take any value greater than or equal to 0. You want to introduce a transformed inequality index that is ordinally equivalent to the original but that will always lie between zero and 1. Which of the following will do?

   $$\frac{I}{I+1}, \sqrt{\frac{I}{I+1}}, \frac{I}{I-1}, \sqrt{I}.$$

4. Methods for analysing inequality of income could be applied to inequality of use of specific health services (Williams and Doessel 2006). What would be the principal problems of trying to apply these methods to inequality of health *status*?
5. After a detailed study of a small village, government experts reckon that the poverty line is 100 rupees a month. In January a joint team from the Ministry of Food and the Central Statistical Office carry out a survey of living standards in the village: the income for each villager (in rupees per month) is recorded. In April the survey team repeats the

exercise. The number of villagers was exactly the same as in January, and villagers' incomes had changed only slightly. An extract from the results is as follows:

January	April
...	...
...	...
92	92
95	92
98	101
104	104
...	...
...	...

(the dots indicate the incomes of all the other villagers for whom income did not change at all from January to April). The Ministry of Food writes a report claiming that poverty has fallen in the village; the Central Statistical Office writes a report claiming that inequality has risen in the village. Can they both be right? (See Thon 1979, 1981, 1983b for more on this.)

6. In Fantasia there is a debate about educational policy. The current situation is that there are two equal-sized groups of people, the Dark-greys who all get an income of \$200, and the Light-greys who all get an income of \$600, as in the top part of the accompanying diagram, labelled 'Parents'. One group of educational experts argue that if the Fantasian government adopts policy A then the future outcome for the next generation will be as shown on the left side of the diagram, labelled

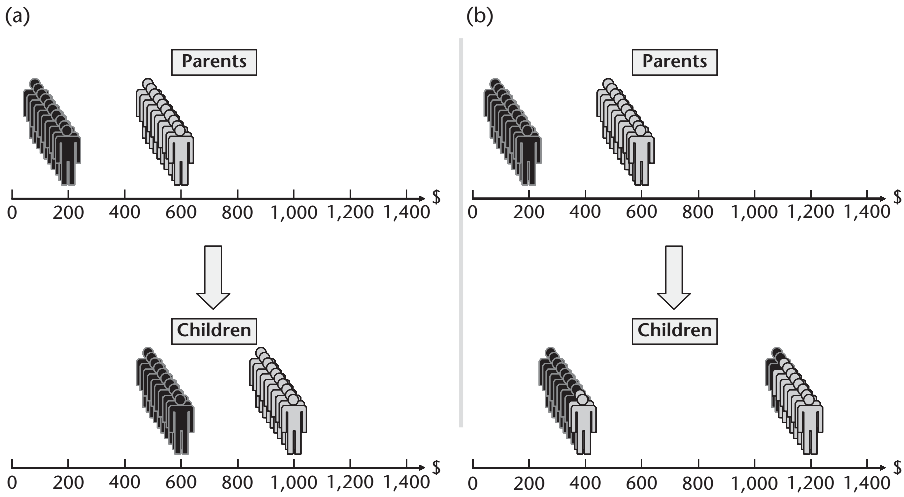

FIG. 1.3. Alternative policies for Fantasia

'Children'; another group of experts argue that if policy B is adopted, the outcome for the next generation will be that on the right side of the diagram (shading shows are used to show whether the children come from Dark-grey families or Light-grey families). According to your view:

- Which of policies A and B would produce lower inequality of outcome?
- Which policy produces higher social mobility?
- Which policy is characterized by lower inequality of opportunity?

2

Charting Inequality

F. Scott Fitzgerald: 'The rich are different from us.'

Ernest Hemingway: 'Yes, they have more money.'

If society really did consist of two or three fairly homogeneous groups, economists and others could be saved a lot of trouble. We could then simply look at the division of income between landlords and peasants, among workers, capitalists, and rentiers, or any other appropriate sections. Naturally we would still be faced with such fundamental issues as how much each group should possess or receive, whether the statistics are reliable, and so on, but questions such as 'what is the income distribution?' could be satisfactorily met with a snappy answer '65 per cent to wages, 35 per cent to profits'. Of course matters are not that simple. As we have argued, we want a way of looking at inequality that reflects both the depth of poverty of the 'have nots' of society and the height of well-being of the 'haves': it is not easy to do this just by looking at the income accruing to, or the wealth possessed by, two or three groups.

So in this chapter we will look at several quite well-known ways of presenting inequality in a large heterogeneous group of people. They are all methods of appraising the sometimes quite complicated information that is contained in an income distribution, and they can be grouped under three broad headings: diagrams, inequality measures, and rankings. To make the exposition easier I shall continue to refer to 'income distribution', but you should bear in mind, of course, that the principles can be carried over to the distribution of any other variable that you can measure and that you think is of economic interest.

2.1 Diagrams

Putting information about income distribution into diagrammatic form is a particularly instructive way of representing some of the basic ideas about inequality. There are several useful ways of representing inequality in pictures; the four that I shall discuss are introduced in the accompanying box. Let us have a closer look at each of them.

- **Parade of Dwarfs**
- **Frequency distribution**
- **Lorenz curve**
- **Log transformation**

PICTURES OF INEQUALITY

Jan Pen's *Parade of Dwarfs* is one of the most persuasive and attractive visual aids in the subject of income distribution. Suppose that everyone in the population had a height proportional to his or her income, with the person on average income being endowed with average height. Line people up in order of height and let them march past in some given time interval—let us say one hour. Then the sight that would meet our eyes is represented by the curve in Fig. 2.1.[1] The whole parade passes in the interval represented by OC. But we do not meet the person with average income until we get to the point B (when well over half the parade has gone by). Divide total income by total population: this gives average or mean income ($\bar{y}$) and is represented by the height OA. We have oversimplified Pen's original diagram by excluding from consideration people with negative reported incomes, which would involve the curve crossing the base line towards its left-hand end. And, in order to keep the diagram on the page, we have plotted the last point of the curve (D) in a position that would be far too low in practice.

This diagram highlights the presence of any extremely large incomes and, to a certain extent, abnormally small incomes. But we may have reservations about the degree of detail that it seems to impart concerning middle income receivers. We shall see this point recur when we use this diagram to derive an inequality measure that informs us about changes in the distribution.

Frequency distributions are well-tried tools of statisticians, and are discussed here mainly for the sake of completeness and as an introduction for those unfamiliar with the concept—for a fuller account see the references cited in

[1] Those with especially sharp eyes will see that the source is more than 20 years old. There is a good reason for using these data—see the notes on page 180.

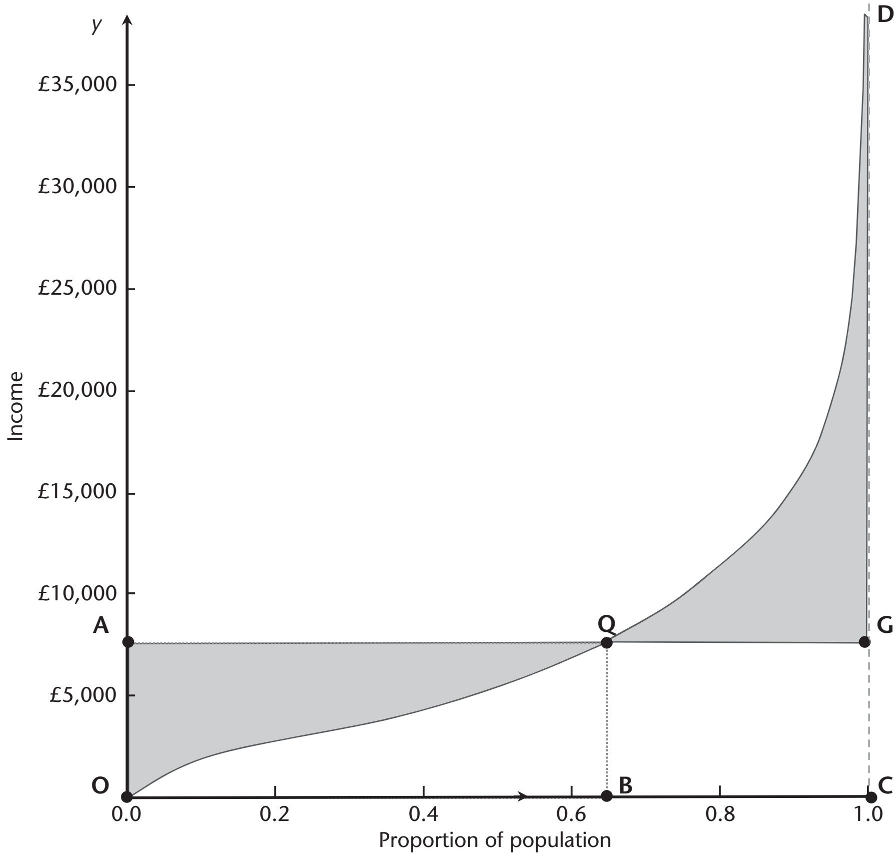

Fig. 2.1. The Parade of Dwarfs UK income before tax, 1984/5
Source: *Economic Trends*, November 1987

the notes to this chapter. An example is found in Fig. 2.2. Suppose you were looking down on a field. On one side, the axis Oy, there is a long straight fence marked off by income categories: the physical distance between any two points along the fence directly corresponds to the income differences they represent. Then, get the whole population to come into the field and line up in the strip of land marked off by the piece of fence corresponding to their income bracket. So the £10,000-to-£12,500-a-year persons stand on the shaded patch. The shape that you get will resemble the stepped line in Fig. 2.2—called a histogram—which represents the frequency distribution. It may be that we regard this as an empirical observation of a theoretical curve which describes the income distribution, for example the smooth curve drawn in Fig. 2.2. The relationship $f(y)$ charted by this curve is sometimes

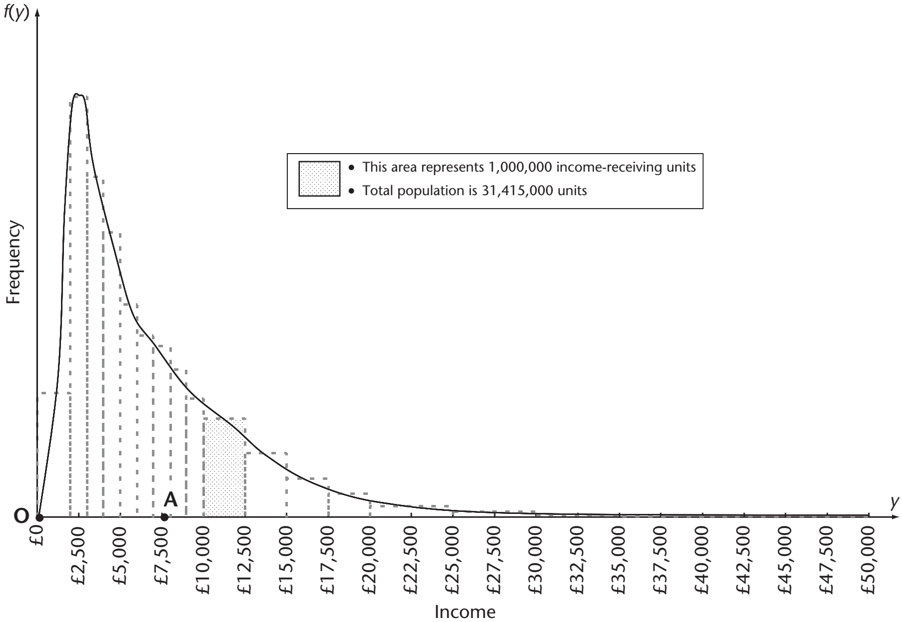

Fig. 2.2. Frequency distribution of income
Source: as for Fig. 2.1

known as a *density function* where the scale is chosen such that the area under the curve and above the line Oy is standardized at unity.

The frequency distribution shows the middle income ranges more clearly. But perhaps it is not so readily apparent what is going on in the upper tail; indeed, in order to draw the figure, we have deliberately made the length of the fence much too short. (On the scale of this diagram it ought to be 100 metres at least!) This diagram and the Parade of Dwarfs are, however, intimately related; and we show this by constructing Fig. 2.3 from Fig. 2.2. The horizontal scale of each figure is identical. On the vertical scale of Fig. 2.3 we plot 'cumulative frequency' written $F(y)$, which is proportional to the area under the curve and to the left of y in Fig. 2.2. If you experiment with the diagram you will see that as you increase y, $F(y)$ usually goes up (it can never decrease)—from a value of zero when you start at the lowest income received, up to a value of one for the highest income. Thus, supposing we consider $y = £30,000$, we plot a point in Fig. 2.3 that corresponds to the proportion of the population with £30,000 or less. And we can repeat this operation for every point on either the empirical curve or on the smooth theoretical curve.

The visual relationship between Figs 2.1 and 2.3 is now obvious. As a further point of reference, the position of mean income has been drawn

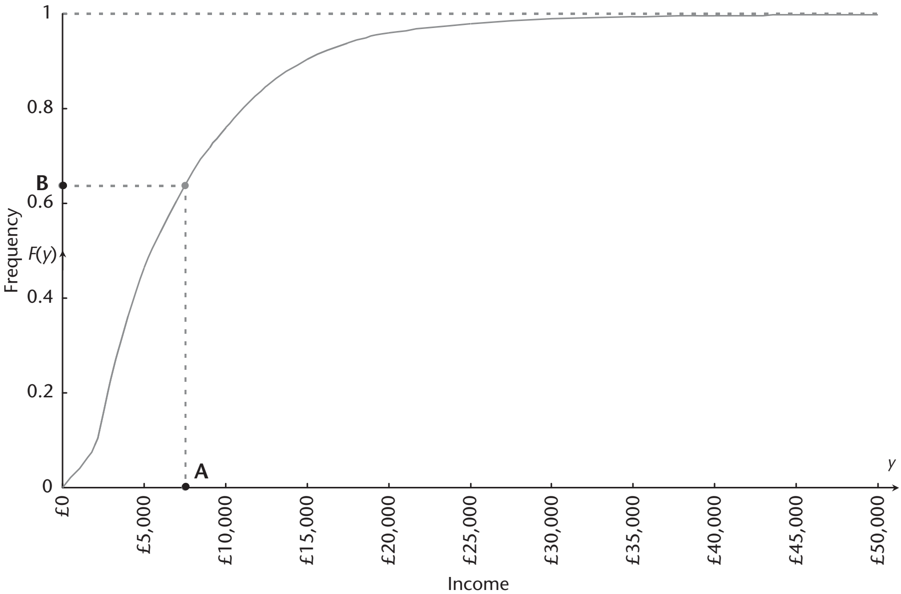

FIG. 2.3. Cumulative frequency distribution
Source: as for Fig. 2.1

in at the point *A* in the two figures. (If you still don't see it, try turning the page round!)

The Lorenz curve was introduced in 1905 as a powerful method of illustrating the inequality of wealth distribution. A simplified explanation of it runs as follows.

Once again, line up everybody in ascending order of incomes and let them parade by. Measure $F(y)$, the proportion of people who have passed by, along the horizontal axis of Fig. 2.4. Once point C is reached everyone has gone by, so $F(y) = 1$. Now as each person passes, hand him his share of the 'cake'—that is, the proportion of total income that he receives. When the parade reaches people with income y, let us suppose that a proportion $\Phi(y)$ of the cake has gone. So of course when $F(y) = 0$, $\Phi(y)$ is also 0 (no cake gone); and when $F(y) = 1$, $\Phi(y)$ is also 1 (all the cake has been handed out). $\Phi(y)$ is measured on the vertical scale in Fig. 2.4, and the graph of Φ plotted against F is the Lorenz curve. Note that it is always convex toward the point C, the reason for which is easy to see. Suppose that the first 10 per cent have filed by ($F(y_1) = 0.1$) and you have handed out 4 per cent of the cake ($\Phi(y_1) = 0.04$); then by the time the next 10 per cent of the people go by ($F(y_2) = 0.2$), you must have handed out at least 8 per cent of the cake ($\Phi(y_2) = 0.08$). Why? Because we arranged the parade in ascending order of cake-receivers. Notice too that if the Lorenz curve lay along OD we would have a state of perfect

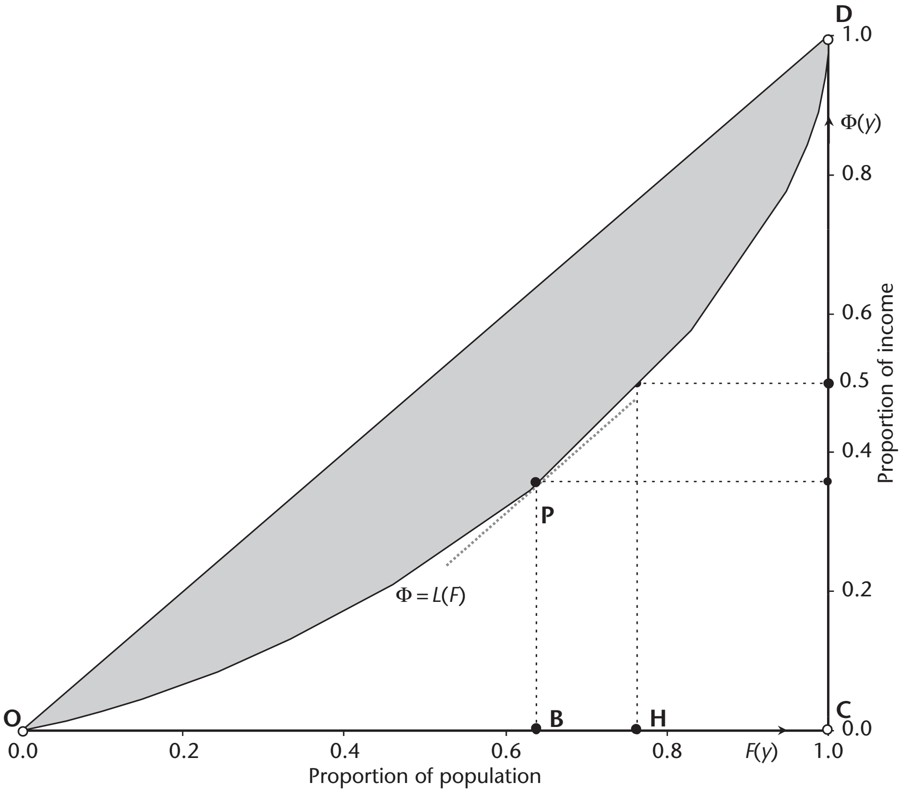

FIG. 2.4. Lorenz curve of income
Source: as for Fig. 2.1

equality, for along that line the first 5 per cent get 5 per cent of the cake, the first 10 per cent get 10 per cent . . . and so on.

The Lorenz curve incorporates some principles that are generally regarded as fundamental to the theory of inequality measurement, as we will see later in this chapter (page 34) and also in Chapter 3 (pages 46 and 62). And again there is a nice relationship with Fig. 2.1. If we plot the slope of the Lorenz curve against the cumulative population proportions, F, then we are back precisely to the Parade of Dwarfs (scaled so that mean income equals unity). Once again, to facilitate comparison, the position where we meet the person with mean income has been marked as point B, although in the Lorenz diagram we cannot represent mean income itself. Note that the mean occurs at a value of F such that the slope of the Lorenz curve is parallel to OD.

Logarithmic transformation. An irritating problem that arises in drawing the frequency curve of Fig. 2.2 is that we must either ignore some of the very large incomes in order to fit the diagram on the page, or put up with a diagram that obscures much of the detail in the middle and lower income ranges. We can avoid this to some extent by drawing a similar frequency distribution, but plotting the horizontal axis on a logarithmic scale as in Fig. 2.5. Equal distances along the horizontal axis correspond to equal proportionate income differences.

Again the point corresponding to mean income, $\bar{y}$, has been marked in as A. Note that the length OA equals $\log(\bar{y})$ and is not the mean of the logarithms of income. This is marked in as the point A′, so that the length OA′ = $\log(y^*)$ where y^* is the geometric mean of the distribution. Assuming incomes are non-negative, the geometric mean, found by taking the mean of the logarithms and then transforming back to natural numbers, can never exceed the conventional arithmetic mean.

We have now seen four different ways of presenting pictorially the same facts about income distribution. Evidently each graphical technique may emphasize quite different features of the distribution: the Parade draws attention to the enormous height of the well-off; the frequency curve presents middle incomes more clearly, the logarithmic transformation captures information from each of the 'tails' as well as the middle, but at the same time sacrifices simplicity and ease of interpretation. This difference

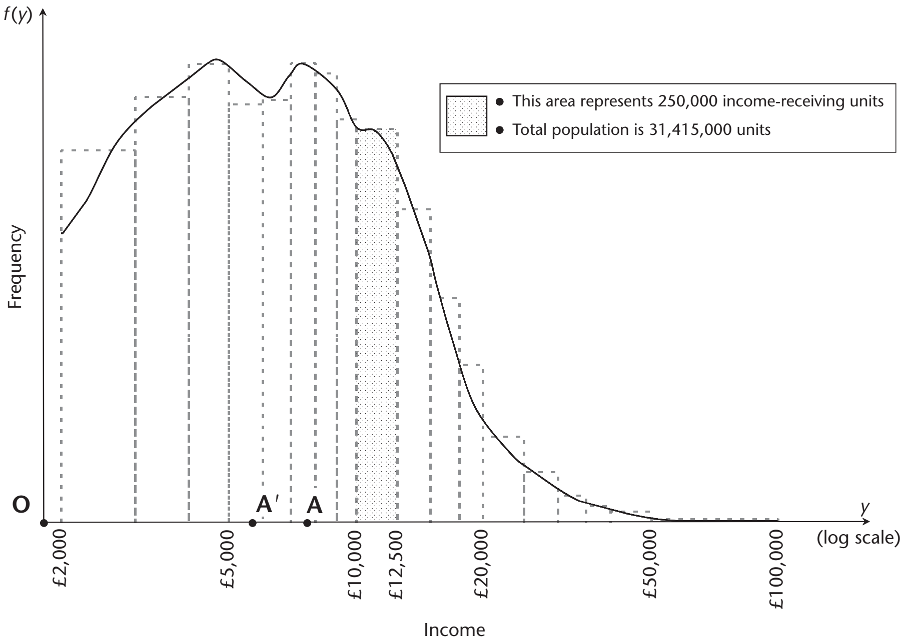

FIG. 2.5. Frequency distribution of income (logarithmic scale)
Source: as for Figure 2.1

in emphasis is partly reflected in the inequality measures derived from the diagrams.

2.2 Inequality measures

We can use Figs 2.1 to 2.5 in order to introduce and illustrate some conventional inequality measures. A few of the more important ones that we shall encounter are listed in the accompanying box. Of course, an inequality measure, like any other tool, is to be judged by the kind of job that it does: is it suitably sensitive to changes in the pattern of distribution? Does it respond appropriately to changes in the overall scale of incomes? As we go through the items in the box we will briefly consider their principal properties: (a proper job must wait until page 67, after we have considered the important analytical points introduced in Chapter 3).

- **Range** R
- **Relative mean deviation** M
- **Variance** V
- **Coefficient of variation** c
- **Gini coefficient** G
- **Log variance** v

INEQUALITY MEASURES

The Parade of Dwarfs suggests the first two of these. First, we have the range, which we define simply as the distance CD in Fig. 2.1 or:

$$R = y_{\max} - y_{\min},$$

where $y_{\max}$ and $y_{\min}$ are, respectively, the maximum and minimum values of income in the parade (we may, of course, standardize by considering $R/y_{\min}$ or $R/\bar{y}$). Plato apparently had this concept in mind when he made the following judgement:

> We maintain that if a state is to avoid the greatest plague of all—I mean civil war, though civil disintegration would be a better term—extreme poverty and wealth must not be allowed to arise in any section of the citizen-body, because both lead to both these disasters. That is why the legislator must now announce the acceptable limits of wealth and poverty. The lower limit of poverty must be the

> value of the holding. The legislator will use the holding as his unit of measure and allow a man to possess twice, thrice, and up to four times its value.
>
> *The Laws*, 745.

The problems with the range are evident. Although it might be satisfactory in a small closed society where everyone's income is known fairly certainly, it is clearly unsuited to large, heterogeneous societies where the 'minimum' and 'maximum' incomes can at best only be guessed. The measure will be highly sensitive to the guesses or estimates of these two extreme values. In practice one might try to get around the problem by using a related concept that is more robust: take the gap between the income of the person who appears exactly at, say, the end of the first three minutes in the Parade, and that of the person exactly at the 57th minute (the bottom 5 per cent and the top 5 per cent of the line of people) or the income gap between the people at the 6th and 54th minute (the bottom 10 per cent and the top 10 per cent of the line of people). However, even if we did that there is a more compelling reason for having doubts about the usefulness of R. Suppose we can wave a wand and bring about a society where the person at position O and the person at position C are left at the same height, but where everyone else in between was levelled to some equal, intermediate height. We would probably agree that inequality had been reduced, though not eliminated. But according to R it is just the same!

You might be wondering whether the problem with R arises because it ignores much of the information about the distribution (it focuses just on a couple of extreme incomes). Unfortunately we shall find a similar criticism in subtle form attached to the second inequality measure that we can read off the Parade diagram, one that uses explicitly the income values of all the individuals. This is the *relative mean deviation*, which is defined as the average absolute distance of everyone's income from the mean, expressed as a proportion of the mean. Take a look at the shaded portions in Fig. 2.1. These portions, which are necessarily of equal size, constitute the area between the Parade curve itself and the horizontal line representing mean income. In some sense, the larger this area, the greater is inequality. (Try drawing the Parade with more giants and more dwarfs.) It is conventional to standardize the inequality measure in unit-free terms, so let us divide by the total income (which equals area OCGA). In terms of the diagram the relative mean deviation is then:[2]

$$M = \frac{\text{area OAQ} + \text{area QGD}}{\text{area OCGA}}.$$

[2] You are invited to check the technical appendix (p. 153) for formal definitions of this and other inequality measures.

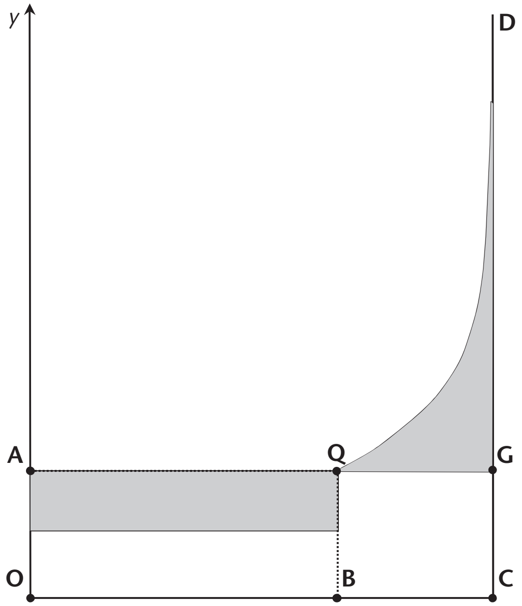

FIG. 2.6. The Parade with partial equalization

But now for the fatal weakness of M. Suppose you think that the stature of the dwarfs to the left of B is socially unacceptable. You arrange a reallocation of income so that everyone with incomes below the mean (to the left of point B) has exactly the same income. The modified parade then looks like Fig. 2.6. But notice that the two shaded regions in Fig. 2.6 are exactly the same area as in Fig. 2.1: so the value of M has not changed. Whatever reallocation you arrange among people to the left of B only, or among people to the right of B only, inequality according to the relative mean deviation stays the same.

The relative mean deviation can be easily derived from the Lorenz curve (Fig. 2.4). From the Technical Appendix, page 156, it can be verified that $M = 2[F(\bar{y}) - \Phi(\bar{y})]$, that is: $M = 2[\text{OB} - \text{BP}]$. However, a more common use of the Lorenz curve diagram is to derive the *Gini coefficient*, G, expressed as the ratio of the shaded area in Fig. 2.4 to the area OCD. There is a variety of equivalent ways of defining G; but perhaps the easiest definition is as the average difference between all possible pairs of incomes in the population, expressed as a proportion of total income: see pages 155 and 156 for a formal definition. The main disadvantage of G is that it places a rather curious implicit relative value on changes that may occur in different parts of the distribution. An income transfer from a relatively rich person to a person with £x less has a much greater effect on G if the two persons are near the middle rather than at either end of the parade.[3] So, consider transferring £1 from a person with £10,100 to a person with £10,000. This has a much greater effect on reducing G than transferring £1 from a person with £1,100 to one with £1,000 or than transferring £1 from a person with £100,100 to a person with £100,000. This valuation may be desirable, but it is not obvious that it is desirable: this point about the valuation of transfers is discussed more fully in Chapter 3 once we have discussed social welfare explicitly.

Other inequality measures can be derived from the Lorenz curve in Fig. 2.4. Two have been suggested in connection with the problem of measuring inequality in the distribution of power, as reflected in voting strength. First, consider the income level y_0 at which half the national cake has been distributed to the parade; that is, $\Phi(y_0) = \frac{1}{2}$. Then define the *minimal majority* inequality measure as $F(y_0)$, which is the distance OH. If Φ is reinterpreted as the proportion of seats in an elected assembly where the votes are spread unevenly among the constituencies, as reflected by the Lorenz curve, and if F is reinterpreted as a proportion of the electorate, then $1 - F(y_0)$ represents the smallest proportion of the electorate that can secure a majority in the elected assembly. Second, we have the *equal shares coefficient*, defined as $F(\bar{y})$: the proportion of the population that has income $\bar{y}$ or less (the distance OB), or the proportion of the population that has 'average voting strength' or less. Clearly, either of these measures as applied to the distribution of income or wealth is subject to essentially the same criticism as the relative mean deviation: they are insensitive to transfers among members of the Parade on the same side of the person with income y_0 (in the case of the minimal majority measure) or $\bar{y}$ (the equal shares coefficient): in effect they measure changes in inequality by only recording transfers between two broadly based groups.

Now let us consider Figs 2.2 and 2.5: the frequency distribution and its log-transformation. An obvious suggestion is to measure inequality in the same way as statisticians measure dispersion of any frequency distribution. In this application, the usual method would involve measuring the distance between the individual's income y_i and mean income $\bar{y}$, squaring this, and then finding the average of the resulting quantity in the whole population. Assuming that there are n people we define the *variance*:

[3] To see why, check the definition of G on page 155 and note the formula for the 'Transfer Effect' (right-hand column). Now imagine persons i and j located at two points y_i and y_j, a given distance x apart, along the fence described on page 19; if there are lots of other persons in the part of the field between those two points then the transfer-effect formula tells us that the impact of a transfer from i to j will be large ($F(y_j) - F(y_i)$ is a large number) and vice versa. It so happens that real-world frequency distributions of income look like that in Fig. 2.2 (with a peak in the mid-income range rather than at either end), so that two income receivers, £100 apart, have many people between them if they are located in the mid-income range, but rather few people between them if located at one end or other.

$$V = \frac{1}{n}\sum_{i=1}^{n}[y_i - \bar{y}]^2. \tag{2.1}$$

However, V is unsatisfactory in that were we simply to double everyone's incomes (and thereby double mean income and leave the shape of the distribution essentially unchanged), V would quadruple. One way round this problem is to standardize V. Define the *coefficient of variation* thus

$$c = \frac{\sqrt{V}}{\bar{y}}. \tag{2.2}$$

Another way to avoid this problem is to look at the variance in terms of the logarithms of income—to apply the transformation illustrated in Fig. 2.5 before evaluating the inequality measure. In fact there are two important definitions:

$$v = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{\bar{y}}\right)\right]^2, \tag{2.3}$$

$$v_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{y^*}\right)\right]^2. \tag{2.4}$$

The first of these we will call the *logarithmic variance*, and the second we may more properly term the *variance of the logarithms* of incomes. Note that v is defined relative to the *logarithm of mean income*; v_1 is defined relative to the *mean of the logarithm of income*. Either definition is invariant under proportional increases in all incomes.

We shall find that v_1 has much to recommend it when we come to examine the lognormal distribution in Chapter 4. However, c, v, and v_1 can be criticized more generally on grounds similar to those on which G was criticized. Consider a transfer of £1 from a person with y to a person with $y - £100$. How does this transfer affect these inequality measures? In the case of c, it does not matter in the slightest where in the parade this transfer is effected: so whether the transfer is from a person with £500 to a person with £400, or from a person with £100,100 to a person with £100,000, the reduction in c is exactly the same. Thus c will be particularly good at capturing inequality among high incomes, but may be of more limited use in reflecting inequality elsewhere in the distribution. In contrast to this property of c, there appears to be good reason to suggest that a measure of inequality has the property that a transfer of the above type carried out in the low income brackets would be quantitatively more effective in reducing inequality than if the transfer were carried out in the high income brackets. The measures v and v_1 appear to go some way towards meeting this objection. Taking the example of the UK in 1984/5 (illustrated in Figs 2.1 to 2.5 where we have $\bar{y} = £7{,}522$), a transfer of £1 from a person with £10,100 to a person with £10,000 reduces v and v_1 less than a transfer of £1 from a person with £500 to a person with £400. But, unfortunately, v and v_1 'overdo' this effect, so to speak. For if we consider a transfer from a person with £100,100 to a person with £100,000 then inequality, as measured by v or v_1, increases! This is hardly a desirable property for an inequality measure to possess, even if it does occur only at high incomes.[4]

Other statistical properties of the frequency distribution may be pressed into service as inequality indices. While these may draw attention to particular aspects of inequality—such as dispersion among the very high or very low incomes, by and large they miss the point as far as general inequality of incomes is concerned. Consider, for example, measures of skewness. For symmetric distributions (such as the Normal distribution, pictured on page 81) these measures are zero; but this zero value of the measure may be consistent with either a very high or a very low dispersion of incomes (as measured by the coefficient of variation). This does not appear to capture the essential ideas of inequality measurement.

Figure 2.2 can be used to derive an inequality measure from quite a different source. Stark (1972) argued that an appropriate practical method of measuring inequality should be based on society's revealed judgements on the definition of poverty and riches. The method is best seen by redrawing Fig. 2.2 as Fig. 2.7. Stark's study concentrated specifically on UK incomes, but the idea it embodies seems intuitively very appealing and could be applied more generally. The distance OP in Fig. 2.7 we will call the range of 'low incomes': P could have been fixed with reference to the income level at which a person becomes entitled to income support, adjusted for need, or with reference to some proportion of average income[5]—this is very similar to the specification of a 'poverty line'. The point R could be determined by the level at which one becomes liable to any special taxation levied on the rich, again adjusted for need.[6] The *high–low* index is then total shaded area between the curve and the horizontal axis.

The high–low index seems imaginative and practical, but it suffers from three important weaknesses. First, it is subject to exactly the same type of criticism that we levelled against M, and against the 'minimal majority' and 'equal share' measures: the measure is completely insensitive to transfers among the 'poor' (to the left of P), among the 'rich' (to the right of R), or

[4] You will always get this trouble if the 'poorer' of the two persons has at least 2.72 times mean income, in this case £20, 447—see the Technical Appendix, page 164.

[5] In Fig. 2.7 it has been located at half median income—check Question 1 on page 37 if you are unsure about how to define the median.

[6] Note that in a practical application the positions of both P and R depend on family composition. This however is a point which we are deferring until later. Figure 2.7 illustrates one type.

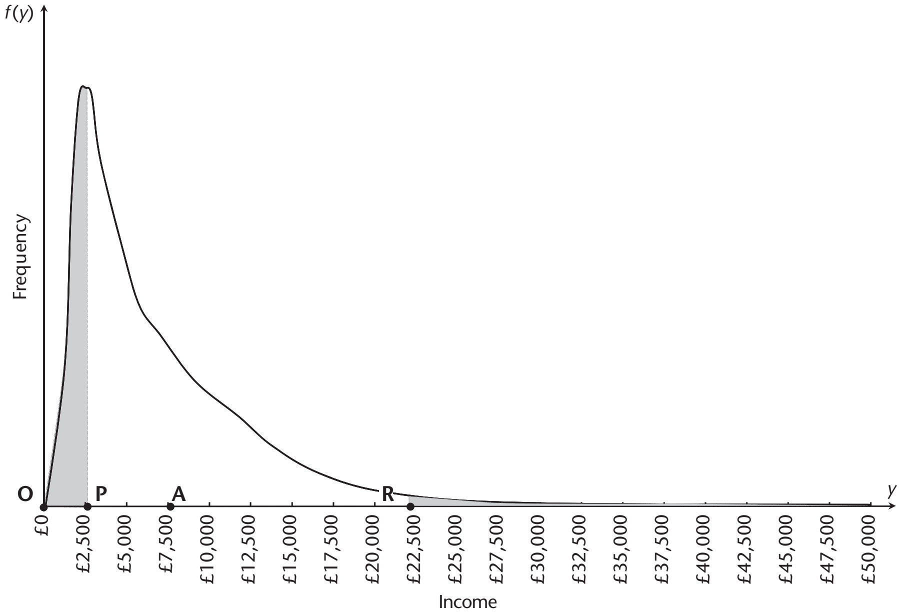

FIG. 2.7. The high–low approach

among the 'middle income receivers'. Second, there is an awkward dilemma concerning the behaviour of points P and R over time. Suppose one leaves them fixed in relative terms, so that OP and OR increase only at the same rate as mean or median income increases over time. Then one faces the criticism that one's current criterion for measuring inequality is based on an arbitrary standard fixed, perhaps, a quarter of a century ago. On the other hand, suppose that OP and OR increase with year-to-year increases in some independent reference income levels (the 'income-support' threshold for point P and the 'higher-rate tax' threshold for point R): then, if the inequality measure shows a rising trend because of more people falling in the 'low income' category, one must face the criticism that this is just an optical illusion created by altering, for example, the definition of 'poor' people; some compromise between the two courses must be chosen and the results derived for a particular application treated with caution.[7] Third, there is the point that in practice the contribution of the shaded area in the upper tail to

[7] There is a further complication in the specific UK application considered by Stark. He fixed point P using the basic national assistance (later supplementary benefit) scale plus a percentage to allow for underestimation of income and income disregarded in applying for assistance (benefit); point R was fixed by the point at which one became liable for surtax. However, national assistance, supplementary benefit, and surtax are no more. Other politically or socially defined P and R points could be determined for other times and other countries; but the basic problem of comparisons over time that I have highlighted would remain. So too, of course, would problems of comparisons between countries.

the inequality measure would be negligible: the behaviour of the inequality measure would be driven by what happens in the lower tail—which may or may not be an acceptable feature—and would simplify effectively to whether people 'fall in' on the right or on the left of point P when we arrange them in the frequency distribution diagram (Figs 2.2 and 2.7). In effect the high–low inequality index would become a slightly modified poverty index.

The use of any one of the measures we have discussed in this section implies certain value judgements concerning the way we compare one person's income against that of another. The detail of such judgements will be explained in the next chapter, although we have already seen a glimpse of some of the issues.

2.3 Rankings

Finally, we consider ways of looking at inequality that may lead to ambiguous results. Let me say straight away that this sort of non-decisive approach is not necessarily a bad thing. As we noted in Chapter 1 it may be helpful to know that over a particular period events have altered the income distribution in such a way that we find offsetting effects on the amount of inequality. The inequality measures that we have examined in the previous section act as 'tie-breakers' in such an event. Each inequality measure resolves the ambiguity in its own particular way. Just how we *should* resolve these ambiguities is taken up in more detail in Chapter 3.

- **Quantiles**
- **Shares**

TYPES OF RANKING

The two types of ranking on which we are going to focus are highlighted in the accompanying box. To anticipate the discussion a little I should point out that these two concepts are not really new to this chapter, because they each have a simple interpretation in terms of the pictures that we were looking at earlier. In fact I could have labelled the items in the box as Parade rankings and Lorenz rankings.

We have already encountered quantiles when we were discussing the incomes of the 3rd and 57th minute people as an alternative to the range, R (page 25). Quantiles are best interpreted using either the Parade diagram or its equivalent, the cumulative frequency distribution (Fig. 2.3). Take the Parade diagram and reproduce it in Fig. 2.8 (the parade of Fig. 2.1 is the solid

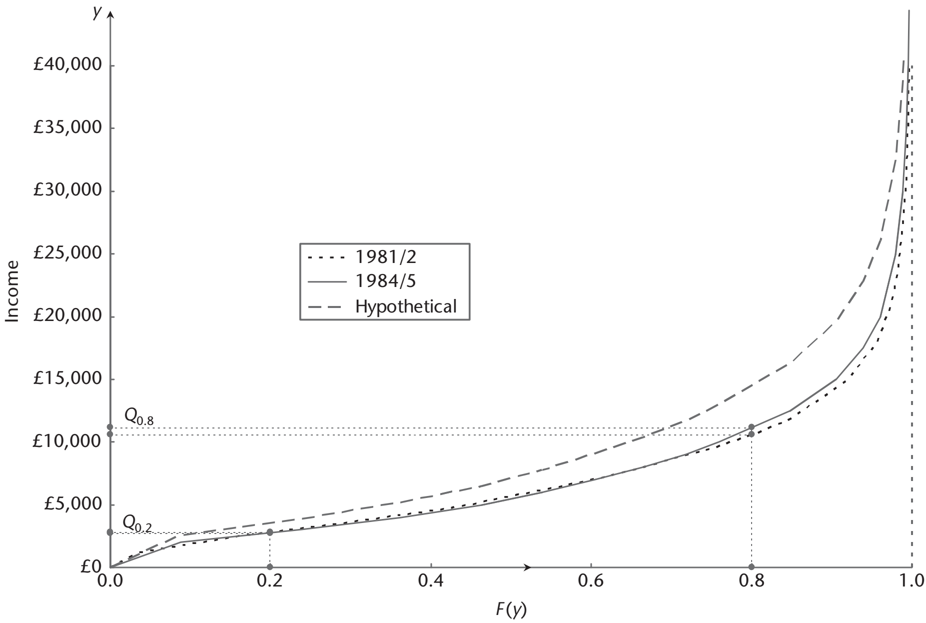

Fig. 2.8. The Parade and the quantile ranking

curve labelled 1984/5; we will come to the other two curves in a moment). Mark the point 0.2 on the horizontal axis, and read off the corresponding income on the vertical axis: this gives the 20 per cent *quantile* (usually known as the first *quintile* just to confuse you): the income at the right-hand end of the first fifth (12 minutes) of the Parade of Dwarfs. Figure 2.8 also shows how we can do the same for the 80 per cent quantile (the top quintile). In general we specify a p-quantile—which I will write Q_p—as follows. Form the Parade of Dwarfs and take the leading proportion p of the Parade (where of course $0 \leq p \leq 1$), then Q_p is the particular income level which demarcates the right-hand end of this section of the Parade.[8]

[8] A note on 'iles'. The generic term is 'quantile'—which applies to any specified population proportion p—but a number of special names for particular convenient cases are in use. There is the *median* $Q_{0.5}$, and a few standard sets such as three quartiles ($Q_{0.25}$, $Q_{0.5}$, $Q_{0.75}$), four quintiles ($Q_{0.2}$, $Q_{0.4}$, $Q_{0.6}$, $Q_{0.8}$) or nine deciles ($Q_{0.1}$, $Q_{0.2}$, $Q_{0.3}$, $Q_{0.4}$, $Q_{0.5}$, $Q_{0.6}$, $Q_{0.7}$, $Q_{0.8}$, $Q_{0.9}$); of course you can specify as many other 'standard' sets of quantiles as your patience and your knowledge of Latin prefixes permits.

I have avoided using the term 'quantile group', that is sometimes found in the literature, which refers to a slice of the population demarcated by two quantiles. For example the slice of the population with incomes at least as great as $Q_{0.1}$ but less than $Q_{0.2}$ could be referred to as the 'second decile group'. I have avoided the term because it could be confusing. However, you may also find references to such a slice of the population as 'the second decile': this usage is not just confusing, it is wrong; the quantiles are the points on the income scale, *not* the slices of the population that may be located between the points.

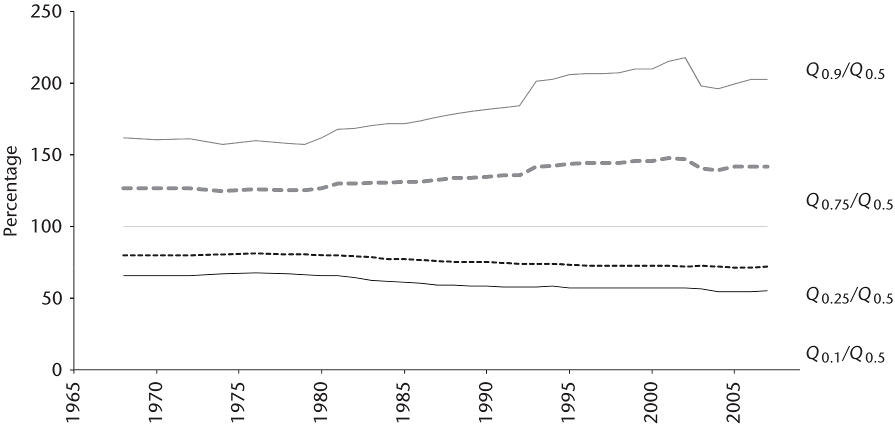

Fig. 2.9. Quantile ratios of earnings of adult men, UK 1968–2007
Source: Annual Survey of Hours and Earnings

How might we use a set of quantiles to compare income distributions? We could produce something like Fig. 2.9, which shows the proportionate movements of the quantiles of the frequency distribution of earnings in the UK in recent years (the diagram has been produced by standardizing the movements of $Q_{0.1}$, $Q_{0.25}$, $Q_{0.75}$, and $Q_{0.9}$, by the median, $Q_{0.5}$). We then check whether the quantiles are moving closer together or farther apart over time. But although the kind of moving apart that we see at the right-hand of Fig. 2.9 appears to indicate greater dispersion, it is not clear that this necessarily means greater inequality: the movement of the corresponding income shares (which we discuss in a moment) could in principle be telling us a different story.[9]

However, we might also be interested in the simple *quantile ranking* of the distributions, which focuses on the absolute values of the quantiles, rather than quantile ratios. For example, suppose that over time all the quantiles of the distribution increase by 30 per cent as shown by the curve labelled 'hypothetical' in Fig. 2.8 (in the jargon we then say that according to the quantile ranking the new distribution *dominates* the old one). Then we might say 'there are still lots of dwarfs about', to which the reply might be 'yes but at least everybody is a bit taller'. Even if we cannot be specific about whether this means that there is more or less inequality as a result, the phenomenon

[9] In case this is not obvious, consider a population with just 8 people in it: in year A the income distribution is (2, 3, 3, 4, 5, 6, 6, 7), and it is fairly obvious that $Q_{0.25} = 3$ and $Q_{0.75} = 6$; in year B the distribution becomes (0, 4, 4, 4, 5, 5, 5, 9) and we can see now that $Q_{0.25} = 4$ and $Q_{0.75} = 5$. Mean income and median income have remained unchanged and the quartiles have narrowed: but has inequality really gone down? The story from the shares suggests otherwise: the share of the bottom 25 per cent has actually fallen (from 5/36 to 4/36) and the share of the top 25 per cent has risen (from 13/36 to 14/36).

of a clear quantile ranking is telling us something interesting about the income distribution which we will discuss more in the next chapter. On the other hand if we were to compare 1981/2 and 1984/5 in Fig. 2.8 we would have to admit that over the three year period the giants became a little taller ($Q_{0.8}$ increased slightly), but the dwarfs became even shorter ($Q_{0.2}$ decreased slightly): the 1984/5 distribution does not dominate that for 1981/2.

Shares by contrast are most easily interpreted in terms of Fig. 2.4. An interesting question to ask ourselves in comparing two income distributions is: does the Lorenz curve of one lie wholly 'inside' (closer to the line of perfect equality) than that of the other? If it does, then we would probably find substantial support for the view that the 'inside' curve represents a more evenly-spread distribution. To see this point take a look at Fig. 2.10, and again do an exercise similar to that which we carried out for the quantiles in Fig. 2.8: for reference let us mark in the share that would accrue to the bottom 20 per cent and to the bottom 80 per cent in distribution B (which is the distribution before tax—the same as the Lorenz curve that we had in Fig. 2.4)—this yields the blobs on the vertical axis. Now suppose we look at the Lorenz curve marked A, which depicts the distribution for after tax

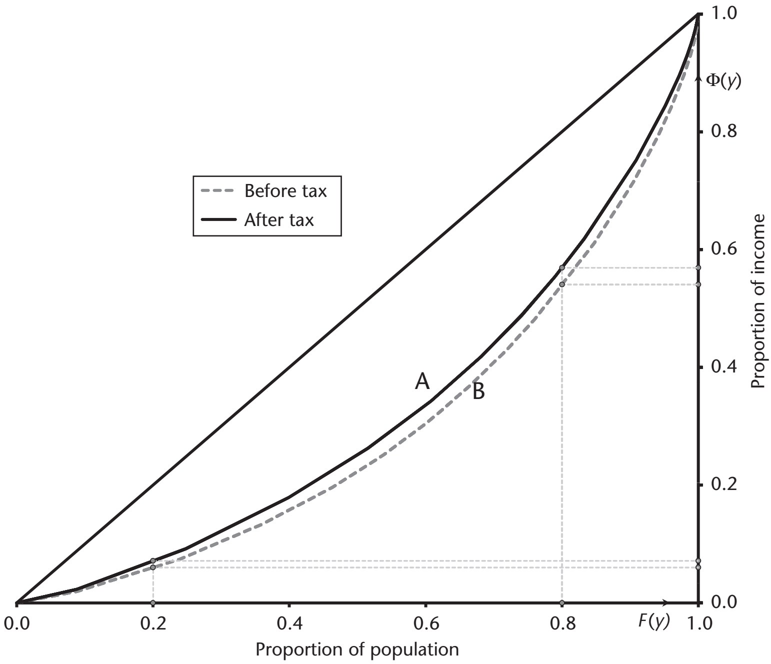

Fig. 2.10. Ranking by shares. UK 1984/5 incomes before and after tax
Source: as for Fig. 2.1

income. As we might have expected, Fig. 2.10 shows that people in the bottom 20 per cent would have received a larger slice of the after-tax cake (curve A) than they used to get in B. So also those in the bottom 80 per cent received a larger proportionate slice of the A-cake than their proportionate slice of the B-cake (which of course is equivalent to saying that the richest 20 per cent gets a *smaller* proportionate slice in A than it received in B). It is clear from the figure that we could have started with any other reference population proportions and obtained the same type of answer: whatever 'bottom proportion' of people $F(y)$ is selected, this group gets a larger share of the cake $\Phi(y)$ in A than in B (according to the shares ranking, A dominates B). Moreover, it so happens that whenever this kind of situation arises all the inequality measures that we have presented (except just perhaps v and v_1) will indicate that inequality has gone down.

However, quite often this sort of neat result does not apply. If the Lorenz curves intersect, then the shares-ranking principle cannot tell us whether inequality is higher or lower, whether it has increased or decreased. Either we accept this outcome with a shrug of the shoulders, or we have to use a tie-breaker. This situation is illustrated in Fig. 2.11, which depicts the way in which the distribution of income after tax changed from 1981/2 to 1984/5. Notice that the bottom 20 per cent of the population did proportionately

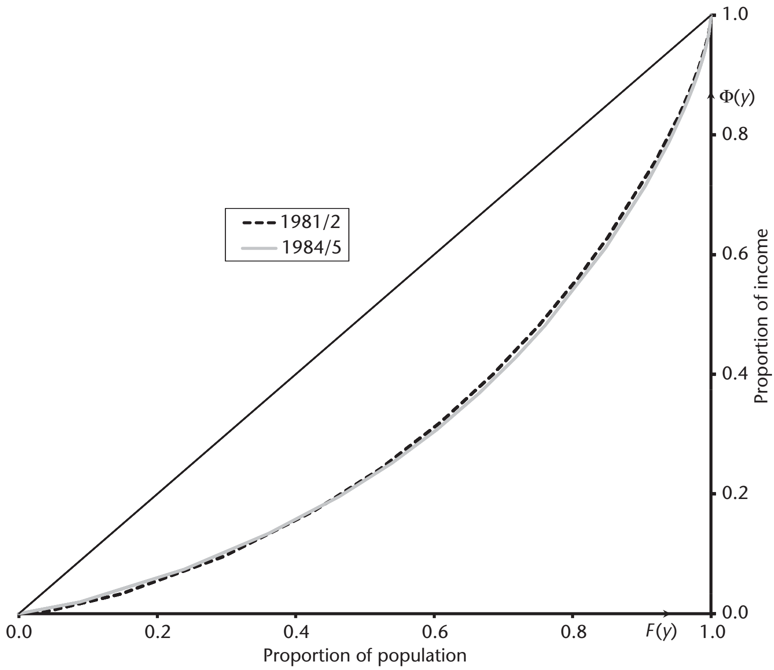

Fig. 2.11. Lorenz curves crossing

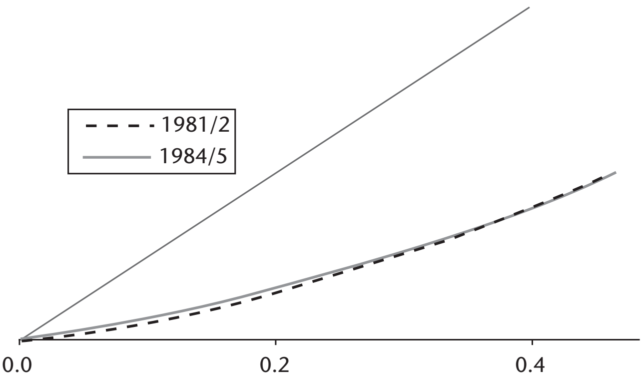

Fig. 2.12. Change at the bottom of the income distribution

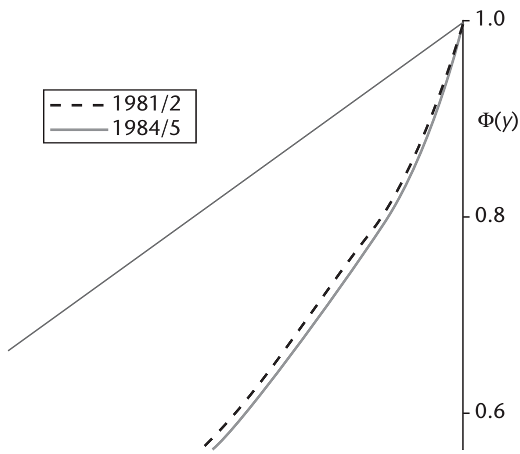

Fig. 2.13. Change at the top of the income distribution

better under 1984/5 than in 1981/2 (see also the close-up in Fig. 2.12), whilst the bottom 80 per cent did better in 1981/2 than in 1984/5 (see also Fig. 2.13). We shall have a lot more to say about this kind of situation in Chapter 3.

2.4 From charts to analysis

We have seen how quite a large number of *ad hoc* inequality measures are associated with various diagrams that chart inequality, which are themselves interlinked. But, however appealing each of these pictorial representations might be, we seem to find important reservations about any of the associated inequality measures. Perhaps the most unsatisfactory aspect of all of these indices is that the basis for using them is indeed *ad hoc*: the rationale for using them was based on intuition and a little graphical serendipity. What we really need is proper theoretical basis for comparing income distributions and for deciding what constitutes a 'good' inequality measure.

This is where the ranking techniques that we have been considering come in particularly useful. Although they are indecisive in themselves, they yet provide a valuable introduction to the deeper analysis of inequality measurement to be found in the next chapter.

2.5 Questions

1. Explain how to represent median income in Pen's Parade. How would you represent the upper and lower quartiles? (See footnote 8).
2. Describe how the following would look:
   (a) Pen's Parade with negative incomes.
   (b) The Lorenz curve if there were some individuals with negative incomes but mean income was still positive.
   (c) The Lorenz curve if there were so many individuals with negative incomes that mean income itself was negative. (See the Technical Appendix, page 169, for more on this.)
3. DeNavas-Walt *et al.* (2008) present a convenient summary of United States' income distribution data based on the Annual Social and Economic Supplement to the 2008 Current Population Survey.
   (a) How would the information in their Table A-1 need to be adapted in order to produce charts similar to Fig. 2.2?
   (b) Use the information in Table A-3 to construct Pen's Parade for 1967, 1987, 2007: how does the Parade appear to have shifted over 40 years?
   (c) Use the information in Table A-3 to construct the Lorenz curves for 1967, 1987, 2007: what has happened to inequality over the period? (Document is available on-line using the link on the website http://darp.lse.ac.uk/MI3)
4. Reconstruct the histogram for the UK 1984/5, before tax income, using the file 'ET income distribution' on the website (see the Technical Appendix page 177 for guidance on how to use the file). Now merge adjacent pairs of intervals (so that, for example, the intervals [£0,£2000] and [£2000,£3000] become [£0,£3000]) and redraw the histogram: comment on your findings.
5. Using the same data source for the UK 1984/5, before-tax income, construct the distribution function corresponding to the histogram drawn in Question 4. Now, instead of assuming that the distribution of income

follows the histogram shape, assume that within each income interval all income receivers get the mean income of that interval. Again draw the distribution function. Why does it look like a flight of steps?

6. Suppose a country's tax and benefit system operates so that taxes payable are determined by the formula

$$t[y - y_0]$$

where y is the person's original (pre-tax) income, t is the marginal tax rate and y_0 is a threshold income. Persons with incomes below y_0 receive a net payment from the government ('negative tax'). If the distribution of original income is $y_1, y_2, \ldots, y_n$, use the formulas given in the Technical Appendix (page 155) to write down the coefficient of variation and the Gini coefficient for after tax income. Comment on your results.

7. Suppose the income distribution before tax is represented by a set of numbers $\{y_{(1)}, y_{(2)}, \ldots, y_{(n)}\}$, where $y_{(1)} \leq y_{(2)} \leq y_{(3)} \ldots$. Write down an expression for the Lorenz curve. If the tax system were to be of the form given in Question 6, what would be the Lorenz curve of disposable (after-tax) income? Will it lie above the Lorenz curve for original income? (For further discussion of the point here, see Jakobsson 1976 and Eichhorn *et al.* 1984.)

8. (a) Ruritania consists of six districts that are approximately of equal size in terms of population. In 2007 per-capita incomes in the six districts were:

   - Rural (\$500, \$500, \$500);
   - Urban (\$20,000, \$28,284, \$113,137).

   What is the mean income for the Rural districts, for the Urban districts, and for the whole of Ruritania. Compute the logarithmic variance, the relative mean deviation, and the Gini coefficient for the Rural districts and the Urban districts separately and for the whole of Ruritania. (You will find that these are easily adapted from the file 'East-West' on the website, and you should ignore any income differences within any one district.)

   (b) By 2008 the per-capita income distribution had changed as follows:

   - Rural: (\$499, \$500, \$501);
   - Urban: (\$21,000, \$26,284, \$114,137).

   Rework the computations of part (a) for the 2008 data. Did inequality rise or fall between 2007 and 2008? (See the discussion on page 65 below for an explanation of this phenomenon.)

3

Analysing Inequality

'He's half a millionaire: he has the air but not the million.'

Jewish Proverb

In Chapter 2 we looked at measures of inequality that came about more or less by accident. In some cases a concept was borrowed from statistics and pressed into service as a tool of inequality measurement. In others a useful diagrammatic device was used to generate a measure of inequality that 'naturally' seemed to fit it, the relative mean deviation and the Parade, for example; or the Gini coefficient and the Lorenz curve.

- **Social welfare**
- **Information theory**
- **Structural approach**

APPROACHES TO INEQUALITY ANALYSIS

However, if we were to follow the austere and analytical course of rejecting visual intuition, and of constructing an inequality measure from 'first principles' , what approach should we adopt? I shall outline three approaches, and in doing so consider mainly special cases that illustrate the essential points easily without pretending to be analytically rigorous. The first method we shall examine is that of making inequality judgements and deriving inequality measures from social welfare functions. The social welfare function itself may be supposed to subsume values of society regarding equality and justice, and thus the derived inequality measures are given a normative basis. The second method is to see the quantification of inequality as an offshoot of the problem of comparing probability distributions: to do this we draw upon a fruitful analogy with information theory. The final—structural—approach is to specify a set of principles or axioms sufficient to determine an inequality measure uniquely; the choice of axioms themselves, of course, will be determined by what we think an inequality measure 'should' look like. Each of these approaches raises some basic questions about the meaning and interpretation of inequality.

3.1 Social welfare functions

One way of introducing social values concerning inequality is to use a social welfare function (SWF) which simply ranks all the possible states of society in the order of (society's) preference. The various 'states' could be functions of all sorts of things—personal income, wealth, size of people's cars—but we usually attempt to isolate certain characteristics which are considered 'relevant' in situations of social choice. We do not have to concern ourselves here with the means by which this social ranking is derived. The ranking may be handed down by parliament, imposed by a dictator, suggested by the trade unions, or simply thought up by the observing economist—the key point is that its characteristics are carefully specified in advance, and that these characteristics can be criticized on their own merits.

In its simplest form, a social welfare function simply orders social states unambiguously: if state A is preferable to state B then, and only then, the SWF has a higher value for state A than that for state B. How may we construct a useful SWF? To help in answering this question I shall list some properties that it may be desirable for the SWF to possess; we shall be examining their economic significance later. First let me introduce a preliminary piece of notation: let y_{iA} be the magnitude of person *i*'s 'economic position' in social state A, where *i* is a label that can be any number between 1 and *n* inclusive. For example, y_{iA} could be the income of Mr Jones of Potter's Bar in the year 1984. Where it does not matter, the A-suffix will be dropped.

Now let us use this device to specify five characteristics of the SWF. The first three are as follows:

- The SWF is *individualistic* and *nondecreasing*, if the welfare level in any state A, denoted by a number W_A, can be written:

  $$W_A = W(y_{1A}, y_{2A}, \ldots, y_{nA})$$

  and, if $y_{iB} \geq y_{iA}$ for all *i* implies, *ceteris paribus*, that $W_B \geq W_A$, which in turn implies that state B is at least as good as state A.

- The SWF is *symmetric* if it is true that, for any state,

$$W(y_1, y_2, \ldots, y_n) = W(y_2, y_1, \ldots, y_n) = \ldots = W(y_n, y_2, \ldots, y_1).$$

  This means that the function W treats individual incomes anonymously: the value of W does not depend on the particular assignment of labels to members of the population.
- The SWF is *additive* if it can be written

$$W(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} U_i(y_i) = U_1(y_1) + U_2(y_2) + \ldots + U_n(y_n), \tag{3.1}$$

  where U_1 is a function of y_1 alone, and so on.

If these three properties are all satisfied then we can write the SWF like this:

$$W(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} U(y_i) = U(y_1) + U(y_2) + \ldots + U(y_n), \tag{3.2}$$

where U is the same function for each person and where $U(y_i)$ increases with y_i. If we restrict attention to this special case the definitions of the remaining two properties of the SWF can be simplified, since they may be expressed in terms of the function U alone. Let us call $U(y_1)$ the *social utility* of, or the *welfare index* for, person 1. The rate at which this index increases is

$$U'(y_1) = \frac{dU(y_1)}{d\,y_1},$$

which can be thought of as the *social marginal utility* of, or the *welfare weight* for, person 1. Notice that, because of the first property, none of the welfare weights can be negative. Then properties 4 and 5 are:

- The SWF is *strictly concave* if the welfare weight always decreases as y_i increases.
- The SWF has *constant elasticity*, or *constant relative inequality* aversion if $U(y_i)$ can be written

$$U(y_i) = \frac{y_i^{1-\varepsilon} - 1}{1 - \varepsilon} \tag{3.3}$$

  (or in a cardinally equivalent form), where ε is the *inequality aversion parameter*, which is non-negative.[1]

I must emphasize that this is a *very* abbreviated discussion of the properties of SWFs. However, these five basic properties—or assumptions about the

[1] Notice that I have used a slightly different cardinalization of U from that employed in the first edition (1977) of this book in order to make the presentation of figures a little clearer. This change does not affect any of the results.

SWF—are sufficient to derive a convenient purpose-built inequality measure, and thus we shall examine their significance more closely.

The first of the five properties simply states that the welfare numbers should be related to individual incomes (or wealth, etc.) so that if any person's income goes up social welfare cannot go down. The term '*individualistic*' may be applied to the case where the SWF is defined in relation to the satisfactions people derive from their income, rather than the incomes themselves. I shall ignore this point and assume that any standardization of the incomes, y_i, (for example to allow for differing needs) has already been performed.[2] This permits a straightforward comparison of the individual levels, and of differences in individual levels, of people's 'economic position'—represented by the y_i and loosely called 'income'. The idea that welfare is non-decreasing in income is perhaps not as innocuous as it first seems: it rules out, for example, the idea that if one disgustingly rich person gets richer still whilst everyone else's income stays the same, the effect on inequality is so awful that social welfare actually goes down.

Given that we treat these standardized incomes y_i as a measure that puts everyone in the population on an equal footing as regards needs and desert, the second property (symmetry) naturally follows—there is no reason why welfare should be higher or lower if any two people simply swapped incomes.

The third assumption is quite strong, and is independent of the second. Suppose you measure $W_B - W_A$, the increase in welfare from state A to state B, where the only change is an increase in person 1's income from £20,000 to £21,000. Then the additivity assumption states that the effect of this change alone (increasing person 1's income from £20,000 to £21,000) is quite independent of what the rest of state A looked like—it does not matter whether everyone else had £1 or £100,000, $W_B - W_A$ is just the same for this particular change. However, this convenient assumption is not as restrictive in terms of the resulting inequality measures as it might seem at first sight—this will become clearer when we consider the concept of 'distance' between income shares later.

We could have phrased the strict concavity assumption in much more general terms, but the discussion is easier in terms of the welfare index U. Note that this is not an ordinary utility function (such as might be used to characterize the benefit that an individual gets from his income), although it may have very similar properties: it represents the valuation given by society of a person's income. One may think of this as a 'social utility function'.

[2] Once again notice my loose use of the word 'person'. In practice incomes may be received by households or families of differing sizes, in which case y_i must be reinterpreted as 'equivalized' incomes: see page 109 for more on this.

In this case, the concept corresponding to 'social marginal utility' is the quantity $U'(y_i)$, which we have called the welfare weight. The reason for the latter term is as follows. Consider a government programme which brings about a (small) change in everyone's income: $\triangle y_1, \triangle y_2, \ldots, \triangle y_n$. What is the change in social welfare? It is simply

$$dW = U'(y_1)\triangle y_1 + U'(y_2)\triangle y_2 + \ldots + U'(y_n)\triangle y_n,$$

so the U'-quantities act as a system of weights when summing the effects of the programme over the whole population. How should the weights be fixed? The strict concavity assumption tells us that the higher a person's income, the lower the social weight he is given. If we are averse to inequality this seems reasonable—a small redistribution from rich to poor should lead to a socially-preferred state.

- **Nondecreasing in incomes**
- **Symmetric**
- **Additive**
- **Strictly concave**

- **Constant elasticity**

SOME PROPERTIES OF THE SOCIAL WELFARE FUNCTION

It is possible to obtain powerful results simply with the first four assumptions—omitting the property that the U-function has constant elasticity. But this further restriction on the U-function—constant relative inequality aversion—turns the SWF into a very useful tool.

If a person's income increases, we know (from the strict concavity property) that his welfare weight necessarily decreases—but by how much? The constant elasticity assumption states that the proportional decrease in the weight U' for a given proportional increase in income should be the same at any income level. So if a person's income increases by 1% (from £100 to £101, or £100,000 to £101,000) his welfare weight drops by ε% of its former value. The higher is ε, the faster is the rate of proportional decline in welfare weight to proportional increase in income—hence its name as the 'inequality aversion parameter'. The number ε describes the strength of our yearning for equality *vis à vis* uniformly higher total income.

Table 3.1. How much should R give up to finance a £1 bonus for P?

value of ε	*maximum amount of sacrifice by R*
0	£1.00
$\frac{1}{2}$	£2.24
1	£5.00
2	£25.00
3	£125.00
5	£3,125.00
...	...

A simple numerical example may help. Consider a rich person R with five times the income of poor person P. Our being inequality averse certainly would imply that we approve of a redistribution of exactly £1 from R to P—in other words a transfer with no net loss of income. But if $\varepsilon > 0$ we might also approve of the transfer even if it were going to cost R more than £1 in order to give £1 to P—in the process of filling up the bucket with some of Mr R's income and carrying it over to Ms P we might be quite prepared for some of the income to leak out from the bucket along the way. In the case where $\varepsilon = 1$ we are in fact prepared to allow a sacrifice of up to £5 by R to make a transfer of £1 to P (£4 leaks out). So, we have the trade-off of social values against maximum sacrifice as indicated in Table 3.1. Furthermore, were we to consider an indefinitely large value of ε, we would in effect give total priority to equality over any objective of raising incomes generally. Social welfare is determined simply by the position of the least advantaged in society.

The welfare index for five constant-elasticity SWFs are illustrated in Fig. 3.1. The case $\varepsilon = 0$ illustrates that of a concave, but not strictly concave, SWF; all the other curves in the figure represent strictly concave SWFs. Figure 3.1 illustrates the fact that as you consider successively higher values of ε the social utility function U becomes more sharply curved (as ε goes up each curve is 'nested' inside its predecessor); it also illustrates the point that for values of ε less than unity, the SWF is 'bounded below' but not 'bounded above': from the $\varepsilon = \frac{1}{2}$ curve we see that with this SWF no one is ever assigned a welfare index lower than -2, but there is no upper limit on the welfare index that can be assigned to an individual. Conversely, for ε greater than unity, the SWF is bounded above, but unbounded below. For example, if $\varepsilon = 2$ and someone's income approaches zero, then we can assign him an indefinitely large negative social utility (welfare index), but no matter how large a person's income is, he will never be assigned a welfare index greater than 1.

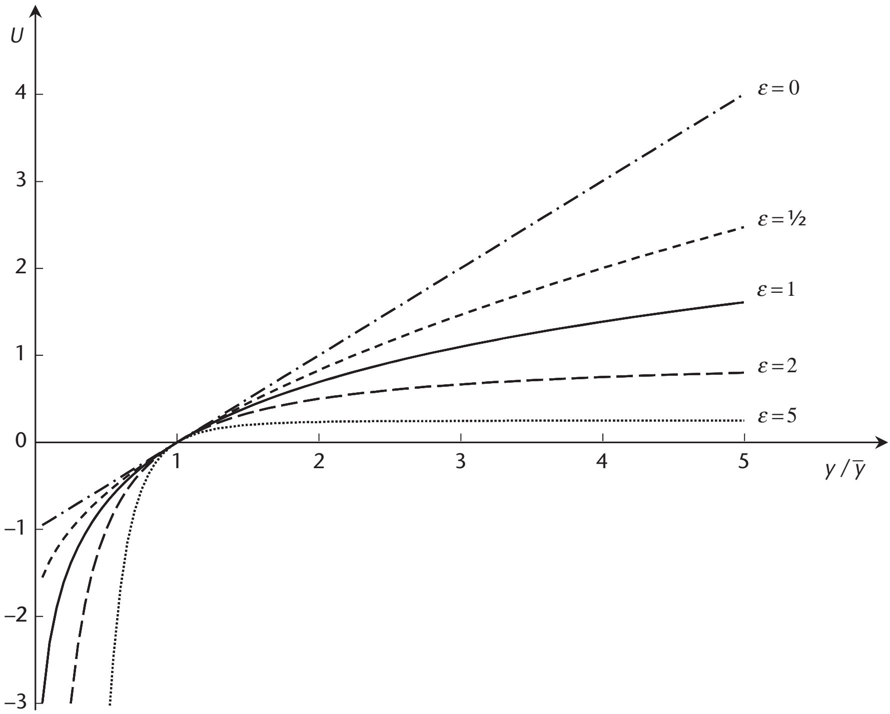

Fig. 3.1. Social utility and relative income

Notice that the vertical scale of this diagram is fairly arbitrary. We could multiply the U-values by any positive number, and add (or subtract) any constant to the U-values without altering their characteristics as welfare indices. The essential characteristic of the different welfare scales represented by these curves is the elasticity of the function $U(y)$ or, loosely speaking, the 'curvature' of the different graphs, related to the parameter ε. For convenience, I have chosen the units of income so that the mean is now unity: in other words, original income is expressed as a proportion of the mean. If these units are changed, then we have to change the vertical scale for each U-curve individually, but when we come to computing inequality measures using this type of U-function, the choice of units for y is immaterial.

The system of welfare weights (social marginal utilities) implied by these U-functions is illustrated in Fig. 3.2. Notice that for every $\varepsilon > 0$, the welfare weights fall as income increases. Notice in particular how rapid this fall is once one reached an ε-value of only 2: evidently one's income has only to be about 45 per cent of the mean in order to be assigned a welfare weight 5 times as great as the weight of the person at mean income.

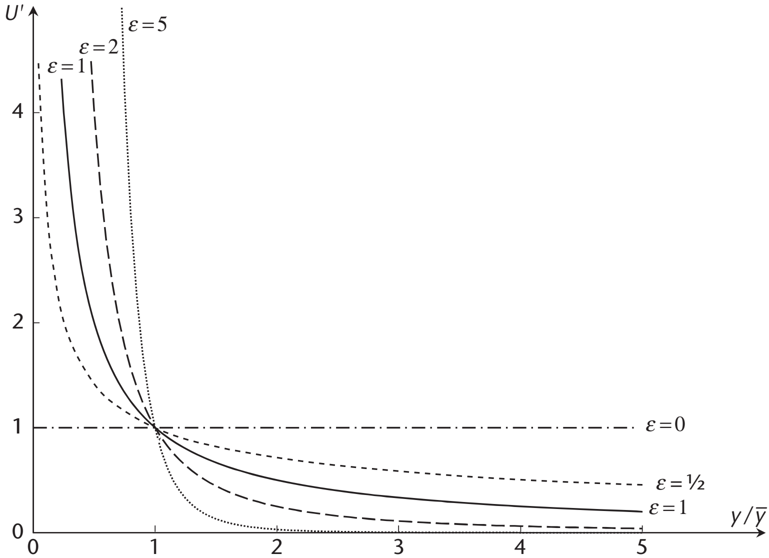

Fig. 3.2. The relationship between welfare weight and income

Let us now put the concept of the SWF to work. First consider the ranking by quantiles that we discussed in connection with Fig. 2.8. The following result does not make use of either the concavity or the constant-elasticity properties that we discussed above.

Theorem 1 *If social state A dominates the state B according to their quantile ranking, then* $W_A > W_B$ *for any individualistic, additive, and symmetric social welfare function W.*

So if the Parade of distribution A lies everywhere above the Parade of distribution B (as in the hypothetical example of Fig. 2.8 on page 32), social welfare must be higher for this class of SWFs. This result is a bit more powerful than it might at first appear. Compare the distribution A = (5, 3, 6) with the distribution B = (2, 4, 6): person 1 clearly gains in a move from B to A, but person 2 is worse off: yet according to the Parade diagram and according to any symmetric, increasing SWF A is regarded better than B. Why? Because the symmetry assumption means that A is equivalent to A′ = (3, 5, 6), and there is clearly higher welfare in A′ than in B.

If we introduce the restriction that the SWF be concave then a further very important result (which again does not use the special constant elasticity restriction) can be established:

Theorem 2 *Let the social state A have an associated income distribution* $(y_{1A}, y_{2A}, \ldots, y_{nA})$ *and social state B have income distribution* $(y_{1B}, y_{2B}, \ldots, y_{nB})$, *where total income in state A and in state B is identical. Then the Lorenz curve for state A lies wholly inside the Lorenz curve for state B if and only if $W_A > W_B$ for any individualistic, increasing, symmetric, and strictly concave social welfare function W.*[3]

This result shows at once the power of the ranking by shares that we discussed in Chapter 2 (the Lorenz diagram), and the relevance of SWFs of the type we have discussed. Re-examine Fig. 2.10. We found that intuition suggested that curve A represented a 'fairer' or 'more equal' distribution than curve B. This may be made more precise. The first four assumptions on the SWF crystallize our views that social welfare should depend on individual economic position, and that we should be averse to inequality. Theorem 2 reveals the identity of this approach with the intuitive method of the Lorenz diagram, subject to the 'constant amount of cake' assumption introduced in Chapter 1. Notice that this does not depend on the assumption that our relative aversion to inequality should be the same for all income ranges—other concave forms of the U-function would do. Also it is possible to weaken the assumptions considerably (but at the expense of ease of exposition) and leave Theorem 2 intact.

Moreover, the result of Theorem 2 can be extended to some cases where the cake does not stay the same size. To do this, define the so-called *generalized Lorenz curve* by multiplying the vertical co-ordinate of the Lorenz curve by mean income (so now the vertical axis runs from 0 to the mean income rather than 0 to 1).

Theorem 3 *The generalized Lorenz curve for state A lies wholly above the generalized Lorenz curve for state B if and only if $W_A > W_B$, for any individualistic, additive increasing, symmetric, and strictly concave social welfare function W.*

For example, we noted in Chapter 2 that the simple shares ranking criterion was inconclusive when comparing the distribution of income after tax in the UK 1981/2 with that for the period 1984/5: the ordinary Lorenz curves intersect (see Figs 2.11–2.13). Now let us consider the generalized Lorenz curves for the same two datasets, which are depicted in Fig. 3.3. Notice that the vertical axis is measured in monetary units, by contrast with that for Figs. 2.4 and 2.10–2.13; notice also that this method of comparing distributions implies a kind of priority ranking for the mean: as is evident from Fig. 3.3 if the mean of distribution A is higher than the distribution B, then the generalized Lorenz curve of B cannot lie above that of A no matter how unequal A may be. So, without further ado, we can assert that *any* SWF that

[3] 'Wholly inside' includes the possibility that the Lorenz curves for A and B may coincide somewhere, but not everywhere.

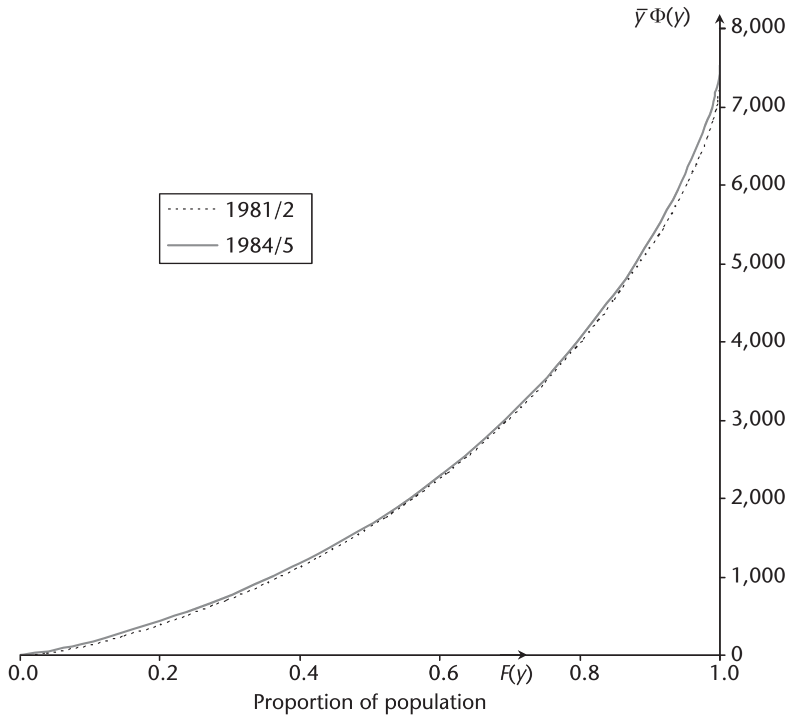

FIG. 3.3. The generalized Lorenz curve comparison: UK income before tax

is additive, individualistic and concave will suggest that social welfare was higher in 1984/5 than in 1981/2.

However, although Theorems 1 to 3 provide us with some fundamental insights on the welfare and inequality rankings that may be inferred from income distributions, they are limited in two ways.

First, the results are cast exclusively within the context of social welfare analysis. That is not necessarily a drawback, since the particular welfare criteria that we have discussed may have considerable intuitive appeal. Nevertheless you might be wondering whether the insights can be interpreted in inequality *without* bringing in the social welfare apparatus: that is something that we shall tackle later in the chapter.

Second, the three theorems are not sufficient for the practical business of inequality measurement. Lorenz curves that we wish to compare often intersect; so too with Parade diagrams and generalized Lorenz curves. Moreover we often desire a unique numerical value for inequality in order to make comparisons of changes in inequality over time or differences in inequality between countries or regions. This is an issue that we shall tackle right away: we use the social welfare function to find measures of inequality.

3.2 SWF-based inequality measures

In fact from (3.2) we can derive two important classes of inequality measure. Recall our piecemeal discussion of ready-made inequality measures in Chapter 2: we argued there that although some of the measures seemed attractive at first sight, on closer inspection they turned out to be not so good in some respects because of the way that they reacted to changes in the income distribution. It is time to put this approach on a more satisfactory footing by building an inequality measure from the groundwork of fundamental welfare principles. To see how this is done, we need to establish the relationship between the frequency distribution of income y—which we encountered in Fig. 2.2—and the frequency distribution of social utility U.

This relationship is actually achieved through the cumulative frequency distribution $F(y)$ (Fig. 2.3). To see the relationship examine Fig. 3.4, which is really three diagrams superimposed for convenience. In the bottom right-hand quadrant we have plotted one of the 'welfare-index', or 'social utility' curves from Fig. 3.1, which of course requires the use of the constant elasticity assumption.

In the top right-hand quadrant you will recognize the cumulative frequency distribution, drawn for income or wealth in the usual way. To construct the curves for the distribution of social utility or welfare index U, pick any income value, let us say y_0; then read off the corresponding proportion of population F_0 on the vertical $0F$ axis, using the distribution function $F(y)$, and also the corresponding U-value (social utility) on $0U$ (bottom right-hand corner). Now plot the F and U-values in a new diagram (bottom left-hand corner)—this is done by using the top left-hand quadrant just to reflect $0F$ axis on to the horizontal $0F$ axis. What we have done is to map the point (y_0, F_0) in the top right-hand quadrant into the point (F_0, U_0) in the bottom left-hand quadrant. If we do this for other y-values and points on the top right-hand quadrant cumulative frequency distribution, we end up with a new cumulative frequency distribution in the bottom left-hand quadrant. (To see how this works, try tracing round another rectangular set of four points like those shown in Fig. 3.4.)

Once we have this new cumulative frequency distribution in terms of social utility, we can fairly easily derive the corresponding frequency distribution itself (this is just the slope of the F-function). The frequency distributions of y and U are displayed in Fig. 3.5: notice that the points y_0 and U_0 correspond to the points y_0 and U_0 in Fig. 3.4 (the shaded area in each case corresponds to F_0).

Now let us derive the inequality measures. For the distribution of income (top half of Fig. 3.5) mark the position of the mean, $\bar{y}$, on the axis $0y$. Do the

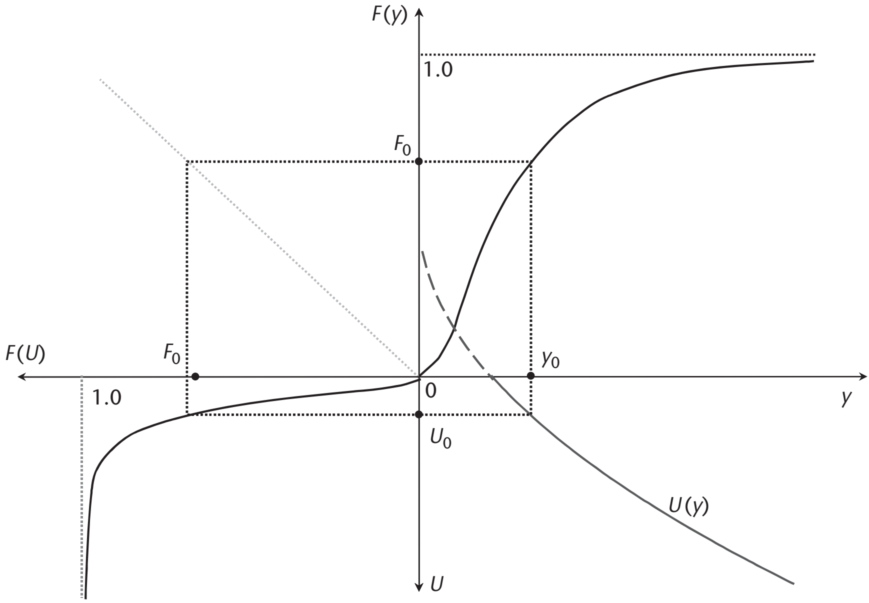

Fig. 3.4. Distribution of income and distribution of social utility

same for the distribution of social utility—the mean is point $\overline{U}$ on the axis $0U$. We can also mark in two other points of interest:

- The social utility corresponding to $\bar{y}$—we do this using the bottom half of Fig. 3.5—point $U(\bar{y})$ on $0U$.
- The income corresponding to average social utility—we do this by a reverse process using the top half of Fig. 3.5 and plotting point y_e on $0y$.

The quantity $U(\bar{y})$ represents the social utility for each person in the community were national income to be distributed perfectly equally. The quantity y_e represents the income which, if received by each member of the community, would result in the same level of overall social welfare as the existing distribution yields. Necessarily $y_e \leq \bar{y}$—we may be able to throw some of the national income away, redistribute the rest equally, and still end up with the same level of social welfare. Notice that we have drawn the diagram for a particular isoelastic utility function in the bottom right-hand quadrant of Fig. 3.4; if ε were changed, then so would the frequency distribution in the bottom half of Fig. 3.5, and of course the positions of $\bar{y}$ and y_e.

Thus we can define a different inequality measure for each value of ε, the inequality aversion parameter. An intuitively appealing way of measuring

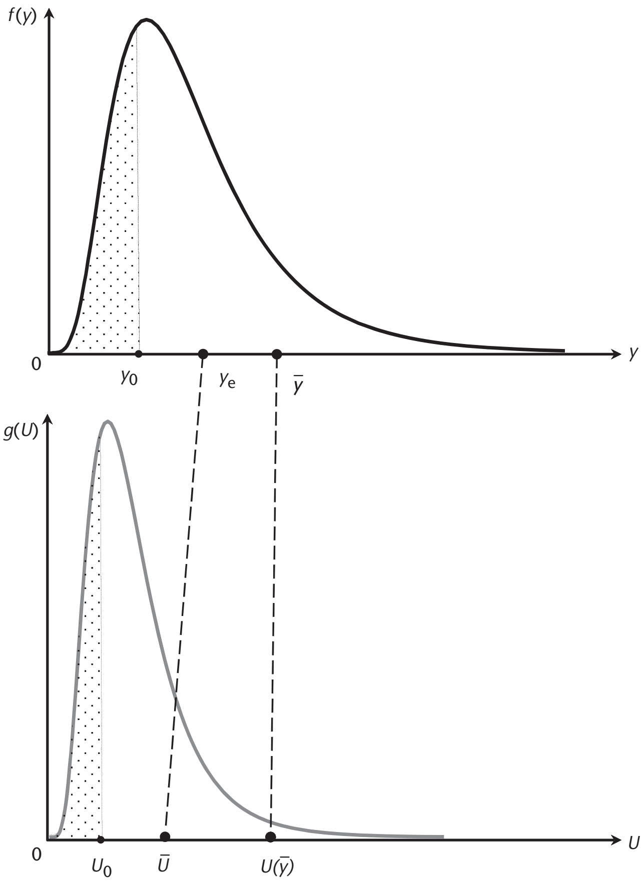

Fig. 3.5. The Atkinson and Dalton indices

inequality seems to be to consider how far actual average social utility falls short of potential average social utility (if all income were distributed equally). So we define *Dalton's inequality index* (for inequality aversion ε) as:

$$D_{\varepsilon} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i^{1-\varepsilon} - 1\right]}{\bar{y}^{1-\varepsilon} - 1},$$

which in terms of the diagram means

$$D_{\varepsilon} = 1 - \frac{\overline{U}}{U(\bar{y})}.$$

We may note that this is zero for perfectly equally distributed incomes (in which case we would have exactly $\overline{U} = U(\bar{y})$. Atkinson (1970) criticizes the use of D_ε on the grounds that it is sensitive to the level from which social utility is measured—if you add a non-zero constant to all the Us, D_ε changes. Now this does not change the ordering properties of D_ε over different distributions with the same mean, but the inequality measures obtained by adding different arbitrary constants to U will not be cardinally equivalent. So Atkinson suggests, in effect, that we perform our comparisons back on the $0y$ axis, not the $0U$ axis, and compare the 'equally distributed equivalent' income, y_e, with the mean $\bar{y}$. To do this, we write U^{-1} for the inverse of the function U (so that $U^{-1}(a)$ gives the income that would yield social utility level a). Then we can define *Atkinson's inequality index* (for inequality aversion ε) as just

$$A_\varepsilon = 1 - \frac{U^{-1}\left(\overline{U}\right)}{\bar{y}},$$

where, as before, $\overline{U}$ is just average social utility $\frac{1}{n}\sum_{i=1}^{n} U(y_i)$. Using the explicit formula (3.3) for the function U we get

$$A_\varepsilon = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}.$$

In terms of the diagram this is:

$$A_\varepsilon = 1 - \frac{y_e}{\bar{y}}.$$

Once again, as for the index D_ε, we find a different value of A_ε for different values of our aversion to inequality.

From the definitions we can check that the following relationship holds for all distributions and all values of ε

$$1 - D_\varepsilon = \frac{U(\bar{y}\,[1 - A_\varepsilon])}{U(\bar{y})},$$

which means that

$$\frac{\partial D_\varepsilon}{\partial A_\varepsilon} = \bar{y}\frac{U'(\bar{y}\,[1 - A_\varepsilon])}{U'(\bar{y})} > 0.$$

Clearly, in the light of this property, the choice between the indices D_ε and A_ε as defined above is only of vital importance with respect to their cardinal properties ('is the reduction in inequality by taxation greater in year A than in year B?'); they are obviously ordinally equivalent in that they produce the same ranking of different distributions with the same mean.[4]

[4] Instead of lying between zero and unity, D_ε lies between 0 and ∞. In order to transform this into an inequality measure that is comparable with others we have used, it would be necessary to look at values of $D_\varepsilon/[D_\varepsilon + 1]$. One might be tempted to suggested that D_ε is thus a suitable choice

Of much greater significance is the choice of the value of ε, especially where Lorenz curves intersect, as in Fig. 2.11. This reflects our relative sensitivity to redistribution from the rich to the not-so-rich *vis à vis* redistribution from the not-so-poor to the poor. The higher the value of ε used, the more sensitive is the index to changes in distribution at the bottom end of the parade—we will come to specific examples of this later in the chapter.

The advantage of the SWF approach is evident. Once agreed on the form of the social welfare function (for example along the lines of assumptions that I have listed above) it enables the analyst of inequality to say, in effect 'you tell me how strong society's aversion to inequality is, and I will tell you the value of the inequality statistic', rather than simply incorporating an arbitrary social weighting in an inequality index that just happens to be convenient.

3.3 Inequality and information theory

Probability distributions sometimes provide useful analogies for income distributions. In this section we shall see that usable and quite reasonable inequality measures may be built up from an analogy with information theory.

In information theory, one is concerned with the problem of 'valuing' the information that a certain event out of a large number of possibilities has occurred. Let us suppose that there are events numbered 1,2,3, . . . , to which we attach probabilities $p_1, p_2, p_3, \ldots$. Each p_i is not less than zero (which represents total impossibility of the event's occurrence) and not greater than one (which represents absolute certainty of the event's occurrence). Suppose we are told that event #1 has occurred. We want to assign a number $h(p_1)$ to the value of this information: how do we do this?

If event #1 was considered to be quite likely anyway (p_1 near to 1), then this information is not fiercely exciting, and so we want $h(p_1)$ to be rather low; but if event #1 was a near impossibility, then this information is amazing and valuable—it gets a high $h(p_1)$. So the implied value $h(p_1)$ should decrease as p_1 increases. A further characteristic which it seems correct that $h(.)$ should have (in the context of probability analysis) is as follows. If event #1 and event #2 are statistically independent (so that the probability that event #1 occurs does not depend on whether or not event #2 occurs,

as A_ε. However, even apart from the fact that D_ε depends on the cardinalization of utility there is another unsatisfactory feature of the relationship between D_ε and ε. For Atkinson's measure, A_ε, the higher is the value of ε, the greater the value of the inequality measure for any given distribution; but this does not hold for D_ε.

and *vice versa*), then the probability that both event #1 and event #2 occur together is $p_1 p_2$. So, if we want to be able to add up the information values of messages concerning independent events, the function h should have the special property

$$h(p_1 p_2) = h(p_1) + h(p_2) \tag{3.4}$$

and the only function that satisfies this for all valid p-values is $h = -\log(p)$.

However, a set of n numbers—the probabilities relating to each of n possible states—is in itself an unwieldy thing with which to work. It is convenient to aggregate these into a single number which describes the 'degree of disorder' of the system. This number will be lowest when there is a probability of 1 for one particular event i and a 0 for every other event: in this case the system is completely orderly and the information that i has occurred is valueless (we already knew it would occur) whilst the other events are impossible; the overall information content of the system is zero. More generally we can characterize the 'degree of disorder'—known technically as the entropy—by working out the average information content of the system. This is the weighted sum of all the information values for the various events; the weight given to event i in this averaging process is simply its probability p_i: In other words we have:

$$\begin{aligned} \text{entropy} &= \sum_{i=1}^{n} p_i h(p_i) \\ &= -\sum_{i=1}^{n} p_i \log(p_i). \end{aligned}$$

Now Theil (1967) has argued that the entropy concept provides a useful device for inequality measurement. All we have to do is reinterpret the n possible events as n people in the population, and reinterpret p_i as the share of person i in total income, let us say s_i. If $\bar{y}$ is mean income, and y_i is the income of person i then:

$$s_i = \frac{y_i}{n\bar{y}},$$

so that, of course:

$$\sum_{i=1}^{n} s_i = 1.$$

Then subtracting the actual entropy of the income distribution (just replace all the p_is with s_is in the entropy formula) from the maximum possible value of this entropy (when each $s_i = 1/n$, everyone gets an even share) we find the following contender for status as an inequality measure.

$$\begin{aligned} T &= \sum_{i=1}^{n} \frac{1}{n} h\left(\frac{1}{n}\right) - \sum_{i=1}^{n} s_i h(s_i) \\ &= \sum_{i=1}^{n} s_i \left[h\left(\frac{1}{n}\right) - h(s_i) \right] \\ &= \sum_{i=1}^{n} s_i \left[\log(s_i) - \log\left(\frac{1}{n}\right) \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \log\left(\frac{y_i}{\bar{y}}\right). \end{aligned}$$

Each of these four expressions is an equivalent way of writing the measure T.

A diagrammatic representation of T can be found in Figs 3.6 and 3.7. In the top right-hand corner of Fig. 3.6, the function $\log(\frac{y}{\bar{y}})$ is plotted (along

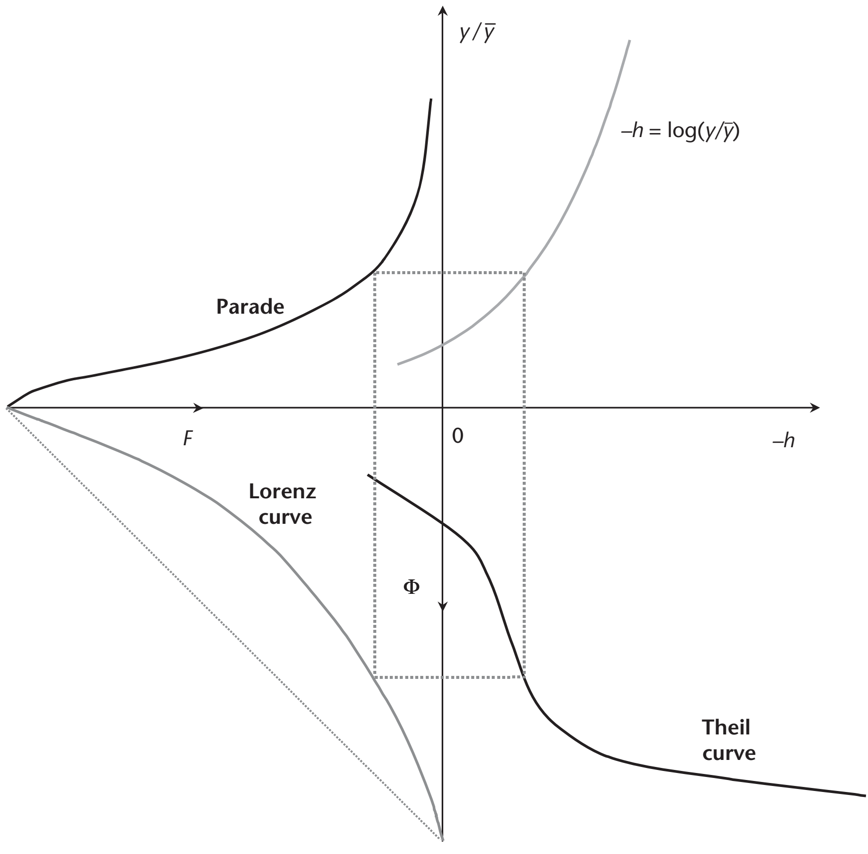

FIG. 3.6. The Theil curve

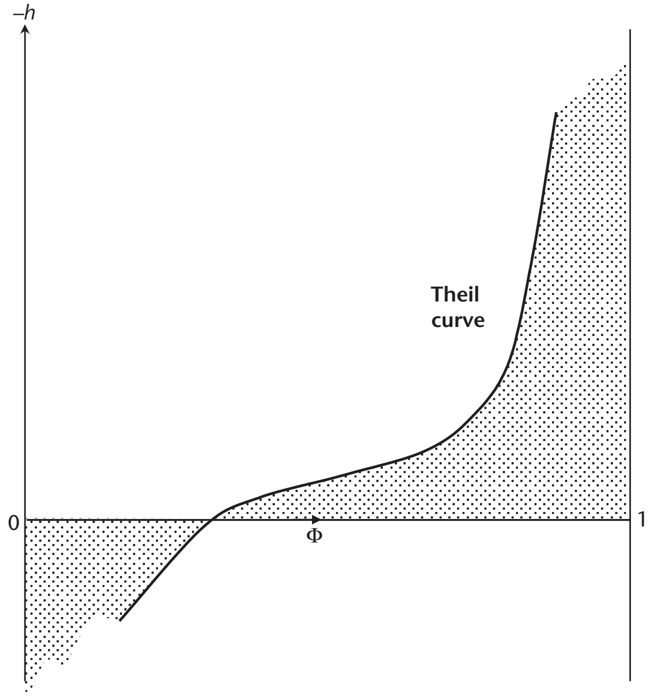

FIG. 3.7. Theil's entropy index

the horizontal axis) against $\frac{y}{\bar{y}}$ (along the vertical axis). In the top left-hand corner we have the Parade, slightly modified from Fig. 2.1, whilst in the bottom left-hand corner we have the Lorenz curve (upside down). We can use these three curves to derive the *Theil curve* in the bottom right hand corner of Fig. 3.6. The method is as follows.

- Pick a particular value of F.
- Use the Parade diagram (top left) to find the corresponding value of $y/\bar{y}$—in other words the appropriate quantile divided by the mean.
- Also use the Lorenz curve (bottom left) to find the corresponding Φ-value for this same F-value—in other words, find the income share of the group in population that has an income less than or equal to y.
- Read off the '$-h$' value corresponding to $\frac{y}{\bar{y}}$ using the log function shown top right.
- You have now fixed a particular point in the bottom right-hand part of the figure as shown by dotted rectangle.
- By repeating this for every other F-value, trace out a curve—the Theil curve—in the bottom right-hand corner.

If you are not yet convinced, you may care to try plotting another set of four points as an exercise. This Theil curve charts the 'information function' against income shares. Unfortunately the entire curve cannot be shown in Fig. 3.6 since it crosses the 0Φ axis; to remedy this I have drawn a fuller picture of the curve in Fig. 3.7 (which is drawn the logical way up, with 0Φ along the horizontal axis). The measure T is then simply the area trapped between this curve and the 0Φ axis—shown as a shaded area.

However, this merely tells us about the mechanics of Theil's measure; we need to look more closely at its implications for the way we look at inequality. To do this, examine what happens to T if the share of a poor person (1) is increased at the expense of a rich person (2). So let the share of person 1 increase from s_1 to a fractionally larger amount $s_1 + \triangle s$ and the share of 2 decrease to $s_2 - \triangle s$. Then, remembering that $h(s) = -\log(s)$, we find (by differentiation) that the resulting change in T is:

$$\begin{aligned}\triangle T &= \triangle s\,[h(s_2) - h(s_1)]\\ &= -\triangle s \log\left(\frac{s_2}{s_1}\right).\end{aligned}$$

As we would expect, the proposed transfer $\triangle s$ results in a negative $\triangle T$, so that the inequality index decreases. But we can say a little more than that. We see that the size of the reduction in T depends only on the ratio of s_2 to s_1. So for any two people with income shares in the same ratio, the transfer s (as above) would lead to the same reduction in inequality T. Thus, for example, a small transfer of from a person with an income share of 2 millionths, to a person with only 1 millionth of the cake, has the same effect on Theil inequality as an identical transfer from a person with 8 millionths of the national cake to one with 4 millionths.

This helps us to complete our analogy between inequality measurement and information theory. It is easy to see that income shares (s_i) serve as counterparts to probabilities (p_i). And now we can interpret the 'social analogue' of the information function h. Evidently, from the formula for $\triangle T$, we can now say under what circumstances s_3 and s_4 are the same 'distance apart' as s_2 and s_1. This would occur if

$$h(s_1) - h(s_2) = h(s_3) - h(s_4),$$

so that a small transfer from s_2 to s_1 has exactly the same effect on inequality as a small transfer from s_4 to s_3. Given this interpretation of $h(s)$ in terms of distance, do we want it to have *exactly* the same properties as $h(p)$ in information theory? There does not seem to be any compelling *a priori* reason why we should do so,[5] although $h(s) = -\log(s)$ gives us a reasonably sensible

[5] Recall that the log-function was chosen in information theory so that $h(p_1 p_2) = h(p_1) + h(p_2)$.

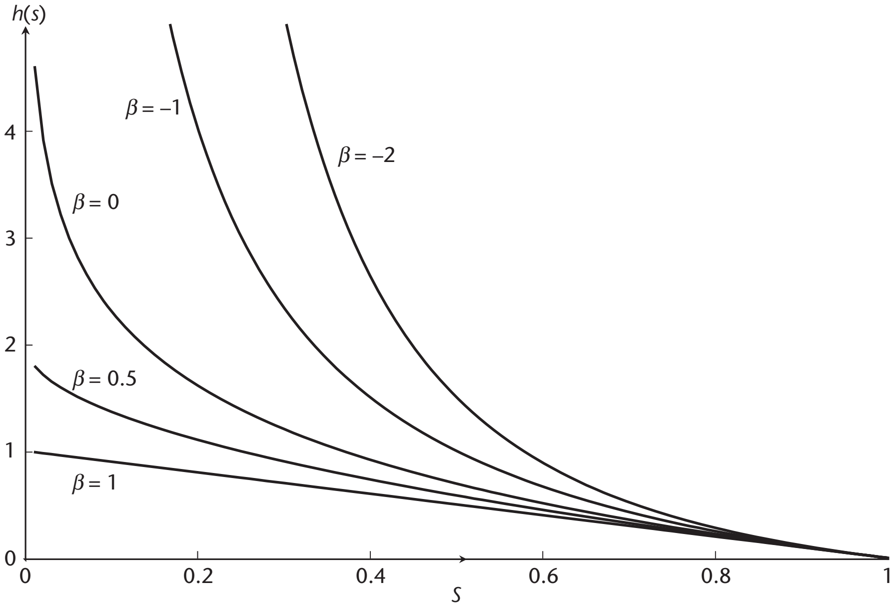

FIG. 3.8. A variety of distance concepts

inequality measure, T. The function, $-\log(s)$ can be seen as a member of a much wider class of functions, illustrated in Fig. 3.8. This figure charts members of the family of curves given by[6]

$$h(s) = \frac{1 - s^{\beta}}{\beta}.$$

Deriving an inequality measure in exactly the same way as before gives us, for any value of β that we choose, a particular inequality measure which may be written in any of the following equivalent ways:

$$\frac{1}{1+\beta}\left[\sum_{i=1}^{n}\frac{1}{n}h\left(\frac{1}{n}\right) - \sum_{i=1}^{n}s_i h\left(s_i\right)\right], \tag{3.5}$$

$$\frac{1}{1+\beta}\sum_{i=1}^{n}s_i\left[h\left(\frac{1}{n}\right) - s_i\right], \tag{3.6}$$

$$\frac{1}{\beta+\beta^2}\sum_{i=1}^{n}s_i\left[s_i^{\beta} - n^{-\beta}\right]. \tag{3.7}$$

[6] Again I have slightly modified the definition of this function from the first edition in order to make the presentation neater, although this reworking does not affect any of the results—see footnote 3.1.

And of course the effect of a small transfer of Δs from rich person 2 to poor person 1 is now

$$-\frac{1}{\beta}\left[s_2^\beta - s_1^\beta\right]\Delta s$$

$$= [h(s_2) - h(s_1)]\,\Delta s.$$

You get the same effect by transferring Δs from rich person 4 to poor person 3 if and only if the 'distance' $h(s_4) - h(s_3)$ is the same as the 'distance' $h(s_2) - h(s_1)$. Let us look at some specific examples of this idea of distance and the associated inequality measures.

- First let us look at the case $\beta = -1$. We obtain the following measure:

$$-\sum_{i=1}^{n} \log(ns_i)\,; \tag{3.8}$$

  this is n times the so-called *mean logarithmic deviation* (MLD)

$$L = \frac{1}{n}\sum_{i=1}^{n}\left[\log(1/n) - \log(s_i)\right]. \tag{3.9}$$

  As the name suggests, L is the average deviation between the log income shares and the log shares that would represent perfect equality (equal to $1/n$). The associated distance concept is given by

$$h(s_1) - h(s_2) = \frac{1}{s_1} - \frac{1}{s_2}.$$

- The special case where $\beta = 0$ simply yields the measure T once again. As we noted, this implies a relative concept of distance: s_2 and s_1 are the same distance apart as s_4 and s_3 if the ratios s_2/s_1 and s_4/s_3 are equal.
- Finally let us consider $\beta = 1$. Then we get the following information theoretic measure:

$$\frac{1}{2}\left[\sum_{i=1}^{n} s_i^2 - \frac{1}{n}\right].$$

  Now *Herfindahl's* index is simply

$$H = \sum_{i=1}^{n} s_i^2,$$

  that is, the sum of the squares of the income shares. So, comparing these two expressions, we see that for a given population, H is cardinally equivalent to the information theoretic measure with a value of $\beta = 1$; and in this case we have the very simple absolute distance measure

$$h(s_1) - h(s_2) = s_1 - s_2.$$

In this case the distance between a person with a 1 per cent share and one with a 2 per cent share is considered to be the same as the distance between a person with a 4 per cent share and one with a 5 per cent share.

Thus, by choosing an appropriate 'distance function', we determine a particular 'information theoretic' inequality measure. In principle we can do this for any value of β. Pick a particular curve in Fig. 3.8: the 'distance' between any two income shares on the horizontal axis is given by the linear distance between their two corresponding points on the vertical axis. The β-curve of our choice (suitably rotated) can then be plugged into the top left-hand quadrant of Fig. 3.6, and we thus derive a new curve to replace the original in the bottom right-hand quadrant, and obtain the modified information theoretic inequality measure. Each distance concept is going to give different weight on the gaps between income shares in different parts of the income distribution. To illustrate this, have a look at the example in Table 3.2: the top part of this gives the income for three (out of many) individuals, poor P, rich R, and quite-well-off Q, and their respective shares in total income (assumed to be £1,000,000); the bottom part gives the

Table 3.2. Is P further from Q than Q is from R?

		income	*share*
	person P	£2,000	0.2%
	...	...	...
	person Q	£10,000	1.0%
	...	...	...
	person R	£50,000	5.0%
	...	...	...
	all:	*£1,000,000*	*100%*
β	$h(s_i) - h(s_j)$	*distance (P, Q)*	*distance (Q, R)*
−1	$\frac{1}{s_i} - \frac{1}{s_j}$	400	80
0	$\log(\frac{s_j}{s_i})$	log(5)	log(5)
1	$s_j - s_i$	0.008	0.04

implied distance from P to Q and the implied distance from Q to R for three of the special values of β that we have discussed in detail. We can see that for $\beta = -1$ the (P,Q)-gap is ranked as greater than the (Q,R)-gap; for $\beta = 1$ the reverse is true; and for $\beta = 0$, the two gaps are regarded as equivalent.

Notice the formal similarity between choosing one of the curves in Fig. 3.8 and choosing a social utility function or welfare index in Fig. 3.1. If we write $\beta = -\varepsilon$, the analogy appears to be almost complete: the choice of 'distance function' seems to be determined simply by our relative inequality aversion. Yet the approach of this section leads to inequality measures that are somewhat different from those found previously. The principal difference concerns the inequality measures when $\beta \geq 0$. As we have seen, the modified information theoretic measure is defined for such values. However, A_ε and D_ε become trivial when ε is zero (since A_0 and D_0 are zero whatever the income distribution); and usually neither A_ε nor D_ε is defined for $\varepsilon < 0$ (corresponding to $\beta > 0$). Furthermore, even for positive values of ε—where the appropriate, modified, information theoretic measure ranks any set of income distributions in the same order as A_ε and D_ε—it is evident that the Atkinson index, the Dalton index, and the information theoretic measure will not be cardinally equivalent. Which forms of inequality measure should we choose then? The remainder of this chapter will deal more fully with this important issue.

3.4 Building an inequality measure

What we shall now do is consider more formally the criteria we want satisfied by inequality measures. You may be demanding to know why this has not been done before. The reason is that I have been anxious to trace the origins of inequality measures already in use and to examine the assumptions required at these origins.

- **Weak principle of transfers**
- **Income scale independence**
- **Principle of population**
- **Decomposability**
- **Strong principle of transfers**

FIVE PROPERTIES OF INEQUALITY MEASURES

However, now that we have looked at *ad hoc* measures, and seen how the SWF and information theory approaches work, we can collect our thoughts on the properties of these measures. The importance of this exercise lies not only in the drawing up of a shortlist of inequality measures by eliminating those that are 'unsuitable'. It also helps to put personal preference in perspective when choosing among those cited in the shortlist. Furthermore, it provides the basis for the third approach of this chapter: building a particular class of mathematical functions for use as inequality measures from the elementary properties that we might think that inequality measures ought to have. It is in effect a structural approach to inequality measurement.

This is a trickier task, but rewarding nonetheless; to assist us there is a check-list of the proposed elementary criteria in the accompanying box. Let us look more closely at the first four of these: the fifth criterion will be discussed a bit later.

Weak Principle of Transfers

In Chapter 2 we were interested to note whether each of the various inequality measures discussed there had the property that a hypothetical transfer of income from a rich person to a poor person reduces measured inequality. This property may now be stated more precisely. We shall say that an inequality measure satisfies the weak principle of transfers if the following is always true. Consider any two individuals, one with income y, the other, a richer person, with income $y + \delta$ where δ is positive. Then transfer a positive amount of income $\triangle y$ from the richer to the poorer person, where $\triangle y$ is less than 2δ. Inequality should then definitely decrease. If this property is true for some inequality measure, no matter what values of y and $y + \delta$ we use, then we may use the following theorem.

Theorem 4 *Suppose the distribution of income in social state A could be achieved by a simple redistribution of income in social state B (so that total income is the same in each case) and the Lorenz curve for A lies wholly inside that of B. Then, as long as an inequality measure satisfies the weak principle of transfers, that inequality measure will always indicate a strictly lower level of inequality for state A than for state B.*

This result is not exactly surprising if we recall the interpretation of the Lorenz curve in Chapter 2: if you check the example given in Fig. 2.10 on page 34 you will see that we could have got to state A from state B by a series of richer-to-poorer transfers of the type mentioned above. However, Theorem 4 emphasizes the importance of this principle for choosing between inequality measures. As we have seen V, c, G, L, T, H, A_ε, D_ε ($\varepsilon > 0$), and the modified information theory indices all pass this test; v and v_1

fail the test in the case of high incomes—it is possible for these to rank B as superior to A. The other measures, R, M, the equal shares coefficient, etc., just fail the test—for these measures it would be possible for state A's Lorenz curve to lie partly 'inside' and to lie nowhere 'outside' that of state B, and yet exhibit no reduction in measured inequality. In other words, we have achieved a situation where there has been some richer-to-poorer redistribution somewhere in the population, but apparently no change in inequality occurs.[7]

I have qualified the definition given above as the weak principle of transfers, because all that it requires is that given the specified transfer, inequality should decrease. But it says nothing about how much it should decrease. This point is considered further when we get to the final item on the list of properties.

Income Scale Independence

This means that the measured inequality of the slices of the cake should not depend on the size of the cake. If everyone's income changes by the same proportion then it can be argued that there has been no essential alteration in the income distribution, and thus that the value of the inequality measure should remain the same. This property is possessed by all the inequality measures we have examined, with the exception of the variance V, and Dalton's inequality indices.[8] This is immediately evident in the case of those measures defined with respect to income shares s_i, since a proportional income change in all incomes leaves the shares unchanged.

Principle of Population

This requires that the inequality of the cake distribution should not depend on the number of cake-receivers. If we measure inequality in a particular economy with n people in it, and then merge the economy with another identical one, we get a combined economy with a population of $2n$, and with the same proportion of the population receiving any given income. If measured inequality is the same for *any* such replication of the economy, then the inequality measure satisfies the principle of population.

However, it is not self-evident that this property is desirable. Consider a two-person world where one person has all the income and the other has none. Then replicate the economy as just explained, so that one now

[7] However, this type of response to a transfer might well be appropriate for poverty measures since these tools are designed for rather different purposes.

[8] Whether a Dalton index satisfies scale independence or not will depend on the particular cardinalization of the function U that is used.

has a four-person world with two destitute people and two sharing income equally. It seems to me debatable whether these two worlds are 'equally unequal'. In fact, nearly all the inequality measures we have considered would indicate this, since they satisfy the principle of population. The notable exceptions are the modified information theoretic indices: if $\beta = 0$ (the original Theil index) the population principle is satisfied, but otherwise as the population is increased the measure will either increase (the case where $\beta < 0$) or decrease (the case where $\beta > 0$, including Herfindahl's index of course). However, as we shall see in a moment, it is possible to adapt this class of measures slightly so that the population principle is always satisfied.

Decomposability

This property implies that there should be a coherent relationship between inequality in the whole of society and inequality in its constituent parts. The basic idea is that we would like to be able to write down a formula giving total inequality as a function of inequality *within* the constituent subgroups, and inequality *between* the subgroups. More ambitiously we might hope to be able to express the within-group inequality as something like an average of the inequality in each individual subgroup. However, in order to do either of these things with an inequality measure, it must have an elementary consistency property: that inequality rankings of alternative distributions in the population as a whole should match the inequality rankings of the corresponding distributions within any of the subgroups of which the population is composed.

This can be illustrated using a pair of examples, using artificial data specially constructed to demonstrate what might appear as a curious phenomenon. In the first we consider an economy of six persons that is divided into two equal-sized parts, East and West. As is illustrated in Table 3.3, the East is much poorer than the West. Two economic programmes (A and B) have been suggested for the economy: A and B each yield the same mean income (7) in the East, but they yield different income distribution amongst the Easterners; the same story applies in the West—A and B yield the same mean income (63.33) but a different income distribution. Taking East and West together, it is clear that the choice between A and B lies exclusively in terms of the impact upon inequality within each region; by construction, income differences between the regions are unaffected by the choice of A or B. Table 3.3 lists the values of four inequality measures—the Gini coefficient, two Atkinson indices and the Theil index—and it is evident that for each of these inequality would be higher under B than it would be under A. This applies to the East, to the West, and to the two parts taken together.

Table 3.3. The breakdown of inequality: poor East, rich West

East A: (6, 7, 8) B: (6, 6, 9)			West A: (30, 30, 130) B: (10, 60, 120)		
	A	B		A	B
$\bar{y}$	7.00	7.00	$\bar{y}$	63.33	63.33
G	0.063	0.095	G	0.351	0.386
A_1	0.007	0.019	A_1	0.228	0.343
A_2	0.014	0.036	A_2	0.363	0.621
T	0.007	0.020	T	0.256	0.290

East and West combined A: (6, 7, 8, 30, 30, 130) B: (6, 6, 9, 10, 60, 120)		
	A	B
$\bar{y}$	35.16	35.16
G	0.562	0.579
A_1	0.476	0.519
A_2	0.664	0.700
T	0.604	0.632

All of this seems pretty unexceptionable: all of the inequality measures would register an increase overall if there were a switch from A to B, and this is consistent with the increase in inequality in each component subgroup (East and West) given the A→B switch. We might imagine that there is some simple formula linking the change in overall inequality to the change in inequality in each of the components. But now consider the second example, illustrated in Table 3.4. All that has happened here is that the East has caught up and overtaken the West: Eastern incomes under A or B have grown by a factor of 10, while Western incomes have not changed from the first example. Obviously inequality *within* the Eastern part and *within* the Western part remains unchanged from the first example, as a comparison of the top half of the two tables will reveal: according to all the inequality measures presented here inequality is higher in B than in A. But now look at the situation in the combined economy after the East's income has grown (the lower half of Table 3.4): inequality is higher in B than in A according to the Atkinson index and the Theil index, but *not* according to the Gini coefficient. So, in this case, in switching from A to B the Gini coefficient in the East would go up, the Gini coefficient in the West would go up, inequality between East and West would be unchanged, and yet . . . the Gini coefficient overall would go *down*. Strange but true.[9]

[9] There is a bit more to the decomposability story and the Gini coefficient, which is explained in the Technical Appendix—see page 165.

Table 3.4. The breakdown of inequality: the East catches up

East
A: (60, 70, 80)
B: (60, 60, 90)

	A	B
$\bar{y}$	70.00	70.00
G	0.063	0.095
A_1	0.007	0.019
A_2	0.014	0.036
T	0.007	0.020

West
A: (30, 30, 130)
B: (10, 60, 120)

	A	B
$\bar{y}$	63.33	63.33
G	0.351	0.386
A_1	0.228	0.343
A_2	0.363	0.621
T	0.256	0.290

East–West combined
A: (60, 70, 80, 30, 30, 130)
B: (60, 60, 90, 10, 60, 120)

	A	B
$\bar{y}$	66.67	66.67
G	0.275	0.267
A_1	0.125	0.198
A_2	0.236	0.469
T	0.126	0.149

Two lessons can be drawn from this little experiment. First, some inequality measures are just not decomposable, in that it is possible for them to register an increase in inequality in every subgroup of the population at the same time as a decrease in inequality overall: if this happens then it is obviously impossible to express the overall inequality change as some consistent function of inequality change in the component subgroups. The Gini coefficient is a prime example of this; other measures which behave in this apparently perverse fashion are the logarithmic variance, the variance of logarithms, and the relative mean deviation. The second lesson to be drawn is that, because decomposability is essentially about consistency in inequality rankings in the small and in the large, if a particular inequality measure is decomposable then so too is any ordinally equivalent transformation of the measure: for example it can readily be checked that the variance V is decomposable, and so is the coefficient of variation c which is just the square root of V.

There is a powerful result that clarifies which inequality measures will satisfy decomposability along with the other properties that we have discussed so far:

Theorem 5 *Any inequality measure that simultaneously satisfies the properties of the weak principle of transfers, decomposability, scale independence, and the population principle must be expressible either in the form*

$$E_\theta = \frac{1}{\theta^2 - \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^\theta - 1\right],$$

or as $J(E_\theta)$, some ordinally-equivalent transformation of E_θ, where θ is a real parameter that may be given any value, positive, zero, or negative.

I have used the symbol 'E' to denote this family of measures, since they have become known in the literature as the *generalized entropy measures*. A quick comparison of this formula with that of the modified information theoretic measures (defined on page 58) shows that the two are very closely related: in fact the generalized entropy measures are just the modified information theoretic family again, now normalized so that they satisfy the population principle, and with the parameter θ set equal to $\beta - 1$.[10] In view of this 'family connection' it is clear that the generalized entropy measures have other connections too: inspection of the generalized entropy formula reveals that the case $\theta = 2$ yields an index that is cardinally equivalent to the Herfindahl index H (and hence ordinally equivalent to V and c); putting $\theta = 1 - \varepsilon$ in the formula we can see that—for values of $\theta < 1$—the measures are ordinally equivalent to the welfare-theoretic indices A_ε and D_ε for distributions with a given mean.

As with our discussion of welfare-based and information-theory-based measures we now have a collection or family of inequality measures that incorporates a set of principles for ranking income distributions. And, as we have just seen, there are close connections between all the indices derived from three approaches. Let us see if we can narrow things down a bit further.

3.5 Choosing an inequality measure

Now that we have seen three approaches to a coherent and comprehensive analysis of inequality, how should we go about selecting an appropriate inequality measurement tool? For a start let us clarify the nature of the choice that we are to make. We need to make the important distinction between choosing a family of inequality measures and choosing a particular member from the family. This sort of distinction would apply to the selection of mathematical functions in other contexts. For example, if we were decorating a piece of paper and wanted to decide on a particular curve or shape

[10] In the first edition (1977) the modified information theoretic measure was denoted I_β and extensively discussed. Since that time the literature has more frequently used the normalization of the generalized entropy family given here as E_θ. Formally one has $E_1 = I_0 = T$, if $\theta = 1$ ($\beta = 0$), and $E_\theta = I_{\beta-1} n^{\beta-1}$ for other values of θ.

to use in the pattern, we might consider first the broader choice between families of curves or shapes—squares, circles, triangles, ellipses, etc.—and then having decided upon ellipses for the design, perhaps we would want to be more specific and pick a particular size and shape of ellipse. Some of the broad principles that we have considered under 'building an inequality measure' are rather like the questions at the level of the 'squares, circles, or ellipses?' stage of designing the decorative pattern. Let us see what guidance we now have in choosing a family of inequality measures.

The first four of the basic properties of inequality measures that we listed earlier—the weak transfer principle, scale independence, the population principle, and decomposability—would probably command wide, although not universal, support. As we have seen they define an extended family of measures: the generalized entropy family and all the measures that are ordinally equivalent to it. It may be worth trying to narrow this selection of measures a bit further, and to do this we should discuss the fifth on the list of the basic principles.

Strong Principle of Transfers

Let us recall the concept of 'distance' between people's income shares, introduced on page 57, to strengthen the principle of transfers. Consider a distance measure given by

$$d = h(s_1) - h(s_2),$$

where s_2 is greater than s_1, and $h(s)$ is one of the curves in Fig. 3.8. Then consider a transfer from rich person 2 to poor person 1. We say that the inequality measure satisfies the principle of transfers in the strong sense if the amount of the reduction in inequality depends only on d, the distance, no matter which two individuals we choose.

For the kind of h-function illustrated in Fig. 3.8, the inequality measures that satisfy this strong principle of transfers belong to the family described by formulas for the modified information theoretic family (of which the Theil index and the Herfindahl index are special cases) or the generalized entropy family which, as we have just seen, is virtually equivalent. Each value of β—equivalently each value of θ—defines a different concept of distance, and thus a different associated inequality measure satisfying the strong principle of transfers.

In effect we have found an important corollary to Theorem 5. Adding the strong principle of transfers to the other criteria means that Theorem 5 can be strengthened a bit: if all five properties listed above are to be satisfied then the only measures which will do the job are the generalized entropy indices E_θ.

Why should we want to strengthen the principle of transfers in this way? One obvious reason is that merely requiring that a measure satisfy the weak principle gives us so much latitude that we cannot even find a method of ranking all possible income distributions in an unambiguous order. This is because, as Theorem 4 shows, the weak principle amounts to a requirement that the measure should rank income distributions in the same fashion as the associated Lorenz curves—no more, no less. Now, the strong principle of transfers by itself does not give this guidance, but it points the way to an intuitively appealing method. Several writers have noted that an inequality measure incorporates some sort of average of income differences. The 'distance' concept, d, allows one to formalize this. For, given a particular d, one may derive a particular inequality measure by using the strong principle as a fundamental axiom.[11] This measure takes the form of the average distance between each person's actual income and the income he would receive in a perfectly equal society, and is closely related to E_θ.[12] The advantage of this is that instead of postulating the existence of a social welfare function, discussing its desired properties, and then deriving the measure, one may discuss the basic idea of distance between income shares and then derive the inequality measure directly.

Most of the *ad hoc* inequality measures do not satisfy the strong principle of transfers as they stand, although some are ordinally equivalent to measures satisfying this axiom. In such cases, the size of a change in inequality due to an income transfer depends not only on the distance between the shares of the persons concerned, but on the measured value of overall inequality as well. It is interesting to note the distance concept implied by these measures. Implicit in the use of the variance and the coefficient of variation (which are ordinally equivalent to H) is the notion that distance equals the absolute difference between income shares. The relative mean deviation implies a very odd notion of distance—zero if both persons are on the same side of the mean, and one if they are on opposite sides. This property can be deduced from the effect of the particular redistribution illustrated in Fig. 2.6. The measures v, v_1, and G are not even ordinally equivalent to a measure satisfying the strong principle. In the case of v and v_1 this is because they do not satisfy the weak principle either; the reason for G's failure is more subtle. Here the size of the change in inequality arising from a redistribution between two people depends on their relative position in the Parade, not on the absolute size of their incomes or their income shares. Hence a redistribution from the 4th to the 5th person (arranged

[11] For the other axioms required see Cowell and Kuga (1981) and the discussion on page 186 which gives an overview of the development of this literature.

[12] This is clear from the second of the three ways in which the information theoretic measure was written down on page 58.

in Parade order) has the same effect as a transfer from the 1,000,004th to the 1,000,005th, whatever their respective incomes. So distance cannot be defined in terms of the individual income shares alone.

A further reason for recommending the strong principle lies in the cardinal properties of inequality measures. In much of the literature attention is focused on ordinal properties, and rightly so. However, sometimes this has meant that because any transformation of an inequality measure leaves its ordering properties unchanged, cardinal characteristics have been neglected or rather arbitrarily specified. For example, it is sometimes recommended that the inequality measure should be normalized so that it always lies between zero and one. To use this as a recommendation for a particular ordinally equivalent variant of the inequality measure is dubious for three reasons.

1. It is not clear that a finite maximum value of inequality, independent of the number in the population, is desirable.
2. There are many ways of transforming the measure such that it lies in the zero-to-one range, each such transformation having different cardinal properties.
3. And, in particular, where the untransformed measure has a finite maximum, the measure can easily be normalized without altering its cardinal properties, simply by dividing by that maximum value.[13]

However, because measures satisfying the strong principle of transfers can be written down as the sum of a function of each income share, they have attractive cardinal properties when one considers either the problem of decomposing inequality by population subgroups (as in the East–West example discussed above), or of quantifying changes in measured inequality. The family E_θ, all members of which satisfy the strong principle, may be written in such a way that changes in inequality overall can easily be related to (a) changes in inequality within given subgroups of the population, and (b) changes in the income shares enjoyed by these subgroups, and hence the inequality between the groups. The way to do this is explained in the Technical Appendix, from which it is clear that a measure such as A_ε, though formally ordinally equivalent to I_β for many values of ε, does not decompose nearly so easily. These cardinal properties are, of course, very important when considering empirical applications, as we do in Chapter 5.

Now let us consider the second aspect of choice: the problem of selecting from among a family of measures one particular index. As we have seen, many, though not all, of the inequality measures that are likely to be of

[13] This assumes that the minimum value is zero; but the required normalization is easy whatever the minimum value.

interest will be ordinally equivalent to the generalized entropy class: this applies, for example, to inequality measures that arise naturally from the SWF method (for example we know that all the measures A_ε are ordinally equivalent to E_θ, for $\theta = 1 - \varepsilon$ where $\varepsilon > 0$). Let us then take the generalized entropy family of measures[14]—extended to include all the measures that are ordinally equivalent—as the selected family and examine the issues involved in picking one index from the family.

If we are principally concerned with the ordering property of the measures, then the key decision is the *sensitivity of* the inequality index to information about different parts of the distribution. We have already seen this issue in our discussion on page 60 of whether the distance between Rich R and quite-well-off Q was greater than the distance between Q and poor P. Different distance concepts will give different answers to this issue. The distance concept can be expressed in terms of the value of the parameter β, or equivalently in terms of the generalized entropy parameter θ (remember that θ is just equal to $1 + \beta$). In some respects we can also express this sensitivity in terms of the SWF inequality-aversion parameter ε since, in the region where it is defined, $\varepsilon = 1 - \theta$ (which in turn equals $-\beta$). We have already seen on page 44 how specification of the parameter ε implies a particular willingness to trade income loss from the leaky bucket against further equalization of income; this choice of parameter ε also determines how the 'tie' will be broken in cases where two Lorenz curves intersect—the problem mentioned in Chapter 2.

To illustrate this point, consider the question of whether or not the Switzerland of 1982 was 'really' more unequal than the USA of 1979, using the data in Fig. 3.9.[15] As we can see from the legend in the figure, the Gini coefficient is about the same for the distributions of the two countries, but the Lorenz curves intersect: the share of the bottom ten per cent in Switzerland is higher than the USA, but so too is the share of the top ten per cent. Because of this property we find that the SWF-based index A_ε will rank Switzerland as more unequal than the USA for low values of inequality aversion ε—see the left-hand end of Fig. 3.10—and will rank the USA as more unequal for high values of ε (where the SWF and its associated distance concept are more sensitive to the bottom of the distribution).

The value of ε or θ that is chosen depends on two things:

- our intrinsic aversion to inequality;
- the discriminatory power of the resulting inequality measure.

[14] Although we could have constructed reasonable arguments for other sets of axioms that would have picked out a different class of inequality measures—see the Technical Appendix for a further discussion.

[15] Source: Bishop *et al*. (1991) based on LIS data.

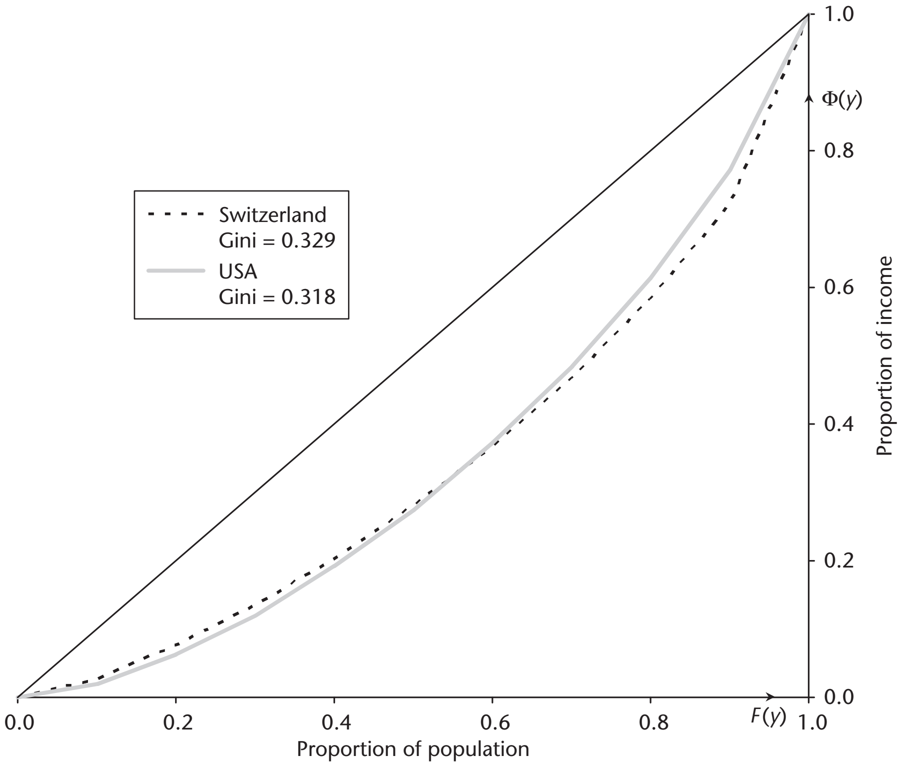

FIG. 3.9. Lorenz curves for equivalized disposable income per person: Switzerland and USA

Of course the first point is just a restatement of our earlier discussion relating ε to our willingness to sacrifice overall income in order to pursue an egalitarian redistribution; a practical example occurs in Chapter 5. The detail of the second point has to be deferred to Chapter 5; however, the main point is that if very high inequality aversion is specified, nearly all income distributions that are encountered will register high measured inequality, so that it becomes difficult to say whether one state is more unequal than another.

3.6 Summary

The upshot of the argument of Chapters 2 and 3 is as follows. If we are interested in dealing with any and every possible income distribution, it may be reasonable to require that a property such as the weak principle of transfers should be satisfied. In choosing a measure that conforms to this

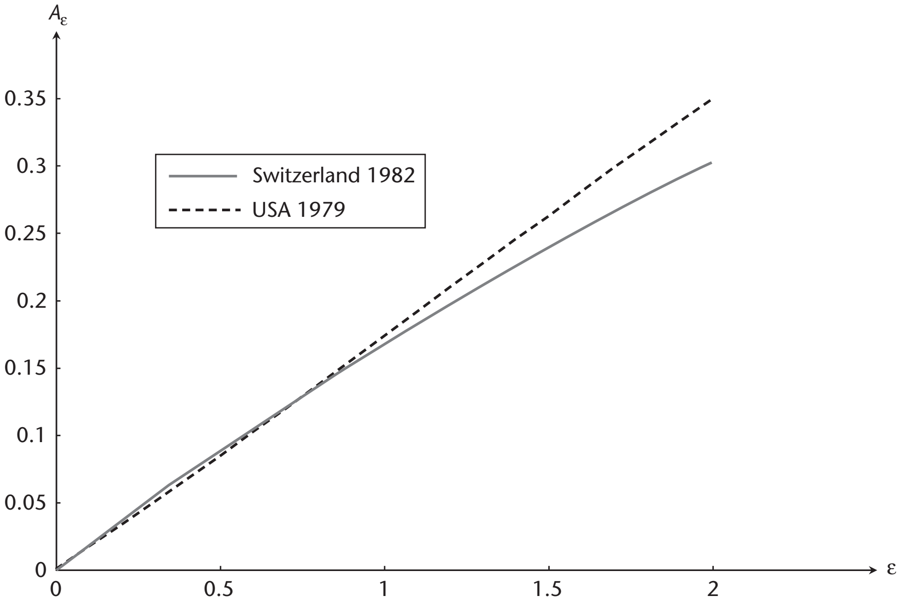

FIG. 3.10. Inequality aversion and inequality rankings, Switzerland and USA
Source: as for Fig. 3.9

principle it is useful to have one that may either be related to an inequality-aversion parameter (such as A_ε or D_ε) or to a concept of distance between income shares (the information theoretic measures or the family E_θ). In order to do this we need to introduce some further assumptions about the measurement tool—such as the decomposability property—which may be more contentious.

Even if these assumptions about building an inequality measure are accepted, this still leaves the question of various cardinal characteristics open. Invariance with respect to proportional changes in all incomes or with respect to increases in the population may be desirable under certain circumstances. Standardization of the measure in a given range (such as 0 to 1) has only a superficial attractiveness to recommend it: it may be well worthwhile sacrificing this in order to put the measure in a cardinal form more useful for analysing the composition of, and changes in, inequality. The way these conclusions relate to the measures we have mentioned is summarized in Table 3.5.

However, these remarks apply to comparisons of all conceivable distributions. You may wonder whether our task could be made easier if our attention were restricted to those distributions that are, in some sense, more likely to arise. The next chapter attempts to deal with this issue.

Table 3.5. Which measure does what?

Index	Principle of transfers	Distance concept	Decomposable?	Independent of income scale & population size?	Range in interval [0,1] ?
Variance, V	strong	Absolute differences	Yes	No: increases with income	No
Coeff. of variation, c	weak	As for variance	Yes	Yes	No
Relative mean deviation, M	just fails	0, if incomes on same side of $\bar{y}$, or 1 otherwise	No	Yes	No: in [0,2]
Logarithmic variance, v	fails	Differences in (log-income)	No	Yes	No
Variance of logarithms, v_1	fails	As for logarithmic variance	No	Yes	No
Equal shares coefficient	just fails	As for relative mean deviation	No	Yes	Yes
Minimal majority	just fails	Similar to M (critical income is y_0, not $\bar{y}$)	No	Yes	Yes
Gini, G	weak	Depends on rank ordering	No	Yes	Yes
Atkinson's index, A_ε	weak	Difference in marginal social utilities	Yes	Yes	Yes
Dalton's index, D_ε	weak	As for Atkinson's index	Yes	No	No
Theil's entropy index, T	strong	Proportional	Yes	Yes	No
MLD index, L	strong	Difference between reciprocal of incomes	Yes	Yes	No
Herfindahl's index, H	strong	As for variance	Yes	No: decreases with population	Yes: but min > 0
Generalized entropy, E_θ	strong	Power function	Yes	Yes	No

Note: 'just fails' means a rich-to-poor transfer may leave inequality unchanged rather than reducing it.

3.7 Questions

1. Show that the inequality aversion parameter ε is the elasticity of social marginal utility defined on page 41.
2. (a) Use the UK 1984/5 data (see file 'ET income distribution' on the website) to compute Atkinson's inequality index with $\varepsilon = 2$, making the same assumptions as in Question 5 of Chapter 2.
   (b) Recompute the index in part (a) after dropping the first income class from the dataset. Why does measured inequality decrease?
   (c) Rework the calculations in (b) for a variety of values of ε so as to verify that measured inequality rises with inequality aversion for a given dataset.
3. Suppose that the assumption of constant relative inequality aversion (page 41) were to be replaced by the assumption of constant absolute inequality aversion, whereby the U-function may be written

$$U(y_i) = -\frac{1}{\kappa} e^{-\kappa y_i}.$$

   (a) Sketch the U-function for different values of κ.
   (b) Write down the corresponding social welfare function, and hence find an expression for the equally-distributed equivalent income.
   (c) Explain what happens to social welfare as y_i goes to zero. Is the social welfare function defined for negative incomes?
4. Consider the following two distributions of income:

$$\begin{aligned} &\text{A} : (1, 4, 7, 10, 13),\\ &\text{B} : (1, 5, 6, 10, 13). \end{aligned}$$

   Which of these appears to be more unequal? Many people when confronted with this question will choose B rather than A. Which fundamental principle does this response violate? (See Amiel and Cowell 1999.)
5. Gastwirth (1974b) proposed the following as an inequality measurement tool:

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{|y_i - y_j|}{y_i + y_j}.$$

   What concept of distance between incomes does it employ? In what way does it differ from the Gini coefficient? For the two distributions (1, 2, 97) and (1, 3, 96) verify that it violates the transfer principle: would it also violate the transfer principle for the distributions (2, 2, 96) and

(1, 3, 96)? (See also Amiel and Cowell 1998 and Nygård and Sandström 1981, p. 264.)

6. Show that the Parade of Dwarfs for a distribution A must lie above that for distribution B if and only if the generalized Lorenz curve of A is steeper than the generalized Lorenz curve of B. (See Thistle 1989b.)

4

Modelling Inequality

> 'I distrust all systematisers and avoid them. The will to a system shows a lack of honesty.'
>
> F. W. Nietzsche, *Maxims and Missiles*

Up till now we have treated information about individual incomes as an arbitrary collection of nuts and bolts which can be put together in the form of an inequality statistic or a graph without any preconception of the general pattern which the distribution ought to take. Any and every logically possible distribution can be encompassed within this analysis, even though we might think it unlikely that we should ever meet any actual situation approximating some of the more abstruse examples. By contrast we might want instead to have a simplified model of the way that the distribution looks. Notice that I am not talking about a model of the *causes* of inequality, although that would be interesting too and might well make use of the sort of models we are going to be handling here. Rather, we are going to examine some important special cases that will enable us to get an easier grip upon particular features of the income distribution. This entails meeting some more specialized jargon, and so it is probably a good idea at the outset to consider in general terms why it is worthwhile becoming acquainted with this new terminology.

The special cases which we shall examine consist of situations in which it is convenient or reasonable to make use of a mathematical formula that approximates the distribution in which we are interested. The inconvenience of having to acquaint yourself with a specific formulation is usually compensated for by a simplification of the problem of comparing distributions in different populations, or of examining the evolution of a distribution over time. The approach can be extremely useful in a variety of applications. You can use it to represent particular parts of the income distribution where a distinctive regularity of form is observed; it can also be used for filling in gaps of information where a dataset is coarse or incomplete (we will be doing just that in Chapter 5); and, as I have mentioned, this technique is often used as a device to characterize the solution to economic models of the income distribution process.

4.1 The idea of a model

At the outset it is necessary to understand the concept of a *functional form*. Typically this is a mathematical formula which defines the distribution function (or the density function, depending on the particular presentation) of not just a single distribution, but of a whole family of such distributions. Each family member has common characteristics and can usually be simply identified within the family by fixing certain numbers known as *parameters*. This can be clarified by an easy example that may be very familiar. Consider the family of all the straight lines that can be drawn on a simple plane diagram. The usual equation that gives the graph of the straight line is:

$$y = mx + c,$$

where y is distance in the 'vertical' direction and x is distance in the 'horizontal' direction. Since this formula defines any straight line in the plane, it can be considered as a general description of the whole family—i.e. as the functional form referred to above. The numbers m and c are, in this case, the parameters. Fix them and you fix a particular straight line as a family member. For example, if you set $m = 1$ and $c = 2$ you get a line with slope 1 (or, a 45° line) that has an intercept on the y-axis at $y = 2$.

When we are dealing with functional forms that are useful in the analysis of inequality, however, we are not of course immediately interested in straight lines, but rather in curves which will look like Figs 2.2 or 2.3. In this case our parameters usually fix things such as the location of the distribution (for example, if one of the parameters is the arithmetic mean) and the dispersion of the distribution (for example, if one of the parameters is the variance).

Now perhaps it is possible to see the advantage of adopting a particular functional form. Let us suppose that you have discovered a formula that fits a particular distribution superbly. We will write down the density function of your fitted formula thus:

$$f = \phi(y; a, b).$$

The notation $\phi(.;.,.)$ simply stands for some expression the details of which we have not troubled to specify; a and b are the parameters. This equation gives you the height f of the smooth curve in the frequency distribution (Fig. 2.2) for any value of income y. Obviously a and b have particular numerical values which give a close fit to the distribution you are examining. However, the empirical distribution that you are considering may be of a very common shape, and it may so happen that your formula will also do for the distribution of income in another population. Then all you have to do is to specify new values of a and b in order to fix a new member of the ϕ-family.

So you could go on using your formula again and again for different distributions (always assuming it *was* a good approximation of course!), each time merely having to reset the two numbers a and b. Let us suppose that the problem in hand is the comparison of the distribution of income in a particular country now with what it was ten years ago, and that it turns out that in each case the ϕ-formula you have discovered very closely fits the observed shape. The comparison is really very easy because you do not have to describe the whole distribution, but you can neatly summarize the whole change by noting the change in the two numbers a and b. No more is required because in specifying a and b you have thus described the whole curve, in the same way that ‘slope’ and ‘intercept’ completely describe an entire straight line.

Because this approach is so convenient it is appropriate to put in some words of warning before going any further. Although this chapter only discusses two functional forms in detail, a great many others have been employed in the social sciences. The properties of some of these are described in the Technical Appendix. However, any such formula is only a convenience. It may turn out that it describes some distributions extremely well, but this should not lull us into expecting it to perform miracles in every situation. Most often we find that such a functional form characterizes certain sections of a distribution. In this case we need to be very aware of its limitations in the less convenient parts—frequently these are the ‘lower tail’ of the distribution. It is usually only fortuitous that a very simple formula turns out to be a highly satisfactory description of the facts in every respect. Finally, in the analysis of economic inequality it is often the case that a simple theoretical caricature of the income- or wealth-generating process leads one to anticipate in theory that a particular functional form of the income or wealth distribution may be realized. Such a conclusion, of course, can only be as sound as the assumptions of the model underlying it. Therefore one is well advised to be suspicious about ‘laws’ of distribution in the sense of claiming that a particular formulation is the one that is somehow metaphysically ‘correct’. In doing so it may be possible to view such formulations in what I believe is the correct perspective: as useful approximations that enable us to describe a lot about different distributions with a minimum of effort.

4.2 The lognormal distribution

In order to grasp the reason for using this apparently unusual distribution with a complicated density function (the mathematical specification is given in the Technical Appendix) it is helpful to come to an understanding of its historical and logical origin. This requires a preliminary consideration of the normal distribution.

The normal distribution itself is of fundamental importance in a vast area of applied statistics, and for an appreciation of its origin and significance reference should be made to sources cited in the notes to this chapter. For our present purposes let us note that since 'the normal curve was, in fact, to the early statisticians what the circle was to the Ptolemaic astronomers' (Yule and Kendall 1950) it is not surprising that scholars have been eager to press it into service in the field of economics and elsewhere. If examination marks, men's height, and errors in experimental observation[1] were supposed to have the normal distribution, then why not look for a 'normal law' governing the distribution of observed quantities in the social sciences?

The term 'normal distribution' describes one family of possible frequency curves, two typical members of which are illustrated in Fig. 4.1. As you can see, the curves are symmetrical about the vertical line through A; point A marks the value μ which is the arithmetic mean of the variable x whose distribution is described by curve (1). This is also the mean of a variable with the distribution of curve (2), which by construction has been drawn with the same mid-value. If curve (2) had a higher mean then it would be displaced bodily to the right of its present position. The higher the variance of the distribution, σ^2, the more 'spread out' will this curve be—compare the values of σ^2 for the two curves. The two numbers μ,σ^2 are the curves' parameters and so completely identify a particular member of the family of normal distributions. If a particular variable x (such as height in a sample of adult males) has the normal distribution with mean μ and variance σ^2, we say that x is distributed $N(x; \mu, \sigma^2)$.

Now it is evident that the distribution of economic quantities such as income do not fit the normal curve (although there are some latter day Ptolemaians who would like to assure us that they 'really' do—see, for example, Lebergott (1959)). As we have seen in Chapter 2, typical income distributions are positively skewed, with a heavy right-hand tail—this is even

[1] It has now been long recognized that the distributions of many such observed characteristics only rarely approximate very closely to the normal distribution. This in no way diminishes the importance of the normal in sampling theory, nor in understanding the historical origin of much of the thought concerning the distribution of incomes.

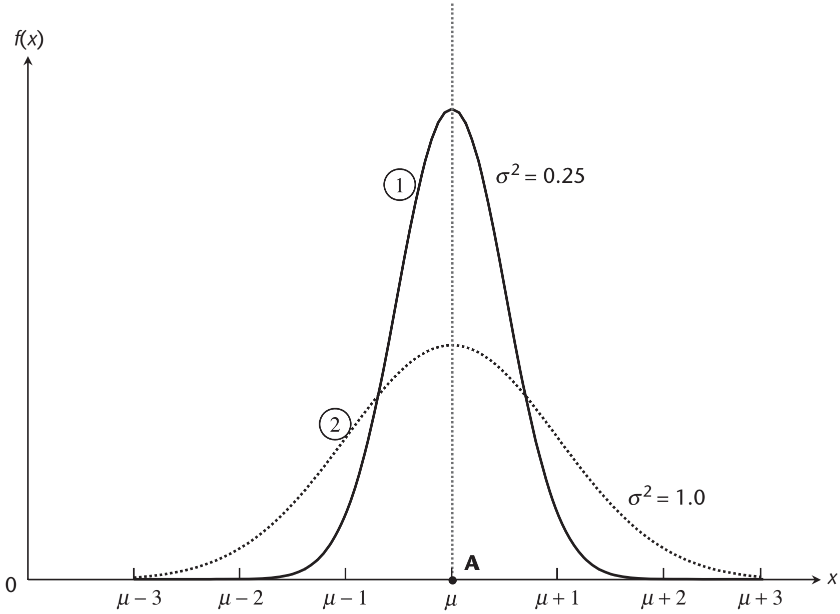

FIG. 4.1. The normal distribution

more noticeable in the case of the distribution of wealth. Is there a simple theoretical distribution that captures this feature?

The lognormal distribution has been suggested as such a candidate, and may be explained in the following manner. Suppose we are considering the distribution of a variable y (income) and we find that the logarithm of y has the normal distribution, then y is said to be *lognormally distributed*. So we transform all our y-values to x-values thus:

$$x = \log(y)$$

(the shape of the curve that describes the relation is given by the $\varepsilon = 1$ curve in Fig. 3.1), we will find that it has the normal distribution like the curves in Fig. 4.1. But what does the distribution of the untransformed variable y itself look like? Two representative members of the lognormal family are illustrated in Fig. 4.2. Notice that, unlike the normal distribution, it is not defined for negative values of the variable y. The reason for this is that as x (the logarithm of y) becomes large and negative, y itself approaches its minimum value of zero, and there is no real number x representing the logarithm of a negative number.

However, the perceptive reader may by now be asking, why choose a logarithmic transformation to produce a distribution of the 'right' shape? There

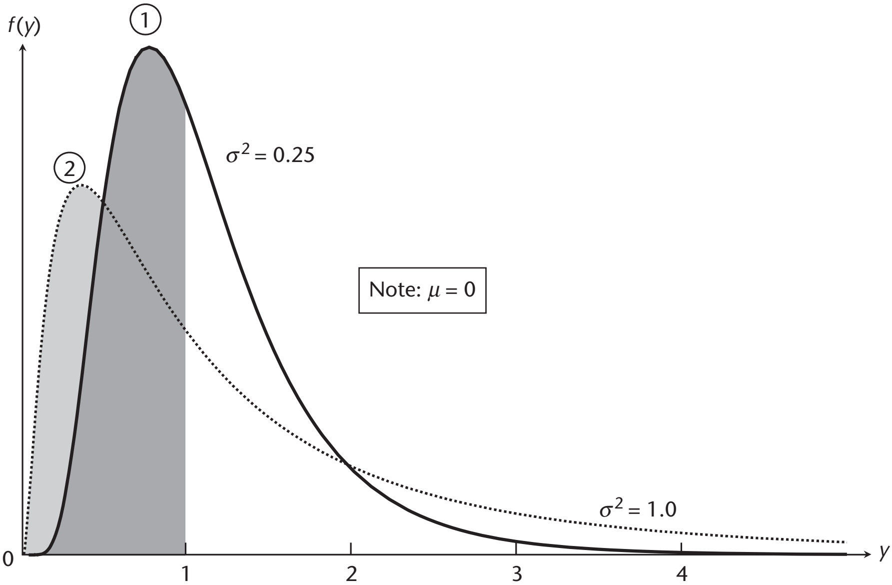

FIG. 4.2. The lognormal distribution

are four reasons. First, the lognormal distribution has a lot of convenient properties, some of which are explained below. Second, it can be shown that under certain kinds of 'random processes' the distribution of incomes eventually turns out to be approximately lognormal. The idea here, roughly speaking, is that changes in people's incomes can be likened to a systematic process whereby, in each moment of time, a person's income increases or decreases by a certain proportion, the exact proportionate increase being determined by chance. If the distribution of these proportionate increments or decrements follows the normal law, then in many cases the overall distribution of income approaches lognormality, provided that you allow enough time for the process to operate.[2] Third, there is still some residual notion of 'individual utility' or 'social welfare' associated with the logarithm of income; it would be nice to claim that although incomes do not follow the normal distribution, 'utility' or 'welfare' do. This will not do, however, for as we have seen in Chapter 3, even if we do introduce a social welfare function, $\log(y)$ is just one among many candidate 'welfare indices'. Fourth, the lognormal provides a reasonable sort of fit to many actual sets of data. This I shall consider later.

[2] Of course, other technical assumptions are required to ensure convergence to the lognormal. In some cases the resulting distribution is similar to, but not exactly equivalent to, the lognormal. This kind of process is also useful in analysing the inequality in the size distribution of firms.

- **Simple relationship to the normal**
- **Symmetrical Lorenz curves**
- **Non-intersecting Lorenz curves**
- **Easy interpretation of parameters**
- **Preservation under loglinear transformations**

THE LOGNORMAL—SPECIAL ATTRACTIONS

Our first reason for using the logarithmic transformation of the normal distribution was, unashamedly, the convenient properties which the resulting distribution possessed. These are now displayed a little more boldly in the accompanying box. Let us look more closely at the 'small print' behind these claims.

The first point, on the relationship with the normal curve, we have already examined in detail. However, it is worth noting that this simple transformation enables the student very easily to obtain the cumulative frequency $F(y)$ corresponding to an income y (the proportion of the population with an income no greater than y):

- find the logarithm of y, say x, from your scientific calculator or a standard computer program;
- 'standardize' this number using the two parameters to calculate $z = \frac{x-\mu}{\sigma}$;
- obtain $F(z)$ from a standard computer program—or look it up in tables of the standard normal distribution.

A further advantage of this close relationship is that a number of common statistical tests which rely on the assumption of normality can be applied straightaway to the logarithm of income, given the lognormal assumption.

The second feature is illustrated in Fig. 4.3: the Lorenz curves are symmetric about the line CQ, where Q is the point on the typical Lorenz curve at which y attains its mean value. This is a little more than a theoretical curiosity since it enables one to see quickly whether there is a *prima facie* case for using the lognormal as an approximation to some given set of data. If the plotted Lorenz curve does not look symmetrical, then it is not very likely that the lognormality assumption will turn out to be satisfactory. The third feature, non-intersecting Lorenz curves, can also be seen in Fig. 4.3.[3] The important conclusion to be derived from this observation is this: *given any two members of the lognormal family of distributions, one will unambiguously*

[3] Please note that this does not follow from the second property. Two arbitrary Lorenz curves, each of which is symmetric, may of course intersect.

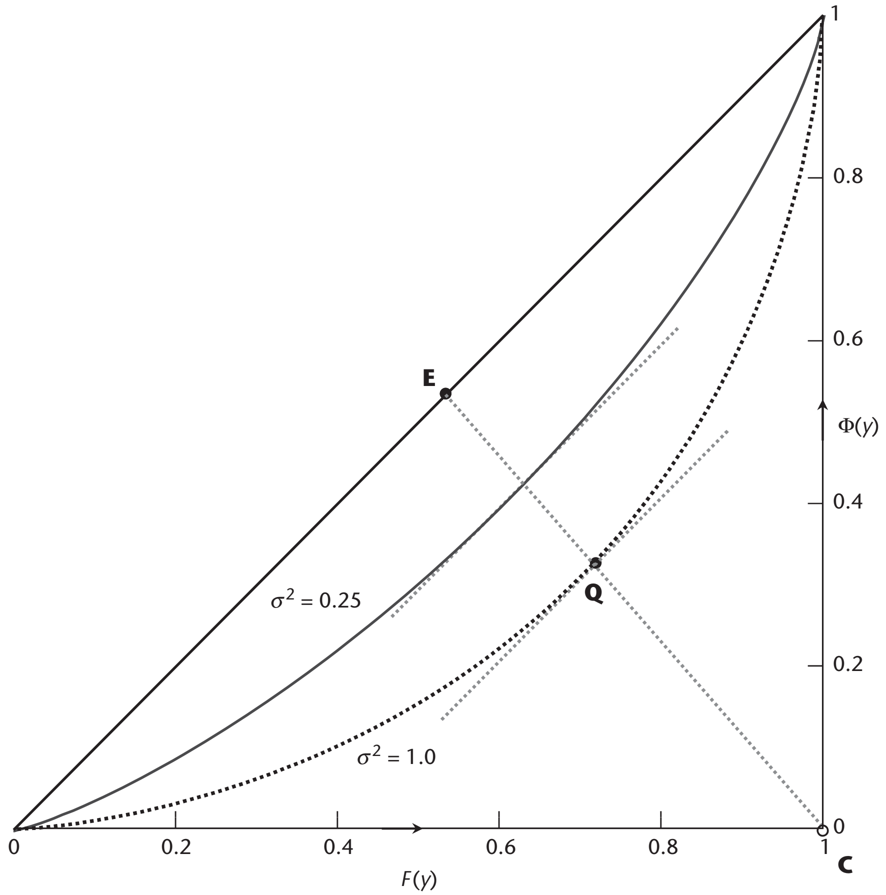

FIG. 4.3. The Lorenz curve for the lognormal distribution

exhibit greater inequality than the other. This remark is to be understood in the sense of comparing the inequality exhibited by the two income distributions using any mean-independent inequality measure that satisfies the weak principle of transfers. It is a direct consequence of Theorem 2, and it is an observation which leads us naturally on to the next feature.

The fourth feature is well-documented. Since there is a simple link with the normal, we may expect a simple link between the parameters μ, σ^2 of the lognormal distribution, written $\Lambda(y; \mu, \sigma^2)$, and the normal distribution. It is evident by definition that μ is the mean of the logarithm of y (or, putting the same point another way, μ is the logarithm of the geometric mean of the values of y). It also happens that μ is the logarithm of the median of y—so that 50 per cent of the distribution lies to the left of the

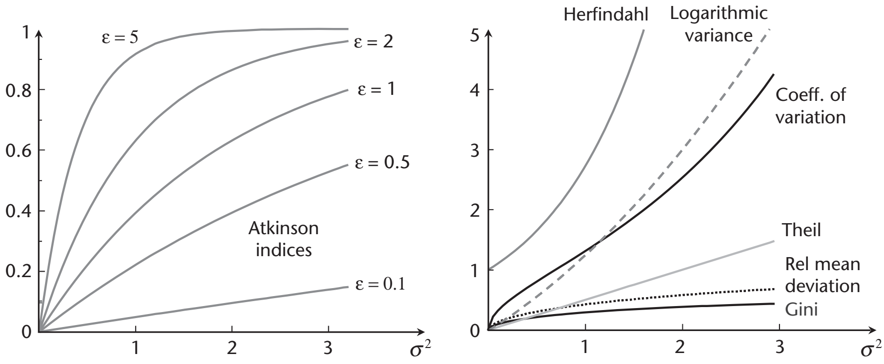

FIG. 4.4. Inequality and the lognormal parameter σ^2

value $y = e^{\mu}$—see the shaded area in Fig. 4.2. Again by definition we see that σ^2 is the variance of the logarithm of y; this is the inequality measure we denoted by v_1 in Chapter 2. As we noted in the last paragraph, if we are comparing members of the two-parameter lognormal family, we never have the problem of intersecting Lorenz curves.[4] Furthermore, since any Lorenz curve is defined independently of the mean, it can be shown that the family of Lorenz curves corresponding to the family of lognormal distributions is independent of the parameter μ. Thus each lognormal Lorenz curve is uniquely labelled by the parameter σ^2. So σ (or σ^2) itself is a satisfactory inequality measure, provided that we restrict our attention to the lognormal family. Of course, if we go outside the family we may encounter the problems noted on page 29.

However, although σ or σ^2 may perform the task of ordinally ranking lognormal curves perfectly well, we may not be attracted by its cardinal properties. Just because the variance of logarithms, σ^2, is a convenient parameter of the lognormal distribution we do not have to use it as an inequality measure. Fortunately, it is very easy to express other inequality measures as simple functions of σ, and a table giving the formula for these is to be found in the Technical Appendix. Some of those which were discussed in the last two chapters are sketched against the corresponding values of σ^2 in Fig. 4.4. Thus to find, say, the value of the Gini coefficient in a population with the lognormal distribution, locate the relevant value of σ^2 on the horizontal axis, and then read off the corresponding value of the inequality measure on the vertical axis from the curve marked Gini.

[4] The problem can arise if one considers more complicated versions of the lognormal curve, such as the three-parameter variant, or if one examines observations from a lognormal population that has been truncated or censored. Consideration of these points is an unnecessary detour in our argument, but you can find out more about this in Aitchison and Brown (1957).

The final point may seem a little mystifying, though it can be useful. It follows from a well-known property of the normal distribution: if a variable x is distributed $N(x; \mu, \sigma^2)$, then the simple transformation $z = a + bx$ has the distribution $N(z; a + b\mu, b^2\sigma^2)$. So the transformed variable also has the normal distribution, but with mean and variance altered as shown.

Let us see how this applies to the lognormal distribution. Now we know that a variable y has the lognormal distribution $\Lambda(y; \mu, \sigma^2)$ if its logarithm $x = \log(y)$ has the normal distribution $N(x; \mu, \sigma^2)$. Suppose we consider any two numbers A, b with the only restriction that A be positive, and write the natural logarithm of A as a. Use these two numbers to transform y into another variable w thus:

$$w = Ay^b,$$

so that by the usual rule of taking logarithms we have

$$\log(w) = a + b\log(y).$$

Denote $\log(w)$ by z and recall the definition that we made above of $x = \log(y)$. Then the last equation can be more simply written

$$z = a + bx.$$

But we know (from above) that because x is distributed $N(x; \mu, \sigma^2)$, z is distributed $N(z; a + b\mu, b^2\sigma^2)$. In other words, the logarithm of w has the normal distribution with mean $a + b\mu$, and variance $b^2\sigma^2$. By definition of the lognormal, therefore, w itself has the lognormal distribution $\Lambda(w; a + b\mu, b^2\sigma^2)$.

To summarize: if y is distributed $\Lambda(x; \mu, \sigma^2)$, then the transformed variable $w = Ay^b$ has the distribution $\Lambda(w; a + b\mu, b^2\sigma^2)$. One of the useful applications of this property is as follows. It has been observed that some countrys' personal tax schedules are approximated reasonably by the formula

$$t = y - Ay^b,$$

where t is individual tax liability and y is income.[5] Then disposable income is given by

$$w = Ay^b.$$

So if the distribution of pre-tax income is approximately lognormal, the distribution of after tax income is also approximately lognormal.

[5] A tax function with this property has been called a 'constant residual progression' tax function after the terminology used by Musgrave and Thin (1948). The parameter b lies between 0 and 1; the smaller is b, the more progressive is the tax schedule; and the smaller is the inequality in the resulting distribution of disposable income.

We will find some very similar properties when we turn to our second special case.

4.3 The Pareto distribution

Although the Pareto formulation has proved to be extremely versatile in the social sciences, in my view the purpose for which it was originally employed is still its most useful application—an approximate description of the distribution of incomes and wealth among the rich and the moderately rich.

Take another look at the frequency distribution of incomes that we first met on page 20. If you cover up the left-hand end of Fig. 2.2 (below about £4,000) you will see that the rest of the underlying curve looks as though it should fit neatly into a simple functional form. Specifically it looks as though this portion of the curve could well be defined by a power function of the form:

$$f(y) = k_1 y^{-k_2},$$

where k_1 and k_2 are constants. With this little exercise you have virtually rediscovered an important discovery by Vilfredo Pareto. In the course of the examination of the upper tails of the income distributions in a number of countries, Pareto found a remarkably close fit to the particular functional form I have just introduced—although in Pareto's standard version the two parameters are specified in a slightly different way from k_1 and k_2, as we shall see below. Since the functional form 'worked' not only for the then current (late nineteenth century) data, but also for earlier periods (as far back as the worthy citizens of Augsburg in 1471), this happy empirical circumstance assumed the status of a Law. Furthermore, since the value of the crucial parameter (now customarily referred to as 'α') seemed to lie within a fairly narrow range, it seemed to Pareto that α might receive the kind of dignification accorded to the gravitational constant in physics.

Unfortunately, I must remind you of the iconoclastic remarks about 'laws' made earlier in this chapter. Although the Paretian functional form provides a good fit for parts of many income or wealth distributions (as well as an abundance of other engaging applications such as the size distribution of cities, the frequency of contribution by authors to learned journals, the frequency of words in the Nootka and Plains Cree languages, the distribution of the length of intervals between repetitions of notes in Mozart's Bassoon Concerto in B♭ Major, and the ranking of the billiards scores by faculty members of Indiana University), the reputation accorded to it by earlier and more naive interpretations has become somewhat tarnished. Neither Davis' mathematical interpretation of history, nor Bernadelli's postulate of the futility of revolutions is comfortably supported by the facts on income distribution.[6] But although the more simplistic hopes (centring on the supposed constancy of Pareto's α) may have been dashed, the underlying distribution remains of fundamental importance for the following reasons.

In the first place, although Pareto's α is not a gravitational constant, as I have pointed out, the functional form still works well for a number of sets of data. Second, the distribution may once again be shown to be related to a simple 'random process' theory of individual income development. The principle is very similar to the process referred to on page 82, the main difference being that a device is introduced to prevent an indefinite increase in dispersion over time, which has the effect of erecting a 'lower barrier' income y below which no one can fall. Third, the Paretian form can be shown to result from simple hypotheses about the formation of individual remuneration within bureaucratic organizations. The idea here is quite simple: given that a hierarchical salary structure exists and that there is a fairly stable relationship between the remuneration of overlord and underling, the resulting frequency distribution of incomes is Paretian. Fourth, the functional form of the Pareto distribution has some remarkably convenient properties in its own right which make it useful for a description of distributional problems and for some technical manipulations, which I discuss in the next chapter.

In order to understand the especially attractive feature of the Pareto distribution you will find it helpful to construct a fresh diagram to present the income distribution data. This will be based on the same facts as were Figs 2.1 to 2.5, but will set out the information in a different manner.

- Along the horizontal axis put income on a logarithmic scale.[7]
- For any income level y transform the cumulative income proportions $F(y)$ by calculating the number $P = 1 - F(y)$.
- Then plot P on the vertical axis also using a logarithmic scale.

What we have done is to plot the proportion of the population with y or more against y itself on a double-logarithmic diagram.

Let us see what the resulting curve must look like. If we look at a low level of income, then the corresponding value of $F(y)$ will be low since there

[6] Curious readers are invited to check the notes to this chapter for details.
[7] This is a scale similar to that used in Fig. 2.5.

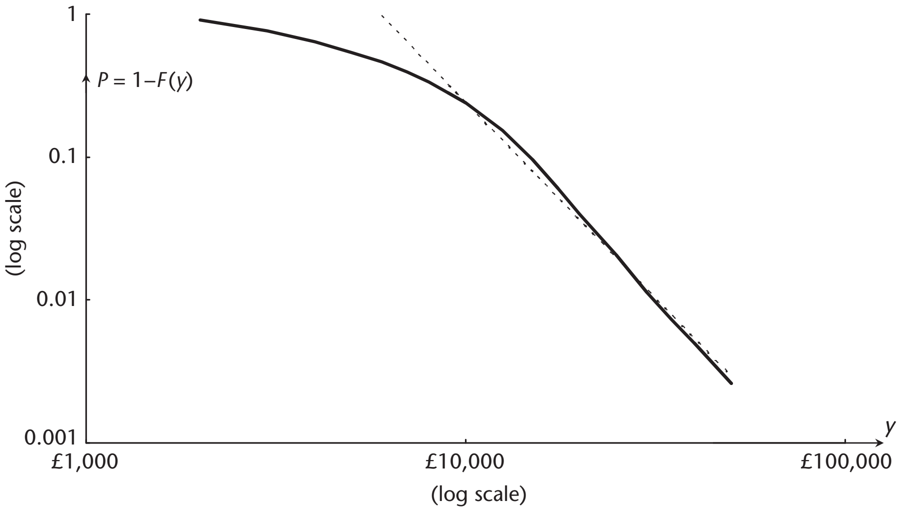

Fig. 4.5. The Pareto diagram
Source: as for Fig. 2.1

will only be a small proportion of the population with that income or less. By the same token the corresponding value of P must be relatively high (close to its maximum value of 1.0). If we look at a much higher level of y, $F(y)$ will be higher (the proportion of the population with that income or less will have risen) and, of course, the number P will be smaller (the proportion of the population with that income or more must have fallen). As we consider larger and larger values of y, the number P dwindles away to its minimum value, zero. Since P is being plotted on a logarithmic scale (and the logarithm of zero is minus infinity) this means that the right-hand end of the curve must go right off the bottom edge of the page. The result is a picture like that of Fig. 4.5. Notice that part of this curve looks as though it may be satisfactorily approximated by a straight line with slope of about $-2\frac{1}{2}$. This gives us the clue to the Pareto distribution.

If the graph we have just drawn turns out to be exactly a straight line throughout its length, then the underlying distribution is known as the Pareto distribution. The slope of the line (taken positively) is one of the parameters of the distribution, usually denoted by α. The income corresponding to the intercept of the line on the horizontal axis gives the other parameter; write this as $\underline{y}$. Two examples of the Pareto family, each with the same $\underline{y}$, but with different values of α are illustrated in Fig. 4.6. The corresponding frequency distributions are drawn in Fig. 4.7. It is apparent from a superficial comparison of this picture with Fig. 2.2 or other frequency distributions based on different datasets that, for income distributions at

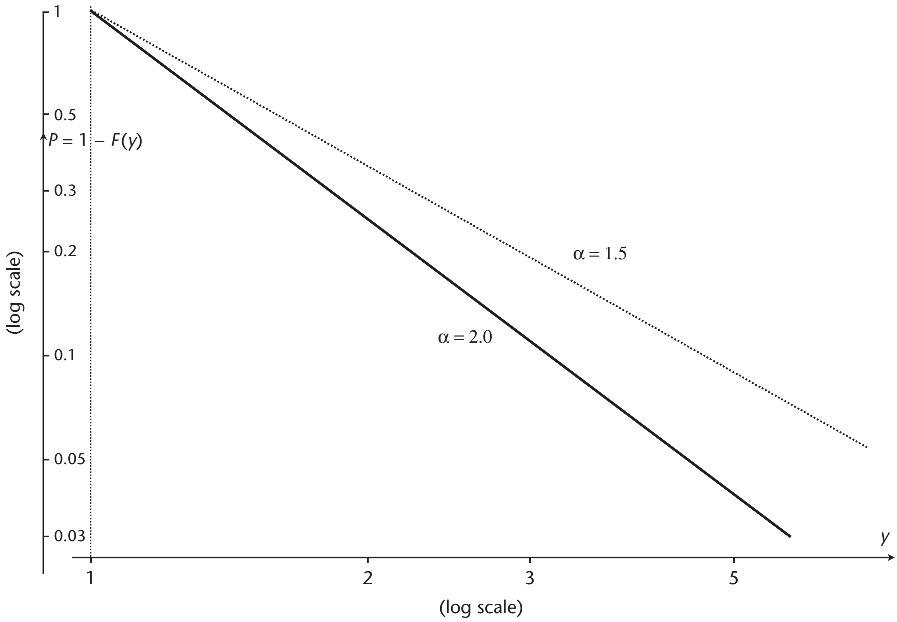

FIG. 4.6. The Pareto distribution in the Pareto diagram

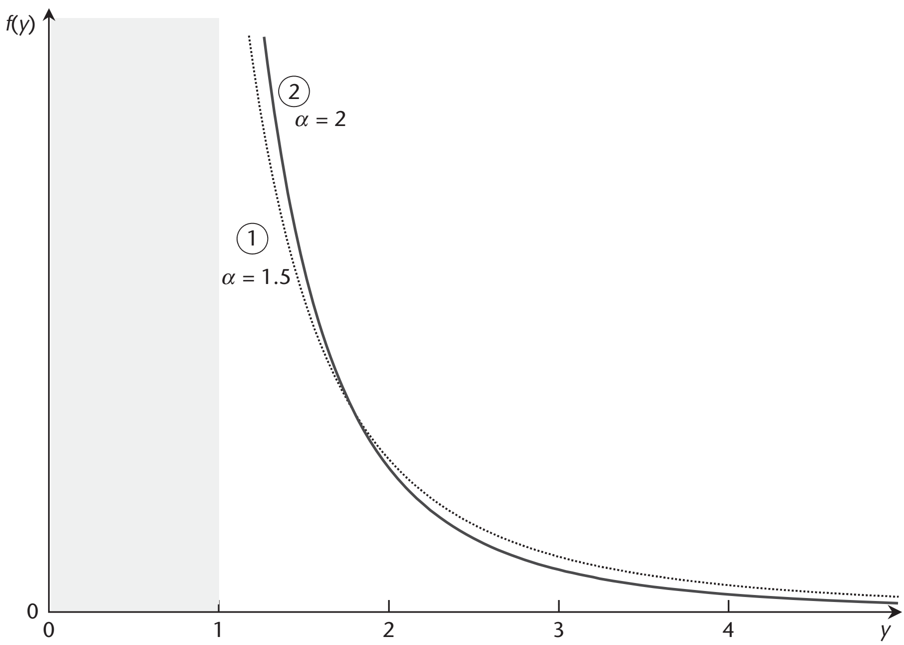

FIG. 4.7. Paretian frequency distribution

least, the Paretian functional form is not likely to be a very good fit in the lower and middle income classes but may work pretty well in the upper ranges, as suggested at the beginning of the section. We shall consider this question of fit further below.

- **Linearity of the Pareto diagram**
- **Van der Wijk's law**
- **Non-intersecting Lorenz curves**
- **Easy interpretation of parameters**
- **Preservation under loglinear transformations**

PARETO—SPECIAL ATTRACTIONS

Let us, then, take a look at some of the special attractions of the Pareto distribution, as advertised, in the accompanying box. Once again we ought to look at the facts behind these assertions.

One particular advantage of the first feature—the simple shape of the Pareto diagram—is that it is easy to work out the distribution function $F(y)$, to calculate the proportion of the population that has y or less. To do this, divide $\underline{y}$ by the required income level y; raise the resulting number to the power α; subtract this result from 1.

On the second point, we find van der Wijk's name attached to a particularly simple law which holds only for the Pareto distribution.[8] Take any income level y as a 'base' income. Then the average income of the subgroup who have an income at least as great as this base income is simply By, where

$$B = \frac{\alpha}{\alpha - 1}.$$

So there is a simple proportionality relationship between this average and the base income level, whatever the chosen value of chosen base income. The constant of proportionality B can itself be seen as a simple inequality measure: 'the average/base' index. Notice that if α increases then B falls: the gap between your own income and the average income of everyone else above you necessarily gets smaller.

The third assertion (of non-intersecting Lorenz curves) is illustrated in Fig. 4.8, and can be readily inferred from the explicit formula for the Lorenz curve of the Pareto distribution given in the Technical Appendix (page 157). From that formula it may be seen that if we choose any value of F in Fig. 4.8

[8] This is true for all continuous distributions. There is a distribution defined for discrete variables (where y takes positive integer values only) which also satisfies the law. See the Technical Appendix, page 161.

Table 4.1. Pareto's α and 'average/base' inequality

Ratio of average income above you to your own income	Pareto coefficient α
1.50	3
1.75	2.333
2.00	2
2.50	1.667
3.00	1.5

(measured along the horizontal axis), then as we choose successively larger values of α, each lying on a new Lorenz curve, the corresponding value of Φ must become progressively larger. In other words, as we choose larger values of α all the points on the relevant Lorenz curve must lie closer to the diagonal. So no two Paretian Lorenz curves can cross.

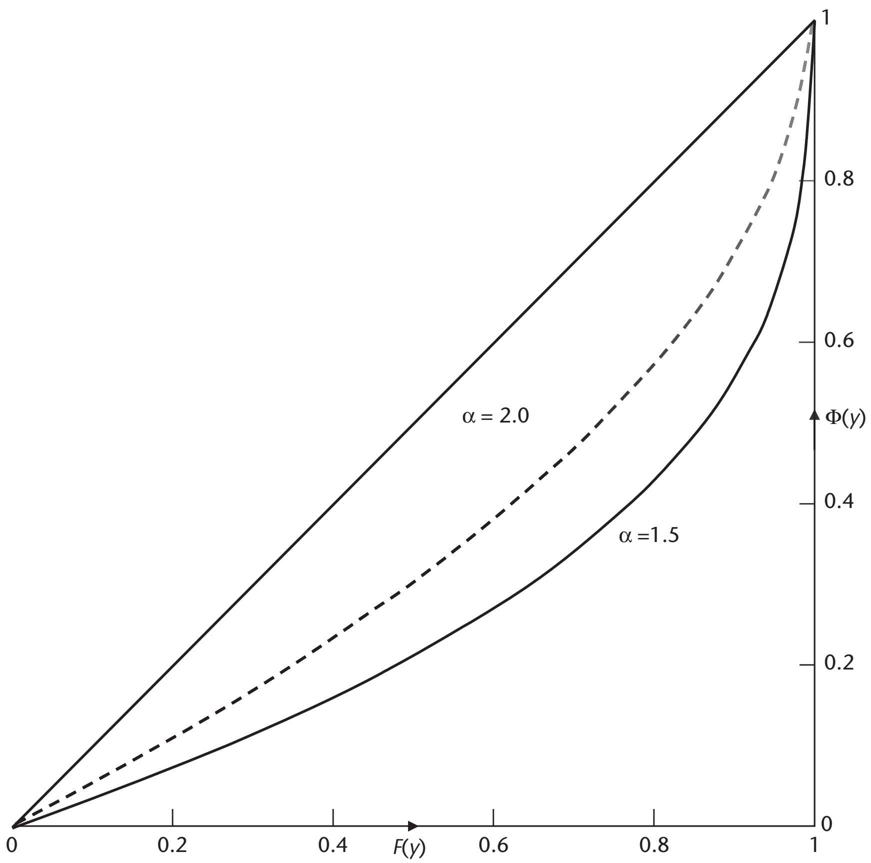

FIG. 4.8. The Lorenz curve for the Pareto distribution

These observations take us naturally on to our fourth point—the interpretation of the parameters. You may already have come to suspect that the parameter α reveals something about the amount of inequality exhibited by a particular Pareto distribution. Since it is evident that, within the Pareto family, Lorenz curves associated with higher values of α are closer to the line of perfect equality, it follows that if we compare two Pareto distributions with the same mean, the one with the higher value of α exhibits the less amount of inequality for all inequality measures satisfying the weak principle of transfers.[9]

Once again, just because the parameter α is convenient in the case of the Pareto distribution, this does not mean that there is any particular merit in using it as a measure of equality. We may prefer the cardinal characteristics of some other measure, in which case we may compute the alternative measure as a function of α using the table in the Technical Appendix, or using Fig. 4.9. This figure is to be interpreted in a manner very similar to that of Fig. 4.4 in the case of the lognormal distribution. The interpretation of the parameter $\underline{y}$ can easily be seen from Fig. 4.9, which has been drawn with $\underline{y}$ set arbitrarily to one. This parameter may assume any positive (but not zero) value, and gives the lower income limit for which the distribution is defined. By a simple application of van der Wijk's law, putting yourself at minimum income $\underline{y}$, it can be seen that mean income for the whole population is

$$\frac{\alpha}{\alpha - 1}\underline{y}.$$

So, average income is proportional to minimum income and is a decreasing function of α.[10]

The formal meaning of the fifth and final point in our list is the same as in the case of the lognormal distribution. A proof is not difficult. Suppose that the quantity y has the Pareto distribution with parameters $\underline{y}$ and α. Then

[9] An intuitive argument can be used here. Using Van der Wijk's law you find the gap between your own income and the average income of everyone above you diminishes the larger is α. Thus the 'unfairness' of the income distribution as perceived by you has diminished.

[10] Another apparently paradoxical result needs to be included for completeness here. Specify any social welfare function that satisfies properties 1 to 3 of Chapter 3 (note that we are not even insisting on concavity of the SWF). Then consider a change from one Pareto distribution to another Pareto distribution with a higher α but the same value of minimum income (for example the two curves in Fig. 4.7). We find that social welfare *decreases* with α although, as we have seen, inequality also decreases for any 'sensible' mean-independent inequality measure. Why does this occur? It is simply that as α is increased (with $\underline{y}$ held constant) mean income $\bar{y}$, which equals $\alpha\underline{y}/[\alpha - 1]$, decreases and this decrease in average income is sufficient to wipe out any favourable effect on social welfare from the reduction in equality. Of course, if α is increased, and minimum income is increased so as to keep $\bar{y}$ constant, social welfare is increased for any individualistic, additive, and concave social welfare function.

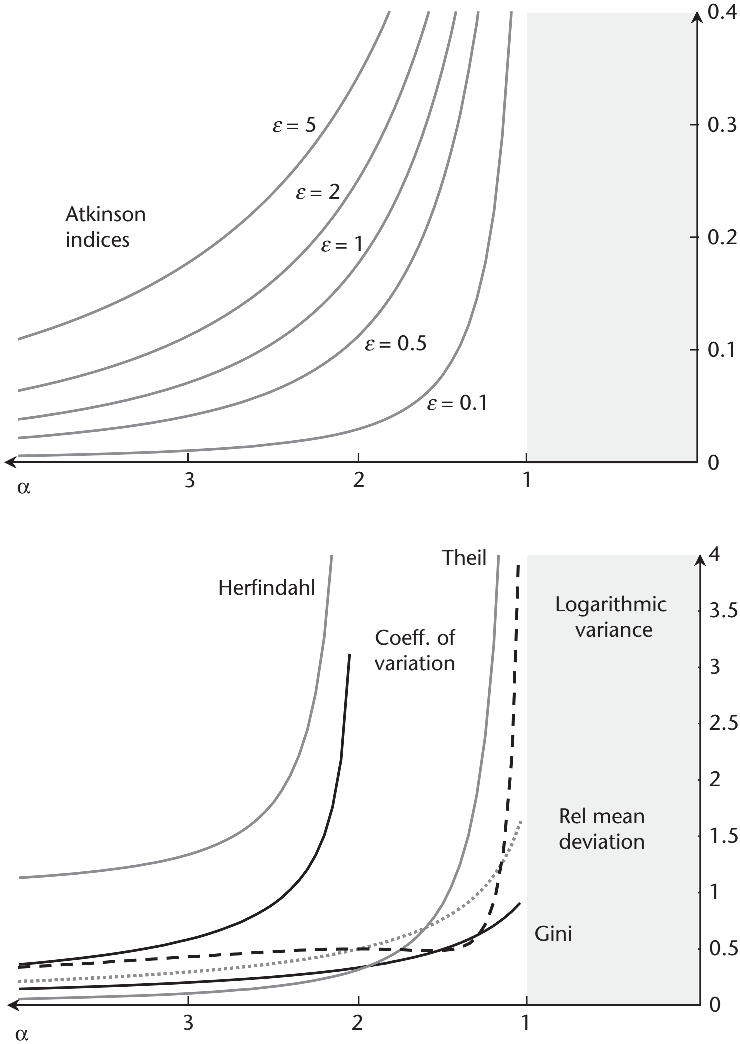

FIG. 4.9. Inequality and Pareto's α

from the Technical Appendix we find that the proportion of the population with income less than or equal to y is given by

$$F(y) = 1 - \left[\frac{y}{\underline{y}}\right]^{-\alpha}.$$

Now consider another quantity w related to y by the formula,

$$w = Ay^b,$$

where of course the minimum value of w is $\underline{w} = A\underline{y}^b$. Then we see that

$$\frac{y}{\underline{y}} = \left[\frac{w}{\underline{w}}\right]^{1/b}.$$

Substituting in the formula for F we find

$$F(w) = 1 - \left[\frac{w}{\underline{w}}\right]^{-\alpha/b}.$$

In other words the transformed variable also has the Pareto distribution with parameters $\underline{w}$ and α/b. Therefore we once again have the simple result that if pre-tax incomes are distributed according to the Pareto law, and if the tax system is closely approximated by the constant residual progression formula, then post-tax incomes are also distributed according to the Pareto law.

4.4 How good are the functional forms?

An obviously important criterion of suitability of a functional form is that it should roughly approximate the facts we wish to examine. It is too much to hope for that one formula is going to fit some of the data all of the time or all of the data some of the time, but if it fits a non-negligible amount of the data a non-negligible amount of the time then the mathematical convenience of the formula may count for a great deal. One immediate difficulty is that the suitability of the functional form will depend on the kind of data being analysed. I shall deal with this by arbitrarily discussing four subject areas which are of particular economic interest. In doing so I am giving a mere sketch of the facts which may provide those interested with a motivation to enquire further.

Aitchison and Brown (1957) argued that the lognormal hypothesis was particularly appropriate for the distribution of earnings in fairly homogeneous sections of the workforce. Thus, for example, in British agriculture in 1950 we find that the distribution of earnings among cowmen, the distribution among horsemen, that among stockmen, and that among market gardeners proves in each case to be close to the lognormal. This evidence is also borne out in other specific sectors of the labour market and in other countries.

When we look at more comprehensive populations a difficulty arises in that the aggregate of several distinct lognormal distributions may not itself be lognormal. Suppose you have a number of different subgroups within the population (for example cowmen, horsemen, stockmen, etc.) and within each subgroup the distribution in the resulting population (all agricultural workers) will only be lognormal if, among other things, the dispersion parameter σ^2 may be taken as uniform throughout the groups. If your lognormal pigmen have a higher σ^2 than your lognormal tractor drivers, then you are in trouble. Possibly because this restrictive condition is not generally satisfied, systematic departures from lognormality are evident in many earnings distributions—although it is interesting to note that Fig. 4.10 illustrates that the lognormal distribution is not a bad approximation for male manual earnings in the UK. Because of this difficulty of aggregation Lydall (1968), in attempting to find a general description of his 'standard distribution' of pretax wages and salaries for all adult non-agricultural workers, makes the following observations. The central part of the distribution (from about the 10th percentile to the 80th percentile) is approximately lognormal. But the observed distribution has more of its population in its tails than a member of the lognormal family should have. In fact the upper tail (about the top 20 per cent of the population) approximates more closely to the Pareto distribution.

If we are going to use current receipts as some surrogate measure of economic welfare then it is clear that a more comprehensive definition of income is appropriate. When we examine the distribution of income (from

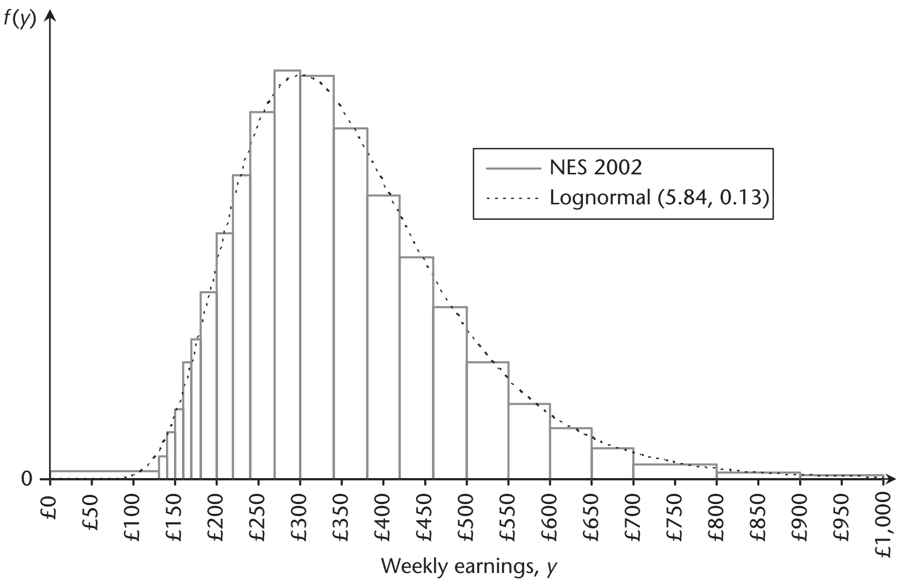

FIG. 4.10. The distribution of earnings. UK male manual workers on full-time adult rates

Source: New Earnings Survey, 2002

all sources) we find that the lognormal assumption is less satisfactory, for reasons similar to those which we discussed when dealing with the aggregation of earnings—compare the logarithmic transformation in Fig. 2.5 with the 'ideal' shape of Fig. 4.1 just above. We are quite likely to find substantial departures at the lower tail, for reasons that are discussed in the next chapter. However, for the middle part of the income distribution, lognormality remains a reasonable assumption in many instances, and the assumption of a Paretian upper tail remains remarkably satisfactory, as the evidence of Fig. 4.5 bears out. This enables us to take a piecemeal approach to modelling inequality, adopting different functional forms for different parts of the income distribution, which may be useful if we just want to focus on one part of the picture of inequality rather than attempting a panoramic view.

As we have seen, it is this close approximation of the upper tail which led to some of the more optimistic conjectures of Pareto's disciples. It is perhaps otiose to mention that since Pareto's data necessarily related to high incomes alone, his law can hardly be expected to apply to the income distribution as a whole. The Paretian upper tail that has emerged from a study of income distributions also works very well for the distribution of wealth. There is a superficial reason to suppose that a curve like Pareto's might be useful in this application. Wealth data are often compiled with any accuracy only for the moderately wealthy and above. Hence—excluding those whose wealth is unrecorded—one often finds a single-tailed distribution. Evidence on the linearity of the Pareto diagram (and hence on the close fit of the Pareto formula) is clear from Fig. 4.11; notice that the straight line approximation

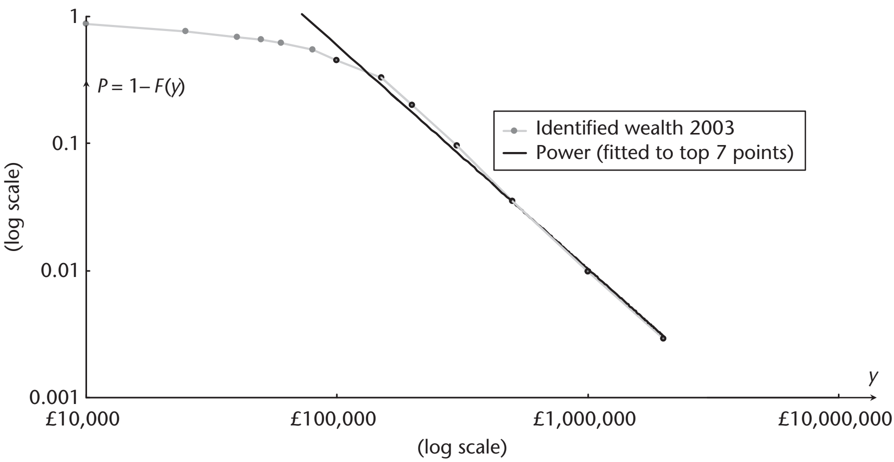

FIG. 4.11. Pareto diagram. UK wealth distribution 2003
Source: Inland Revenue Statistics

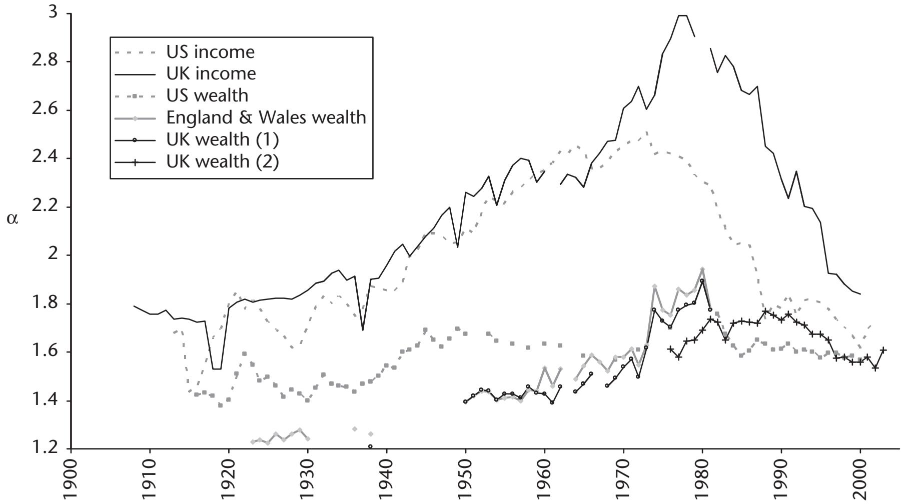

FIG. 4.12. Pareto's α: USA and UK
Source: see text

Table 4.2. Pareto's α for income distribution in the UK and the USA

UK		USA	
1688	1.58	1866–71	1.40–1.48
1812	1.31	1914	1.54
1843	1.50	1919	1.71
1867	1.47	1924	1.67
1893	1.50	1929	1.42
1918	1.47	1934	1.78
1937–38	1.57	1938	1.77
1944–45	1.75	1941	1.87
		1945	1.95

Source: Bronfenbrenner (1971), p. 46

is particularly good if we drop the first few observations rather than trying to fit a line to all the points.

Figure 4.12 illustrates the history of Pareto's α from the early twentieth century to the early twenty-first century, for both income and wealth;[11] Table 4.2 gives some elements of the incomes story from earlier times. It

[11] The series are based on tax data and focus solely on upper incomes (before tax) and substantial wealth. Sources are as follows. US income: Atkinson and Piketty (2007), Chapter 5. UK income: Atkinson and Piketty (2007), Chapter 4. US wealth: Kopczuk and Saez (2004). England and Wales wealth: Atkinson *et al*. (1989). UK wealth (1): Atkinson *et al*. (1989). UK wealth (2): HMRC Series C. For the data and methods see the file 'Pareto Example' on the website; see also Question 8 below.

is clear that, in the case of incomes, the values of α are typically in the range 1.5 to 2.5 and that the value for wealth is somewhat lower than that. It is also clear that α had been rising for much of the twentieth century (in other words inequality was *falling*) but that that in the last 25 years or so there has been a marked reversal of this trend.

For our final application, the analysis of the distribution of firms by size, succinct presentation of the evidence and comparison with the special functional forms can be found in Hart and Prais (1956) (for the UK) and in Steindl (1965) (for the USA and Germany). The Pareto law only works for a small number of firms that happen to be very large—but, as Steindl (1965) points out, although this represents a small proportion of individual business units, it accounts for a large proportion of total corporate assets. You typically find α in the (rather low) 1.0 to 1.5 range. However, the lognormal functional form fits a large number of distributions of firms by size—where size can variously be taken to mean corporate assets, turnover, or number of employees. These approximations work best when industries are taken in broad groupings rather than individually.

This perfunctory glimpse of the evidence is perhaps sufficient to reinforce three conclusions which may have suggested themselves earlier in the discussion.

- Neither the Pareto nor the lognormal hypothesis provides a 'law' of distribution in the strict sense that one particular member of either family is an exact model of income or distribution in the long run. In particular it is nonsense to suppose that the Pareto curve (where applicable) should remain stable over long periods of history. As it happens, α had been increasing nearly everywhere until recently.
- However, interpreting the Pareto or the lognormal 'law' as a description of the shape of particular distributions is more promising. Neither hypothesis usually works very well,[12] since the real world is too complicated for this, unless we look at a very narrow and well-defined piece of the real world such as the earnings of cowmen or the wealth of people with more than £100,000.
- Nevertheless one or other functional form is a reasonable approximation in a heartening number of cases. The short cuts in empirical analysis that are thus made possible amply repay the trouble of understanding the mechanics of the mathematical formulas in the first place. In some cases one may be able to make much better approximations using more

[12] See the next chapter for a brief discussion of the criteria of fit.

sophisticated functional forms—a discussion of these is provided on pages 158 onwards.

This simplification will perhaps be more readily appreciated when we come to wrestle with some of the difficulties that arise in the next chapter.

4.5 Questions

1. Suppose $\{u_1, u_z, , \ldots, u_t, \ldots\}$ is a sequence of independently and identically distributed normal variables. If u_t is distributed $N(0, v^2)$ what is the distribution of λu_t where λ is a positive constant? Now suppose that successive values of the variable x_t are determined by the following process:

$$x_t = \lambda x_{t-1} + \mu_t,$$

for $t = 1, 2, 3, \ldots$ where u_t satisfies the assumptions just described and is independent of x_t. Write x_t as a function of the initial value x_0 and the sequence $\{u_1, u_1, , \ldots, u_t, \ldots\}$. Show that

$$\text{var}(x_t) = \lambda^{2t}\text{var}(x_0) + v^2 \frac{1 - \lambda^{2t}}{\lambda^{2t} - 1}.$$

2. Suppose income at time 0, y_0, is distributed lognormally. Over a sequence of periods $t = 1, 2, 3, \ldots$ the logarithm of income x_t then follows the above process. Give a simple economic interpretation of what is happening. What will be the distribution of income in period t? Under what conditions will the distribution of income converge in the long run? If there is convergence what is the long-run value of the Gini coefficient?
3. Using the data for the UK 2003 earnings distribution ('NES' on the website) compute the mean and the coefficient of variation (i) directly from the raw data and (ii) using the fitted lognormal distribution illustrated in Fig. 4.10 (use the relevant formula in Table A.2 on page 156).
4. Show that the 'first guess' at the Pareto distribution given by the formula for the frequency distribution on page 87 really does correspond to the formula for the distribution function F on page 157 of Appendix A. What is the relationship of the constants k_1 and k_2 to the parameters $\underline{y}$ and α?
5. Use the formulas given in the Table A.2 and on page 157 to:
   (a) derive the generalized Lorenz curve for the Pareto distribution;

(b) sketch the relationship between the coefficient of variation c and α in Fig. 4.9;

(c) show why is c undefined for $\alpha \leq 2$.

6.

(a) Using the data for the UK wealth distribution 2003 (see the file 'IR wealth' on the website) compute the Gini coefficient on the assumptions (i) that persons not covered by the wealth table are simply excluded from the calculation, and (ii) individuals in a given wealth interval class possess the mean wealth of that interval.

(b) Rework the calculation in part (a) on the alternative assumption that the group excluded by assumption (i) actually consists of n persons each with a wealth y_0, where n and y_0 are positive numbers (chosen by you). What would be reasonable ranges of possible values for these numbers? How does the computed Gini coefficient vary with n and y_0?

7.

(a) Using the same source on the website as in Question 6, for the lower bound of each wealth interval y, compute P (as defined on page 88) and then use ordinary least squares to fit the equation

$$\log(P) = \beta_0 + \beta_1 \log(y);$$

then find the estimate of Pareto's α. Use this estimate to compute the Gini coefficient on the hypothesis that the underlying distribution is Paretian.

(b) Repeat part (a) after dropping the first three intervals.

(c) Compare your answers with those for Question 6.

8.

(a) Suppose the Pareto-type density given on page 87 applies only to a bounded income interval $[a, b]$ rather than to the whole range of incomes. Compute the mean and the variance of this distribution, and compare them with the results for the standard Pareto type I distribution given on page 157.

(b) Suppose that in a set of official income data you are told the upper and lower boundaries of a particular income interval, the numbers of incomes in the interval, and the total amount of income in the interval. Show how you could use the formula derived in part (a) for the mean to derive an estimate of the value of Pareto's α in the interval (see also the discussion on page 128 in Chapter 5 and page 175 in the Technical Appendix).

(c) Suppose that you are given the following information about top incomes in a case where you believe the underlying distribution to be Paretian:

Group	Income share
top 0.01%	3.21%
top 0.05%	6.58%
top 0.1%	8.68%
top 0.5%	15.46%
top 1%	19.24%
top 5%	30.35%
top 10%	37.03%

Show how you could use this information to provide a simple estimate of Pareto's α (see the file 'Pareto example' on the website).

5

From Theory to Practice

'What would life be without arithmetic, but a scene of horrors?'

Rev. Sydney Smith (1835)

So where do we go now? One perfectly reasonable answer to this would be to return to some of the knotty theoretical issues to which we accorded only scant attention earlier.

Were we to follow this course, however, we should neglect a large number of problems which must be wrestled with before our ideas on inequality can be applied to numbers culled from the real world. In this chapter we shall review these problems in a fairly general way, since many of them arise in the same form whatever concept of income, wealth, or other personal attribute is examined, and whatever the national or international source from which the data are drawn.

- **Data**
- **Computation**
- **Appraising the results**
- **Special functional forms**
- **Interpretation**

A CATALOGUE OF PROBLEMS

It is expedient to subdivide the practical problems that we shall meet into five broad groups: those that are associated with getting and understanding the original data; those arising from computations using the data; those involved in an appraisal of the significance of these calculations; the problems connected with the use of special functional forms for income distribution; and the interpretation of results. Of course many of these problems interact. But we shall try to deal with them one at a time.

5.1 The data

The primary problem to be dealt with by anyone doing quantitative research into inequality is that of defining the variable y which we have loosely called 'income', and then getting observations on it. In this section we deal with some of these conceptual and practical issues.

For certain specific problem areas the choice of variable and of source material is usually immediately apparent. For example, if one is interested in the inequality of voting power in a political system, the relevant variable is the number of seats allocated per thousand of the population (the fraction of a representative held by a voting individual); in this situation it is a straightforward step to impute an index of voting power to each member of the electorate. However, in a great many situations where inequality measures are applied, a number of detailed preliminary considerations about the nature of the 'income' variable, y, and the way it is observed in practice are in order. The reasons for this lie not only in the technique of measurement itself, but also in the economic welfare connotations attached to the variable y. For in such cases we typically find that a study of the distribution of income or wealth is being used as a surrogate for the distribution of an index of individual well-being. We shall consider further some of the problems of interpreting the data in this way once we have looked at the manner in which the figures are obtained.

There are basically two methods of collecting this kind of information:

- You can ask people for it.
- You can make them give it to you.

Neither method is wholly satisfactory since, in the first case, some people may choose not to give the information, or may give it incorrectly and, in the second case, the legal requirement for information may not correspond exactly to the data requirements of the social analyst. Let us look more closely at what is involved.

Method 1: Asking People

This approach is commonly used by those organizations that desire the raw information for its own sake. It involves the construction of a carefully stratified (and thus representative) sample of the population, and then requesting the members of this sample to give the information that is required about their income, wealth, types of asset-holding, spending patterns, household composition, etc. This method is used in the UK's Family Resources Survey, and in the Current Population Surveys conducted by the US Bureau of the Census. Obviously a principal difficulty is, as I mentioned, that of non-response or misinformation by those approached in the survey. A common presumption is that disproportionately many of those refusing to cooperate will be among the rich, and thus a potentially significant bias may be introduced into the results. However, the response rate in some of the major surveys is surprisingly good (typically some 60 per cent to 80 per cent), and usually the raw data are weighted in order to mitigate the effect of non-response bias. A manifest advantage of this method of data collection is that if a person volunteers to take part in a survey, it may be possible to secure much more detailed and diverse information than could be arranged under a method involving compulsion, thus potentially broadening the scope of social enquiry.

Method 2: Compulsion

Useful information on income and wealth is often obtained as a by-product of such tiresome official obligations as making tax returns. The advantages of this conscript data over the volunteered survey data are obvious. Except where the tax administration is extremely informal (as is commonly supposed to be true in some Mediterranean countries) such that evasion introduces a substantial bias, it is usually possible to obtain a larger and more representative sample of the population. Non-response bias is less important, and it may be that in some countries legal penalties act as a suitable guarantee to ensure the minimum of misinformation.

However, the drawbacks of such data are equally evident. In the study of income distributions, the income concept is that for which it is expedient for the authorities to define the tax base, rather than a person's 'net accretion of economic power between two points in time' (Royal Commission on the Taxation of Profits and Income 1955), which is considered to be ideal for the purposes of the economist. Hence many components of a comprehensive definition of income—such as capital gains, fringe benefits, home production, the imputed value of leisure time and of owner-occupancy—may be imperfectly recorded, if recorded at all. Indeed, one may suppose that frequently both the rich and the not-so-rich will have taken steps legally to avoid the tax by transforming some part of their income into non-taxable—and unpublished—forms. These warnings apply with increased emphasis in the case of wealth. Furthermore the sample population whose income or wealth is reported in the official figures often inaccurately represents the poor, since those with income or wealth below the tax exemption limit may either be excluded, or be recorded in insufficient detail.

The picture of inequality that would emerge from this sort of study is seen in Fig. 5.1, which illustrates the UK distribution of income before and after tax in 2005/6, based on tax returns. It is tempting to contrast this with the picture that we have already seen based on the more comprehensive *Economic Trends* data for 1984/5 (compare the broken curve in Fig. 5.1 here with Fig. 2.2 on page 20 above). Of course this is not an entirely satisfactory comparison between the distributions to be obtained from the two data sources; after all the diagrams refer to periods that are years apart. However, if we try to bring the comparison up-to-date we encounter a difficulty that is common even in countries with well-developed statistical services: the *Economic Trends* series no longer exists.

To make a reasonable comparison of the pictures of income distribution that would emerge from the two principal methods of data-gathering, we could use a more recently published source that is now the UK's official income distribution series. *Households Below Average Income* (HBAI) provides estimates of disposable income based on the UK's Family Resources Survey, the results of which are summarized in Fig. 5.2, using the same income groupings as in Fig. 5.1. In comparing this figure with the Inland

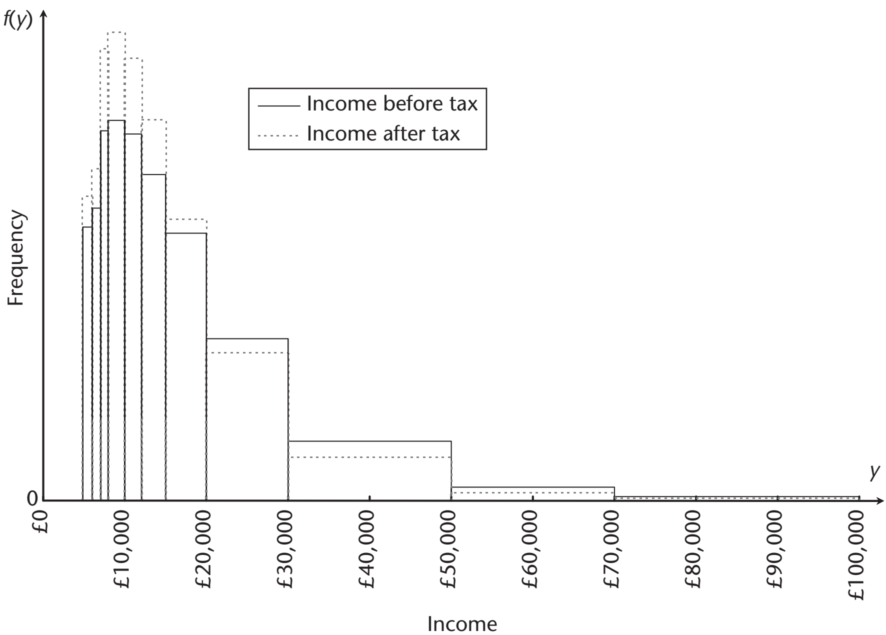

Fig. 5.1. Frequency distribution of income, UK 2005/6, before and after tax
Source: Inland Revenue Statistics

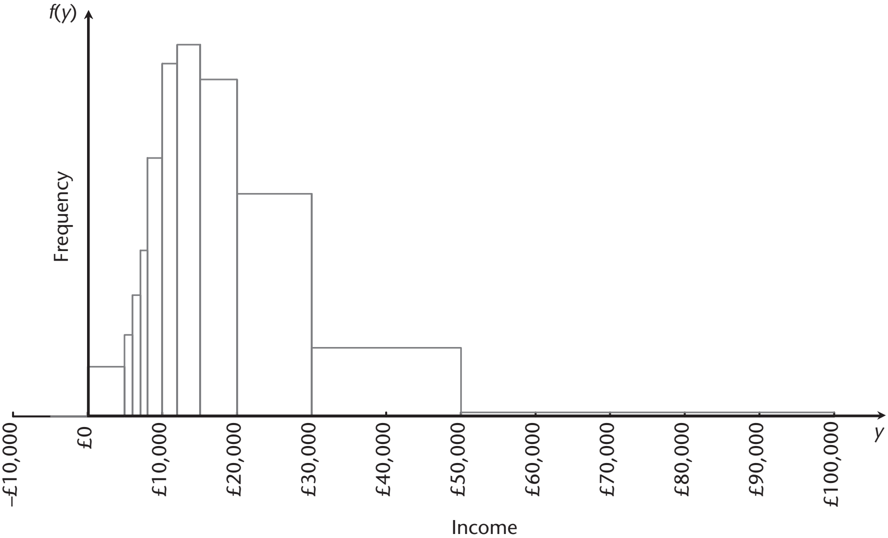

Fig. 5.2. Disposable income (before housing costs) UK 2006/7
Source: *Households Below Average Income*, 2008

Revenue Statistics distribution of income after tax (the dotted line in Fig. 5.1) we immediately notice the interesting shape of the lower tail in Fig. 5.2 by contrast to the manifestly incomplete picture of the lower tail in Fig. 5.1.

- **What is included?**
- **Which heads are counted, and who shares in the cake?**
- **To what time period does it relate?**
- **What valuation procedure has been used?**
- **Which economic assumptions have been made?**

THE VARIABLE *y*: A USER'S GUIDE

With little mental effort, then, we see that the practical definition of the variable *y*—and hence the picture of its distribution—is only going to be as good as the way in which the information on it is compiled. So if you, as a student of inequality, are being asked to 'buy' a particular set of data on income or wealth, what should you watch out for? For a quick assessment, try the check-list in the accompanying box. Let us briefly examine each of these five items in turn.

WHAT IS INCLUDED?

Recalling the argument of Chapter 1, if we concern ourselves with a narrowly defined problem there is relatively little difficulty: an inquiry into, say, the inequality in earnings in some particular occupation will probably require a simple definition of the income variable. I shall use this approach later in the chapter when we look at inequality in the income reported to the tax authorities in the USA. For a wide interpretation of inequality, of course, you obviously need to reflect on whether the definition of income is as all-embracing as it was suggested on page 105 that it should be. Furthermore, if you want to arrive at people's disposable incomes, then careful consideration must be given to the adjustment that has been made for direct and indirect taxes, for social security benefits and other money transfer incomes, and for benefits received 'in kind' from the state, such as education.

This point raises issues that deserve a chapter—if not a book—to themselves. However, we can get a feel for the practical impact of an adjustment in the concept of income by referring again to the data source used for Fig. 5.2. Some have argued that, because of the way in which housing expenditures are sometimes treated as a kind of committed expenditure component in the UK they should be treated as though they were a tax, and should therefore be deducted to get a truer picture of disposable income. Irrespective of the economic merits of this argument, it is interesting to note the impact of this on the apparent inequality of the income distribution—see Fig. 5.3 which

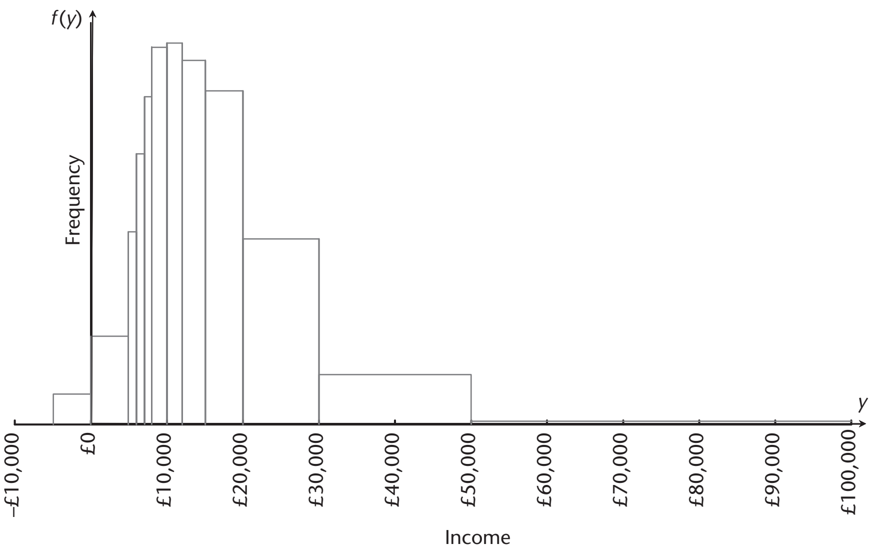

FIG. 5.3. Disposable income (after housing costs) UK 2006/7
Source: *Households Below Average Income*, 2008

presents the after-housing-cost (AHC) distribution using the same income groupings as for Figs 5.1 and 5.2 (note that the AHC distribution has a number of negative incomes).

WHICH HEADS ARE COUNTED?

The answer is obvious in some cases—for example in a study of the distribution of voting power one considers each enfranchized person. In other cases, such as those where tax returns are used, the choice of 'heads' is made for us—they are the 'tax units', which sometimes means all men and women individually, but often refers to nuclear families and to unrelated individuals. For wealth data, the unit is in general a single 'estate', the benefits of which may be enjoyed by one person, or by a number in a family group. Unfortunately, detailed information such as family composition of the income- or wealth-holding tax units is available for few countries, whereas this detail can usually be obtained from voluntary sample surveys. Where this detail is available one may allow for differing family size by taking two distinct steps:

	Modified OECD		McClements	
	BHC	**AHC**	**BHC**	**AHC**
First adult	0.67	0.58	0.61	0.55
Spouse	0.33	0.42	0.39	0.45
Other second adult	0.33	0.42	0.46	0.45
Third adult	0.33	0.42	0.42	0.45
Subsequent adults	0.33	0.42	0.36	0.40
Children aged under 14 yrs	0.20	0.20	0.20	0.20
Children aged 14 yrs and over	0.33	0.42	0.32	0.34

Source: Department of Work and Pensions (2008) Appendix 2

EQUIVALENCE SCALES

- Adjusting each family's income to allow for differences in needs between different types of families. The process—known in the jargon as 'equivalizing' the incomes—involves dividing the income by an index. The first column in the accompanying box is a modified version of the widely used OECD equivalence scale, where the scale is normalized so that a couple—i.e. two adults living together—has an index equal to 1 (for example taking the BHC version a couple with two children under 14 and a nominal income of £40,000 would have an equivalized income of £40,000/(0.67+0.33+0.20+0.20) = £28,571.43); the second column is the counterpart scale that would be applied to AHC data. The HBAI data now uses this method of adjusting for needs as standard, but it used to use the scale presented in the third and fourth columns (McClements 1977). As we can see, the two conventional scales will produce the same results for a family consisting of a couple and young children, but they would give different results for single adults living alone. The equivalence scale could in principle be derived in a number of ways: by using expert assessments of budgets required to meet minimum standards, by comparing the actual expenditure patterns of different types of family on particular categories of goods, or by taking the relative needs implicit in official income support scales, for example.

- Weighting each family's representation in the sample so that the income distribution is amongst persons rather than arbitrary family units. This is usually done by weighting in proportion to the number of persons in the family (so the above imaginary family of a married couple and two children would be weighted by a factor of four).

There is a variety of alternative assumptions that could be made about each of these two steps, and you should be warned that these adjustments can significantly affect the picture of inequality that emerges (see Question 2 for an example of this).

You may well conclude that big enough problems are raised in dealing with the heterogeneous people who are there in the sample population; but an even bigger problem is posed by those who aren't there. This remark applies generally to tax-based data, and particularly to wealth. Only those estates that are sufficiently extensive to attract the attention of the tax authorities are usually included in the data, and hence there is a large proportion of the population which, although not destitute, does not appear in the published figures. Basically you have to do one of three things: leave these people out altogether (and so underestimate the amount of inequality); include them, but with zero wealth (and so overestimate inequality); or make some estimate of the wealth to be imputed *per capita*, by using information from alternative sources on total wealth, or—more ambitiously—by guessing at the distribution among these excluded persons.

WHAT TIME PERIOD?

Income—as opposed to wealth—is defined relative to a particular time unit, and you will generally find that measured inequality is noticeably lower if the personal income concept relates to a relatively long period than if quite a short time interval, such as a week or a month, is considered. The reason is simply that people's incomes fluctuate, and the longer you make the time unit, the more you 'average out' this volatility. As we noted in Chapter 1, the ultimate extension of this is to examine the distribution of lifetime average income. However, apart from the conceptual difficulties involved, there may be practical problems too. In some cases *longitudinal* datasets are available that track the individual incomes over more than one period: these may be used to derive estimates of the interpersonal distribution of a lifetime average, although fairly sophisticated techniques may be required; in some cases sufficiently detailed data are just not available.

WHAT VALUATION PROCEDURE HAS BEEN USED?

As we have seen, there are substantial problems of incorporating non-monetary items into the income or wealth aggregate, such as income in kind or assets for which no easily recognized market price exists. In addition to these problems, the question of the valuation procedure arises particularly when analysing trends of inequality over time, or in making comparisons between countries. For, when looking at time trends, we must recognize that changes in consumer goods' prices will affect the purchasing power of the poor and of the rich in different ways if the spending patterns of these two groups are significantly different. In some advanced economies during the recent past, price increases happen to have affected necessities disproportionately more than luxuries, and as a consequence looking at inequality purely in money-income terms conceals an increasing trend in inequality of real purchasing power. If we want to compare inequality within different countries, or to examine inequality among countries in *per capita* income, then even worse trouble lies ahead: one must wrestle with diverse definitions of income, differing relative prices (as in the time trend problem), different levels and forms of public expenditure, and artificial exchange rates—which collectively are giants barring the way to comparability in income- or wealth-valuation.

WHICH ECONOMIC ASSUMPTIONS HAVE BEEN MADE?

To procure certain versions of the income or wealth variable some economic sleight-of-hand is essential, and it is important to grasp the legitimate tricks involved. Let us briefly consider two of the most frequently encountered issues.

First, how are we to allow for people's reactions to price and income changes? Taxation generally involves distortion of prices—those of commodities, and the value of time available for work. Now people's choices of the amount they work and the amount they save may be affected by changes in these prices, which means in turn that the income distribution itself is affected. So if you want to infer from the published figures what the shape of the income distribution would be without government intervention, you must allow for this income response, which in practice usually means flatly ignoring it. This remark applies to the effects of indirect taxation as well as to income tax.

The second issue concerns the assumptions about markets. Time and again one has to sum unlike components in an income or wealth aggregate. To get an overall measure of net worth one adds a person's current wealth (in terms of marketable assets) to a present valuation of future income receipts from other sources. To evaluate a family's disposable income after all forms of intervention one must include the value of non-monetary government transfers along with money income. Either exercise involves not only the selection of prices, as we discussed above, but usually a tacit assumption about the existence of efficiently-operating markets for capital and for government-provided goods. To see this, note that a person with high future income but low current wealth can only be said to be as well-off as a person with high current wealth but low income prospects if it is possible to borrow from the capital market on the strength of one's anticipated high earnings. Taking your cue from the Rev. S. Smith, you might think that enough 'horrors' had been met in just examining the data. But we must press on.

5.2 Computation of the inequality measures

Let us assume that you have decided on the variable y that you wish to use, and the source from which you are going to extract the data. As we shall see, there are some potentially significant problems associated with the arithmetic involved in proceeding from a table of raw data to a number giving the realized value of an inequality measure. We proceed by describing a number of inequality measures that were introduced in Chapters 2 and 3 in a formal but economical manner, and then using this presentation to explore the practical difficulties.

Suppose that for a particular population you know the theoretical density function $f(y)$, which gives the proportion of the population that has an income in the infinitesimal interval y to $y + dy$.[1] This function is defined

[1] For those who are uneasy about integration an intuitive description may help. Suppose that you have a diagram of a smooth curve $\phi(y)$, drawn with y measured 'horizontally' and ϕ 'vertically'. Then $\int_a^b \phi(y)dy$ means the area under the curve, above the horizontal axis and bounded on either side by the vertical lines $y = a$ and $y = b$. Thus in Fig. 2.2 $\int_{10\,000}^{12\,500} \phi(y)dy$ means the area between the smooth curve and the line OF that also lies between the points marked 10,000 and 12,500. Instead of working out just the one single shaded rectangle, it is as though we calculated the area of lots of rectangles of tiny base width made to fit under the curve along this small interval. The '$\int$' sign can be taken as something quite similar to the summation sign 'Σ'.

so that if it is summed over the entire income range the result is exactly 1; formally:

$$\int_0^\infty f(y)dy = 1.$$

Now let us suppose that the desired inequality measure, or an ordinally equivalent transformation of the desired inequality measure, can be written in the following way, which we shall refer to as the *basic form*:

$$J = \int_0^\infty h(y)\, f(y)dy,$$

where $h(.)$ is an *evaluation function*—some function of y that we have yet to specify and which may also depend on mean income. It so happens that nearly every inequality measure that is of interest, except the Gini coefficient, can be shown to be ordinally equivalent to something that can be written in the basic form—mathematically inclined readers are invited to check this from Tables A.1 and A.2 in the Technical Appendix. Some can be written exactly in the basic form—for example the relative mean deviation, for which we would have the following evaluation function

$$h(y) = \left| \frac{\underline{y}}{y} - 1 \right|$$

or Theil's inequality measure, for which we find

$$h(y) = \frac{\underline{y}}{y} \log\left(\frac{\underline{y}}{y}\right).$$

Others are related to the basic form by a simple transformation—for example if we specify

$$h(y) = \left[\frac{\underline{y}}{y}\right]^{1-\varepsilon}$$

and then consider the transformation $1 - J^{1/[1-\varepsilon]}$ we find that we have A_ε, Atkinson's inequality index with inequality aversion parameter ε. It is worth re-emphasizing that, as long as we have defined a sensible inequality measure, the exact specification of the evaluation function $h(.)$ does not matter at all, and the basic form is just a neat way of describing a large number of measures.

However, the basic form gives the inequality measure in theoretical terms using a continuous distribution function. One might specify one particular such continuous function (for example, the lognormal or the Pareto) as a rough and ready approximation to the facts about the distribution of income, wealth, etc.; the problems associated with this procedure are taken up later. But in practice we may not wish to use such approximating devices, and we would then want to know what modifications need to be made to the basic form in order to use it directly with actual data.

- $\int_0^\infty h(y)\, f(y)dy$
- **Density function:** $f(y)$
- **Evaluation function:** $h(y)$
- **Lower bound of y-range:** 0
- **Upper bound of y-range:** ∞

THE MEASURE J: BASIC FORM

First of all, let us note that if we are presented with n actual observations $y_1, y_2, y_3, \ldots, y_n$ of all n people's incomes, some of our problems appear to be virtually over. It is appropriate simply to replace the theoretical basic form of J with its discrete equivalent:

$$J = \frac{1}{n}\sum_{i=1}^{n} h(y_i).$$

What this means is that we work out the evaluation function $h(y)$ for Mr Jones and add it to the value of the function for Ms Smith, and add it to that of Mr Singh, . . . and so on.

It is a fairly simple step to proceed to the construction of a Lorenz curve and to calculate the associated Gini coefficient. There are several ways of carrying out the routine computations, but the following is straightforward enough. Arrange all the incomes into the 'Parade' order, and let us write the observations ordered in this fashion as $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$, (so that $y_{(1)}$ is the smallest income, $y_{(2)}$ the next, and so on up to person n). For the Lorenz curve, mark off the horizontal scale (the line OC in Fig. 2.4) into n equal intervals. Plot the first point on the curve just above the endpoint of the first interval at a 'height' of $y_{(1)}/n$; plot the second at the end of the second interval at a height of $[y_{(1)} + y_{(2)}]/n$; the third at the end of the third interval at a height $[y_{(1)} + y_{(2)} + y_{(3)}]/n$; . . . and so on. You can calculate the Gini coefficient from the following easy formula:

$$G = \frac{2}{n^2\overline{y}}\left[y_{(1)} + 2y_{(2)} + 3y_{(3)} + \ldots + ny_{(n)}\right] - \frac{n+1}{n}.$$

This observation-by-observation approach will usually work well for all the methods of depicting and measuring inequality that we considered in Chapters 2 and 3 with just two exceptions: the frequency distribution and the log frequency distribution. To see what the problem is here imagine

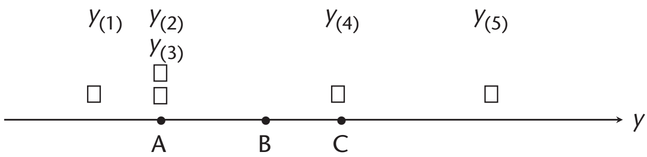

FIG. 5.4. Income observations arranged on a line

setting out the n observations in order along the income line as represented by the little blocks in Fig. 5.4. Obviously we have a count of two incomes exactly at point A ($y_{(2)}$ and $y_{(3)}$) and one exactly at point C (income $y_{[4]}$), but there is a count of zero at any intermediate point such as B. This approach is evidently not very informative: there is a problem of filling in the gaps. In order to get a sensible estimate of the frequency distribution, we could try a count of the numbers of observations that fall within each of a series of small fixed-width intervals, rather than at isolated points on the income line in Fig. 5.4. This is in fact how the published HBAI data are presented—see Fig. 5.5. Of course the picture that emerges will be sensitive to the arbitrary width that is used in this exercise (compare Fig. 5.5 with the deliberately coarse groupings used for the same data in Fig. 5.3); more seriously this method is going to yield a jagged discontinuous frequency distribution that appears to be an unsatisfactory representation of the underlying density

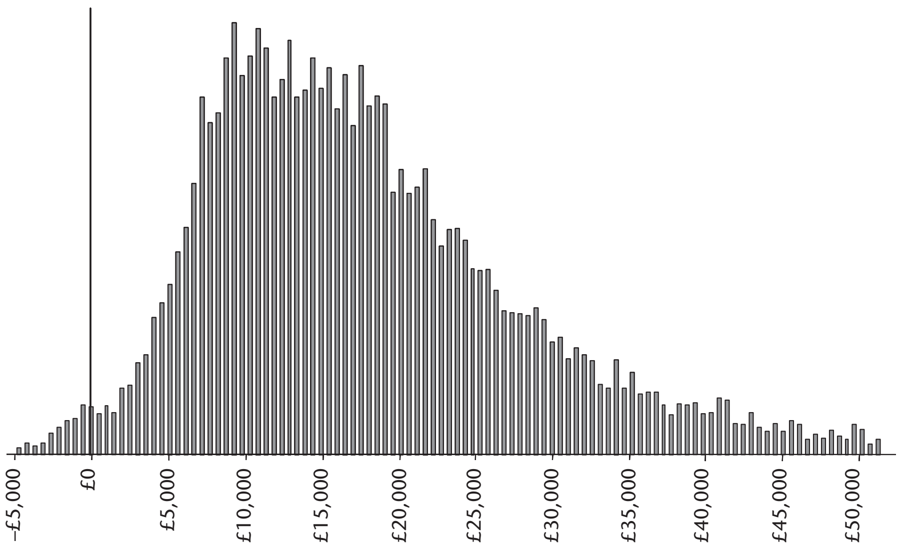

FIG. 5.5. Frequency distribution of disposable income, UK 2006/7, (after housing costs), unsmoothed

Source: as for Fig. 5.3

function. It may be better to estimate the density function by allowing each observation in the sample to have an influence upon the estimated density at neighbouring points on the income line (a strong influence for points that are very close, and a weaker influence for points that are progressively further away); this typically yields a curve that is smoothed to some extent. An illustration of this on the data of Fig. 5.5 is provided in Figs 5.6 and 5.7—the degree of smoothing is governed by the 'bandwidth' parameter (the greater the bandwidth the greater the influence of each observation on estimates of the density at distant points), and the method is discussed in detail on pages 172ff in the Technical Appendix.

Unfortunately, in many interesting fields of study, the procedures that I have outlined so far are not entirely suitable for the lay investigator. One reason for this is that some of the published and accessible data on incomes, wealth, etc. is presented in grouped form, rather than made available as individual records.

However, there is a second reason. Some of the important sets of ungrouped data that are available are not easily manipulated by the layman, even a layman with a state-of-the-art personal computer. The problem derives not from mathematical intractability—the computational techniques would be much as I have just described—but from the vast quantity

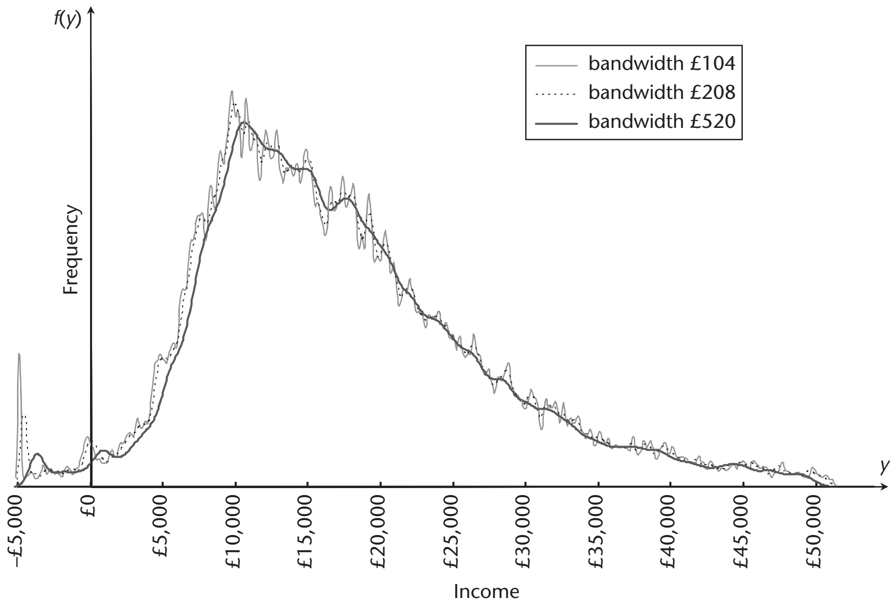

Fig. 5.6. Estimates of distribution function. Disposable income, UK 2006/7, (after housing costs), moderate smoothing

Source: as for Fig. 5.3

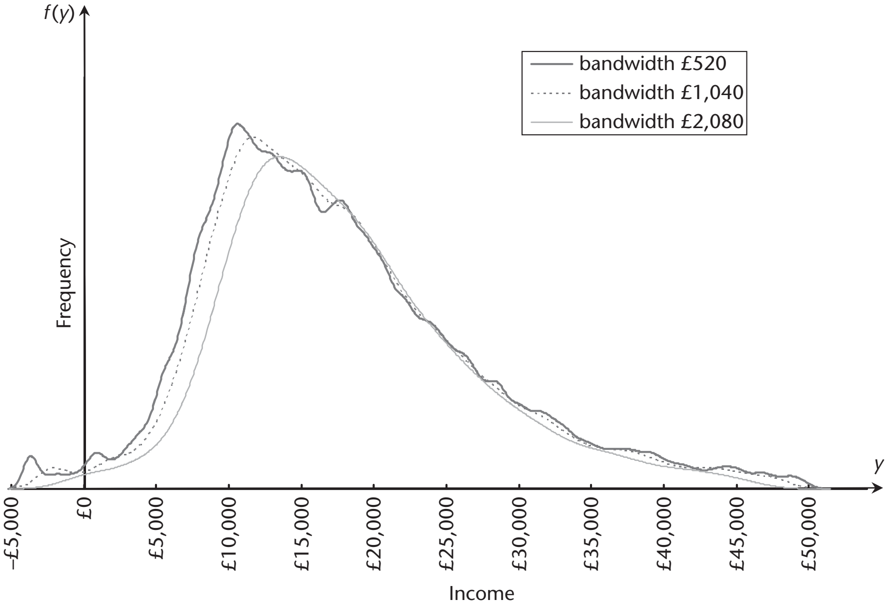

FIG. 5.7. Estimates of distribution function. Disposable income, UK 2006/7, (after housing costs), high smoothing
Source: as for Fig. 5.3

of information typically involved. An 'important' study with ungrouped data usually involves the coverage of a large and heterogeneous population, which means that n may be a number of the order of tens of thousands. Such data-sets are normally obtained from computerized records of tax returns, survey interviews, and the like, and the basic problems of handling and preparing the information require large-scale data-processing techniques. Of course it is usually possible to download extracts from large datasets on to storage media that will make it relatively easy to analyse on a microcomputer: from then on you can apply the formulas given here and in the Technical Appendix using even simple spreadsheet tools (see the website). Nevertheless if you are particularly concerned with easy availability of data, and wish to derive simple reliable pictures of inequality that do not pretend to moon-shot accuracy, you should certainly consider the use of published data, which means working with grouped distributions. Let us look at what is involved.

Were we to examine a typical source of information on income or wealth distributions, we should probably find that the facts are presented in the following way. 'In the year in question, n_1 people had at least \$$a_1$ and less than \$$a_2$; n_2 people had at least \$$a_2$ and less than \$$a_3$; n_3 people had at least \$$a_3$ and less than \$$a_4$,. . . .' In addition we may be told that the average income of

Table 5.1. Distribution of income before tax, USA 2006

Lower boundary of income range	Number in groups ('000)	Group mean income	Relative freq		Cumulative freq	
			pop	inc	pop	inc
(1)	(2)	(3)	(4)	(5)	(6)	(7)
(<$1,000)	2,676	-$34,006		-0.011	0.000	0.000
$1,000	11,633	$2,665	0.086	0.004	0.086	0.004
$5,000	11,787	$7,466	0.087	0.011	0.173	0.015
$10,000	11,712	$12,466	0.086	0.018	0.259	0.033
$15,000	10,938	$17,462	0.081	0.024	0.339	0.056
$20,000	9,912	$22,498	0.073	0.027	0.412	0.084
$25,000	8,750	$27,429	0.064	0.030	0.477	0.113
$30,000	14,152	$34,765	0.104	0.061	0.581	0.174
$40,000	10,687	$44,821	0.079	0.059	0.660	0.233
$50,000	18,855	$61,416	0.139	0.143	0.799	0.375
$75,000	11,140	$86,266	0.082	0.118	0.881	0.494
$100,000	12,088	$132,859	0.089	0.198	0.970	0.692
$200,000	3,121	$286,767	0.023	0.110	0.993	0.802
$500,000	589	$679,117	0.004	0.049	0.997	0.851
$1,000,000	150	$1,213,333	0.001	0.022	0.998	0.873
$1,500,000	64	$1,718,750	0.000	0.014	0.999	0.887
$2,000,000	99	$2,979,798	0.001	0.036	1.000	0.923
$5,000,000	25	$6,840,000	0.000	0.021	1.000	0.944
$10,000,000	16	$28,250,000	0.000	0.056	1.000	1.000
all ranges	*138,394*					
(positive inc.)	*135,718*	*$59,830*				

Source: Internal Revenue Service

people in the first group ($\$a_1$ to $\$a_2$) was reported to be $\$\mu_1$, average income in the second group ($\$a_2$ to $\$a_3$) turned out to be $\$\mu_2$, and so on. Columns 1–3 of Table 5.1 are an example of this kind of presentation. Notice the difference between having the luxury of knowing the individual incomes $y_1, y_2, y_3, \ldots, y_n$ and of having to make do with knowing the numbers of people falling between the arbitrary income-class boundaries $a_1, a_2, a_3, \ldots$ which have been set by the compilers of the official statistics.

Suppose that these compilers of statistics have chopped up the income range into a total of k intervals:

$$(a_1, a_2)\ (a_2, a_3)\ (a_3, a_4) \ldots (a_k, a_{k+1}).$$

If we assume for the moment that $a_1 = 0$ and $a_{k+1} = \infty$, then we have indeed neatly subdivided our entire theoretical range, zero to infinity (these assumptions will not do in practice as we shall soon see). Accordingly, the inequality measure in basic form may be modified to:

$$\int_{a_1}^{a_2} h(y)\, f(y) dy + \int_{a_2}^{a_3} h(y)\, f(y) dy + \ldots + \int_{a_k}^{a_{k+1}} h(y)\, f(y) dy$$

which can be written more simply:

$$\sum_{i=1}^{k}\left[\int_{a_i}^{a_{i+1}} h(y)\, f(y)dy\right].$$

It may be worth repeating that this is *exactly* the same mathematical formula as the 'basic form' given above, the only notational difference being that the income range has been subdivided into k pieces. However, although we have observations on the average income and the number of people in each class (a_i, a_{i+1}), we probably have not the faintest idea what the distribution $F(y)$ looks like within each class. How can we get round this problem?

In the illustrations of income distribution datasets used earlier in the book (for example Fig. 5.1 above) we have already seen one way of representing the distribution within each class, namely that $f(y)$ should be constant within each class. If we used the same assumption of uniformity within each income class for the US income distribution data in Fig. 5.1 we would get a picture like Fig. 5.8. However, in practice this is not a very good assumption. In order to get the height of each bar in the histogram you just divide the number of persons in the income class n_i by the number in the total population n to give the *relative frequency* in class i (columns 2 and 4 in

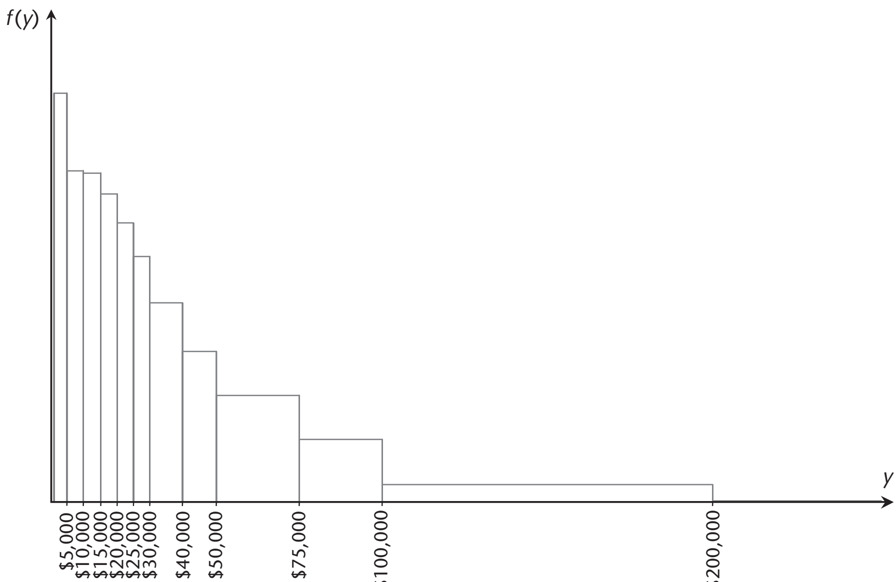

FIG. 5.8. Frequency distribution of income before tax. US 2006
Source: Internal Revenue Service

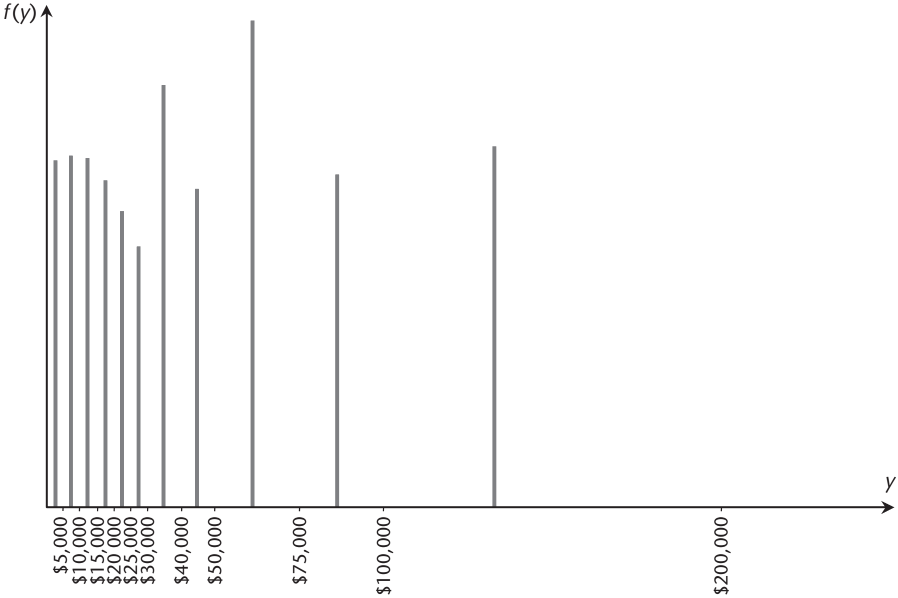

Fig. 5.9. Lower bound inequality, distribution of income before tax. US 2006
Source: Internal Revenue Service

Table 5.1), and then divide the relative frequency n_i/n by the width of the income class $a_{i+1} - a_i$ (column 1). But this procedure does not use any of the information about the mean income in each class μ_i (column 3), and that information is important, as we shall see.

A better—and simple—alternative first step is to calculate from the available information lower and upper limits on the unknown theoretical value J. That is, we compute two numbers J_L and J_U such that it is certain that

$$J_L \leq J \leq J_U$$

even though the true value of J is unknown.

The lower limit J_L is found by assuming that everyone in the first class gets the average income in that class, $\$\mu_1$, and everyone in the second class gets the average income in that class, $\$\mu_2, \ldots$ and so on. So, to compute J_L one imagines that there is no inequality within classes (a_i, a_{i+1}) for every $i = 1, 2, \ldots, k$, as depicted in Fig. 5.9. Given that the population relative frequency in income class i is n_i/n (column 4 in Table 5.1) and the class mean is μ_i (column 3) we then have:

$$J_L = \sum_{i=1}^{k} \frac{n_i}{n} h(\mu_i).$$

Notice that if we are given the average income in each class, $\mu_1, \mu_2, \mu_3, \ldots, \mu_k$, we do not need to know the income-class boundaries $a_1, a_2, a_3, \ldots, a_{k+1}$, in order to calculate J_L.

By contrast, the upper limit J_U is found by assuming that there is maximum inequality within each class, subject to the condition that the assumed average income within the class tallies with the observed number μ_i. So we assume that in class 1 everyone gets either \$$a_1$ or \$$a_2$, but that no one actually receives any intermediate income. If we let a proportion

$$\lambda_1 = \frac{a_2 - \mu_1}{a_2 - a_1}$$

of the class 1 occupants be stuck at the lower limit, \$$a_1$, and a proportion $1 - \lambda_1$ of class 1 occupants receive the upper limit income \$$a_2$, then we obtain the right answer for average income within the class, namely \$$\mu_1$. Repeating this procedure for the other income classes and using the general definition

$$\lambda_i = \frac{a_{i+1} - \mu_i}{a_{i+1} - a_i},$$

we may now write:

$$J_U = \sum_{i=1}^{k} \frac{n_i}{n} \left[\lambda_i h(a_i) + [1 - \lambda_i]\, h(a_{i+1})\right].$$

A similar procedure can be carried out for the Gini coefficient. We have:

$$G_L = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{n_i n_j}{n^2 \bar{y}} \left|\mu_i - \mu_j\right|$$

and

$$G_U = G_L + \sum_{i=1}^{k} \frac{n_i^2}{n^2 \bar{y}} \lambda_i \left[\mu_i - a_i\right].$$

The upper-bound distribution is illustrated in Fig. 5.10.

We now have our two numbers J_L, J_U which will meet our requirements for lower and upper bounds. The strengths of this procedure are that we have not had to make any assumption about the underlying theoretical distribution $f(y)$ and that the calculations required in working out formulas for J_L and J_U in practice are simple enough to be carried out on a pocket calculator: there is an example of this in the 'Inequality calculator' file on the website.

The practical significance of the divergence between J_L and J_U is illustrated for six inequality measures (c, G, T, $A_{0.5}$, A_1, and A_2) in Table 5.2: this has been constructed from the data of Table 5.1, on the basis of a variety of

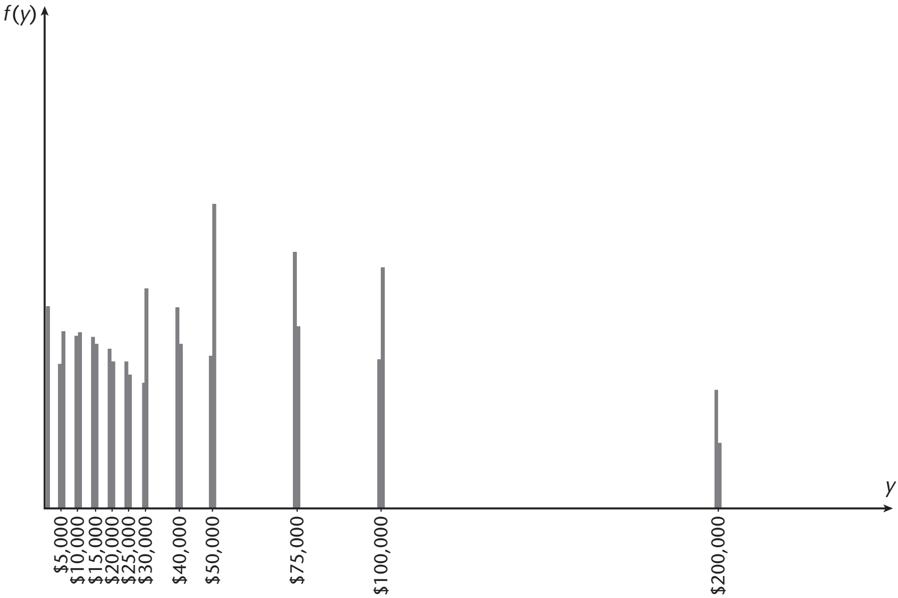

FIG. 5.10. Upper bound inequality, distribution of income before tax. US 2006
Source: Internal Revenue Service

alternative assumptions about the underlying distribution of income. Because of the negative mean in the first interval the computations have been performed only for the distribution of incomes of $1,000 or more. For each inequality measure the columns marked 'Lower Bound' and 'Upper Bound' correspond to the cases J_L and J_U above (see Figs. 5.9 and 5.10 respectively); the 'Compromise' value and the term in parentheses will be discussed a little later. Likewise the rows marked (1), (2), (3) correspond to three alternative assumptions about what happens to the income distribution in the upper and lower tails. Let us take first the simplest—though not necessarily the best—of these: the central case (2) which amounts to assuming that the lowest possible income, a_1, was $1,000 and that the highest possible income a_{k+1}, was $40,000,000. It is obvious from the values of the six inequality measures recorded that the size of the Upper–Lower gap as a proportion of the compromise value varies a great deal from one measure to another. While this gap is just 1.3 per cent for the Gini coefficient, 3.5 per cent for Atkinson ($A_{0.5}$), and 4.7 per cent for Theil, it is as much as 11.3 per cent for the coefficient of variation.[2]

[2] Recall that c is not written exactly in the 'basic form'. However, the Herfindahl index $H = [c^2 + 1]/n$ can be written in this way. The proportionate gap between J_L and J_U for H would be 22.3 per cent.

Table 5.2. Values of inequality indices under a variety of assumptions about the data. US Internal Revenue Service 2006

	Lower	Compromise	Upper		Lower	Compromise	Upper	
		c				$A_{0.5}$		
(1)	5.684	***	***		0.324	0.329	0.336	
(2)	**5.684**	5.915	**6.352**	(0.346)	**0.324**	0.328	**0.336**	(0.334)
(3)	5.448	5.670	6.091	(0.346)	**0.290**	0.294	**0.301**	(0.337)
		G				A_1		
(1)	0.594	0.600	0.602		0.514	0.523	0.537	
(2)	**0.594**	0.600	**0.602**	(0.667)	**0.514**	0.522	**0.537**	(0.324)
(3)	0.563	0.568	0.571	(0.667)	**0.442**	0.447	**0.455**	(0.336)
		T				A_2		
(1)	1.003	1.060	1.086		0.760	0.784	0.828	
(2)	**1.003**	**1.019**	**1.051**	(0.335)	**0.760**	0.784	**0.828**	(0.351)
(3)	0.933	0.949	0.980	(0.335)	**0.626**	0.633	0.647	(0.335)

(1) Top interval is a Pareto tail, bottom interval included
(2) Top interval closed at \$40m, bottom interval included
(3) Top interval closed at \$40m, bottom interval excluded

Of course, the lower- and upper-bound estimates of inequality measures may be sensitive to the assumptions made about the two extreme incomes a_1, (\$1,000), and a_{k+1}, (\$40,000,000). To investigate this let us first look at the lower tail of the distribution. Consider the calculations after all income-receivers below \$3,000 have been eliminated (metaphorically speaking)—see row (3) for each of the measures presented in Table 5.2. As we expect, for all the measures the amount of inequality is less for the distribution now truncated at the lower end. But the really significant point concerns the impact upon the Upper–Lower gap that we noted in the previous paragraph: it is almost negligible for every case except A_2 which, as we know, is sensitive to the lower tail of the income distribution (see page 53). Here the proportionate gap is dramatically cut from 8.6 per cent to 3.3 per cent. This suggests that the practical usefulness of a measure such as this will depend crucially on the way lower incomes are treated in grouped distributions—a point to which we return in the next section when considering SWF-based measures.

Now consider the upper tail. It is no good just putting $a_{k+1} = \infty$, because for several inequality measures this results in J_U taking on the 'complete inequality' value, whatever the rest of the distribution looks like.[3] If the average income in each class is known, the simplest solution is to make a sensible guess as we have done in row (2) for each measure in Table 5.2. To see how important this guess is, suppose that instead of closing off the last interval at an arbitrary upper boundary a_{k+1} we assumed that the distribution

[3] A similar problem can also arise for some inequality measures if you put $a_1 \leq 0$.

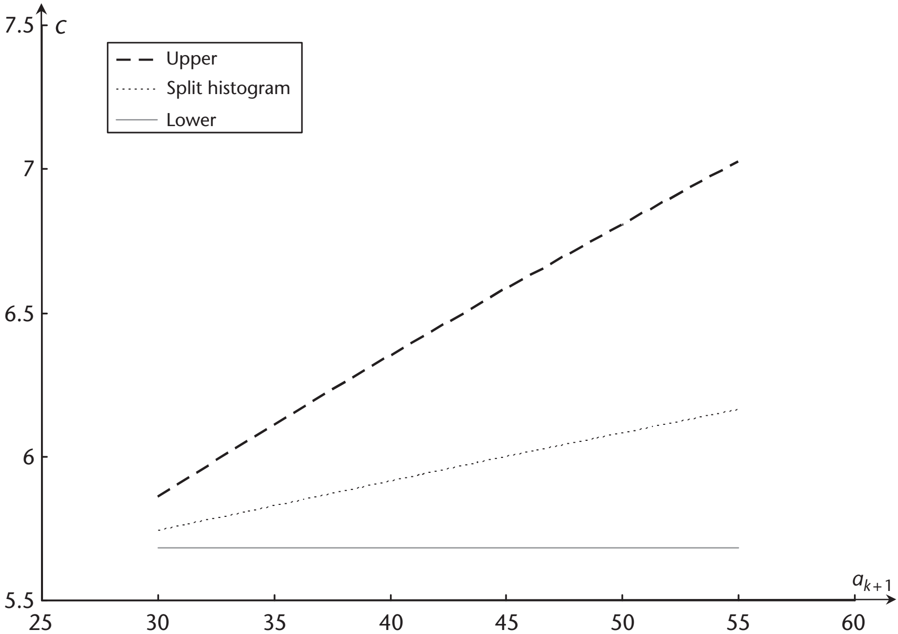

FIG. 5.11. The coefficient of variation and the upper bound of the top interval

in the top interval k was Paretian: this would then yield the results in row (1) of Table 5.2. Comparing rows (1) and (2) we can see that for measures such as A_1 or A_2 there is little discernible effect: this comes as no surprise since we noted (page 53 again) that indices of this sort would be mainly sensitive to information at the bottom end of the distribution rather than the top.[4] By contrast the impact upon T of changing the assumption about the top interval is substantial; and for the coefficient of variation c—which is particularly sensitive to the top end of the distribution—the switch to the Pareto tail is literally devastating: what has happened is that the estimate of α for the fitted Pareto distribution is about 1.55, and because this is less than 2, the coefficient of variation is effectively infinite: hence the asterisks in Table 5.2. All this confirms that estimates of c—and of measures that are ordinally equivalent to c—are sensitive to the precise assumption made about the top interval. To illustrate this further, the results reported in Table 5.2 were reworked for a number of values of a_{k+1}: the only measure whose value changes significantly was the coefficient of variation, for which the results are plotted in Fig. 5.11; the two outer curves represent the lower- and upper-bound assumptions, and the curve in the middle represents a

[4] There would be no effect whatsoever upon the relative mean deviation M: the reason for this is that noted in Fig. 2.6: rearranging the distribution on one side of the mean had no effect on M.

possible compromise assumption, about which we shall say more in just a moment.

Let us now see how to draw a Lorenz curve. From column 5 of Table 5.1 construct column 6 in an obvious way by calculating a series of running totals. Next calculate the percentage of total income accounted for in each interval by multiplying each element of column 5 by the corresponding number in column 4 and dividing by the population mean; calculate the cumulative percentages as before by working out running totals—this gives you column 7. Columns 6 (population shares) and 7 (income shares) form a set of observed points on the Lorenz curve for the US Internal Revenue Service data relating to 2006. These points are plotted in Fig. 5.12. We now have a problem similar to those which used to occur so frequently in my sons' playbooks—join up the dots.

However, this is not as innocuous as it seems, because there are infinitely many curves that may be sketched in, subject to only three restrictions, mentioned below. Each such curve drawn has associated with it an implicit assumption about the way in which income is distributed within the income classes, and hence about the 'true' value of the inequality measure that we wish to use. If the dots are joined by straight lines, then we are assuming that there is no inequality within income classes—in other words, this

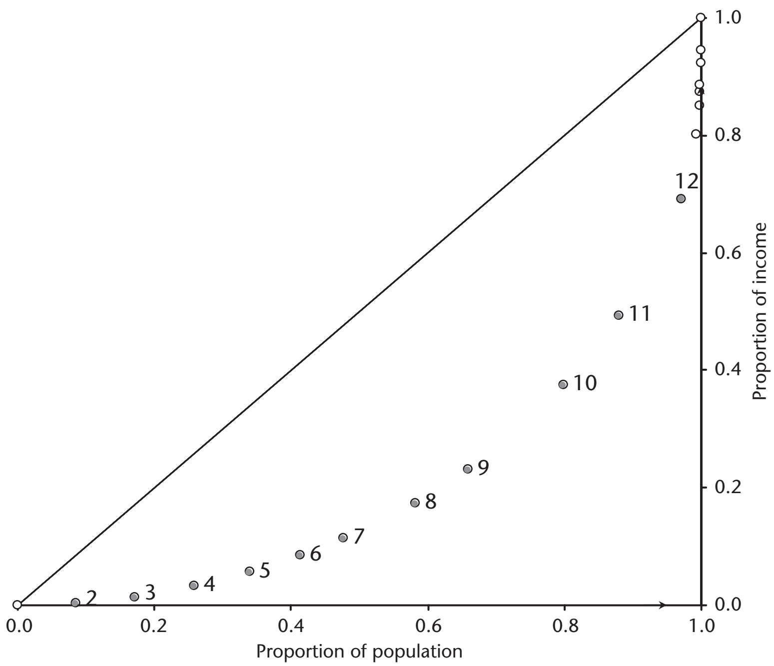

FIG. 5.12. Lorenz co-ordinates for Table 5.1

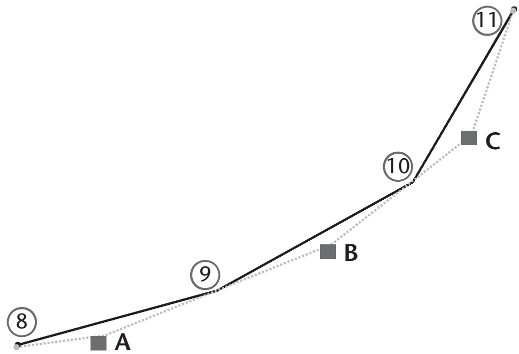

FIG. 5.13. Upper and lower bound Lorenz curves

corresponds to the use of J_L, the lower bound on the calculated inequality measure, (also illustrated by the distribution in Fig. 5.9). This method is shown in detail by the solid lines connecting vertices (8), (9), (10), (11) in Fig. 5.13 which is an enlargement of the central portion of Fig. 5.12. By contrast you can construct a maximum inequality Lorenz curve by drawing a line of slope $a_i/\bar{y}$ through the ith dot, repeating this for every dot, and then using the resulting 'envelope' of these lines. This procedure is illustrated by the dashed line connecting points A, B, C in Fig. 5.13 (in turn this corresponds to J_U and Fig. 5.10). Now we can state the three rules that any joining-up-the-dot procedure must satisfy:

- Any curve must go through all the dots, including the two vertices (0,0) and (1,1) in Fig. 5.12.
- It must be convex.
- It must not pass below the maximum inequality curve.

Notice that the first two of these rules ensure that the curve does not pass above the minimum-inequality Lorenz curve.

One of the reasons for being particularly interested in fitting a curve satisfying these requirements is that the observed points on the Lorenz curve in Table 5.1 (columns 6 and 7) only give us the income shares of the bottom 8.6 per cent, the bottom 17.2 per cent, ... and so on, whereas we would be more interested in the shares of, say, the bottom 10 per cent, the bottom 20 per cent, and to get these we must interpolate on a curve between the points. Presumably the interpolation should be done using neither the extreme upper- or lower-bound assumptions, but rather according to some 'compromise' Lorenz curve. One suggestion for this compromise method is to use the basic Pareto interpolation formula (A.3) (given on page 157 in the Technical Appendix), which is much less fearsome than it looks, because you do not have to compute the parameters α, along the way. All you need

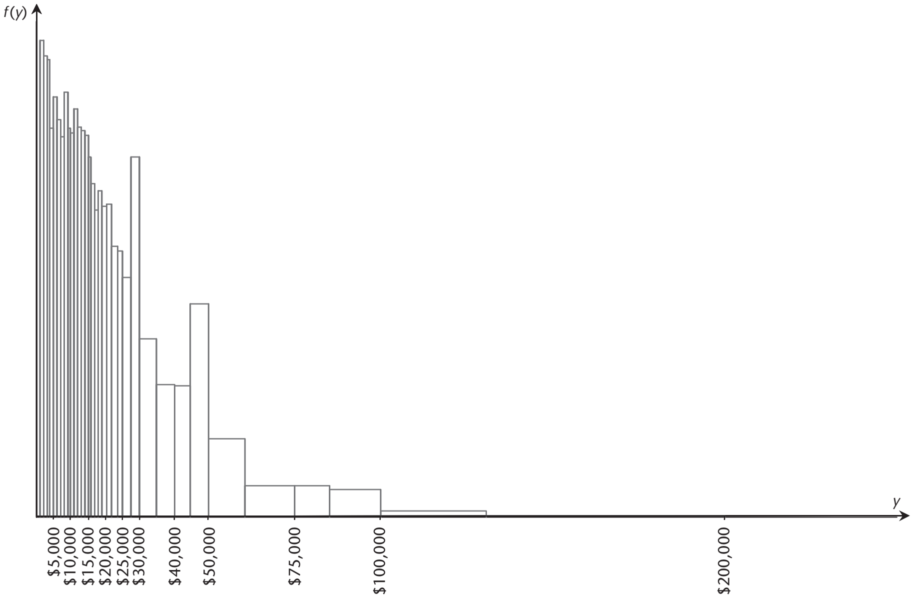

FIG. 5.14. The 'split histogram' compromise

are the population and income shares. Unfortunately this simplicity is also its weakness. Because the formula does not use information about the μ_is the resulting curve may violate the third condition cited above (the same problem would arise if we used a Lorenz curve based on the simple histogram density function illustrated in Fig. 5.14).

An alternative method—which may be implemented so that all three conditions are satisfied—is to fit a theoretical frequency distribution within each interval in Fig. 5.14, and work out the Lorenz curve from that. What frequency distribution? In fact it does not matter very much what type is used: all the standard 'compromise' interpolation methods[5] produce inequality estimates that are remarkably similar. These methods (which are more easily explained using the associated density function) include:

- a 'split histogram' density function in each interval. This is illustrated in Fig. 5.14: contrasting this with Fig. 5.8 you will note that in each interval there are two horizontal 'steps' rather than a single step in the case of the regular histogram; this simple device enables one to use all the information about the interval and is the procedure that was used for the 'compromise' column in Table 5.2;

[5] A minimal requirement is that the underlying density function be well-defined and piecewise continuous (Cowell and Mehta 1982).

- a separate straight line density function fitted to each interval;[6]
- loglinear interpolation in each interval. This is in effect a separate Pareto distribution fitted to each interval (a_i, a_{i+1}), using all the available information;
- a quadratic interpolation in each interval.

The details of all of these—and of how to derive the associated Lorenz curve for each one—are given in the Technical Appendix.

It is reasonably straightforward to use any of these methods to compute a compromise value for an inequality measure. But if you do not need moonshot accuracy, then there is another delightfully simple method of deriving a compromise inequality estimate. The clue to this is in fact illustrated by the columns in parentheses in Table 5.2: this column gives, for each inequality measure, the relative position of the compromise estimate in the interval (J_L, J_U) (if the compromise estimate were exactly halfway between the lower and the upper bound, for example, then this entry would be 0.500). For most inequality measures that can be written in the standard form, a good compromise estimate can be found by taking $\frac{2}{3}$ of the lower bound and adding it to $\frac{1}{3}$ of the upper bound (see for example the results on the Atkinson and Theil indices). One notable exception is the Gini coefficient: for this measure, the compromise can be approximated by $\frac{1}{3}G_L + \frac{2}{3}G_U$ which works extremely well for most distributions, and may also be verified from Table 5.2. Given that it requires nothing more than simple arithmetic to derive the lower and upper bound distributions from a set of grouped data, this $\frac{1}{3} - \frac{2}{3}$ rule (or $\frac{2}{3} - \frac{1}{3}$ rule) evidently provides us with a very handy tool for getting good estimates from grouped data.

5.3 Appraising the calculations

We have now seen how to calculate the indices themselves, or bounds on these indices from the raw data. Taking these calculations at face value, let us see how much significance should be attached to the numbers that emerge.

The problem may be introduced by way of an example. Suppose that you have comparable distribution data for two years, 1985, 1990, and you want to know what has happened to inequality between the two points in time. You compute some inequality indices for each dataset, let us say the coefficient of variation, the relative mean deviation, Theil's index, and the Gini coefficient, so that two sets of numbers result: $\{c_{1985}, M_{1985}, T_{1985}, G_{1985}\}$ and $\{c_{1990}, M_{1990}, T_{1990}, G_{1990}\}$, each set giving a picture of inequality in the

[6] A straight line density function implies that the corresponding Lorenz curve is a quadratic.

appropriate year. You now have another play-book puzzle—spot the difference between the two pictures. This is, of course, a serious problem; we may notice, say, that c_{1990} is 'a bit' lower than c_{1985}—but is it noticeably lower, or are the two numbers 'about the same'? Readers trained in statistical theory will have detected in this a long and imprecise way round to introducing tests of significance.

However, this thought experiment reveals that the problem at issue is a bit broader than just banging out some standard statistical significance tests. Given that we are looking at the difference between the observed value of an inequality measure and some base value (such as an earlier year's inequality) there are at least three ways in which the word 'significance' can be interpreted, as applied to this difference:

- statistical significance in the light of variability due to the sampling procedure;
- statistical significance in view of the arbitrary grouping of observations;
- social or political significance.

The last of these three properly belongs to the final section of this chapter. As far as the first two items are concerned, since space is not available for a proper discussion of statistical significance, I may perhaps be forgiven for mentioning only some rough guidelines—further reference may be made to the Technical Appendix and the notes to this chapter (page 193).

Let us suppose that we are dealing with sampling variability in an ungrouped distribution (unfortunately, rigorous analysis with grouped data is more difficult). The numbers $y_1, y_2, y_3, \ldots, y_n$ are regarded as a sample of independent random observations. We perform the calculations described earlier and arrive at a number J. An essential piece of equipment for appraising this result is the standard error[7] of J which, given various assumptions about the underlying distribution of y and the manner of drawing the sample, can be calculated from the observations $y_1, \ldots, y_n$. Since the ys are assumed to be random, the number J must also be taken to be an observation on a random variable. Given the theoretical distribution of the ys it is possible to derive in principle the distribution of the values of the computed number J. The standard deviation or square-root-of-variance of this derived distribution is known as the standard error of J. Given this standard error an

[7] A couple of technical words of warning should be noted. Firstly, in an application we ought to examine carefully the character of the sample. If it is very large by comparison with the whole finite population, the formulas in the text must be modified; this is in fact the case in my worked example—although the qualitative conclusions remain valid. If it is non-random, the formulas may be misleading. Secondly for some of the exercises carried out we should really use standard error formulas for differences in the Js; but this is a complication which would not affect the character of our results.

answer can be provided to the kind of question raised earlier in this section: if the difference $c_{1990} - c_{1985}$ is at least three times the standard error for c, then it is 'quite likely' that the change in inequality is not due to sampling variability alone and that thus this drop is significant.

Some rule-of-thumb formulas for the standard errors are readily obtainable if the sample size, n, is assumed to be large, and if you are prepared to make some pretty heroic assumptions about the underlying distribution from which you are sampling. Some of these are given in Table 5.3, but I should emphasize that they are rough approximations intended for those who want to get an intuitive feel for the significance of numbers that may have been worked out by hand.

I would like to encourage even those who do not like formulas to notice from the above expressions that in each case the standard error will become very small for a large sample size n. Hence for a sample as large as that in Table 5.1, the sampling variability is likely to be quite small in comparison with the range of possible values of the inequality measure on account of the grouping of the distribution. A quick illustration will perhaps suffice. Suppose for the moment that the compromise value of $c = 5.915$ given in Table 5.2 were the actual value computed from ungrouped data. What would the standard error be? Noting that the sample size is about 136 million, the standard error is about

$$5.915 \times \sqrt{\frac{1 + 2 \times 5.915^2}{136 \times 10^6}} = 4.273 \times 10^{-3}.$$

We can be virtually certain that sampling variability introduces an error of no more than three times this on the ungrouped value of c. Contrast this with the gap between the upper bound and lower bound estimates found from Table 5.2 as 6.352 – 5.684 = 0.668. Hence for this kind of distribution,

Table 5.3. Approximation formulas for standard errors of inequality measures

Inequality measure	Standard error approximation	Assumed underlying distribution*
coefficient of variation c	$c\sqrt{\frac{1+2c^2}{n}}$	normal
relative mean deviation M	$\sqrt{\frac{c^2-M^2}{n}}$	normal
Gini coefficient G	$G\sqrt{\frac{0.8086}{n}}$	symmetrical
variance of logarithms v_1	$v_1\sqrt{\frac{2}{n}}$	lognormal

* See Kendall *et al.* (1994), sec 10.5.

Table 5.4. Atkinson index and coefficient of variation: Internal Revenue Service 1987 to 2006

	c			A_2		
	Lower Bnd	Compromise	Upper Bnd	Lower Bnd	Compromise	Upper Bnd
1987	2.046	2.222	2.538	0.646	0.659	0.687
1995	2.281	2.481	2.838	0.690	0.702	0.725
2001	2.612	2.821	3.199	0.717	0.726	0.746
2006	5.684	5.915	6.352	0.760	0.784	0.828

the grouping error may be of the order of five hundred times as large as the sampling variability.

As we have noted, the grouping variability may be relatively large in comparison to the value of the measure itself. This poses an important question. Can the grouping variability be so large as to make certain inequality measures useless? The answer appears to be a qualified 'yes' in some cases. To see this, let us look at two inequality measures.

First, take the coefficient of variation that we have just been discussing. From Table 5.2 we know that the value of c in 2006 lies in the range (5.684, 6.352)—see the bottom row of Table 5.4. The corresponding values of c for some earlier years are also shown on the left-hand side of Table 5.4. It is immediately clear that, even though the (c_L, c_U) gap is large in every year, inequality in 2006 was unquestionably higher than in any of the other three years shown.[8] Figure 5.15 illustrates why this is so: it is clear to the naked eye that the Lorenz curve for 1987 dominates that for 2006; even if we drew in the upper-bound and lower-bound Lorenz curves for each year this conclusion will not go away. By contrast we cannot immediately say that, say, c was higher in 1995 than 1987: the lower-bound, compromise, and upper-bound estimates of c all grew by about 11.5 per cent over the period but, in either year, the (c_L, c_U) gap (as a proportion of the compromise value) is over 22 per cent.

Next consider Atkinson's measure A_ε for the same data. The lower and upper bounds and compromise value are represented pictorially in Fig. 5.16. We can see that the upper–lower gap increases with ε but that it stays relatively modest in size. So it is unsurprising to see that the Atkinson index (with $\varepsilon = 2$), with one minor exception, provides an unambiguous comparison between any pair of the years given in Table 5.4. So it is still

[8] The lower-bound and compromise values of c more than double from 2001 to 2006. Here I am making the assumption that the top interval is closed as in case (2) of Table 5.2. Had we assumed that the top interval had a Pareto tail (case 1) then we would have found that c was unbounded in each of the four years—see the 'Inequality calculator' on the website.

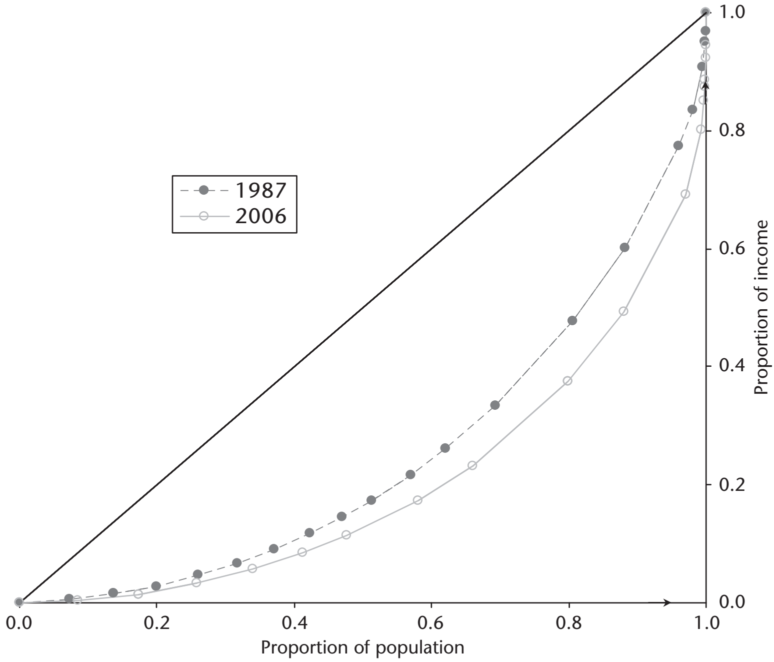

FIG. 5.15. Lorenz curves—Income before tax. USA 1987 and 2006
Source: Internal Revenue Service

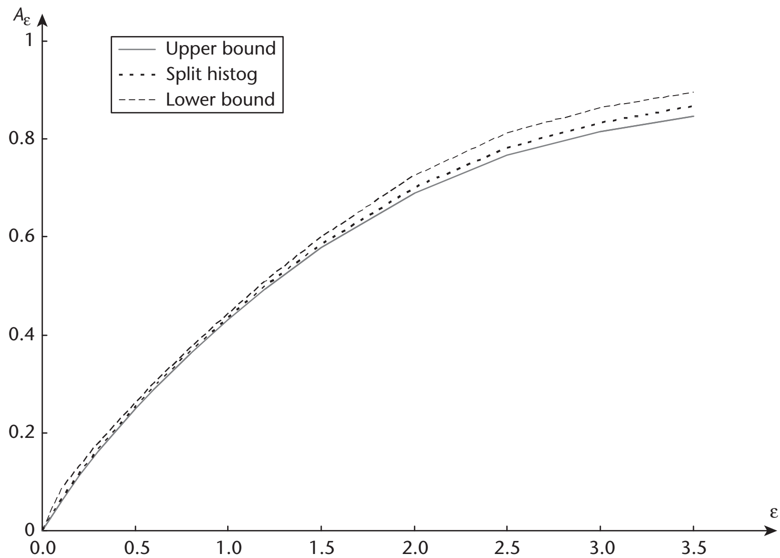

FIG. 5.16. The Atkinson index for grouped data, US 2006
Source: as for Table 5.1

Table 5.5. Individual distribution of household net *per capita* annual income. Czechoslovakia 1988

Income range (crowns)	Number of persons	Mean
1–9 600	176,693	8,421
9 601–10 800	237,593	10,290
10 801–12 000	472,988	11,545
12 001–13 200	640,711	12,638
13 201–14 400	800,156	13,845
14 401–15 600	1,003,174	15,036
15 601–16 800	1,160,966	16,277
16 801–18 000	1,257,160	17,420
18 001–19 200	1,277,633	18,610
19 201–20 400	1,104,486	19,814
20 401–21 600	974,158	21,008
21 601–22 800	871,624	22,203
22 801–24 000	738,219	23,406
24 001–25 200	665,495	24,603
25 201–26 400	579,495	25,810
26 401–27 600	490,502	26,998
27 601–28 800	434,652	28,217
28 801–30 000	367,593	29,419
30 001–31 200	315,519	30,616
31 201–32 400	280,371	31,804
32 401–32 400	245,630	32,976
33 601–34 800	206,728	34,176
34 801–36 000	163,851	35,418
36 001–38 400	257,475	37,154
38 401& over	605,074	48,338
All ranges	*15,327,946*	*21,735*

Source: see Appendix B

true to say that the IRS income distribution of 2006 is more unequal than that of 1987, just as we found for c and just as we saw in Fig. 5.15.

However, Fig. 5.16 in some respects under-represents the problem: the principal reason for this is that in analysing the inequality represented by the data in Table 5.1 we had to drop the first interval which contained a negative mean, so that only incomes over $1,000 were left in the data. Consider instead the Czechoslovakian data presented in Table 5.5.[9] Notice that the first interval is quite wide and has a lower limit of 1 crown per year. If we plot the Atkinson index for these data and drop the first interval (as we did for the American data) it appears that inequality is quite low—see the picture in Fig. 5.17—and this picture is in fact borne out by other inequality measures as well as A_ε. But if we attempt to take account of all the data—including the first interval—then the picture of Fig. 5.18 emerges. Notice that not only is the upper-bound estimate of inequality seriously affected for $\varepsilon > 1$ (which

[9] Taken from Atkinson and Micklewright (1992) Table CSI1.

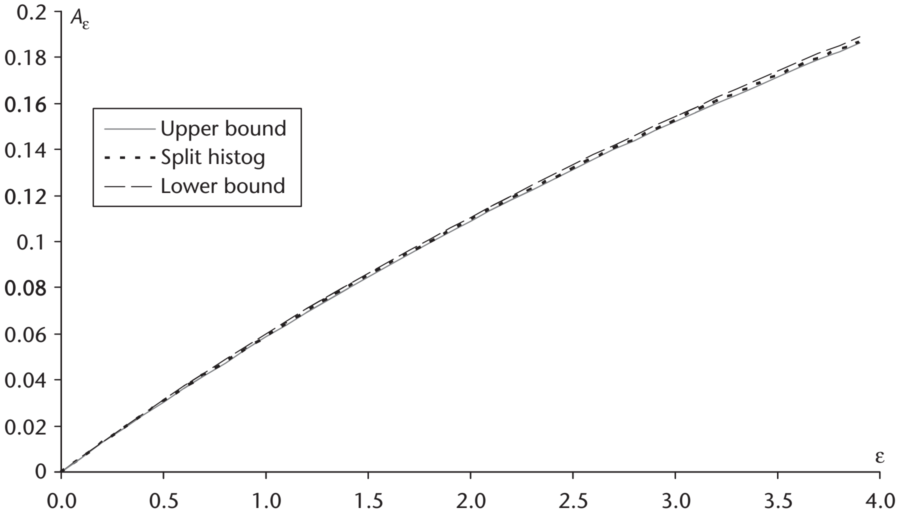

FIG. 5.17. The Atkinson index for grouped data: First interval deleted. Czechoslovakia 1988

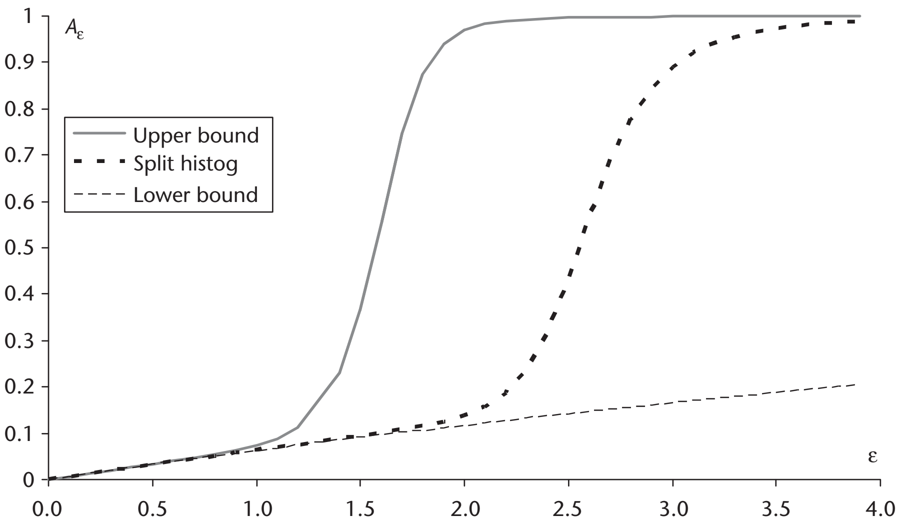

FIG. 5.18. The Atkinson index for grouped data: All data included. Czechoslovakia 1988

we might have guessed) but so too is the compromise value. Obviously truncating the data (or manipulating in some other way the assumption about a_1 which is causing all the trouble) is convenient, but in one sense this is to avoid the problem, since we are deliberately ignoring incomes in the range where our inequality measure is designed to be particularly sensitive. The unpalatable conclusion is that because of grouping error (and perhaps sampling error too) either we shall have to discard certain sensitive measures of inequality from our toolkit on empirical grounds, or the distribution must provide extremely detailed information about low incomes so that measures with high inequality aversion can be used, or the income distribution figures will have to be truncated or doctored at the lower end in a way which may reduce their relevance in the particular area of social enquiry.

5.4 Shortcuts: fitting functional forms[10]

And now for something completely different. Instead of attempting to work out inequality statistics from empirical distribution data directly, it may be expedient to fit a functional form to the raw data, and thus compute the inequality statistics by indirect means. The two steps involved are as follows.

- Given the family of distributions represented by a certain functional form, estimate the parameter values which characterize the particular family member appropriate to the data.
- Given the formula for a particular inequality measure in terms of the family parameters (see the Technical Appendix), calculate the inequality statistics from the parameter estimates obtained in step 1.

For the Pareto distribution, the first step involves estimation of the parameter α from the data, and the second step might be to write down the value of the Gini coefficient, which for the Pareto is simply

$$G = \frac{1}{2\alpha - 1}$$

(see page 156).

For the lognormal distribution, the first step involves estimation of σ^2. Since the second step is simple once you have the formula (it usually involves merely an ordinally equivalent transformation of one of the parameters), I shall only consider in detail methods relating to the first step—the estimation of the parameters.

[10] This section contains material of a more technical nature which can be omitted without loss of continuity.

Two words of warning. Up to now we have used symbols such as $\bar{y}$, V, etc. to denote the theoretical mean, variance, etc., of some distribution. From now on, these symbols will represent the computed mean, variance, etc., of the set of observations that we have under consideration. Although this is a little sloppy, it avoids introducing more symbols. Also, note that often there is more than one satisfactory method of estimating a parameter value, yielding different results. Under such circumstances it is up to the user to decide on the relative merits of the alternative methods.

Let us move straightaway on to the estimation of the parameters of the lognormal distribution for ungrouped and for grouped data.

If the data are in ungrouped form—that is we have n observations, $y_1, y_2, \ldots, y_n$—then on the assumption that these come from a population that is lognormal, it is easy to use the so-called method of moments to calculate estimates $\tilde{\mu}$, $\tilde{\sigma}^2$ for the lognormal distribution. Calculate the mean, and the Herfindahl index (the sum of the squares of the shares—see page 59) for these n incomes:

$$H = \sum_{i=1}^{n} \left[\frac{y_i}{n\bar{y}} \right]^2.$$

Then we find:

$$\tilde{\sigma}^2 = \log(nH)$$

$$\tilde{\mu} = \log(\bar{y}) - \frac{1}{2}\tilde{\sigma}^2.$$

While this is very easy, it is not as efficient[11] as the following method.

An alternative procedure that is fairly straightforward for ungrouped data is to derive the maximum likelihood estimates, $\hat{\mu}$, $\hat{\sigma}^2$. To do this, transform all the observations $y_1, y_2, \ldots, y_n$ to their logarithms $x_1, x_2, \ldots, x_n$. Then calculate:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} [x_i - \hat{\mu}]^2.$$

It is evident that $\hat{\mu}$ is simply $\log(y^*)$—the logarithm of the geometric mean, and that $\hat{\sigma}^2$ is v_1, the variance of the logarithms defined relative to y^*.

In the case of grouped data, maximum likelihood methods are available, but they are too involved to set out here. However, the method of moments

[11] The standard errors of the estimates will be larger than those for the maximum likelihood procedure (which is the most efficient in this case).

can be applied similarly to the way it was in the ungrouped case, provided that in the computation of H an appropriate correction is made to allow for the grouping of observations.

We shall go straight on now to consider the estimation of the parameters of the Pareto distribution, once again dealing first with ungrouped data.

For the method of moments, once again arrange the n observations $y_1, y_2, \ldots, y_n$ in Parade order $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$, (as on page 114). It can be shown that the expected value of the lowest observation $y_{(1)}$, given the assumption that the sample has been drawn at random from a Pareto distribution with parameters α, is $\alpha n\underline{y}/[\alpha n - 1]$. Work out the observed mean income $\overline{y}$. We already know (from page 93) the expected value of this, given the Pareto assumption: it is $\alpha\underline{y}/[\alpha - 1]$. We now simply equate the sample observations ($y_{(1)}$ and $\overline{y}$) to their expected values:

$$y_{(1)} = \frac{\alpha n\underline{y}}{\alpha n - 1}$$

$$\overline{y} = \frac{\alpha\underline{y}}{\alpha - 1}.$$

Solving these two simple equations in two unknowns α, $\underline{y}$, we find the method of moments estimates for the two parameters:

$$\tilde{\alpha} = \frac{\overline{y} - \frac{y_{(1)}}{n}}{\overline{y} - y_{(1)}}$$

$$\tilde{\underline{y}} = \left[1 - \frac{1}{\tilde{\alpha}}\right]\overline{y}.$$

However, this procedure is not suitable for grouped data. By contrast, the ordinary least squares method for estimating α can be applied whether the data are grouped or not. Recall the point in Chapter 4 that if y is any income level, and P is the proportion of the population with that income or more, then under the Pareto distribution, a linear relationship exists between $\log(P)$ and $\log(y)$, the slope of the line being α. Indeed we may write this as

$$p = z - \alpha x$$

where p represents $\log(P)$, x represents $\log(y)$, and z gives the intercept of the straight line.

Given a set of ungrouped observations $y_1, y_2, \ldots, y_n$ arranged say in ascending size order, it is easy to set up the estimating equation for α. For the first observation, since the entire sample has that income or more ($P = 1$), the relevant value of p is

$$p_1 = \log(1) = 0.$$

For the second observation, we have

$$p_2 = \log\left(1 - \frac{1}{n}\right)$$

and for the third

$$p_3 = \log\left(1 - \frac{2}{n}\right)$$

and for the very last we have

$$p_n = \log\left(1 - \frac{n-1}{n}\right) = \log\left(\frac{1}{n}\right)$$

which gives a complete set of transformed values of the dependent variable.[12] Given the values of the independent variable $x_1, x_2, \ldots, x_n$ (calculated from the y-values) we may then write down the following set of regression equations:

$$\begin{aligned} p_1 &= z - \alpha x_1 + e_1 \\ p_2 &= z - \alpha x_2 + e_2 \\ \ldots &= \ldots \quad \ldots \quad \ldots \\ p_n &= z - \alpha x_n + e_n \end{aligned}$$

where $e_1, e_2, \ldots, e_n$ are error terms. One then proceeds to obtain least squares estimates of α and z in the usual way by minimizing the sum of the squares of the es.

Of course you are at liberty to fit a lognormal, Pareto or some other function to any set of data you like, but this is only a useful occupation if a 'reasonable' fit is obtained. What constitutes a 'reasonable' fit?

An answer that immediately comes to mind if you have used a regression technique is to use the correlation coefficient R^2. However, taking a high value of R^2 as a criterion of a satisfactory fit can be misleading when fitting a curve to a highly skewed distribution, since a close fit in the tail may mask substantial departures elsewhere. This caution applies also to line-of-eye judgements of suitability, especially where a log-transformation

[12] In the case of grouped data, let f_1 be the observed proportion of the population lying in the ith income interval, and take x_1 to be $\log(\alpha_1)$, that is the logarithm of the lower bound of the interval, for every interval $i = 1, 2, 3, \ldots k$. The p_is are then found by cumulating the f_is upwards from interval i and taking logarithms, thus:

$$\begin{aligned} p_1 &= \log(1) = 0 \\ p_2 &= \log(f_2 + f_3 + f_4 + \ldots + f_{k-1} + f_k) \\ p_2 &= \log(f_3 + f_4 + \ldots + f_{k-1} + f_k) \\ p_3 &= \log(f_4 + \ldots + f_{k-1} + f_k) \\ \ldots &= \ldots \\ p_{k-1} &= \log(f_{k-1} + f_k) \\ p_k &= \log(f_k). \end{aligned}$$

has been used, as in the construction of Fig. 4.11. For small samples, standard 'goodness-of-fit' tests such as the χ^2-criterion may be used, although for a large sample size you may find that such tests reject the suitability of your fitted distribution even though on other grounds it may be a perfectly reasonable approximation.

An easy alternative method of discovering whether a particular formula is 'satisfactory' can be found using an inequality measure. Let us look at how it is done with grouped data and the Gini coefficient—the argument is easily extended to other inequality measures and their particular concept of 'distance' between income shares. Work out G_{L} and G_{U}, the lower and upper limits on the 'true' value of the Gini. Given the fitted functional form, the Pareto let us say, we can calculate G_{Π}, the value of the Gini coefficient on the supposition that the data actually follow the Pareto law. If

$$G_{\mathrm{L}} \leq G_{\Pi} \leq G_{\mathrm{U}}$$

then it is reasonable to accept the Pareto functional form as a close approximation. What we are saying is that according to the concept of 'distance between incomes' implied by this inequality measure, it is impossible to distinguish the theoretical curve from the 'true' distribution underlying the observations. Of course, a different concept of distance may well produce a contradictory answer, but we have the advantage of specifying in advance the inequality measure that we find appropriate, and then testing accordingly. In my opinion this method does not provide a definitive test; but if the upper-and-lower-limit criterion is persistently violated for a number of inequality measures, there seems to be good reason for doubting the closeness of fit of the proposed functional form.

Let us apply this to the Internal Revenue Service (IRS) data of Table 5.1 and examine the Pareto law. Since we expect only higher incomes to follow this law, we shall truncate incomes below \$25,000. First of all we work out from column 6 of Table 5.1 the numbers p_i as (the transformed values of the dependent variable) by the methods just discussed, and also the logarithms of the lower bounds a_i given in column 1 of Table 5.1, in order to set up the regression equations. Using ordinary least squares on these last 13 intervals we find our estimate of α as 1.496 with a standard error of 0.0072, and $R^2 = 0.996$. Figure 5.19 is the Pareto diagram for this problem; the solid line represents the regression for the top 13 intervals and the broken line represents the regression obtained using all the data. Using the formula for the Gini coefficient on the hypothesis of the Pareto distribution (see page 135 above) we find

$$G_{\Pi} = \frac{1}{2 \times 1.496 - 1} = 0.502.$$

Now, noting that the lower and upper bounds on the Gini for incomes over \$25,000 are $G_L = 0.472$ and $G_U = 0.486$ respectively, it is clear that G_Π lies outside these limits; further experimentation reveals that the same conclusion applies if we choose any point of truncation other than \$25,000. So, according to the Gini criterion, the Pareto distribution is in fact a poor representation of the upper tail in this case, even though it looks as though it 'should be' a good fit in Fig. 5.19. But for other years, the hypothesized Pareto tail looks quite good. Consider the situation in 1987, depicted in Fig. 5.20. Following the same procedure as before we truncate the data to use only the top 13 intervals (incomes above \$15,000 at 1987 prices, which works out at incomes above \$26,622 in 2006 prices). Now we find the estimated α to be 1.879 (s.e. = 0.0092, $R^2 = 0.992$) so that

$$G_\Pi = \frac{1}{2 \times 1.879 - 1} = 0.3625.$$

In this case $G_L = 0.3554$ and $G_U = 0.3628$ and G_Π lies within these bounds. So the Pareto distribution certainly seems to be an acceptable fit for the top 13 income classes in 1987. In passing, it is interesting to see how dramatically inequality increased over the period 1987 to 2006—a point which we had already noted from Fig. 4.12.

Two points should be noted from this exercise. First, just relying on judgement by eye may be unsatisfactory—the Pareto tail yielded a misleading estimate of the Gini coefficient in 2006. Second, had we relied on the R^2 criterion alone, however, we would also have been seriously led astray. If we

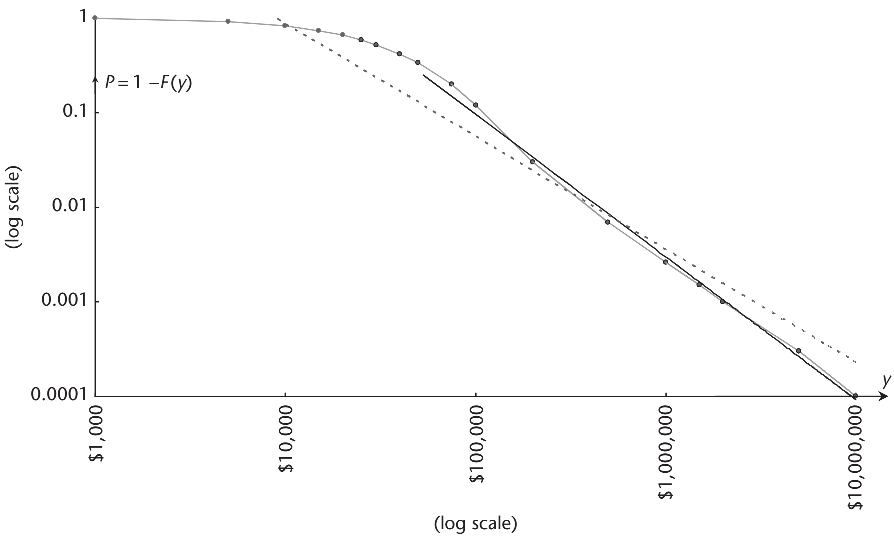

FIG. 5.19. Fitting the Pareto diagram for the data in Table 5.1

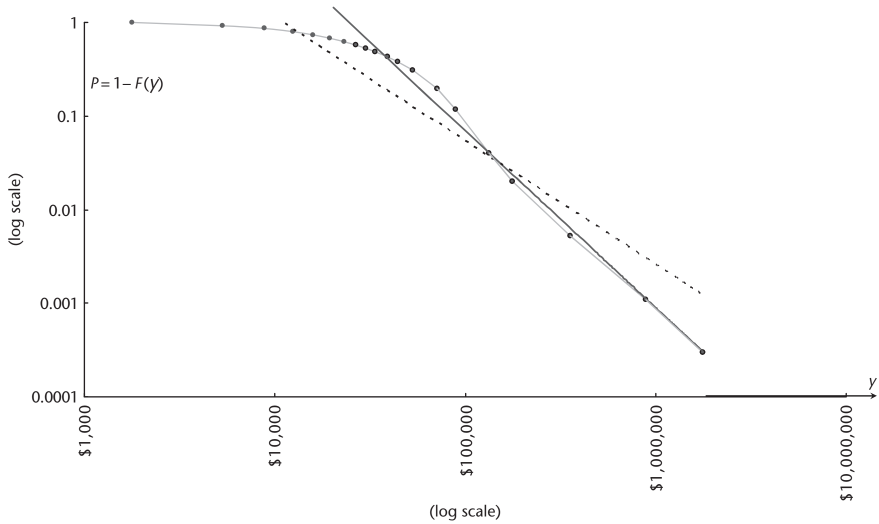

FIG. 5.20. Fitting the Pareto diagram for IRS data in 1987 (values in 2006 dollars)

reworked the 2006 calculations for all incomes above $1,000 we would still have a high R^2 (0.918) but a much lower value of α (1.185); the implied value of $G_{\Pi} = 0.616$ lies well above the upper bound $G_{\mathrm{U}} = 0.602$ recorded for this group of the population in Table 5.2, thus indicating that the Pareto distribution is in practice a bad fit for all incomes above $1,000. It is easy to see what is going on in the Pareto diagram, Fig. 5.19: as we noted the solid regression line depicts the fitted Pareto distribution for $15,000 on which we based our original calculations; if we were to fit a straight line to all the data (the broken line), we would still get an impressive R^2 because of the predominance of the points at the right-hand end, but it is obvious that the straight line assumption would now be rather a poor one. (This is in fact characteristic of income distribution data: Compare the results for IRS 1987 in Fig. 5.20 and for the UK data in Fig. 4.5.)

It seems that we have discovered three main hazards in the terrain covered by this section.

- We should inspect the statistical properties of the estimators involved in any fitting procedure.
- We should check which parts of the distribution have had to be truncated in order to make the fit 'work'.
- We must take care over the 'goodness-of-fit' criterion employed.

However, in my opinion, none of these three is as hard as the less technical problems which we encounter next.

5.5 Interpreting the answers

Put yourself in the position of someone who is carrying out an independent study of inequality, or of one examining the summary results of some recent report on the subject. To fix ideas, let us assume that it appears that inequality has decreased in the last five years. But presumably we are not going to swallow any story received from a computer print-out or a journal article straightaway. In this final and important puzzle of 'colour the picture', we will do well to question the colouring instructions which the presentation of the facts suggests.

- **What cardinal representation has been used?**
- **Has the cake shrunk?**
- **Is the drop in inequality an optical illusion?**
- **How do we cope with problems of non-comparability?**
- **Is the trend toward equality large enough to matter?**

INEQUALITY CHANGE: A CHECK-LIST

Although the queries that you raise in the face of the evidence may be far more penetrating than mine, I should like to mention some basic questions that ought to be posed, even if not satisfactorily resolved. In doing so I shall take as understood two issues that we have already laboured to some extent:

- that agreement has been reached on the definition of 'income' and other terms and on the choice of inequality measure(s);
- that we are satisfied that the observed changes in inequality are 'significant' in a statistical or formal sense as discussed in this chapter.

Each of these questions is of the sort that merits several journal articles in its own right. That being said, I am afraid that you will not find that they are asked often enough.

What Cardinal Representation Has Been Used?

The retentive reader will recall from the first chapter that two inequality measures, although ordinally equivalent (so that they always rank any list of social states in the same order), might not have equivalent cardinal properties, so that percentage changes in inequality could appear different according to the two measures.

As examples of this, take the Herfindahl index H and the coefficient of variation c. Since

$$H = \frac{c^2 + 1}{n}$$

for the same population size H and c will always rank any pair of states in the same order. However, the relative size of any difference in inequality will be registered differently by H and by c. To see this, re-examine Table 5.1 where we noted that the minimum and maximum values of c were 5.684 and 6.352, which means that there is a difference in measured inequality of about 12 per cent which is attributable to the effect of grouping. If we did the same calculation for H, we would find that the gap appeared to be much larger, namely 22 per cent. The measure H will always register larger proportional changes in inequality than c, as long as c lies above one (exactly the reverse is true for c less than one).

What this implies more generally is that we should not be terribly impressed by a remark such as 'inequality has fallen by x per cent according to inequality measure J' unless we are quite clear in our own minds that according to some other sensible and ordinally equivalent measure the quantitative result is not substantially different.[13]

Has the Cake Shrunk?

Again you may recollect that in Chapter 1 we noted that for much of the formal work it would be necessary to take as axiomatic the existence of a fixed total of income or wealth to be shared out. This axiom is implicit in the definition of many inequality measures so that they are insensitive to changes in mean income and, insofar as it isolates a pure distribution problem, seems quite reasonable. However, presuming that society has egalitarian preferences,[14] the statement 'inequality has decreased in the last five years' cannot by itself imply 'society is now in a better state' unless one is quite sure that the total to be divided has not drastically diminished also. Unless society is very averse to inequality, a mild reduction in inequality accompanied by a significant drop in average income may well be regarded as a definitely retrograde change.

We can formulate this readily in the case of an inequality measure that is explicitly based upon a social welfare function: by writing down the social welfare function in terms of individual incomes $y_1, y_2, \ldots, y_n$ we

[13] A technical note. It is not sufficient to normalize so that the minimum value of J is 0, and the maximum value 1. For, suppose J does have this property, then so does J^m where m is any positive number, and of course, J and J^m are ordinally, but not cardinally, equivalent.

[14] This is implied in the use of any inequality measure that satisfies the weak principle of transfers.

are specifying both an inequality ranking and a trade-off between average income and an inequality index consistent with this ranking.[15] Atkinson's measure A_ε and the social welfare function specified on page 41 form a good example of this approach: by the definition of A_ε, social welfare is an increasing function of $[1 - A_\varepsilon]$. Hence a fall in inequality by one per cent of its existing value will be exactly offset (in terms of this social welfare function) if average income also falls by an amount

$$g_{\min} = \frac{A_\varepsilon}{1 - A_\varepsilon}.$$

Likewise, a rise in inequality by one per cent of its existing value will be wiped out in social welfare terms if average income grows by at least this same amount. Call this minimum income growth rate $g_{\min}$: obviously $g_{\min}$ increases with A_ε which in turn increases with ε. So, noting from Fig. 5.16 that for $\varepsilon = \frac{1}{2}$, $A_\varepsilon = 0.25$, we find that on this criterion $g_{\min} = 0.33$: a one per cent reduction in inequality would be exactly wiped out by a 0.33 per cent reduction in income per head. But if $\varepsilon = 3$, $A_\varepsilon = 0.833$, and a one per cent reduction in inequality would need to be accompanied by a 5 per cent reduction in the cake for its effect on social welfare to be eliminated. Obviously all the remarks of this paragraph apply symmetrically to a growing cake accompanied by growing inequality.

I should perhaps stress again that this is a doubly value-laden exercise: first the type of social welfare function that is used to compute the equality-mean income trade-off is itself a judgement; then the choice of ε along the horizontal axis in Fig. 5.21 is obviously a matter of social values too.

Is the Drop in Inequality an Optical Illusion?

Unfortunately this may very well be so if we have not taken carefully into consideration demographic, social, and occupational shifts during the period. Some of these shifts you may want to include within the ambit of inequality anyway, but the treatment of others is less clear. Let us follow through two examples.

[15] Actually, this requires some care. Notice that the same inequality measure can be consistent with a variety of social welfare functions. For example, if we do not restrict the SWF to be additive, the measure A_ε could have been derived from any SWF of the form:

$$\phi(\bar{y}) \sum_{i=1}^{n} \frac{y_i^{1-\epsilon} - 1}{1 - \epsilon}$$

which means that virtually any trade-off between equality and income can be obtained, depending on the specification of ϕ. Pre-specifying the SWF removes this ambiguity, for example, if we insist on the additivity assumption for the SWF then ϕ = constant, and there is the unique trade-off between equality and mean income.

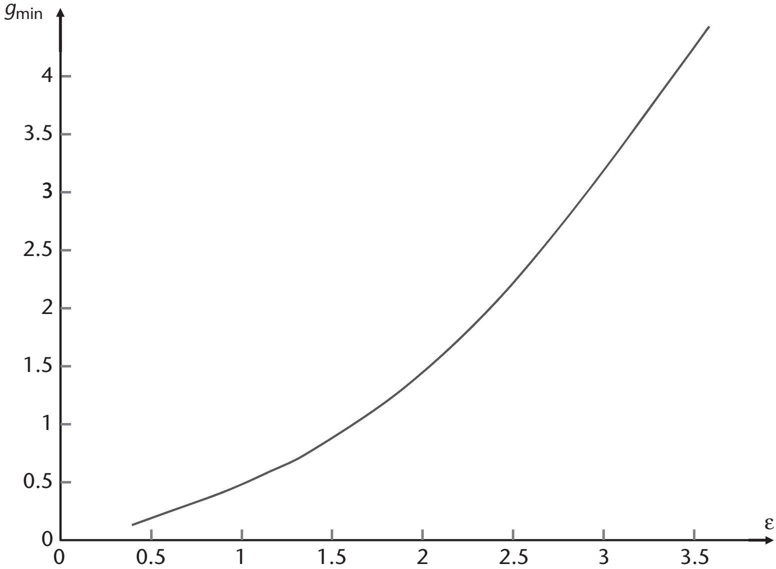

FIG. 5.21. The minimum income growth to offset a 1 per cent growth in inequality

First, suppose there is higher inequality of earnings among doctors than among dockers, that relative remuneration and inequality within occupations have not altered over time, but that there are now relatively more dockers. Inequality in the aggregate will have decreased, although the inequality of earnings opportunity facing a new entrant to either occupation will have remained unchanged. Whether or not one concludes that inequality has 'really' gone down is in large part a matter of interpretation, though my opinion is that it has done so.

However, I would not be so confident in the case of the second example: Suppose income inequality within age groups increases with the age of the group (this is very often true in practice). Now imagine that the age distribution is gradually shifting in favour of the young, either because the birth rate has been rising, or because pensioners are dying earlier, but that inequality within age groups remains unaltered. It will appear that inequality is falling, but this is due entirely to the demographic change. Indeed, if your chances of physical survival are closely linked to your income, the appearance that inequality is decreasing can be quite misleading, since the death rate may have been substantially boosted by the greater inequality among the old.

There are obviously several social and economic factors that ought to be considered in a similar way. Among these are changes in the frequency of marriage and divorce, shifts (possibly cyclical) of the numbers of wives, children, and other part-time or temporary workers in the labour force, and price changes that affect people's real incomes in different ways depending on their position in the Parade of incomes.

How Do we Cope with Problems of Non-Comparability?

This question follows naturally from the last and can be approached in two ways: non-comparability of types of income, and non-comparability of groups of income recipients. In the first case we may well want to examine, say, the inequality of labour earnings, of income from property and the relationship of these quantities to overall inequality. We evidently need to have a detailed breakdown of the income distribution by both income type and recipient—information that is usually hard to come by. Furthermore the mechanics of the relationship between inequality of components of income and inequality of income as a whole are by no means straightforward—see the Technical Appendix.

In the second case, while examining the effect of demographic and other shifts, we may conclude that crudely lumping together different groups of the population and thus treating them as comparable in every way is unwarranted. In order to handle this difficulty, it helps to have an inequality measure that can be conveniently decomposed into a component representing inequality within groups, and a component giving inequality between groups. It would look something like this:

$$I_{\text{total}} = w_1 I_1 + w_2 I_2 + \ldots + w_k I_k + I_{\text{between}}$$

where I_{total} is the value of inequality in the aggregate, $I_1, I_2, \ldots, I_k$ is the value of inequality within subgroup $1, 2, \ldots, k$ respectively, $w_1, w_2, \ldots, w_k$ form a set of weights, and I_{between} is the between-group inequality, found by assuming that everyone within a particular group has the same income. The details of this decomposition, and in particular the specification of the weights for different inequality measures, can be found in the Technical Appendix. Given different problems of non-comparability of income recipients there are, broadly speaking, two courses of action open to us, each of which I shall illustrate by an example.

First, suppose that each group corresponds to a particular family-size class, with the family taken as the fundamental income-receiving unit. Then we may be able to avoid the problem of non-comparability between groups by adjusting incomes to an 'adult-equivalent' basis, as mentioned earlier. If the weights w depend on the shares of each group in total income, then such an adjustment will involve increasing the weights for a group containing small families, decreasing the w for a group of large families. The value of I_{between} would have to be recomputed for average 'per-adult equivalent' income in each group. A similar procedure can be adopted in the case of an aggregation of economically diverse nations within a political grouping such as the European Union; because of artificiality of exchange rates and other reasons listed on page 111, average income in each nation and thus the weights for each nation may have to be adjusted.

In the second place, there may be little point in trying to adjust I_{between} since 'between-group' inequality may be intrinsically meaningless. A case can be made for this in examining income distributions that are differentiated by age group. Although the measured inequality within an age group can be seen as reflecting a genuine disparity among people's economic prospects, the between-group component merely reflects, for the most part, the fact that people's incomes are not uniform over their lives. The expression I_{between} may thus not reflect inequality in the conventional sense at all. This being so, the problem of non-comparability of people at different points in the lifecycle can be overcome by dropping the I_{between} component and adopting some alternative weighting scheme that does not involve income shares (perhaps, for example, population shares instead) so as to arrive at an average value of inequality over the age groups.

Is the Trend toward Equality Large Enough to Matter?

The discussion of significance in its formal, statistical sense leaves some unsettled questions. All that we glean from this technical discussion are guidelines as to whether an apparent change in inequality could be accounted for simply by sampling variability or by the effect of the grouping of observations in the presentation. Whether a reduction in inequality that passes such significance tests is then regarded as 'important' in a wider economic or social sense is obviously a subjective matter—it depends on the percentage change that you happen to find personally exciting or impressive. However, I do not think that we have to leave the matter there. In the case of economic inequality there are at least two ways of obtaining a crude independent check.

The first method is to contrast the historical change with some other easily measured inequality difference. An interesting exercise is to compare the magnitude of the reduction in inequality in the population as a whole during a number of years with the change in inequality over the lifecycle as observed for the age groups in any one year. Alternatively, we might consider the secular change in inequality alongside the apparent[16] redistribution

[16] The qualification 'apparent' is included because, as we noted on page 112, the observed distribution of income before tax is not equivalent to the theoretical distribution of income 'without the tax'.

achieved in any one year by a major government policy instrument, such as the income tax. Neither of these comparisons yields an absolute standard of economic significance, of course, but each can certainly put a historical trend into a clear current perspective.

The second device is applicable to measures based on social welfare functions, and may be taken as an extension of the earlier shrinking cake question. We noted there that a 1 per cent reduction in A_ε is equivalent in social welfare terms to a $A_\varepsilon/[1 - A_\varepsilon]$% increase in income per head. So let us suppose that, for some value of ε, at the beginning of the period $A_\varepsilon = 0.5$ (so that $A_\varepsilon/[1 - A_\varepsilon] = 1$). Then if economic growth during the period raised per capita income by 10 per cent, an accompanying fall of A_ε to say 0.45 would be quite impressive, since the gain to society through reduction in inequality would be as great as the benefit to society of the increase in average living standards. However, the procedure in general obviously depends on your acceptance of the social welfare function approach, and the particular result depends on the inequality aversion which you are prepared to impute to society.

5.6 A sort of conclusion

Finding and asking the right questions is an irksome task. But it is evidently a vital one too, since our brief enquiry has revealed several pitfalls which affect our understanding of the nature of inequality and the measurement of its extent and change. It has been persuasively argued by some writers that inequality is what economics should be all about. If this is so, then the problem of measurement becomes crucial, and in my opinion handling numbers effectively is what measuring inequality is all about.

Technical progress in computing hardware and statistical software has greatly alleviated the toil of manipulation for layman and research worker alike. So the really awkward work ahead of us is not the mechanical processing of figures. It is rather that we have to deal with figures which, instead of being docile abstractions, raise fresh challenges as we try to interpret them more carefully. However, the fact that the difficulties multiply the more closely we examine the numbers should reassure us that our effort at inequality measurement is indeed worthwhile.

'Problems worthy
Of Attack
Prove their worth
By hitting back.'

Piet Hein.

5.7 Questions

1. (a) Use the file 'World Bank data' on the website to provide an inequality ranking of countries according to (i) the share of the bottom 20 per cent (ii) the share of the top 20 per cent and (iii) the Gini coefficient.
   (b) Use the information on shares in the file to compute an estimate of the Gini coefficient: why would one expect this estimate to be different from that provided in the file?
   (c) Some of the datasets in this compilation (taken from World Bank 2004) are from income surveys and some from surveys of expenditure: which type of survey would you expect to result in higher inequality? (See World Bank 2005, page 38.)
2. The data in Table 5.6 (taken from Jones 2008) show the distribution by decile groups according to five different concepts of income corresponding to five successive notional stages of government intervention. Draw the Lorenz curves and generalized Lorenz curves. What effect on income inequality does each tax or benefit component appear to have? Does the distribution of final income dominate the distribution of original income according to the principles in Theorem 3 on page 47? (See 'taxes and benefits' on the website for a copy of the data and a hint at the answers; see Hills 2004, pp. 90–94 for a discussion of the practical issues relating to these data. See Wolff and Zacharias 2007 for the corresponding issue in the USA.)
3. Consider an income distribution in which there are two families. Family 1 contains one person with an income of $\$10,000$; family 2 contains two persons with a combined income of $\$15,000$. Assume that the formula for the number of equivalent adults in a family of size s is given by s^{η} where η is an index of sensitivity to size. What situations do the cases $\eta = 0$ and $\eta = 1$ represent?
   (a) Compute the generalized entropy measure ($\theta = -1$) for this economy on the assumption that each family is given an equal weight and that income is family income per equivalent adult. Do this for a range of η-values from 0 to 1 and plot the results on a graph.
   (b) Repeat the exercise for the cases $\theta = 0.5$ and $\theta = 2$. Do you get the same relationship between measured inequality and η?
   (c) Repeat the exercise for the case where each family is weighted according to the number of individuals in it. Does the re-weighting affect the results? (See the file 'Equivalence scales and weighting' on the website for the answers. See also Coulter *et al.* 1992b for further discussion.)
4. Suppose you have income data which has been grouped into three intervals: (\$0,\$2,000), (\$2,000,\$4,000), (\$4,000,\$6,000). There are 1,000

Table 5.6. Average income, taxes, and benefits by decile groups of all households. UK 1998–9

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
No of h'holds	2464	2465	2469	2463	2471	2463	2465	2469	2465	2470
								Average per household, £ per year		
Original income	£2,119	£3,753	£5,156	£9,365	£14,377	£18,757	£23,685	£29,707	£36,943	£65,496
Cash benefits	£4,262	£5,351	£5,552	£4,794	£3,907	£2,979	£2,506	£1,551	£1,252	£990
Gross income	£6,381	£9,104	£10,708	£14,159	£18,284	£21,736	£26,191	£31,258	£38,195	£66,486
Dir taxes	-£803	-£1,029	-£1,269	-£2,144	-£3,185	-£4,183	-£5,414	-£6,855	-£8,760	-£16,559
Disp income	£5,578	£8,075	£9,439	£12,016	£15,099	£17,554	£20,777	£24,403	£29,434	£49,927
Indirect taxes	-£2,238	-£2,150	-£2,365	-£2,940	-£3,587	-£4,055	-£4,611	-£5,065	-£5,527	-£7,153
Post-tax income	£3,340	£5,925	£7,074	£9,076	£11,511	£13,498	£16,166	£19,338	£23,908	£42,774
Benefits in kind	£4,604	£3,771	£3,501	£3,294	£3,457	£3,219	£2,787	£2,468	£2,187	£2,015
Final income	£7,945	£9,696	£10,575	£12,370	£14,969	£16,717	£18,953	£21,806	£26,095	£44,789

Source: Office for National Statistics individuals in each interval and the mean of each interval is at the midpoint. Draw the lower-bound and upper-bound Lorenz curves as described on page 125.

5. Compute the mean and variance for a split histogram distribution over an interval $[a, b]$: i.e. a distribution for which the density is a constant f_1 for $a \leq y < \bar{y}$ and f_2 for $\bar{y} \leq y < b$. Given the US data in Table 5.1 (see file 'IRS income distribution' on the website) find the numbers f_1 and f_2 for each interval.
6. Show that you can write the formula for Gini coefficient on page 114 as $G = \sum_{i=1}^{n} w_i y_{(i)}$ where the $w_1, w_2, \ldots, w_n$ are weights for each income from the lowest ($i = 1$) to the highest ($i = n$).
   (a) What is the formula for w_i?
   (b) If there is a small income transfer of $\triangle y$ from person i to person j what is the impact on G according to this formula?
   (c) Suppose a six-person economy has income distribution A given in Table 3.3 (page 65). Use your solution for $\triangle y$ to evaluate the effect on Gini of switching to distribution B for the East, for the West, and for the economy as a whole.
   (d) Suppose a six-person economy has income distribution A given in Table 3.4. Again use your solution for $\triangle y$ to evaluate the effect on Gini of switching to distribution B for the East, for the West, and for the economy as a whole. Why do you get a rather different answer from the previous case?

Table 5.7. Observed and expected frequencies of household income per head, Jiangsu, China

1980			1983			1986		
y	Obs	Exp	y	Obs	Exp	y	Obs	Exp
0	12	3.5	*0*	5	0.3	*0*	3	1.1
80	33	30.3	*100*	21	10.9	*100*	16	16
100	172	184.8	*150*	81	65.8	*150*	73	65.3
150	234	273.8	*200*	418	385.2	*200*	359	355.4
200	198	214.1	*300*	448	463.6	*300*	529	561.9
250	146	133.3	*400*	293	305.1	*400*	608	598.4
300	190	145.2	*500*	212	247.8	*500*	519	503.2
			800	15	16	*600*	657	672.8
			1,000	5	3.3	*800*	346	330.4
						1,000	237	248.3
						1,500	40	38.4
						2,000	13	8.8
						5,000		
all ranges	*985*	*985*		*1,498*	*1,498*		*3,400*	*3,400*

y: lower limit of income interval (yuan pa)

Source: Statistical Office, Jiangsu Province, Rural Household Budget Survey.

7. For the same dataset as in Question 5 verify the lower-bound and the upper-bound estimates of the Atkinson index $A_{0.5}$ given in Table 5.2.
8. Apply a simple test to the data in Table 5.7 (also available in file 'Jiangsu' on the website) to establish whether or not the lognormal model is appropriate in this case. What problems are raised by the first interval here? (Kmietowicz and Ding 1993).

APPENDIX A

Technical Appendix

A.1 Overview

This appendix assembles some of the background material for results in the main text and covers some important related points that are of a more technical nature. The topics covered, section by section, are as follows:

- Standard properties of inequality measures both for general income distributions (discrete and continuous) and for specific distributions.
- The properties of some important standard functional forms of distributions, focusing mainly upon the lognormal and Pareto families.
- Interrelationships amongst important specific inequality measures.
- Inequality decomposition by population subgroup.
- Inequality decomposition by income components.
- Negative incomes.
- Estimation problems for (ungrouped) microdata.
- Estimation problems for grouped data, where the problem of interpolation within groups is treated in depth.
- Using the website to work through practical examples.

A.2 Measures and their properties

This section reviews the main properties of standard inequality indices; it also lists the conventions in terminology and notation used throughout this appendix. Although all the definitions could be expressed concisely in terms of the distribution function F, for reasons of clarity I list first the terminology and definitions suitable specifically for discrete distributions with a finite population, and then present the corresponding concepts for continuous distributions.

Discrete Distributions

The basic notation required is as follows. The population size is n, and the income of person i is y_i, $i = 1, \ldots, n$. The arithmetic mean and the geometric mean are defined, respectively, as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

$$y^* = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log y_i\right) = [y_1 y_2 y_3 \ldots y_n]^{1/n}.$$

Using the arithmetic mean we may define the share of person i in total income as $s_i = y_i / [n\bar{y}]$.

Table A.1 lists the properties of many inequality measures mentioned in this book, in the following format:

- A general definition of inequality measure given a discrete income distribution.
- The maximum possible value of each measure on the assumption that all incomes are non-negative. Notice in particular that for $\varepsilon \geq 1$ the maximum of A_ε is 1 and the maximum of D_ε is ∞, but not otherwise. The minimum value of each measure is zero, with the exception of the Herfindahl index for which the minimum is $\frac{1}{n}$.
- The transfer effect for each measure: the effect of the transfer of an infinitesimal income transfer from person i to person j.

Continuous Distributions

The basic notation required is as follows. If y is an individual's income, $F(y)$ denotes the proportion of the population with income less than or equal to y. The operator $\int$ implies that integration is performed over the entire support of y; i.e. over $[0, \infty)$ or, equivalently for F, over the interval $[0, 1]$. The arithmetic mean and the geometric mean are defined as

$$\bar{y} = \int y\, dF,$$

$$y^* = \exp\left(\int \log y\, dF\right).$$

If the density function $f(\cdot)$ is everywhere well-defined then these definitions can be written equivalently as

$$\bar{y} = \int y f(y)\, dy,$$

$$y^* = \exp\left(\int \log y \;\, f(y)\, dy\right).$$

From the above we may define the proportion of total income received by persons who have an income less than or equal to y as

$$\Phi(y) = \frac{1}{\bar{y}}\int_0^y z dF(z).$$

The Lorenz curve is the graph of (F, Φ).

Table A.1. Inequality measures for discrete distributions

Name	Definition	Maximum	Transfer effect
Variance	$V = \frac{1}{n}\sum_{i=1}^{n}[y_i - \bar{y}]^2$	$\bar{y}^2[n-1]$	$\frac{2}{n}[y_j - y_i]$
Coefficient of variation	$c = \frac{\sqrt{V}}{\bar{y}}$	$\sqrt{n-1}$	$\frac{y_j - y_i}{n\bar{y}\sqrt{V}}$
Range	$R = y_{\max} - y_{\min}$	$n\bar{y}$	2 if $y_i = y_{\min}$ and $y_j = y_{\max}$, 1 if $y_i = y_{\min}$ or $y_j = y_{\max}$, 0 otherwise
Rel. mean deviation	$M = \frac{1}{n}\sum_{i=1}^{n}\left\vert\frac{y_i}{\bar{y}} - 1\right\vert$	$2 - \frac{2}{n}$	$\frac{2}{n\bar{y}}$ if $[y_i - \bar{y}][y_j - \bar{y}] < 0$ 0 otherwise
Logarithmic variance	$v = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{\bar{y}}\right]^2$	∞	$\frac{2}{ny_j}\log\frac{y_j}{\bar{y}} - \frac{2}{ny_i}\log\frac{y_i}{\bar{y}}$
Variance of logarithms	$v_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{y^*}\right]^2$	∞	$\frac{2}{ny_j}\log\frac{y_j}{y^*} - \frac{2}{ny_i}\log\frac{y_i}{y^*}$
Gini	$\frac{1}{2n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\vert y_i - y_j\vert$	$\frac{n-1}{n}$	$\frac{2F(y_j) - F(y_i)}{n\bar{y}}$
Atkinson	$A_\varepsilon = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$	$1 - n^{\frac{-\varepsilon}{1-\varepsilon}}$ or 1^*	$\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{n\bar{y}^{1-\varepsilon}[1 - A_\varepsilon]^{-\varepsilon}}$
Dalton	$D_\varepsilon = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} y_i^{1-\varepsilon} - 1}{\bar{y}^{1-\varepsilon} - 1}$	$\frac{1 - n^{-\varepsilon}}{1 - \bar{y}^{\varepsilon-1}}$ or 1^{**}	$\frac{1-\varepsilon}{n}\,\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{\bar{y}^{1-\varepsilon} - 1}$
Generalized entropy	$E_\theta = \frac{1}{\theta^2 - \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^\theta - 1\right],\ \theta \neq 0, 1$	$\frac{n^{\theta-1} - 1}{\theta^2 - \theta}$ or ∞^{***}	$\frac{y_j^{\theta-1} - y_i^{\theta-1}}{[\theta - 1]n\bar{y}^\theta}$
MLD	$L = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\bar{y}}{y_i}\right) = -\frac{1}{n}\sum_{i=1}^{n}\log(ns_i) = E_0$	∞	$\frac{1}{n}\left[\frac{1}{y_i} - \frac{1}{y_j}\right]$
Theil	$T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\bar{y}}\log\left(\frac{y_i}{\bar{y}}\right) = \sum_{i=1}^{n} s_i\log(ns_i) = E_1$	$\log n$	$\frac{1}{n\bar{y}}\log\frac{y_j}{y_i}$
Herfindahl	$H = \frac{1}{n}[c^2 + 1] = \sum_{i=1}^{n} s_i^2$	1	$\frac{2}{n^2\bar{y}^2}[y_j - y_i]$

Notes: * 1 if $\varepsilon \geq 1$; ** ∞ if $\varepsilon \geq$; *** $\propto$ if $\theta \leq 0$.

Table A.2 performs the rôle of Table A.1 for the case of continuous distributions as well as other information: to save space not all the inequality measures have been listed in both tables. The maximum value for the inequality measures in this case can be found by allowing $n \to \infty$ in column 3 of Table A.1. In order to interpret Table A.2 you also need the standard normal distribution function

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}u^2}du,$$

Table A.2. Inequality measures for continuous distributions

Name	Definition	$\Lambda(y; \mu, \sigma^2)$	$\Pi(y; \underline{y}, \alpha)$
Variance	$V = \int [y_i - \bar{y}]^2\, dF$	$e^{2\mu+\sigma^2}\left[e^{\sigma^2} - 1\right]$	$\frac{\alpha \underline{y}^2}{[\alpha-1]^2[\alpha-2]}$
Coefficient of variation	$c = \frac{\sqrt{V}}{\bar{y}}$	$\sqrt{e^{\sigma^2} - 1}$	$\sqrt{\frac{1}{\alpha[\alpha-2]}}$
Rel. mean deviation	$M = \int \left\vert \frac{y}{\bar{y}} - 1 \right\vert dF$	$2\left[2N\left(\frac{\sigma}{2}\right) - 1\right]$	$2\frac{[\alpha-1]^{\alpha-1}}{\alpha^{\alpha}}$
Logarithmic variance	$v = \int \left[\log \frac{y}{\bar{y}}\right]^2 dF$	$\sigma^2 + \frac{1}{4}\sigma^4$	$\log \frac{\alpha-1}{\alpha} + \frac{1}{\alpha} + \frac{1}{\alpha^2}$
Variance of logarithms	$v_1 = \int \left[\log \frac{y}{y^*}\right]^2 dF$	σ^2	$\frac{1}{\alpha^2}$
Equal shares	$F(\bar{y})$	$N\left(\frac{\sigma}{2}\right)$	$1 - \left[\frac{\alpha-1}{\alpha}\right]^{\alpha}$
Minimal majority	$F\left(\Phi^{-1}(0.5)\right)$	$N(\sigma)$	$1 - 2^{\frac{\alpha}{\alpha-1}}$
Gini	$G = 1 - 2\int \Phi dF$	$2N\left(\frac{\sigma}{\sqrt{2}}\right) - 1$	$\frac{1}{2\alpha-1}$
Atkinson	$A_\varepsilon = 1 - \left[\int \left[\frac{y}{\bar{y}}\right]^{1-\varepsilon} dF\right]^{\frac{1}{1-\varepsilon}}$	$1 - e^{-\frac{1}{2}\varepsilon\sigma^2}$	$1 - \left[\frac{\alpha-1}{\alpha}\right]\left[\frac{\alpha}{\alpha+\varepsilon-1}\right]^{\frac{1}{1-\varepsilon}}$
Generalized entropy	$E_\theta = \frac{1}{\theta^2-\theta}\left[\int \left[\frac{y}{\bar{y}}\right]^{\theta} dF - 1\right],\ \theta \neq 0, 1$	$\frac{e^{\left[\theta^2-\theta\right]\frac{\sigma^2}{2}} - 1}{\theta^2-\theta}$	$\frac{1}{\theta^2-\theta}\left[\left[\frac{\alpha-1}{\alpha}\right]^{\theta}\frac{\alpha}{\alpha-\theta} - 1\right]$
MLD	$L = \int \log\left(\frac{\bar{y}}{y}\right) dF = E_0$	$\frac{\sigma^2}{2}$	$\log\left(\frac{\alpha}{\alpha-1}\right) - \frac{1}{\alpha}$
Theil	$T = \int \frac{y}{\bar{y}} \log\left(\frac{y}{\bar{y}}\right) dF = E_1$	$\frac{\sigma^2}{2}$	$\log\left(\frac{\alpha-1}{\alpha}\right) + \frac{1}{\alpha-1}$

provided in most spreadsheet software and tabulated in Lindley and Miller (1966) and elsewhere; $N^{-1}(\cdot)$ denotes the inverse function corresponding to $N(\cdot)$. In summary Table A.2 gives:

- A definition of inequality measures for continuous distributions.
- The formula for the measure given that the underlying distribution is lognormal.
- The formula given that the underlying distribution is Pareto (type I).

A.3 Functional forms of distribution

We begin this section with a simple listing of the principal properties of the lognormal and Pareto distributions in mathematical form. This is deliberately brief since a full verbal description is given in Chapter 4, and the formulas of inequality measures for these distributions are in Table A.2.

The lognormal distribution
The basic specification is:

$$\begin{aligned}
F(y) &= \Lambda\left(y;\ \mu, \sigma^2\right) = \int_0^y \frac{1}{\sqrt{2\pi}\,\sigma x} \exp\left(-\frac{1}{2\sigma^2}\left[\log x - \mu\right]^2\right) dx,\\
\Phi(y) &= \Lambda\left(y;\ \mu + \sigma^2, \sigma^2\right),\\
\bar{y} &= e^{\mu + \frac{1}{2}\sigma^2},\\
y^* &= e^{\mu}.
\end{aligned}$$

where μ and σ are parameters; and the Lorenz curve is given by:

$$\Phi = N\left(N^{-1}(F) - \sigma\right)$$

The Pareto distribution (type I)
The basic specification is:

$$\begin{aligned}
F(y) &= \Pi(y;\ \underline{y}, \alpha) = 1 - \left[\underline{y}/y\right]^{\alpha},\\
\Phi(y) &= \Pi(y;\ \underline{y}, \alpha - 1),\\
\bar{y} &= \frac{\alpha}{\alpha - 1}\underline{y},\\
y^* &= e^{1/\alpha}\underline{y}.
\end{aligned}$$

where α and $\underline{y}$ are parameters. The density function is

$$f(y) = \frac{\alpha \underline{y}^{\alpha}}{y^{\alpha+1}}$$

and the Lorenz curve is given by:

$$\Phi = 1 - [1 - F]^{\frac{\alpha-1}{\alpha}}.$$

The last equation may be used to give a straightforward method of interpolation between points on a Lorenz curve. Given two observed points (F_0, Φ_0), (F_1, Φ_1), then for an arbitrary intermediate value F (where $F_0 < F < F_1$), the corresponding intermediate Φ-value is:

$$\Phi(y) = 1 - [1 - \Phi_0] \exp\left(\frac{\log \frac{1-F(y)}{1-F_0} \log \frac{1-\Phi_1}{1-\Phi_0}}{\log \frac{1-F_1}{1-F_0}}\right)$$

However, if this formula is used to interpolate between observed points when the underlying distribution is not Pareto type I, then the following difficulty may arise. Suppose the class intervals used in grouping the data $\{a_1, a_2, a_3, \ldots, a_k, a_{k+1}\}$, the proportions of the population in each group $\{f_1, f_2, f_3, \ldots, f_k\}$, and the average income of each group $\{\mu_1, \mu_2, \mu_3, \ldots, \mu_k\}$, are all known. Then, as described on page 125, a 'maximum inequality' Lorenz curve may be drawn through the observed points using this information. But the above Pareto interpolation formula does not use the information on the as, and the resulting interpolated Lorenz curve may cross the maximum inequality curve. Methods that use all the information about each interval are discussed below in the section 'Estimation problems' on page 175.

VAN DER WIJK'S LAW

As mentioned in Chapter 4, the Pareto type I distribution has an important connection with van der Wijk's law. First we shall derive the average income $z(y)$ of everyone above an income level y. This is

$$z(y) = \frac{\int_y^\infty u f(u) du}{\int_y^\infty f(u) du} = \bar{y}\frac{1 - \Phi(y)}{1 - F(y)}.$$

From the above, for the Pareto distribution (type I) we have

$$\begin{aligned} z(y) &= \frac{\alpha}{\alpha - 1}\underline{y}\left[\underline{y}/y\right]^{\alpha-1}\left[\underline{y}/y\right]^{-\alpha} \\ &= \frac{\alpha}{\alpha - 1}y. \end{aligned}$$

Hence the average income above the level y is proportional to y itself.

Now let us establish that this result is true only for the Pareto (type I) distribution within the class of continuous distributions. Suppose for some distribution it is always true that $z(y) = \gamma y$ where γ is a constant; then, on rearranging, we have

$$\int_y^\infty u f(u) du = \gamma y \int_y^\infty f(u) du,$$

where $f(\cdot)$ is unknown. Differentiate this with respect to y:

$$-yf(y) = -\gamma y f(y) + \gamma\left[1 - F(y)\right].$$

Define $\alpha = \gamma/[\gamma - 1]$; then, rearranging this equation, we have

$$yf(y) + \alpha F(y) = \alpha.$$

Since $f(y) = dF(y)/dy$, this can be treated as a differential equation in y. Solving for F, we have

$$y^\alpha F(y) = y^\alpha + B,$$

where B is a constant. Since $F\left(\underline{y}\right) = 0$ when we have $B = -\underline{y}^\alpha$. So

$$F(y) = 1 - \left[\underline{y}/y\right]^\alpha.$$

Other Functional Forms

As noted in Chapter 4, many functional forms have been used other than the lognormal and the Pareto. Since there is not the space to discuss these in the same detail, the remainder of this section simply deals with the main types; indicating family relationships, and giving the moments about zero where possible. (If you have the rth moment about zero, then many other inequality measures are easily calculated; for example,

$$A_{\varepsilon} = 1 - \frac{[\mu_r']^{1/r}}{\mu_1'},$$

where μ_r' is the rth moment about zero, $r = 1 - \varepsilon$ and $\mu_1' = \bar{y}$.)

We deal first with family relations of the Pareto distribution. The distribution function of the general form, known as the type III Pareto distribution, may be written as

$$F(y) = 1 - \frac{e^{-\beta y}}{[\gamma y + \delta]^{\alpha}},$$

where $\alpha > 0, \beta \geq 0, \gamma > 0$ and $\gamma y + \delta \geq 0$. By putting $\beta = 0$ and $\delta = 1$ in the above equation we obtain the Pareto type II distribution (see below). By putting $\beta = \delta = 0$ and $\gamma = 1/\underline{y}$ in the type III distribution we get the Pareto type I distribution, $\Pi(y;\ \underline{y}, \alpha)$.

Rasche *et al.* (1980) suggested a functional form for the Lorenz curve as follows:

$$\Phi = [1 - [1 - F]^{a}]^{\frac{1}{b}}.$$

Clearly this expression also contains the Pareto type I distribution as a special case. The Rasche *et al.* (1980) form is somewhat intractable, and so in response Gupta (1984) and Rao and Tam (1987) have suggested the following:

$$\Phi = F^{a} b^{F-1},\ a \geq 1, b > 1.$$

(Gupta's version has $a \equiv 1$.) A comparative test of these and other forms is also provided by Rao and Tam (1987).

Singh and Maddala (1976) suggested as a useful functional form the following:

$$F(y) = 1 - \frac{1}{[1 + \gamma y^{\beta}]^{\alpha}},$$

where α, β, γ are positive parameters. From this we can derive the following special cases.

- If $\beta = 1$ we have the Pareto type II distribution.
- If $\gamma = [1/\alpha]\, k^{\beta}$ and $\alpha \to \infty$ then the *Weibull distribution* is generated: $F(y) = 1 - \exp\left(-[ky]^{\beta}\right)$. The rth moment about zero is given by $\mu_r' = k^{-r}\,\Gamma(1 + r/\beta)$, where $\Gamma(\cdot)$ is the Gamma function defined by $\Gamma(x) = \int_0^{\infty} uxe^{-u}du$.
- A special case of the Weibull may be found when $\beta = 1$, namely the exponential distribution $F(y) = 1 - \exp(-ky)$. Moments are given by $k^{-r}\,\Gamma(1 + r)$ which for integral values of r is simply $k^{-r}r!$.

- If $\alpha = 1$ and $\gamma = \underline{y}^{-\beta}$, then we find Fisk's *sech*^{2}*-distribution*:

$$F(y) = 1 - \left[1 + \left[\frac{y}{\underline{y}}\right]^{\beta}\right]^{-1},$$

with the rth moment about zero given by

$$\mu_r' = r\underline{y}^r \frac{\pi}{\beta \sin\left(\frac{r\pi}{\beta}\right)}, \quad -\beta < r < \beta.$$

Furthermore the upper tail of the distribution is asymptotic to a conventional Pareto type I distribution with parameters $\underline{y}$ and β (for low values of y the distribution approximates to a reverse Pareto distribution—see Fisk (1961), p. 175). The distribution gets its name from the transformation $\left[y/\underline{y}\right]^{\beta} = e^x$, whence the transformed density function is $f(x) = e^x/[1+e^x]^2$, a special case of the logistic function.

The sech^2-distribution can also be found as a special case of the Champernowne distribution:

$$F(y) = 1 - \frac{1}{\theta}\tan^{-1}\left(\frac{\sin\theta}{\cos\theta + \left[y/\underline{y}\right]^{\beta}}\right)$$

where θ is a parameter lying between $-\pi$ and π (see Champernowne 1953, Fisk 1961). This likewise approximates the Pareto type I distribution in its upper tail and has the following moments about zero:

$$\mu_r' = \underline{y}^r \frac{\pi}{\theta} \frac{\sin\left(\frac{r\theta}{\beta}\right)}{\sin\left(\frac{r\pi}{\beta}\right)}, \quad -\beta < r < \beta.$$

The required special case is found by letting $\theta \to 0$.

The *Yule distribution* can be written either in general form with density function

$$f(y) = AB_\nu(y, \rho + 1)$$

where $B_\nu(y, \rho+1)$ is the incomplete Beta function $\int_0^\nu u^{y-1}[1-u]^\rho du$, $\rho > 0$ and $0 < \nu \leq 1$, or in its special form with $\nu = 1$, where the frequency is then proportional to the complete Beta function $B(y, \rho+1)$.[1] Its moments are

$$\mu_r' = \sum_{i=1}^{n} \frac{\rho n!}{\rho - n} \Delta_{n,r}\ , \quad \rho > r$$

[1] The analytical properties of the Beta and Gamma functions are discussed in many texts on statistics, for example Berry and Lindgren (1996), Freund (2003).

where

$$\Delta_{n,r} = \begin{cases} [-1]^{r-n} \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=j}^{n} \ldots [ijk \ldots]}_{r-n \text{ terms}} & \text{if } n < r \\ 1 & \text{if } n = r \end{cases}$$

The Yule distribution in its special form approximates the distribution $\Pi(y;\ \Gamma(\rho)^{1/\rho}, \rho)$ in its upper tail. A further interesting property of this special form is that for a discrete variable it satisfies van der Wijk's law.

We now turn to a rich family of distributions of which two members have been used to some extent in the study of income distribution—the Pearson curves. The Pearson type I is the *Beta distribution* with density function:

$$f(y) = \frac{y^{\xi} [1 - y]^{\eta}}{B(\xi, \eta)}$$

where $0 < y < 1$,[2] and $\xi, \eta > 0$. The rth moment about zero can be written $B(\xi + r, \eta)/B(\xi, \eta)$ or as $\Gamma(\xi + r)\Gamma(\xi + \eta)/[\Gamma(\xi)\Gamma(\xi + \eta + r)]$. The *Gamma distribution* is of the type III of the Pearson family:

$$f(y) = \frac{\lambda^{\phi}}{\Gamma(\phi)} y^{\phi - 1} e^{-\lambda y},$$

where $\lambda, \phi > 0$. The moments are given by

$$\mu_r' = \lambda^{-r} \frac{\Gamma(\phi + r)}{\Gamma(\phi)}.$$

Three interesting properties of the Gamma function are as follows. Firstly, by putting $\phi = 1$, we find that it has the exponential distribution as a special case. Secondly, suppose that $\lambda = 1$, and that y has the Gamma distribution with $\phi = \phi_1$ while w has the Gamma distribution with $\phi = \phi_2$. Then the sum $w + y$ also has the Gamma distribution with $\phi = \phi_1 + \phi_2$: a property that is obviously useful if one is considering, say, the decomposition of income into constituent parts such as earned and unearned income. Thirdly, a Beta distribution with a high parameter η looks very similar to a Gamma distribution with high values of parameters λ, ϕ. This can be seen from the formula for the moments. For high values of x and any constant k it is the case that $\Gamma(x)/\Gamma(x + k) \simeq x^{-k}$. Hence the moments of the B-distribution approximate to $[\xi + \eta]^{-r}\Gamma(\xi + r)/\Gamma(\xi)$.

[2] This restriction means that y must be normalized by dividing it by its assumed maximum value.

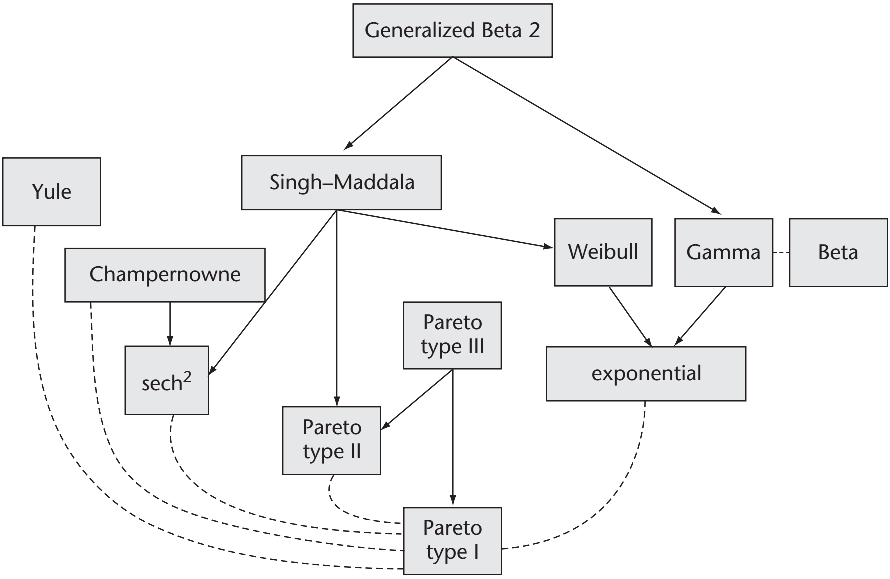

FIG. A.1. Relationships between functional forms

The two-parameter Gamma distribution and the three-parameter Singh–Maddala distribution can each be shown to be a particular case of the four-parameter generalized Beta distribution of the second kind for which the density is:

$$f(y) = \frac{\beta\gamma^{\delta} y^{\beta\delta-1}}{B\left(\delta, \alpha+1\right)\left[1+\gamma y^{\beta}\right]^{\alpha+\delta+1}}$$

Putting $\delta = 1$ in this expression produces the Singh–Maddala density; putting $\alpha = k/\gamma - 1$, $\beta = 1$ and letting $\gamma \to 0$ yields the Gamma density.

The relationships mentioned in the previous paragraphs are set out in Fig. A.1. Solid arrows indicate that one distribution is a special case of another. Broken lines indicate that for high values of the income variable or for certain parameter values, one distribution closely approximates another.

Finally let us look at distributions related to the lognormal. The most obvious is the three-parameter lognormal which is defined as follows. If $y - \tau$ has the distribution $\Lambda(\mu, \sigma^2)$ where τ is some parameter, then y has the three-parameter lognormal distribution with parameters τ, μ, σ^2. The moments about zero are difficult to calculate analytically, although the moments about $y = \tau$ are easy: $\int [y - \tau]^r dF(y) = \exp(r\mu + \frac{1}{2}r^2\sigma^2)$. Certain inequality measures can be written down without much difficulty—see Aitchison and Brown (1957), p. 15. Also note that the lognormal distribution is related indirectly to the Yule distribution: a certain class of stochastic processes which is of interest in several fields of economics has as its limiting distribution either the lognormal or the Yule distribution, depending on the restrictions placed upon the process. On this see Simon and Bonini (1958).

A.4 Interrelationships between inequality measures

In this section we briefly review the properties of particular inequality measures which appear to be fairly similar but which have a number of important distinguishing features.

Atkinson (A_ε) and Dalton (D_ε) Measures
As we have seen, the Atkinson index may be written

$$A_\varepsilon = 1 - \left[\sum_{i=1}^{n} \left[\frac{y_i}{\bar{y}} \right]^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}.$$

Rearranging this and differentiating with respect to ε, we may obtain:

$$\log(1 - A_\varepsilon) + \frac{1-\varepsilon}{1-A_\varepsilon}\frac{\partial A_\varepsilon}{\partial \varepsilon} = \frac{\sum_{i=1}^{n} \left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon} \log\left(\frac{y_i}{\bar{y}}\right)}{\sum_{i=1}^{n} \left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}}.$$

Define $x_i = [y_i/\bar{y}]^{1-\varepsilon}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Noting that $y_i \geq 0$ implies $x_i \geq 0$ and that $\bar{x} = [1 - A_\varepsilon]^{1-\varepsilon}$ we may derive the following result:

$$\frac{\partial A_\varepsilon}{\partial \varepsilon} = \frac{1 - A_\varepsilon}{\bar{x}\,[1-\varepsilon]^2} \left[\frac{1}{n} \sum_{i=1}^{n} x_i \log(x_i) - \bar{x}\log\bar{x} \right].$$

The first term on the right-hand side cannot be negative, since $\bar{x} \geq 0$ and $0 \leq A_\varepsilon \leq 1$. Now $x \log x$ is a convex function so we see that the second term on the right-hand side is non-negative. Thus $\partial A_\varepsilon/\partial\varepsilon \geq 0$: the index A_ε never decreases with ε for any income distribution.

However, the result that inequality increases with inequality aversion for any given distribution does not apply to the related Dalton family of indices. Let us consider D_ε for the cardinalization of the social utility function U used in Chapter 3 and for the class of distributions for which $\bar{y} \neq 1$ (if $\bar{y} = 1$ we would have to use a different cardinalization for the function U—a problem that does not arise with the Atkinson index). We find that if $\varepsilon \neq 1$:

$$D_\varepsilon = 1 - \frac{\bar{y}^{1-\varepsilon}\,[1 - A_\varepsilon]^{1-\varepsilon} - 1}{\bar{y}^{1-\varepsilon} - 1}$$

and in the limiting case $\varepsilon = 1$:

$$D_\varepsilon = 1 - \frac{\log(\bar{y}\,[1 - A_1])}{\log(\bar{y})}$$

As ε rises, $\bar{y}^{1-\varepsilon}$ falls, but A_ε rises, so the above equations are inconclusive about the movement of D_ε. However, consider a simple income distribution given by $y_1 = 1$, $y_2 = Y$ where Y is a constant different from 1. A simple experiment with the above formulas will reveal that D_ε rises with ε if $Y > 1$ (and hence $\bar{y} > 1$) and falls with ε otherwise.

The Logarithmic Variance (v) and the Variance of Logarithms (v_1)
First, note from Table A.1 that $v = v_1 + [\log(y^*/\bar{y})]^2$. Consider the general form of inequality measure

$$\frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{a}\right)\right]^2$$

where a is some arbitrary positive number. The change in inequality resulting from a transfer of a small amount of income from person j to person i is:

$$\frac{2}{ny_i}\log\left(\frac{y_i}{a}\right) - \frac{2}{ny_j}\log\left(\frac{y_j}{a}\right) + \frac{2}{na}\left[\frac{\partial a}{\partial y_i} - \frac{\partial a}{\partial y_j}\right]\sum_{k=1}^{n}\log\left(\frac{y_k}{a}\right)$$

If $a = \bar{y}$ (the case of the measure v) then $\partial a/\partial y_i = \partial a/\partial y_j$ and so the last term is zero. If $a = y^*$ (the case of the measure v_1), then $\sum\log(y_k) = n\log a$, and once again the last term is zero. Hence we see that for v or v_1 the sign of the above expression depends entirely on the behaviour of the function $[1/x]\log x$, which occurs in the first two terms. Now the first derivative of this function is $[1 - \log x]/x^2$, which is positive or negative as x □ $e = 2.71828\ldots$. Suppose $y_i > y_j$. Then, as long as $y_i \leq ae$, we see that because $(1/x)\log x$ is an increasing function under these conditions, the effect of the above transfer is to increase inequality (as we would require under the weak principle of transfers). However, if $y_j \geq ae$, then exactly the reverse conclusions apply—the above transfer effect is negative. Note that in this argument a may be taken to be $\bar{y}$ or y^* according as the measure under consideration is v or v_1 (see also Foster and Ok 1999).

A.5 Decomposition of inequality measures

By Subgroups
As discussed in Chapter 3, some inequality measures lend themselves readily to an analysis of inequality within and between groups in the population. Let there be k such groups so arranged that every member of the population belongs to one and only one group, and let the proportion of the population falling in group j be f_j;[3] by definition we have $\sum_{j=1}^{k} f_j = 1$. Write mean income in group j as $\bar{y}_j$, and the share of group j in total income as g_j (which you get by adding up the income shares of all the members of group j), so that $g_j = f_j\bar{y}_j/\bar{y}$, $\sum_{j=1}^{k} f_j\bar{y}_j = \bar{y}$ and $\sum_{j=1}^{k} g_j = 1$. For some specified inequality measure, let inequality in group j (in other words the inequality measures calculated for group j as if it were a population in its own right) be denoted I_j and let inequality for the total population be I_{total}.

[3] This is equivalent to the term 'relative frequency' used in Chapter 5.

An inequality index I is then considered to be decomposable if there can be found some aggregation function Ξ possessing the following basic property: for any arbitrary income distribution we may write

$$I_{\text{total}} = \Xi\,(I_1, I_2, \ldots, I_k; \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k; n_1, n_2, \ldots, n_k)\,.$$

In other words, total inequality should be a specific function Ξ of inequality in each subgroup, this function depending perhaps on group mean incomes and the population in each group, but nothing else. The principal points to note about decomposability are as follows:

- Some inequality measures simply will not let themselves be broken up in this way: for them no such Ξ exists. As Chapter 3 discussed, the relative mean deviation, the variance of logarithms, and the logarithmic variance cannot be decomposed in a way that depends only on group means and population shares; the Gini coefficient can only be decomposed if the constituent subgroups are 'non-overlapping' in the sense that they can be strictly ordered by income levels. In this special case we have

$$G_{\text{total}} = \underbrace{\frac{n_1^2\bar{y}_1}{n^2\bar{y}}G_1 + \frac{n_2^2\bar{y}_2}{n^2\bar{y}}G_2 + \ldots + \frac{n_k^2\bar{y}_k}{n^2\bar{y}}G_k}_{\text{within}} \;+\; G_{\text{between}}. \qquad \text{(A.1)}$$

  where G_{between} is the value of the Gini coefficient that you would obtain if all individuals in group j receive $\bar{y}_j$.
- On the other hand, there is a large class of measures which will work, and the allocation of inequality between and within groups is going to depend on the inequality aversion, or the appropriate notion of 'distance' which characterizes each measure. The prime example of this is the generalized entropy class E_θ introduced on page 66 for which the scale independence property also holds. Another important class is that of the *Kolm indices* which take the form

$$\frac{1}{\kappa}\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\kappa[\bar{y}-y_i]}\right)$$

  where κ is a parameter that may be assigned any positive value. Each member of this family has the property that if you add the same absolute amount to every y_i then inequality remains unaltered (by contrast to the proportionate invariance of E_θ).
- The cardinal representation of inequality measures—not just the ordinal properties—matters, when you break down the components of inequality.

Let us see how these points emerge in the discussion of the generalized entropy family E_θ and the associated Atkinson indices. For the generalized entropy class E_θ

the inequality aggregation result can be expressed in particularly simple terms. If we define[4]

$$I_{\text{between}} = \frac{1}{\theta^2 - \theta}\left[\sum_{j=1}^{k} f_j \left[\frac{\bar{y}_j}{\bar{y}}\right]^{\theta} - 1\right]$$

and

$$I_{\text{within}} = \sum_{j=1}^{k} w_j I_j, \text{ where } w_j = g_j^{\theta} f_j^{1-\theta}$$

then for any generalized entropy measure we have:

$$I_{\text{total}} = I_{\text{between}} + I_{\text{within}}.$$

From these three equations we may note that in the case of the generalized entropy class:

- Total inequality is a simple additive function of between-group and within-group inequality.
- The between-group component of inequality is found simply by assuming that everyone within a group receives that group's mean income: it is independent of redistribution within any of the j groups.
- Within-group inequality is a weighted average of inequality in each sub-group, although the weights w_j do not necessarily sum to one.
- The within-group component weights will only sum to one if $\theta = 0$ (the case of the mean logarithmic deviation L) or if $\theta = 1$ (the case of Theil's index T).

The Atkinson index A_ε is ordinally equivalent to E_θ for $\varepsilon = 1 - \theta > 0$ (they will always rank any set of Lorenz curves in the same order, as we noted in Chapter 3); in fact we have

$$A_\varepsilon = \begin{cases} 1 - \left[\left[\theta^2 - \theta\right] E_\theta + 1\right]^{1/\theta} & \text{for } \theta \neq 0 \\ A_1 = 1 - e^{-E_0} & \text{for } \theta = 0 \end{cases}.$$

However, because this relationship is nonlinear, we do not have cardinal equivalence between the two indices; as a result we will get a different relationship between total inequality and its components. We can find this relationship by substituting the last formula into the decomposition formula for the generalized entropy measure above. If we do this then—for the case where I is the Atkinson index with parameter ε—we get the following:

[4] Notice that this is the same as the expression given for the generalized entropy measure in Table A.1 for the case where $f_i = 1/n$: in other words you can imagine the whole population of size n as being composed of n groups each of size 1.

$$I_{\text{between}} = 1 - \left[\sum_{j=1}^{k} f_j \left[\frac{\bar{y}_j}{\bar{y}} \right]^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}$$

$$I_{\text{total}} = 1 - \left[\sum_{i=1}^{n} \frac{1}{n} \left[\frac{y_i}{\bar{y}} \right]^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}$$

$$[1 - I_{\text{total}}]^{1-\varepsilon} = \left[[1 - I_{\text{between}}]^{1-\varepsilon} + [1 - I_{\text{within}}]^{1-\varepsilon} - 1 \right]$$

and from these we can deduce

$$I_{\text{within}} = 1 - \left[1 - \sum_{j=1}^{k} f_j^{\varepsilon} g_j^{1-\varepsilon} \left[[1 - I_j]^{1-\varepsilon} - 1 \right] \right]^{\frac{1}{1-\varepsilon}}.$$

To restate the point: the decomposition formula given here for the Atkinson index is different from that given on page 166 for the generalized entropy index because one index is a nonlinear transformation of the other. Let us illustrate this further by taking a specific example using the two inequality measures, A_2 and E_{-1}, which are ordinally but not cardinally equivalent. We have the following relationship:

$$A_2 = 1 - \frac{1}{2E_{-1} + 1}.$$

Applying this formula and using a self-explanatory adaptation of our earlier notation the allocation of the components of inequality is as follows:

$$E_{-1[\text{within}]} = \sum_{j=1}^{k} \frac{f_j^2}{g_j} E_{-1[j]}$$

$$E_{-1[\text{between}]} = \frac{1}{2} \left[\sum_{j=1}^{k} \frac{f_j^2}{g_j} - 1 \right]$$

$$E_{-1[\text{total}]} = E_{-1[\text{between}]} + E_{-1[\text{within}]}$$

whereas

$$A_{2[\text{total}]} = \frac{A_{2[\text{between}]} + A_{2[\text{within}]} - A_{2[\text{between}]} A_{2[\text{within}]}}{1 - A_{2[\text{between}]} A_{2[\text{within}]}}.$$

Now let us consider the situation in China represented in Table A.3. The top part gives the mean income, population, and inequality for each of the ten regions, and for urban and rural groups within each region. The bottom part of the table gives the corresponding values for A_2 and E_{-1} broken down into within- and between-group components (by region) for urban and regional incomes. Notice that:

- The proportion of total inequality 'explained' by the interregional inequality differs according to whether we use the generalized entropy measure or its ordinally equivalent Atkinson measure.

Table A.3. Decomposition of inequality in Chinese provinces, rural and urban subpopulations

	Urban				Rural			
	Pop	*Mean*	A_2	E_{-1}	*Pop*	*Mean*	A_2	E_{-1}
Beijing	463	93	0.151	0.089	788	58	0.135	0.078
Shanxi	564	65	0.211	0.134	1394	29	0.197	0.123
Heilongjiang	506	79	0.160	0.095	1566	33	0.178	0.108
Gansu	690	73	0.153	0.090	1345	19	0.220	0.141
Jiangsu	403	89	0.118	0.067	1945	39	0.180	0.110
Anhui	305	70	0.129	0.074	2284	33	0.112	0.063
Henan	402	75	0.195	0.121	2680	26	0.226	0.146
Hubei	764	81	0.118	0.067	2045	34	0.171	0.103
Guangdong	546	82	0.159	0.095	1475	34	0.211	0.134
Sichuan	1126	84	0.205	0.129	2698	30	0.148	0.087
All	*5769*				*18220*			
Inequality breakdown:								
total			0.175	0.106			0.222	0.142
within			0.168	0.101			0.190	0.118
			(96.2%)	(95.5%)			(86.0%)	(82.7%)
between			0.009	0.005			0.047	0.025
			(5.4%)	(4.5%)			(21.2%)	(17.3%)

Source: The Institute of Economics, the Chinese Academy of Social Sciences.
Incomes: Yuan/month

- The between-group and within-group components sum exactly to total inequality in the case of the generalized entropy measure, but not in the case of the Atkinson measure (these satisfy the more complicated decomposition formula immediately above).

Finally, a word about V, the ordinary variance, and v_1, the variance of logarithms. The ordinary variance is ordinally equivalent to E_2 and is therefore decomposable in the way that we have just considered. In fact we have:

$$V_{[\text{total}]} = \sum_{j=1}^{k} f_j V_{[j]} + V_{[\text{between}]}$$

where $V_{[j]}$ is the variance in group j. Now in many economic models where it is convenient to use a logarithmic transformation of income, one often finds a 'decomposition' that looks something like this:

$$v_{1[\text{total}]} = \sum_{j=1}^{k} f_j v_{1[j]} + v_{1[\text{between}]}.$$

However, this is not a true inequality decomposition. To see why, consider the meaning of the between-group component in this case. We have

$$v_{1[\text{between}]} = \sum_{j=1}^{k} f_j \left[\log y_j^* - \log y^*\right]^2.$$

But, unlike the between-group component of the decomposition procedure we outlined earlier, this expression is not independent of the distribution within each group: for example if there were to be a mean-preserving income equalization in group j, both the within-group geometric mean (y_j^*) and the overall geometric mean (y^*) will be affected. As mentioned above, you cannot properly disentangle the within-group and between-group inequality components for the variance of logarithms.

By Income Components

By contrast to the problem of decomposition by population subgroups, there are relatively few inequality measures that will allow a convenient breakdown by component of income. However, the coefficient of variation c and measures that are ordinally equivalent to it (such as V and H) can be handled relatively easily. Nothing is lost by simplifying to a pairwise decomposition: let income be made up of two components, A and B so that for any person: $y_i = y_{i\mathrm{A}} + y_{i\mathrm{B}}$. Further, let c, c_A, c_B be, respectively, the value of the coefficient of variation for total income, component A income and component B, let λ be the overall amount of component A as a proportion of total income, and let ρ be the correlation coefficient between component A and component B of income. Then:

$$c^2 = \lambda^2 c_{\mathrm{A}}^2 + [1 - \lambda]^2 c_{\mathrm{B}}^2 + 2\lambda [1 - \lambda] c_{\mathrm{A}} c_{\mathrm{B}} \rho.$$

Note that this is well-defined even in the presence of negative income components.

A.6 Negative incomes

For a great many applications in economics it is convenient and reasonable to assume that incomes are non-negative. In fact most of the material in this book has proceeded on this basis. However, there are some important exceptions to this: for example personal wealth (net worth) may be negative at various points of the lifecycle, individuals' incomes may contain substantial losses from self-employed or unincorporated business activity.

The possibility that even a few observations may be negative raises some issues of principle for inequality measurement. Many of the standard inequality measures are simply undefined for negative incomes; furthermore there is a substantial class of these measures that will not work even for zero incomes.

However, the standard 'ranking' tools, such as quantiles and shares, are well defined for all incomes—positive, zero or negative—although they may need to be interpreted with some care. For example the Parade diagram will probably look much the same as that depicted in Fig. 2.1, but the axes will have been shifted vertically.

To see how the shape of the Lorenz curve and the generalized Lorenz curve is affected by the presence of negative incomes, recall that the slope of the Lorenz curve is given by $y/\bar{y}$, and the slope of the generalized Lorenz curve by y. So, if there are some negative incomes, but the mean is still strictly positive, then both curves will initially pass below the horizontal axis (they will be downward-sloping for as long as incomes are negative), will be horizontal at the point where zero income is encountered, and then will adopt a fairly conventional shape over the rest of the diagram. If mean income is actually negative, then the Lorenz curve will appear to be 'flipped vertically' (the generalized Lorenz curve is not affected in this way).

In fact, the use of the conventional Lorenz curve is somewhat problematic in the presence of negative incomes. For this reason it is sometimes convenient to use the absolute Lorenz curve (Moyes 1987), which may be described as follows. The ordinary (or relative) Lorenz curve can be thought of as the generalized Lorenz curve of the distribution $\left(\frac{y_1}{\bar{y}}, \frac{y_2}{\bar{y}}, \ldots, \frac{y_n}{\bar{y}}\right)$ and the absolute Lorenz curve is the generalized Lorenz curve of the distribution $(y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y})$.

The reason that many conventional inequality tools will not work in the presence of negative incomes can be seen from the 'evaluation function' $h(\cdot)$ introduced on page 113. Recall that many inequality measures can be defined in terms of the evaluation function. Consider, for example, the generalized entropy family which will have an evaluation function of the form

$$h(y) = Ay^{\theta}.$$

This function—and hence the associated inequality measures—will be well defined for all negative incomes for the special case where θ is a positive integer greater than 1. However, this severely restricts the choice of θ, because measures with even moderately large values of θ prove to be extraordinarily sensitive to incomes in the upper tail. This means, for example, that in estimating inequality from a sample of microdata, one or two large incomes will drive the estimates of inequality by themselves. The coefficient of variation ($\theta = 2$) is the only member of the generalized entropy class that is likely to be of practical use.

By contrast, all the Kolm indices work with negative incomes; the h function here is

$$h(y) = Ae^{-\kappa y}$$

($\kappa > 0$) which is well-defined for all values of y. Finally, measures that are based on absolute differences—such as the Gini coefficient and relative mean deviation—will also be able to cope with negative incomes.

A.7 Estimation problems

Microdata

As noted in Chapter 5, point estimates of inequality measures from a sample can be obtained just by plugging in the observations to the basic formulas given in Chapters 3 and 4. For computation of point estimates of inequality using unweighted microdata there are very few operations involved:

- Transformation of the income variable $h(y_i)$—the calculation of each term in the formula for J on page 114; the function $h(\cdot)$ typically involves taking logs or raising to a power and normalizing by the mean; so the calculation can usually be performed by standard built-in functions in spreadsheet software.
- Calculating mean income and the mean of the transformed variables (as in the formula for J on page 114)—another standard spreadsheet operation.

- Sorting the data if you want to compute the Gini coefficient (page 114) or plot Lorenz curves—again this is standard for spreadsheets.

The only further qualification that ought to be made is that in practice one often has to work with weighted data (the weights could be sampling weights for example). In this case, associated with each observed income y_i there is a non-negative weight w_i; let us suppose the weights have been normalized so that they sum to 1. Then, instead of the J-formula on page 114, one computes

$$J = \sum_{i=1}^{n} w_i h(y_i)$$

and instead of the formula for G on page 114 one computes[5]

$$G = \frac{1}{\bar{y}} \sum_{i=1}^{n} \kappa_i w_{(i)} y_{(i)}$$

where

$$\kappa_i = 2 \sum_{j=1}^{i} w_{(j)} - w_{(i)} - 1.$$

This requires a little more care, of course, but is still within the capability of standard spreadsheets.

Now consider the standard errors of inequality estimates in the case of unweighted data. As we noted on page 159, inequality measures can be expressed in terms of standard statistical moments. Correspondingly, in situations where we are working with a sample $\{y_1, y_2, \ldots y_n\}$ of n observations from a target population, we will be interested in the sample moment about zero:

$$m'_r = \frac{1}{n} \sum_{i=1}^{n} y_i^r .$$

Standard results give the expected value (mean) and variance of the sample statistic m'_r:

$$\mathcal{E}\left(m'_r\right) = \mu'_r$$

$$\operatorname{var}\left(m'_r\right) = \frac{1}{n}\left[\mu'_{2r} - \left[\mu'_r\right]^2\right]$$

and an unbiased estimate of the sample variance of m is

$$\widehat{\operatorname{var}}\left(m'_r\right) = \frac{1}{n-1}\left[m'_{2r} - \left[m'_r\right]^2\right].$$

[5] In line with our previous usage $w_{(i)}$ and $y_{(i)}$ denote the weight and income for observation i after the observations have been sorted in ascending order of incomes. Note that in each case you can recover the original formulas for J and G for the unweighted case by setting $w_i = 1/n$ everywhere.

If the mean of the distribution is known and you have unweighted data, then this last formula gives you all you need to set up a confidence interval for the generalized entropy measure E_θ. Writing $r = \theta$ and substituting we get (in this special case):

$$E_\theta = \frac{1}{\theta^2 - \theta}\left[\frac{m'_\theta}{\bar{y}^\theta} - 1\right]$$

where $\bar{y}$ is the known mean (μ'_1).

However, if the mean also has to be estimated from the sample (as m'_1), or if we wish to use a nonlinear transformation of m'_θ, then the derivation of a confidence interval for the inequality estimate is a bit more complicated. Applying a standard result (Rao 1973) we may state that if ψ is a differentiable function of m'_r and m'_1, then the expression $\sqrt{n}\left[\psi\left(m'_r, m'_1\right) - \psi\left(\mu'_r, \mu'_1\right)\right]$ is asymptotically normally distributed thus:

$$N\left(0, \left[\frac{\partial\psi}{\partial m'_r}\right]^2 \operatorname{var}\left(m'_r\right) + 2\frac{\partial\psi}{\partial m'_r}\frac{\partial\psi}{\partial m'_{1r}}\operatorname{cov}\left(m'_r, m'_1\right) + \left[\frac{\partial\psi}{\partial m'_1}\right]^2 \operatorname{var}\left(m'_1\right)\right).$$

Again, if one has to work with weighted data the formulas for the standard errors will need to be modified to take into account the weighting. A crucial point here is whether the weights themselves also should be treated as random variables—see the notes (page 193) for further discussion of this point.

Finally, let us consider the problem of estimating the density function from a set of n sample observations. As explained on page 114 in Chapter 5, a simple frequency count is unlikely to be useful. An alternative approach is to assume that each sample observation gives some evidence of the underlying density within a 'window' around the observation. Then you can estimate $F(y)$, the density at some income value y, by specifying an appropriate kernel function K (which itself has the properties of a density function) and a window width (or 'bandwidth') w and computing the function

$$\hat{f}(y) = \frac{1}{w}\sum_{i=1}^{n} K\left(\frac{y - y_i}{w}\right)$$

—the individual terms in the summation on the right-hand side can be seen as contributions of the observations y_i to the density estimate $\hat{f}(y)$. The simple histogram is an example of this device—see, for example, Fig. 5.5. All the sample observations that happen to lie on or above a_j and below a_{j+1} contribute to the height of the horizontal line-segment in the interval (a_j, a_{j+1}). In the case where all the intervals are of uniform width so that $w = a_{j+1} - a_j$, we would have

$$K\left(\frac{y - y_i}{w}\right) = \begin{cases} 1 \text{ if } & \begin{array}{l} a_j \leq y < a_{j+1} \text{ and} \\ a_j \leq y_i < a_{j+1}. \end{array} \\ 0 \text{ otherwise} & \end{cases}$$

However, this histogram rule is crude: each observation makes an 'all or nothing' contribution to the density estimate. So it may be more useful to take a kernel

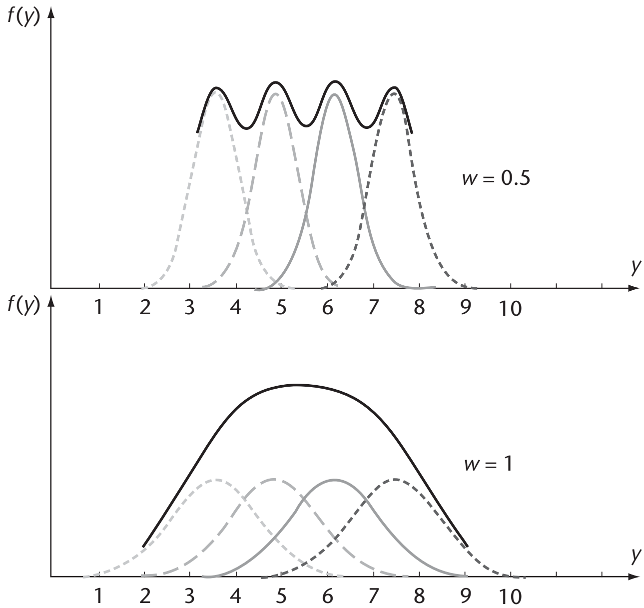

FIG. A.2. Density estimation with a normal kernel

function that is less drastic. For example K is often taken to be the normal density so that

$$K\left(\frac{y-y_i}{w}\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{y-y_i}{w}\right]^2}.$$

The effect of using the normal kernel is illustrated in Fig. A.2 for the case where there are just four income observations. The upper part of Fig. A.2 illustrates the use of a fairly narrow bandwidth, and the lower part the case of a fairly wide window: the kernel density for each of the observations $y_1 \dots y_4$ is illustrated by the lightly-drawn curves: the heavy curve depicts the resultant density estimate. There is a variety of methods for specifying the kernel function K and for specifying the window width w (for example, so as to make the width of the window adjustable to the sparseness or otherwise of the data): these are discussed in Silverman (1986) and Simonoff (1996). Of course associated with each kernel point estimate $\hat{f}(y)$ there will also be a sampling variance, but that takes us beyond the scope of this book.

Grouped Data

Now let us suppose that you do not have microdata to hand, but that it has been presented in the form of income groups. There are three main issues to be discussed.

- *How much information do you have?* Usually this turns on whether you have three pieces of information about each interval (the interval boundaries a_i, a_{i+1}, the relative frequency within the interval f_i, and the interval mean μ_i) or two (the interval boundaries, and the frequency). We will briefly consider both situations.

- *What assumption do you want to make about the distribution within each interval?* You could be interested in deriving lower and upper bounds on the estimates of the inequality measure, consistent with the available information, or you could derive a particular interpolation formula for the density function $\phi_i(y)$ in interval i.
- *What do you want to assume about the distribution across interval boundaries?* You could treat each interval as a separate entity, so that there is no relationship between $\phi_i(y)$ and $\phi_{i+1}(y)$; or you could require that at the boundary between the two intervals (a_{i+1} in this case) the frequency distribution should be continuous, or continuous and smooth, etc. This latter option is more complicated and does not usually have an enormous advantage in terms of the properties of the resulting estimates. For this reason I shall concentrate upon the simpler case of independent intervals.

Given the last remark, we can estimate each function ϕ_i solely from the information in interval i. Having performed this operation for each interval, then to compute an inequality measure we may, for example, write the equation on page 114 as

$$J = \sum_{i=1}^{k} \int_{a_i}^{a_{i+1}} h(y)\phi_i(y)dy.$$

Interpolation on the Lorenz curve may be done as follows. Between the observations i and $i+1$ the interpolated values of F and Φ are

$$F(y) = F_i + \int_{a_i}^{y} \phi_i(x)dx,$$
$$\Phi(y) = \Phi_i + \int_{a_i}^{y} x\phi_i(x)dx.$$

So, to find the share of the bottom 20 per cent, let us say, you set $F(y) = 0.20$ on the left-hand side of the first equation, substitute in the appropriate interpolation formula and then find the value of y on the right-hand side that satisfies this equation; you then substitute this value of y into the right-hand side of the second equation and evaluate the integral.

Interval Means Unknown

In the interpolation formulas presented for this case there is, in effect, only one parameter to be computed for each interval. The *histogram density* is found as the following constant in interval i.

$$\phi_i(y) = \frac{f_i}{a_{i+1} - a_i}, \quad a_i \leq y < a_{i+1}.$$

Using the formulas given on page 157 above, we can see that the *Paretian density* in any closed interval is given by

$$\phi_i(y) = \frac{\alpha a_i^\alpha}{y^{\alpha+1}}, \; a_i \le y < a_{i+1}.$$

$$\alpha = \frac{\log(1+f_i)}{\log \frac{a_i}{a_{i+1}}}$$

We can use a similar formula to give an estimate of Pareto's α for the top (open) interval of a set of income data. Suppose that the distribution is assumed to be Paretian over the top two intervals. Then we may write:

$$\frac{f_k}{f_{k-1}} = \frac{-a_k^{-\alpha}}{a_k^{-\alpha} - a_{k-1}^{-\alpha}}$$

from which we obtain

$$\frac{\log\left(1 + \frac{f_{k-1}}{f_k}\right)}{\log \frac{a_k}{a_{k-1}}}$$

as an estimate of α in interval k.

Interval Means Known

Let us begin with methods that will give the bounding values J_{L} and J_{U} cited on page 120. Within each interval the principle of transfers is sufficient to give the distribution that corresponds to minimum and maximum inequality:[6] for a minimum all the observations must be concentrated at one point, and to be consistent with the data this one point must be the interval mean μ_i; for a maximum all the observations must be assumed to be at each end of the interval.

Now let us consider interpolation methods: in this case they are more complicated because we also have to take into account the extra piece of information for each interval, namely μ_i, the within-interval mean income.

The *split histogram density* is found as the following pair of constants in interval i

$$\phi_i(y) = \begin{cases} \dfrac{f_i}{a_{i+1} - a_i} \dfrac{a_{i+1} - \mu_i}{\mu_i - a_i}, & a_i \le y < \mu_i, \\ \dfrac{f_i}{a_{i+1} - a_i} \dfrac{\mu_i - a_i}{a_{i+1} - \mu_i}, & \mu_i \le y < a_{i+1}. \end{cases}$$

This method is extremely robust, and has been used, unless otherwise stated, to calculate the 'compromise' inequality values in Chapter 5.

The *log-linear interpolation* is given by

$$\phi_i(y) = \frac{c}{y^{\alpha+1}}, \; a_i \le y < a_{i+1}$$

where

$$c = \frac{\alpha f_i}{a_i^{-\alpha} - a_{i+1}^{-\alpha}}$$

[6] Strictly speaking we should use the term 'least upper bound' rather than 'maximum' since the observations in interval i are strictly less than (not less than or equal to) a_{i+1}.

and α is the root of the following equation:

$$\frac{\alpha}{\alpha-1}\frac{a_i^{1-\alpha}-a_{i+1}^{1-\alpha}}{a_i^{-\alpha}-a_{i+1}^{-\alpha}}=\mu_i$$

which may be solved by standard numerical methods. Notice the difference between this and the Pareto interpolation method used in the case where the interval means are unknown: here we compute *two* parameters for each interval, α and c which fixes the height of the density function at a_i, whereas in the other case c was automatically set to $a_i^{-\alpha}$. The last formula can be used to compute the value of α in the upper tail. Let $i = k$ and $a_{k+1} \to \infty$: then, if $\alpha > 1$, we have $a_{k+1}^{-\alpha} \to 0$ and $a_{k+1}^{1-\alpha} \to 0$. Hence we get:

$$\frac{\alpha}{\alpha-1}a_k=\mu_k$$

from which we may deduce that for the upper tail $\alpha = 1/[1 - a_k/\mu_k]$.

Warning: If the interval mean μ_i happens to be equal to, or very close to, the midpoint of the interval $\frac{1}{2}[a_i + a_{i+1}]$, then this interpolation formula collapses to that of the histogram density (see above) and $\alpha \to \infty$. It is advisable to test for this first rather than letting a numerical algorithm alert you to the presence of an effectively infinite root.

The *straight line density* is given by

$$\phi_i(y) = b + cy,\ \ a_i \le y < a_{i+1}$$

Table A.4. Source files for tables and figures

Table		Figure		Figure	
3.3	East-West	2.1	ET Income Distribution	5.1	IR income
3.4	East-West	2.2	ET Income Distribution	5.2	HBAI
5.2	IRS Income Distribution	2.3	ET Income Distribution	5.3	HBAI
5.4	IRS Income Distribution	2.4	ET Income Distribution	5.5	HBAI
5.5	Czechoslovakia	2.5	ET Income Distribution	5.6	HBAI
5.6	Taxes and Benefits	2.9	Earnings Quantiles	5.7	HBAI
5.7	Jiangsu	2.10	ET Income Distribution	5.8	IRS Income Distribution
A.3	Decomp	2.11	ET Income Distribution	5.9	IRS Income Distribution
		2.12	ET Income Distribution	5.10	IRS Income Distribution
		2.13	ET Income Distribution	5.11	IRS Income Distribution
		3.1	Atkinson SWF	5.12	IRS Income Distribution
		3.2	Atkinson SWF	5.14	IRS Income Distribution
		3.9	LIS comparison	5.15	IRS ineq
		3.10	LIS comparison	5.16	IRS ineq
		4.5	ET Income Distribution	5.17	Czechoslovakia
		4.10	NES	5.18	Czechoslovakia
		4.11	IR wealth	5.19	IRS Income Distribution
		4.12	Pareto example	5.20	IRS Income Distribution

where

$$b = \frac{12\mu_i - 6\left[a_{i+1} - a_i\right]}{\left[a_{i+1} - a_i\right]^3} f_i$$
$$c = \frac{f_i}{a_{i+1} - a_i} - \frac{1}{2}\left[a_{i+1} + a_i\right] b.$$

Warning: this formula has no intrinsic check that $\phi_i(y)$ does not become negative for some y in the interval. If you use it, therefore, you should always check that $\phi_i(a_i) \geq 0$ and that $\phi_i(a_{i+1}) \geq 0$.

A.8 Using the website

To get the best out of the examples and exercises in the book it is helpful to run through some of them yourself: the data files make it straightforward to do that. The files are accessed from the website at http://darp.lse.ac.uk/MI3 in Excel 2003 format.

You may find it helpful to be able to recreate the tables and figures presented in this book using the website: the required files are summarized in Table A.4. Individual files and their provenance are cited in detail in Appendix B.

APPENDIX B

Notes on Sources and Literature

This appendix describes the datasets that have been used for particular examples in each chapter, cites that which has been used for the discussion in the text, and provides a guide for further reading. In addition some more recondite supplementary points are mentioned. The arrangement follows the order of the material in the five chapters and Technical Appendix.

B.1 Chapter 1

For a general discussion of terminology and the approach to inequality you could go to Chapter 1 of Atkinson (1983), Cowell (2008b, 2008c), and Chapter 2 of Thurow (1975); reference may also be made to Bauer and Prest (1973). For a discussion of the relationship between income inequality and broader aspects of economic inequality see Sen (1997). For other surveys of inequality measurement issues see Jenkins and Van Kerm (2008) and for a more technical treatment, Cowell (2000) and Lambert (2001).

Inequality of What?

This key question is explicitly addressed in Sen (1980, 1992). The issue of the measurability of the income concept is taken up in a very readable contribution by Boulding (1975), as are several other basic questions about the meaning of the subject which were raised by the nine interpretations cited in the text Rein and Miller (1974). For an introduction to the formal analysis of measurability and comparability, see Sen and Foster (1997 [Sen 1973], pp. 43–46), and perhaps then try going on to Sen (1974) which, although harder, is clearly expounded. There are several studies which use an attribute other than income or wealth, and which provide interesting material for comparison: Jencks (1973) puts income inequality in the much wider context of social inequality; Addo (1976) considers international inequality in such things as school enrolment, calorie consumption, energy consumption and numbers of physicians; Alker (1965) discusses a quantification of voting power; Russet (1964) relates inequality in land ownership to political instability. The problem of the size of the cake depending on the way it is cut has long been implicitly recognized (for example, in the optimal taxation literature) but does not feature prominently in the works on inequality measurement. For a general treatment read Tobin (1970), reprinted in Phelps (1973). On this see also the Okun (1975, Chapter 4) illustration of 'leaky bucket' income transfers.

The issue of rescaling nominal incomes so as to make them comparable across families or households of different types—known in the jargon as 'equivalization'—and its impact upon measured inequality is discussed in Coulter *et al*. (1992a, 1992b)—see also page 191 below. Alternative approaches to measuring inequality in the presence of household heterogeneity are discussed in Cowell (1980), Ebert (1995, 2004), Glewwe (1991), Jenkins and O'Higgins (1989), Jorgenson and Slesnick (1990). The issues of measuring inequality when the underlying 'income' concept is something that is not cardinally measurable—for example measuring the inequality of health status—is discussed in Abul Naga and Yalcin (2008), Allison and Foster (2004).

Inequality Measurement, Justice, and Poverty

Although inequality is sharply distinct from mobility, inequality measures have been used as a simple device for characterizing income mobility—after covering the material in Chapter 3 you may find it interesting to check Shorrocks (1978). The application of inequality-measurement tools to the analysis of *inequality of opportunity* is addressed in Lefranc *et al*. (2008) and Pistolesi (2009).

On the desirability of equality *per se* see Broome (1988). Some related questions and references are as follows: Why care about inequality? (Milanovic 2007) Does it make people unhappy? (Alesina *et al*. 2004) Why measure inequality? Does it matter? (Bénabou 2000, Elliott 2009, Kaplow 2005) Do inequality measures really measure inequality? (Feldstein 1998)

On some of the classical principles of justice and equality, see Rees (1971), Chapter 7 and Wilson (1966). The idea of basing a model of social justice upon that of economic choice under risk is principally associated with the work of Harsanyi (1953, 1955)—see also Bosmans and Schokkaert (2004), Amiel *et al*. (2009), and Cowell and Schokkaert (2001). Hochman and Rodgers (1969) discuss concern for equality as a consumption externality. A notable landmark in modern thought is Rawls (1971) which, depending on the manner of interpretation of the principles of justice there expounded, implies most specific recommendations for comparing unequal allocations. Bowen (1970) introduces the concept of 'minimum practicable inequality', which incorporates the idea of special personal merit in determining a just allocation.

Stark's (1972) approach to an equality index is based on a head-count measure of poverty and is discussed in Chapter 2; Batchelder (1971, p. 30) discusses the 'poverty gap' approach to the measurement of poverty. The intuitive relationships between inequality and growth (or contraction) of income are set out in a novel approach by Temkin (1986) and are discussed further by Amiel and Cowell (1994b) and Fields (2007). The link between a measure that captures the depth of poverty and the Gini coefficient of inequality (see Chapter 2) was analysed in a seminal paper by Sen (1976a), which unfortunately the general reader will find quite hard; the huge literature which ensued is surveyed by Foster (1984), Hagenaars (1986), Ravallion (1994), Seidl (1988), and Zheng (1997). The relationship between inequality and poverty measures is discussed in some particularly useful papers by Thon (1981, 1983a). An appropriate approach to poverty may require a measure of economic status that is richer than income—see Anand and Sen (2000).

Inequality and the Social Structure

The question of the relationship between inequality in the whole population and inequality in subgroups of the population with reference to heterogeneity due to age is tackled in Paglin (1975) and in Cowell (1975). The rather technical paper of Champernowne (1974) explores the relationship between measures of inequality as a whole and measures that are related specifically to low incomes, to middle incomes, or to high incomes.

B.2 Chapter 2

The main examples here are from the tables in *Economic Trends*, November 1987 (based on the Inland Revenue's Survey of Personal Incomes augmented by information from the Family Expenditure Survey), which are reproduced on the website in the file 'ET income distribution': the income intervals used are those that were specified in the original tables. If you open this file you will also see exactly how to construct the histogram for yourself: it is well worth running through this as an exercise. The reason for using these data to illustrate the basic tools of inequality analysis is that they are based on reliable data sources, have an appropriate definition of income, and provide a good coverage of the income range providing some detail for both low incomes and high incomes. Unfortunately this useful series has not been maintained: we will get to the issue of what can be done with currently available datasets in Chapter 5.

The example in Fig. 2.9 is taken from the Annual Survey of Hours and Earnings (formerly the New Earnings Survey) data—see file 'Earnings quantiles' on the website. The reference to Plato as an early precursor of inequality measurement is to be found in Saunders (1970), pp. 214–15.

Diagrams

One often finds that technical apparatus or analytical results that have become associated with some famous name were introduced years before by someone else in some dusty journals, but were never popularized. So it is with Pen's Parade, set out in Pen (1974), which had been anticipated by Schutz (1951), and only rarely used since—cf. Budd (1970). As we have seen, the Parade is simply related to the cumulative frequency distribution if you turn the piece of paper over once you have drawn the diagram: for more about this concept, and also frequency distributions and histograms, consult a good statistics text such as Berry and Lindgren (1996), Casella and Berger (2002), or Freund and Perles (2007); for an extensive empirical application of Pen's parade see Jenkins and Cowell (1994a). The log-representation of the frequency distribution is referred to by Champernowne (1973, 1974) as the 'people curve'.

The Lorenz curve originally appeared in Lorenz (1905). Its convex shape (referred to on page 170) needs to be qualified in one very special case: where the mean of the thing that you are charting is itself negative—see page 170 in the Technical Appendix and Amiel *et al.* (1996). For a formal exposition of the Lorenz curve and proof of the assertions made in the text see Levine and Singer (1970) and Gastwirth (1971). Lorenz transformations are used to analyse the impact of income redistributive policies—see Arnold (1990), Fellman (2001), and the references in Question 7 on page 38. On using a transformation of the Lorenz curve to characterize income distributions see Aaberge (2007); see Fellman (1976) and Damjanovic (2005) for general results on the effect of transformations on the Lorenz curve. Lam (1986) discusses the behaviour of the Lorenz curve in the presence of population growth.

The relationship between the Lorenz curve and Pen's parade is also discussed by Alker (1970). The Lorenz curve has further been used as the basis for constructing a segregation index (Duncan and Duncan 1955; Cortese *et al.* 1976). For more on the Lorenz curve see also Blitz and Brittain (1964), Crew (1982), Hainsworth (1964), Koo *et al.* (1981), and Riese (1987).

Inequality Measures

The famous concentration ratio Gini (1912) also has an obscure precursor. Thirty-six years before Gini's work, Helmert (1876) discussed the ordinally equivalent measure known as Gini's mean difference—for further information see David (1968, 1981). Some care has to be taken when applying the Gini coefficient to indices of data where the number of individuals n is relatively small (Allison 1978, Jasso 1979): the problem is essentially whether the term n^2 or $n[n-1]$ should appear in the denominator of the definition—see the Technical Appendix page 155. A convenient alternative form of the standard definition is given in Dorfman (1979):

$$G = 1 - \frac{1}{\bar{y}} \int_0^\infty P(y)^2 dy \text{ where } P(y) = 1 - F(y).$$

For an exhaustive treatment of the Gini coefficient see Yitzhaki (1998).

The process of rediscovering old implements left lying around in the inequality analyst's toolshed continues unabated, so that often several labels and descriptions exist for essentially the same concept. Hence M, the relative mean deviation, used by Schutz (1951), Dalton (1920), and Kuznets (1959), reappears as the maximum equalization percentage, which is exactly $2M$ (United Nations Economic Commission for Europe 1957), and as the 'standard average difference' (Francis 1972). Eltetö and Frigyes (1968) produce three measures which are closely related to M, and Addo's 'systemic inequality measure' is essentially a function of these related measures; see also Kondor (1971). Gini-like inequality indices have been proposed by Basmann and Slottje (1987), Basu (1987), Berrebi and Silber (1987), Chakravarty (1988), and Yitzhaki (1983), and generalizations and extensions of the Gini are discussed by Barrett and Salles (1995), Bossert (1990), Donaldson and Weymark (1980), Kleiber and Kotz (2002), Moyes (2007), Weymark (1981), and Yaari (1988); see also Lin (1990). The Gini coefficient has also been used as the basis for regression analysis (Schechtman and Yitzhaki 1999) and for constructing indices of relative deprivation (Bishop *et al.* 1991, Chakravarty and Chakraborty 1984, Cowell 2008a, Yitzhaki 1979).

The properties of the more common *ad hoc* inequality measures are discussed at length in Atkinson (1970, pp. 252–57; 1983, pp. 53–58), Champernowne (1974, p. 805), Foster (1985), Jenkins (1991), and Sen and Foster (1997, pp. 24–36). Berrebi and Silber (1987) show that for all symmetric distributions $G < 0.5$: a necessary condition for $G > 0.5$ is that the distribution be skewed to the right. Chakravarty (2001) considers the use of the variance for the decomposition of inequality and Creedy (1977) and Foster and Ok (1999) discuss the properties of the variance of logarithms. The use of the skewness statistic was proposed by Young (1917); this and other statistical moments are considered further by Champernowne (1974); Butler and McDonald (1989) discuss the use of incomplete moments in inequality measurement (the ordinates of the Lorenz curve are simple examples of such incomplete moments—see the expressions on page 114). On the use of the moments of the Lorenz curve as an approach to characterizing inequality see Aaberge (2000). Further details on the use of moments may be found in texts such as Casella and Berger (2002) and Freund (2003). For more on the minimal majority coefficient (sometimes known as the Dauer–Kelsay index of malapportionment) see Alker and Russet (1964), Alker (1965), and Davis (1954, pp. 138–43). Some of the criticisms of Stark's high–low measure were originally raised in Polanyi and Wood (1974). Another such practical measure with a similar flavour is the Wiles (1974) semi-decile ratio: (minimum income of top 5 per cent)/(maximum income of bottom 5 per cent). Like R, M, 'minimal majority', 'equal shares', and 'high–low', this measure is insensitive to certain transfers, notably in the middle income ranges (you can redistribute income from a person at the sixth percentile to a person at the ninety-fourth without changing the semi-decile ratio). In my opinion this is a serious weakness, but Wiles recommended the semi-decile ratio as focusing on the essential feature of income inequality.

Rankings

Wiles and Markowski (1971) argued for a presentation of the facts about inequality that captures the whole distribution, since conventional inequality measures are a type of sophisticated average, and 'the average is a very uninformative concept' (1971, p. 351). In this respect[1] their appeal is similar in spirit to that of Sen and Foster (1997, Chapter 3) who suggest using the Lorenz curve to rank income distributions in a 'quasi-ordering'—in other words a ranking where the arrangement of some of the items is ambiguous. An alternative approach to this notion of ambiguity is the use of 'fuzzy' inequality discussed in Basu (1987) and Ok (1995).

The method of percentiles was used extensively by Lydall (1959) and Polanyi and Wood (1974); for recent applications to trends in the earnings distribution and the structure of wages see Atkinson (2007a) and Harvey and Bernstein (2003). The

[1] But only in this respect, since they reject the Lorenz curve as an 'inept choice', preferring to use histograms instead.

formalization of this approach as a 'comparative function' was suggested by Esberger and Malmquist (1972).

B.3 Chapter 3

The dataset used for the example on page 72 is given in the file 'LIS comparison' on the website. The artificial data used for the example in Tables 3.3 and 3.4 are in the file 'East West'.

Social Welfare Functions

The traditional view of social welfare functions is admirably and concisely expounded in Graaff (1957). One of the principal difficulties with these functions, as with the physical universe, is—where do they come from? On this technically difficult question, see Boadway and Bruce (1984, Chapter 5), Gaertner (2006), and Sen (1970, 1977). If you are sceptical about the practical usefulness of SWFs you may wish to note some other areas of applied economics where SWFs similar to those discussed in the text have been employed. They are introduced to derive interpersonal weights in applications of cost-benefit analysis, and in particular into project appraisal in developing countries—see Layard (1994), Little and Mirrlees (1974, Chapter 3), and Salanié (2000, Chapters 1, 2). Applications of SWF analysis include taxation design (Atkinson and Stiglitz 1980, Salanié 2003, Tuomala 1990), the evaluation of the effects of regional policy (Brown 1972, pp. 81–84), the impact of tax legislation (Mera 1969), and measures of national income and product (Sen 1976b).

As we noted when considering the basis for concern with inequality (pages 12 and 179) there is a connection between inequality and risk. This connection was made explicit in Atkinson's seminal article (Atkinson 1970) where the analogy between risk aversion and inequality aversion was also noted. However, can we just read across from private preferences on risk to social preferences on inequality? Amiel *et al.* (2008) show that the phenomenon of preference reversals may apply to social choice amongst distributions in a manner that is similar to that observed in personal choice amongst lotteries. However, experimental evidence suggests that individuals' attitude to inequality (their degree of inequality aversion ε) is sharply distinguished from their attitude to risk as reflected in their measured risk aversion—Kroll and Davidovitz (2003), Carlsson *et al.* (2005); estimates of inequality aversion have been made using classroom experiments (Amiel *et al.* 1999) and from representative sample survey evidence (Pirttilä and Uusitalo 2010). Estimates of inequality aversion across country (based on data from the World Bank's World Development Report) are discussed in Lambert *et al.* (2003); an interesting study on changes over time in attitudes to inequality in one country is to be found in Grosfeld and Senik (2010). If we were to interpret U as individual utility derived from income we would then interpret ε as the elasticity of marginal utility of income, and then one could perhaps estimate this elasticity directly from surveys of subjective happiness: this is done in Layard *et al.* (2008). Cowell and Gardiner (2000) survey methods for estimating this elasticity and HM Treasury (2003), page 94, provides a nice example of how such estimates can be used to underpin policy making. Ebert and Welsch (2009) examine the extent to which conventional inequality measures can be used to represent rankings of income distributions as reflected in survey data on subjective well-being.

The dominance criterion associated with quantile ranking (or Parade ranking) on page 33 and used in Theorem 1 is known as *first-order dominance*. The concept of *second-order dominance* refers to the ranking by generalized Lorenz curves used in Theorem 3 (the shares dominance used in Theorem 2 can be seen as a special case of second-order dominance for a set of distributions that all have the same mean). First-order dominance, principles of social welfare, and Theorem 1 are discussed in Saposnik (1981, 1983). The proofs of Theorems 2 and 4, using slightly more restrictive assumptions than necessary, were established in Atkinson (1970) who drew heavily on an analogy involving probability theory; versions of these two theorems requiring weaker assumptions but rather sophisticated mathematics are found in Dasgupta *et al.* (1973), Kolm (1969), and Sen and Foster (1997, pp. 49–58). In fact a lot of this work was anticipated by Hardy *et al.* (1934, 1952); Marshall and Olkin (1979) develop this approach and cover in detail relationships involving Lorenz curves, generalized Lorenz curves, and concave functions: readers who are happy with an undiluted mathematical presentation may find this the most useful single reference on this part of the subject (see also Arnold 1987).

Shorrocks (1983) introduced the concept of the generalized Lorenz curve and proved Theorem 3. As a neat logical extension of the idea Moyes (1989) showed that if you take income and transform it by some function ϕ (for example by using a tax function, as in the exercises on page 38) then the generalized Lorenz ordering of distributions is preserved if and only if ϕ is concave—see also page 181 above. Iritani and Kuga (1983) and Thistle (1989a, 1989b) discuss the interrelations between the Lorenz curve, the generalized Lorenz curve, and the distribution function. A further discussion and overview of these topics is to be found in Lambert (2001).

Where Lorenz curves intersect we know that unambiguous inequality comparisons cannot be made without some further restriction, such as imposing a specific inequality measure. However, it is also possible to use the concept of *third-order dominance* discussed in Atkinson (2008) and Davies and Hoy (1995). For corresponding results concerning generalized Lorenz curves see Dardanoni and Lambert (1988).

SWF-Based Inequality Measures

For the relationship of SWFs to inequality measurement, either in general form, or the specific type mentioned here, see Atkinson (1974, p. 63; 1983, pp. 56–57), Blackorby and Donaldson (1978, 1980), Champernowne and Cowell (1998), Dagum (1990), Dahlby (1987), Schwartz and Winship (1980), Sen (1992), and Sen and Foster (1997). The formal relationships between inequality and social welfare are discussed in Ebert (1987) and Dutta and Esteban (1992). For a general discussion of characterizing social welfare orderings in terms of degrees of inequality aversion see Bosmans (2007a). The association of the Rawls (1971) concept of justice (where society gives priority to improving the position of the least advantaged person) with a social welfare function exhibiting extreme inequality aversion is discussed in Arrow (1973), Hammond (1975), Sen (1974, pp. 395–98), and Bosmans (2007b). Lambert (1980) provides an extension of the Atkinson approach using utility shares rather than income shares. Inequality measures of the type first suggested by Dalton (1920) are further discussed by Aigner and Heins (1967) and Bentzel (1970). Kolm (1976a) suggests a measure based on an alternative to assumption 5, namely constant absolute inequality aversion (see page 165 above), so that as we increase a person's income y by one unit (pound, dollar, etc.) his welfare weight U' drops by κ% where κ is the constant amount of absolute inequality aversion: this approach leads to an inequality measure which does not satisfy the principle of scale independence. He also suggests a measure generalizing both this and Atkinson's measure. See also Bossert and Pfingsten (1990) and Yoshida (1991). The implications of using absolute rather than relative measures in analysing world income distribution are examined in Atkinson and Brandolini (2009a). The SWF method is interpreted by Meade (1976, Chapter 7 and appendix) in a more blatantly utilitarian fashion; his measure of 'proportionate distributional waste' is based on an estimation of individual utility functions. Ebert (1999) suggests a decomposable inequality measure that is a kind of 'inverse' of the Atkinson formula.

An ingenious way of extending dominance results to cases where individuals differ in their needs as well as their incomes is the concept known as sequential dominance (Atkinson and Bourguignon 1982, 1987). Further discussion of multidimensional aspects of inequality are to be found in Diez *et al*. (2007), Maasoumi (1986, 1989), Rietveld (1990), Savaglio (2006), and Weymark (2006); multidimensional inequality indices are discussed by Tsui (1995).

Inequality and Information Theory

The types of permissible 'distance' function, and their relationship with inequality are discussed in Cowell and Kuga (1981); Love and Wolfson (1976) refer to a similar concept as the 'strength-of-transfer effect'. The special relationship of the Herfindahl index and the Theil index to the strong principle of transfers was first examined in Kuga (1973). Krishnan (1981) (see also reply by Allison 1981) discusses the use of the Theil index as a measure of inequality interpreted in terms of average distance. Kuga (1980) shows the empirical similarities of the Theil index and the Gini coefficient, using simulations.

The Herfindahl (1950) index (closely related to c^2, or to Francis' standard average square difference) was originally suggested as a measure of concentration of individual firms—see Rosenbluth (1955). Several other inequality measures can be used in this way, notably other members of the E_θ family. The variable corresponding to income y may then be taken to be a firm's sales. However, one needs to be careful about this analogy since inequality among persons and concentration among firms are rather different concepts in several important ways: (i) the definition of a firm is often unclear, particularly for small production units; (ii) in measuring concentration we may not be very worried about the presence of tiny sales shares of many small firms, whereas in measuring inequality we may be considerably perturbed by tiny incomes received by a lot of people—see Hannah and Kay (1977). The relationship between the generalized entropy measures and the Lorenz curves is examined further in Rohde (2008) and the problem of capturing Lorenz orderings by a small number of inequality measures is considered by Shorrocks and Slottje (2002).

A reworking of the information theory analogy leads us to a closely related class of measures that satisfy the strong principle of transfers, but where the average of the distance of actual incomes from inequality is found by using population shares rather than income shares as weights, thus:

$$\frac{1}{\beta}\sum_{i=1}^{n}\frac{1}{n}\left[h(s_i)-h\left(\frac{1}{n}\right)\right]$$

—compare equation (3.6) on page 58. The special case $\beta = 0$ which becomes $\sum_{i=1}^{n}\log(\bar{y}/y_i)/n$ (the MLD index) was already discussed in Theil (1967, Chapter 4, p. 126 and appendix). An ordinally equivalent variant of Theil's index is used in Marfels (1971); see also Gehrig (1988). Jasso (1980) suggests that an appropriate measure of justice evaluation for an individual is log(actual share / just share). From this it is easy to see that you will get a generalized entropy measure with parameter $\theta = 0$ (equivalently Atkinson index with $\varepsilon = 1$).

Building an Inequality Measure

The social value judgements implied by the use of the various *ad hoc* inequality measures in Chapters 2 and 3 are analysed in Kondor (1975) who extends the discussion in the works of Atkinson, Champernowne, and Sen cited in the notes to Chapter 2. The question of what happens to inequality measures when all incomes are increased or when the population is replicated or merged with another population is discussed in Aboudi *et al.* (2010), Frosini (1985), Eichhorn and Gehrig (1982), Kolm (1976a, 1976b), and Salas (1998). Shorrocks and Foster (1987) examine the issue of an inequality measure's sensitivity to transfers in different parts of the distribution and Barrett and Salles (1998) discuss classes of inequality measures characterized by their behaviour under income transfers; Lambert and Lanza (2006) analyse the effect on inequality of changing isolated incomes. The Atkinson and generalized entropy families are examples of the application of the concept of the quasi-linear mean, which is discussed in Hardy *et al.* (1934, 1952) and Chew (1983).

Distributional principles that can be applied when households are not homogeneous are discussed in Ebert (2007) and Shorrocks (2004). The axiomatic approach to inequality measurement discussed on page 66 is not of course restricted to the generalized entropy family; with a suitable choice of axiom the approach can be extended to pretty well any inequality measure you like: for example see Thon's (1982) axiomatization of the Gini coefficient, or Foster (1983) on the Theil index. The validity of standard axioms when viewed in the light of people's perceptions of inequality is examined in Amiel and Cowell (1992, 1994a, 1999) and Cowell (1985a); for a discussion and survey of this type of approach see Amiel (1999) and Kampelmann (2009). The problematic cases highlighted in the examples on page 38 and 65 are based on Cowell (1988a). Ebert (1988) discusses the principles on which a generalized type of the relative mean deviation may be based and Ebert (2009) addresses ways of axiomatizing inequality that will be consistent with the apparently heterodox views illustrated in Question 4 on page 75.

The normative significance of decomposition is addressed by Kanbur (2006). Examples of approaches to inequality measurement that explicitly use criteria that may conflict with decomposability include basing social welfare on income satisfaction in terms of ranks in the distribution (Hempenius 1984), the use of income gaps (Preston 2007), the use of reference incomes to capture the idea of individual 'complaints' about income distribution (Cowell and Ebert 2004, Devooght 2003, Temkin 1993)—see also the discussion on page 195.

B.4 Chapter 4

The Idea of a Model

For an excellent coverage of the use of functional forms in modelling income distributions see Kleiber and Kotz (2003).

The Lognormal Distribution

Most texts on introductory statistical theory give a good account of the normal distribution—for example Berry and Lindgren (1996), Casella and Berger (2002), or Freund and Perles (2007). The standard reference on the lognormal and its properties (Aitchison and Brown 1957) also contains a succinct account of a simple type of random process theory of income development. A summary of several such theories can be found in Bronfenbrenner (1971) and in Brown (1976). On some of the properties of the lognormal Lorenz curve, see also Aitchison and Brown (1954).

The Pareto Distribution

An excellent introduction to Pareto's law is provided by Persky (1992). Pareto's original work can be consulted in Pareto (1896, 1965, 2001) or in Pareto (1972), which deals in passing with some of Pareto's late views on the law of income distribution; the development of Pareto's thought on inequality is discussed in Maccabelli (2009). Tawney (1964) argues forcefully against the strict interpretation of Pareto's law:

> It implies a misunderstanding of the nature of economic laws in general, and of Pareto's laws in particular, at which no one, it is probable, would have been more amused than Pareto himself, and which, indeed, he expressly repudiated in a subsequent work. It is to believe in economic Fundamentalism, with the New Testament left out, and the Books of Leviticus and Deuteronomy inflated to unconscionable proportions by the addition of new and appalling chapters. It is to dance naked, and roll on the ground, and cut oneself with knives, in honour of the mysteries of Mumbo Jumbo.

However, I do not find his assertion of Pareto's recantation convincing—see Pareto (1972); see also Pigou (1952, pp. 650 ff). Oversimplified interpretations of the law have persisted—Adams (1976) suggested a 'golden section' value of $\alpha = 2/[\sqrt{5} - 1]$ as a cure for inflation. Van der Wijk's (1939) law is partially discussed in Pen (1974, Chapter 6); in a sense it is a mirror image of the Bonferroni index (Bonferroni 1930) which is formed from an average of 'lower averages'—see Chakravarty (2007). Several of the other results in the text are formally proved in Chipman (1974). Nicholson

(1969, pp. 286–92) and Bowman (1945) give a simple account of the use of the Pareto diagram. The discussion of a random process model leading to a Pareto distribution is presented in Champernowne (1953, 1973) and the non-technical reader will find a simple summary in Pen (1971, 1974). The Pareto distribution as an equilibrium distribution of a wealth model is treated in Wold and Whittle (1957) and Champernowne and Cowell (1998), Chapter 10. A recent overview of Pareto-type distributions in economics and finance is provided by Gabaix (2008).

How Good Are the Functional Forms?

The example of earnings displayed on page 96 can be reproduced from file 'NES' on the website; the income example of page 88 is taken from the website file 'ET income distribution' again, and the wealth example on page 97 is based on file 'IR wealth'. Evidence on the suitability of the Pareto and lognormal distributions as approximations to actual distributions of earnings and of income can be found in the Royal Commission on the Distribution of Income and Wealth (1975, Appendix C; 1976, Appendix E).

In discussing the structure of wages in Copenhagen in 1953 Bjerke (1970) showed that the more homogenous the occupation, the more likely it would be that the distribution of earnings within it was lognormal. Weiss (1972) shows the satisfactory nature of the hypothesis of lognormality for graduate scientists' earnings in different areas of employment—particularly for those who were receiving more than $10,000 a year. Hill (1959) shows that merging normal distributions with different variances leads to 'leptokurtosis' (more of the population in the 'tails' than expected from a normal distribution)—a typical feature of the distribution of the logarithm of income. Other useful references on the lognormal distribution in practice are Fase (1970), Takahashi (1959), and Thatcher (1968). Evidence for lognormality is discussed in the case of India (Rajaraman 1975), Kenya (Kmietowicz and Webley 1975), Iraq (Kmietowicz 1984), and China (Kmietowicz and Ding 1993). Kmietowicz (1984) extends the idea of lognormality of the income distribution to bivariate lognormality of the joint distribution of income and household size. Battistin *et al.* (2009) demonstrate that consumption is 'more lognormal' than income and explain the economic reasons for this phenomenon.

Atkinson (1975) and Soltow (1975) produce evidence on the Pareto distribution and the distribution of wealth in the UK and the USA of the 1860s respectively. Klass *et al.* (2006) do this using the Forbes 400; Clementi and Gallegati (2005) examine Pareto's law for Germany, the UK, and the USA. For further evidence on the variability of Pareto's α in the USA, see Johnson (1937), a cautious supporter of Pareto. The Paretian property of the tail of the wealth distribution is also demonstrated admirably by the Swedish data examined by Steindl (1965) where α is about 1.5 to 1.7.

Some of the less orthodox applications of the Pareto curve are associated with 'Zipf's law' (Zipf 1949) which has been fruitfully applied to the distribution of city size (Nitsch 2005). Harold T. Davis, who has become famous for his theory of the French Revolution in terms of the value of Pareto's α under Louis XVI, produces further evidence on the Pareto law in terms of the distribution of wealth in the pre-Civil War southern states (wealth measured in terms of number of slaves) and of the distribution of income in England under William the Conqueror—see Davis (1954). For the latter example (based on the Domesday Book, 1086) the fit is surprisingly good, even though income is measured in 'acres'—i.e. that area of land which produces 72 bushels of wheat per annum. The population covered includes Cotters, Serfs, Villeins, Sokemen, Freemen, Tenants, Lords and Nobles, Abbots, Bishops, the Bishop of Bayeux, the Count of Mortain, and of course King William himself.

However, Davis's (1941) interpretation of these and other intrinsically interesting historical excursions as evidence for a 'mathematical theory of history' seems mildly bizarre: supposedly if α is too low or too high a revolution (from the left or the right, respectively) is induced. Although there is clearly a connection between extreme economic inequality and social unrest, seeking the mainspring of the development of civilization in the slope of a line on a double-log graph does not appear to be a rewarding or convincing exercise. There is a similar danger in misinterpreting a dynamic model such as of Champernowne (1953), in which a given pattern of social mobility always produces, eventually, a unique Pareto distribution, independent of the income distribution originally prevailing. Bernadelli (1944) postulates that a revolution having redistribution as an aim will prove futile because of such a mathematical process. Finding the logical and factual holes in this argument is left as an exercise for you.

Other Distributions

Finally, a mention of other functional forms that have been claimed to fit observed distributions more or less satisfactorily (see the Technical Appendix page 158). Some of these are generalizations of the lognormal or Pareto forms, such as the three-parameter lognormal (Metcalf 1969), or the generalized Pareto–Levy law, which attempts to take account of the lower tail (Arnold 1983, Mandelbrot 1960). Indeed, the formula we have described as the Pareto distribution was only one of many functions suggested by Pareto himself; it may thus be more accurately described as a 'Pareto type I' distribution (Hayakawa 1951, Quandt 1966). Champernowne (1952) provides a functional form which is close to the Pareto in the upper tail and which fits income distributions quite well; some technical details on this are discussed in Harrison (1974), with empirical evidence in Thatcher (1968)—see also Harrison (1979, 1981) and Sarabia *et al.* (1999).

Other suggestions are Beta distribution (Slottje 1984, Thurow 1970), the Gamma distribution (Salem and Mount 1974, McDonald and Jensen 1979), the sech^2-distribution, which is a special case of the Champernowne (1952) distribution (Fisk 1961), and the Yule distribution (Simon 1955, 1957; Simon and Bonini 1958); see also Campano (1987) and Ortega *et al.* (1991). Evans *et al.* (1993) and Kleiber and Kotz (2003) provide a very useful summary of the mathematical properties of many of the above. The Singh and Maddala (1976) distribution is discussed further in Cramer (1978), Cronin (1979), McDonald and Ransom (1979), Klonner (2000) (first-order dominance), and Wilfling and Krämer (1993) (Lorenz curves); cf also the closely related model by Dagum (1977). A generalized form of the Gamma distribution has been used by Esteban (1986), Kloek and Van Dijk (1978), and Taille (1981). An overview of several of these forms and their interrelationships is given in McDonald

(1984) as part of his discussion of the generalized Beta distribution of the second kind; on this distribution see also Bordley *et al*. (1996), Jenkins (2009), Majumder and Chakravarty (1990), McDonald and Mantrala (1995), Parker (1999), Sarabia *et al*. (2002), Wilfling (1996), and for an implementation with Chinese data see Chotikapanich *et al*. (2007). Alternative approaches to parameterizing the Lorenz curve are discussed in Basmann *et al*. (1990, 1991), and Kakwani and Podder (1973).

Other functional forms based on the exponential distribution are considered in Jasso and Kotz (2007). Some of the Lorenz properties noted for the lognormal and for the Pareto hold for more general functional forms—see Arnold *et al*. (1987) and Taguchi (1968).

B.5 Chapter 5

The Data

The UK data used for Fig. 5.1 are from Inland Revenue Statistics (see file 'IR income' on the website), and the US data in Table 5.1 from Internal Revenue Service, *Statistics of Income: Individual Tax Returns* (see file 'IRS Income Distribution'). The UK data used for Figs 5.2–5.7 are taken from the *Households Below Average Income* dataset (HBAI), which is now the principal data source for UK income distribution; summary charts and results are published in Department of Work and Pensions (2009).

For a general introduction to the problem of specifying an income or wealth variable see Atkinson (1983). The quality of the *administrative data* on personal incomes—derived from tax agencies or similar official sources—depends crucially on the type of tax administration and government statistical service for the country in question. On the one hand extremely comprehensive and detailed information about income and wealth (including cross-classifications of these two) is provided, for example, by the Swedish Central Statistical Bureau, on the basis of tax returns. On the other, one must overcome almost insuperable difficulties where the data presentation is messy, incomplete, or designedly misleading. An excellent example of the effort required here is provided by the geometric detective work of Wiles and Markowski (1971) and Wiles (1974) in handling Soviet earnings distribution data. Fortunately for the research worker, some government statistical services modify the raw tax data so as to improve the concept of income and to represent low incomes more satisfactorily. Stark (1972) gives a detailed account of the significance of refinements in the concepts of income using the UK data; for an exhaustive description of these data and their compilation see Stark in Atkinson *et al*. (1978) and for a quick summary, the Royal Commission on the Distribution of Income and Wealth (1975, Appendices F and H). For a discussion of the application of tax data to the analysis of top incomes see Atkinson (2007b). As for *survey data* on incomes, the HBAI in the UK draws on the Family Resources Survey and Family Expenditure Survey—see Frosztega (2000) for a detailed consideration of the underlying income concept: UK datasets are available from the UK Data Archive (http://www.data-archive.ac.uk). Summary charts and results for HBAI are published in Department of Work and Pensions (2009) and Brewer *et al*. (2008) provide a useful critique of this source. On the widely used Current Population Survey (CPS) data (see Question 3 in Chapter 2) in the USA see Burkhauser *et al.* (2004) and Welniak (2003). A general overview of inequality in the USA is provided in Bryan and Martinez (2008), Reynolds (2006), and Ryscavage (1999). On US data and the quality of sample surveys in particular it is worth checking the two classic references Budd and Radner (1975) and Ferber *et al.* (1969). Since publication of the first edition of this book, large comprehensive datasets of individual incomes have become much more readily available and it is impossible to do justice to them. One that deserves attention from the student of inequality are the early example based on data from the Internal Revenue Service and Survey of Economic Opportunity discussed in Okner (1972, 1975); an extremely useful source of internationally comparable microdata in incomes (and much else) is the Luxembourg Income Study (http://www.lis-project.org). An early and comprehensive source of US longitudinal data is the Panel Study of Income Dynamics (http://psidonline.isr.umich.edu/) described in Hill (1992); more recent European examples of longitudinal data are the British Household Panel Survey (http://www.iser.essex.ac.uk/survey/bhps) and the German Socio-Economic Panel (http://www.diw-berlin.de/de/soep). The classic reference on wealth data in the UK is Atkinson and Harrison (1978) and an important resource for international comparisons of wealth distributions is provided by the Luxembourg Wealth Study (Sierminska *et al.* 2006), OECD (2008) Chapter 10.

A good statement of principles concerning the income concept is provided by the Canberra Group (2001) report. Several writers have tried to combine theoretical sophistication with empirical ingenuity to extend income beyond the conventional definition. Notable among these are the income-cum-wealth analysis of Weisbrod and Hansen (1968), and the discussion by Morgan *et al.* (1962) of the inclusion of the value of leisure time as an income component. An important development for international comparisons is the Human Development Index which has income as just one component (Anand and Sen 2000); Fleurbaey and Gaulier (2009) in similar spirit propose a measure of living standards for international comparisons based on GDP per capita adjusted for personal and social characteristics including inequality; perhaps unsurprisingly the ranking of countries by this measure differs substantially from the conventional GDP ranking. Goodman and Oldfield (2004) contrast income inequality and expenditure inequality in the UK context. Stevenson and Wolfers (2008) examine the way inequality in happiness has changed in the USA.

In Morgan (1962), Morgan *et al.* (1962), and Prest and Stark (1967) the effect of family grouping on measured inequality is considered. For a fuller discussion of making allowance for income sharing within families and the resulting problem of constructing 'adult equivalence' scales, consult Abel-Smith and Bagley (1970); the internationally standard pragmatic approach to equivalization is the OECD scale (see, for example, Atkinson *et al.* 1995) although many UK studies use a scale based on McClements (1977); the idea that equivalence scales are revealed by community expenditures is examined in Olken (2005). The relationship between equivalence scales and measured inequality is examined in Buhmann *et al.* (1988), Coulter *et al.* (1992b), and Jenkins and Cowell (1994b): for a survey see Coulter *et al.* (1992a). The fact that averaging incomes over longer periods reduces the resulting inequality statistics emerges convincingly from the work of Hanna *et al.* (1948) and Benus and

Morgan (1975). The key reference on the theoretical and empirical importance of price changes on measured inequality is Muellbauer (1974); see also Crawford and Smith (2002), Hobijn and Lagakos (2005), and Slottje (1987). A further complication which needs to be noted from Metcalf (1969) is that the way in which price changes affect low-income households may depend on household composition; whether there is a male bread-winner present is particularly important. On the effect of non-response on income distribution and inequality refer to Korinek *et al.* (2006).

International comparisons of datasets on inequality and poverty are provided by Ferreira and Ravallion (2009); an early treatment of the problems of international comparison of data is found in Kuznets (1963, 1966) and Atkinson and Brandolini (2009b) provide an excellent general introduction to issues of data quality. Appropriate price adjustments to incomes can be especially problematic when making international comparisons. A standard approach is to use an index of *Purchasing Power Parity* (PPP) rather than converting incomes at nominal exchange rates. The issues involved in constructing PPP are treated in Heston *et al.* (2001); the method of imputation of PPP can have a substantial impact on estimates of between-country inequality and hence on the picture of global inequality; the topic is treated exhaustively in Anand and Segal (2008), Kravis *et al.* (1978a, 1978b), and Summers and Heston (1988, 1991). The issue of international comparability of income distribution data is one of the main reasons for the existence of the Luxembourg Income Study: see Smeeding *et al.* (1990) for an introduction and a selection of international comparative studies; Lorenz comparisons derived from this data source are in the website file 'LIS comparison'. On the use of data in OECD countries see Atkinson and Brandolini (2001) and on international comparisons of earnings and income inequality refer to Gottschalk and Smeeding (1997). Atkinson and Micklewright (1992) compare the income distributions in Eastern European economies in the process of transition. Other important international sources for studying inequality are Deininger and Squire (1996) and also UNU-WIDER (2005) which provides Gini indices drawn from a large number of national sources.

Beckerman and Bacon (1970) provide a novel approach to the measurement of world (i.e. inter-country) inequality by constructing their own index of 'income per head' for each country from the consumption of certain key commodities. Becker *et al.* (2005) examine the effect on trends in world inequality of trying to take into account people's quality of life.

Computation of the Inequality Measures

For detail on computation of point estimates of inequality go to the Technical Appendix. For an excellent general text on empirical methods including computation of inequality measures and other welfare indicators see Deaton (1997). For a discussion of how to adapt standard methodology to estimation problems in small areas with few observations see Tarozzi and Deaton (2009).

Decomposition techniques have been widely used to analyse spatial inequality (Shorrocks and Wan 2005) including China (Yu *et al.* 2007) and Euroland (Beblo and Knaus 2001) and for the world as a whole (Novotný 2007). For a systematic analysis of world inequality using (fully decomposable) generalized entropy indices see Berry *et al*. (1983a, 1983b), Bourguignon and Morrisson (2002), Sala-i-Martin (2006), Ram (1979, 1984, 1987, 1992), and Theil (1979b, 1989); Milanovic and Yitzhaki (2002) use the (not fully decomposable) Gini coefficient.

Appraising the Calculations

An overview of many of the statistical issues is to be found in Cowell (1999) and Nygård and Sandström (1981, 1985). If you are working with data presented in the conventional grouped form, then the key reference on the computation of the bounds J_{L}, J_{U} is Gastwirth (1975). Now, in addition to the bounds on inequality measures that we considered in the text, Gastwirth (1975) shows that if one may assume 'decreasing density' over a particular income interval (i.e. the frequency curve is sloping downwards to the right in the given income bracket) then one can calculate bounds J'_{L}, J'_{U} that are sharper—i.e. the bounds J'_{L}, J'_{U} lie within the range of inequality values (J_{L}, J_U) which we computed: the use of these refined bounds leaves the qualitative conclusions unchanged, though the proportional gap is reduced a little. The problem of finding such bounds is considered further in Cowell (1991). The special case of the Gini coefficient is treated in Gastwirth (1972) and McDonald and Ransom (1981); the properties of bounds for grouped data are further discussed in Gastwirth *et al.* (1986); Mehran (1975) shows that you can work out bounds on G simply from a set of sample observations on the Lorenz curve without having to know either mean income or the interval boundaries $a_1, a_2, \ldots, a_{k+1}$ and Hagerbaumer (1977) suggests the upper bound of the Gini coefficient as an inequality measure in its own right. In Gastwirth (1972, 1975) there are also some refined procedures for taking into account the open-ended interval forming the top income bracket—an awkward problem if the total amount of income in this interval is unknown. Ogwang (2003) discusses the problem of putting bounds on Gini coefficient when data are sparse. As an alternative to the methods discussed in the Technical Appendix (using the Pareto interpolation, or fitting Paretian density functions), the procedure for interpolating on Lorenz curves introduced by Gastwirth and Glauberman (1976) works quite well.

Cowell and Mehta (1982) investigate a variety of interpolation methods for grouped data and also investigate the robustness of inequality estimates under alternative grouping schemes. Aghevli and Mehran (1981) address the problem of optimal choice of the income interval boundaries used in grouping by considering the set of values $\{a_1, a_2, \ldots, a_k\}$ which will minimize the Gini coefficient; Davies and Shorrocks (1989) refine the technique for larger datasets.

For general information on the concept of the standard error see Berry and Lindgren (1996) or Casella and Berger (2002). On the sampling properties of inequality indices generally see Victoria-Feser (1999). Formulas for standard errors of specific inequality measures can be found in the following references: Kendall *et al.* (1994, sec. 10.5) (relative mean deviation, coefficient of variation), David (1968, 1981), Nair (1936) (Gini's mean difference), Gastwirth (1974a) (relative mean deviation), Aitchison and Brown (1957, p. 39) (variance of logarithms). For more detailed analysis of the Gini coefficient see Davidson (2009), Deltas (2003), Gastwirth *et al.* (1986), Giles (2004), Glasser (1962), Lomnicki (1952), Modarres and Gastwirth (2006), Nygård and Sandström (1989), Ogwang (2000, 2004), and Sandström *et al.* (1985, 1988). Allison (1978) discusses issues of estimation and testing based on microdata using the Gini coefficient, coefficient of variation, and Theil index. The statistical properties of the generalized entropy and related indices are discussed by Cowell (1989) and Thistle (1990). A thorough treatment of statistical testing of Lorenz curves is to be found in Beach and Davidson (1983), Beach and Kaliski (1986), Beach and Richmond (1985), and Davidson and Duclos (2000); for generalized Lorenz estimation refer to Bishop *et al.* (1989), and Bishop *et al.* (1989). See also Hasegawa and Kozumi (2003) for a Bayesian approach to Lorenz estimation and Schluter and Trede (2002) for problems of inference concerning the tails of Lorenz curves. For a treatment of the problem of estimation with complex survey design go to Biewen and Jenkins (2006), Cowell and Jenkins (2003), Binder and Kovacevic (1995), Bhattacharya (2007), and Kovacevic and Binder (1997). Cowell and Victoria-Feser (2003) treat the problem of estimation and inference when the distribution may be censored or truncated and Cowell and Victoria-Feser (2007, 2008) discuss the use of a Pareto tail in a 'semi-parametric' approach to estimation from individual data. The effect of truncation bias on inequality judgements is also discussed in Fichtenbaum and Shahidi (1988) and Bishop *et al.* (1994); the issue of whether 'top-coding' (censoring) of the CPS data makes a difference to the estimated trends in US income inequality is analysed in Burkhauser *et al.* (2008). So-called 'bootstrap' or resampling methods are dealt with by Biewen (2002), Davidson and Flachaire (2007), and Van Kerm (2002)—see also Davison and Hinkley (1997). For an interesting practical example of the problem of ranking distributions by inequality when you take into account sampling error, see Horrace *et al.* (2008).

On the robustness properties of measures in the presence of contamination or outliers see Cowell and Victoria-Feser (1996, 2002, 2006) and for the way inequality measures respond to extreme values go to Cowell and Flachaire (2007). Chesher and Schluter (2002) discuss more generally the way measurement errors affect the comparison of income distributions in welfare terms.

Fitting Functional Forms

Refer to Chotikapanich and Griffiths (2005) on the problem of how to choose a functional form for your data and to Maddala and Singh (1977) for a general discussion of estimation problems in fitting functional forms. Ogwang and Rao (2000) use hybrid Lorenz curves as a method of fit. If you want to estimate lognormal curves from grouped or ungrouped data, you should refer to Aitchison and Brown (1957, pp. 38–43, 51–54) first. Baxter (1980), Likes (1969), Malik (1970), and Quandt (1966) deal with the estimation of Pareto's α for ungrouped data. Now the ordinary least squares method, discussed by Quandt, despite its simplicity has some undesirable statistical properties, as explained in Aigner and Goldberger (1970). In the latter paper you will find a discussion of the difficult problem of providing maximum likelihood estimates for α from grouped data. The fact that in estimating a Pareto distribution a curve is fitted to cumulative series which may provide a misleadingly good fit was noted in Johnson (1937), while Champernowne (1956) provided the warning about uncritical use of the correlation coefficient as a criterion of suitability of fit. The suggestion of using inequality measures as an alternative basis for testing goodness-of-fit was first put forward by Gastwirth and Smith (1972), where they test the hypothesis of lognormality for United States IRS data; see also Gail and Gastwirth (1978b, 1978a). To test for lognormality one may examine whether the skewness and the kurtosis ('peakedness') of the observed distribution of the logarithms of incomes are significantly different from those of a normal distribution; for details consult Kendall *et al.* (1999). Hu (1995) discusses the estimation of Gini from grouped data using a variety of specific functional forms.

B.6 Technical Appendix

For a general technical introduction see Duclos and Araar (2006) and Cowell (2000); functional forms for distributions are discussed in Kleiber and Kotz (2003) and Evans *et al.* (1993).

The formulas in the Technical Appendix for the decomposition of inequality measures are standard—see Bourguignon (1979), Cowell (1980), Das and Parikh (1981, 1982), and Shorrocks (1980).

For a characterization of some general results in decomposition, see Bosmans and Cowell (2010), Chakravarty and Tyagarupananda (1998, 2000), Cowell (2006), Foster and Shneyerov (1999), Kakamu and Fukushige (2009), Toyoda (1980), Shorrocks (1984, 1988), and Zheng (2007). Establishing the main results typically requires the use of functional equations techniques, on which see Aczél (1966). For applications of the decomposition technique, see the references on spatial and world inequality in Chapter 5 (page 192) and also Anand (1983), Borooah *et al.* (1991), Ching (1991), Cowell (1984, 1985b), Frosini (1989), Glewwe (1986), Mookherjee and Shorrocks (1982), and Paul (1999).

Decomposition by income components is discussed by Satchell (1978), Shorrocks (1982), and Theil (1979a). Applications to Australia are to be found in Paul (2004), to New Zealand in Podder and Chatterjee (2002), and to UK in Jenkins (1995). The issues underlying an application of the *Shapley value* to decomposition analysis are examined in Sastre and Trannoy (2002). The use of partitions into subgroups as a method of 'explaining' the contributory factors to inequality is dealt with in Cowell and Jenkins (1995) and Elbers *et al.* (2008). Alternative pragmatic approaches to accounting for changes in inequality are provided by Bourguignon *et al.* (2008), Morduch and Sicular (2002), Fields (2003), and Jenkins and Van Kerm (2005); Cowell and Fiorio (2009) reconcile these alternatives with conventional decomposition analysis.

The relationship between decomposition of inequality and the measurement of poverty is examined in Cowell (1988b). As noted in Chapter 3 the decomposition of the Gini coefficient presents serious problems of interpretation. However, Pyatt (1976) tackles this by 'decomposing' the Gini coefficient into a component that represents within-group inequality, one that gives between-group inequality, and one that depends on the extent to which income distributions in different groups overlap one another. The properties of the Gini when 'decomposed' in this way are further discussed by Lambert and Aronson (1993), Lerman and Yitzhaki (1984, 1989), Yitzhaki and Lerman (1991), and Sastry and Kelkar (1994). Braulke (1983) examines the Gini decomposition on the assumption that within-group distributions are Paretian. Silber (1989) discusses the decomposition of the Gini coefficient by subgroups of the population (for the case of non-overlapping partitions) and by income components.

The data in Table A.3 is based on Howes and Lanjouw (1994) and Hussain *et al.* (1994). For recent decomposition analysis of China, see Kanbur and Zhang (1999, 2005), Lin *et al.* (2008), and Sicular *et al.* (2007).

Bibliography

Aaberge, R. (2000). Characterizations of Lorenz curves and income distributions. *Social Choice and Welfare* **17**, 639–53.

Aaberge, R. (2007). Gini's nuclear family. *Journal of Economic Inequality* **5**, 305–22.

Abel-Smith, B. and C. Bagley (1970). The problem of establishing equivalent standards of living for families of different composition. In P. Townsend (Ed.), *The Concept of Poverty*. London: Heinemann.

Aboudi, R., D. Thon, and S. W. Wallace (2010). Inequality comparisons when the populations differ in size. *Journal Of Economic Inequality* **8**, 47–70.

Abul Naga, R. H. and T. Yalcin (2008). Inequality measurement for ordered response health data. *Journal of Health Economics* **27**, 1614–25.

Aczél, J. (1966). *Lectures on Functional Equations and their Applications*. Number 9 in Mathematics in Science and Engineering. New York: Academic Press.

Adams, K. (1976). Pareto's answer to inflation. *New Scientist* **71**, 534–37.

Addo, H. (1976). Trends in international value-inequality 1969–1970: an empirical study. *Journal of Peace Research* **13**, 13–34.

Aghevli, B. B. and F. Mehran (1981). Optimal grouping of income distribution data. *Journal of the American Statistical Association* **76**, 22–26.

Aigner, D. J. and A. S. Goldberger (1970). Estimation of Pareto's law from grouped observations. *Journal of the American Statistical Association* **65**, 712–23.

Aigner, D. J. and A. J. Heins (1967). A social welfare view of the measurement of income equality. *Review of Income and Wealth* **13**(3), 12–25.

Aitchison, J. and J. A. C. Brown (1954). On the criteria for descriptions of income distribution. *Metroeconomica* **6**, 88–107.

Aitchison, J. and J. A. C. Brown (1957). *The Lognormal Distribution*. London: Cambridge University Press.

Alesina, A., R. Di Tella, and R. MacCulloch (2004). Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics* **88**, 2009–42.

Alker, H. R. (1965). *Mathematics and Politics*. New York: Macmillan.

Alker, H. R. J. (1970). Measuring inequality. In E. R. Tufte (Ed.), *The Quantitative Analysis of Social Problems*. Reading, Massachusetts: Addison-Wesley.

Alker, H. R. J. and B. Russet (1964). On measuring inequality. *Behavioral Science* **9**, 207–18.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review* **43**, 865–80.

Allison, P. D. (1981). Inequality measures for nominal data. *American Sociological Review* **46**, 371–72. Reply to Krishnan.

Allison, R. A. and J. E. Foster (2004). Measuring health inequality using qualitative data. *Journal of Health Economics* **23**, 505–52.

Amiel, Y. (1999). The subjective approach to the measurement of income inequality. In J. Silber (Ed.), *Handbook on Income Inequality Measurement*. Dewenter: Kluwer.

Amiel, Y. and F. A. Cowell (1992). Measurement of income inequality: experimental test by questionnaire. *Journal of Public Economics* **47**, 3–26.

Amiel, Y. and F. A. Cowell (1994a). Income inequality and social welfare. In J. Creedy (Ed.), *Taxation, Poverty and Income Distribution*, pp. 193–219. Edward Elgar.

Amiel, Y. and F. A. Cowell (1994b). Inequality changes and income growth. In W. Eichhorn (Ed.), *Models and Measurement of Welfare and Inequality*, pp. 3–26. Berlin, Heidelberg: Springer-Verlag.

Amiel, Y. and F. A. Cowell (1998). Distributional orderings and the transfer principle: a re-examination. *Research on Economic Inequality* **8**, 195–215.

Amiel, Y. and F. A. Cowell (1999). *Thinking about Inequality*. Cambridge: Cambridge University Press.

Amiel, Y., F. A. Cowell, L. Davidovitz, and A. Polovin (2008). Preference reversals and the analysis of income distributions. *Social Choice and Welfare* **30**, 305–30.

Amiel, Y., F. A. Cowell, and W. Gaertner (2010). To be or not to be involved: a questionnaire-experimental view on Harsanyi's utilitarian ethics. *Social Choice and Welfare*, forthcoming.

Amiel, Y., F. A. Cowell, and A. Polovin (1996). Inequality amongst the kibbutzim. *Economica* **63**, S63–S85.

Amiel, Y., J. Creedy, and S. Hurn (1999). Attitudes towards inequality. *The Scandinavian Journal of Economics* **101**, 83–96.

Anand, S. (1983). *Inequality and Poverty in Malaysia*. London: Oxford University Press.

Anand, S. and P. Segal (2008). What do we know about global income inequality? *Journal of Economic Literature* **46**, 57–94.

Anand, S. and A. K. Sen (2000). The income component of the human development index. *Journal of Human Development* **1**, 83–106.

Arnold, B. C. (1983). *Pareto Distributions*. Fairland, MD: International Cooperative Publishing House.

Arnold, B. C. (1987). *Majorization and the Lorenz Order: A Brief Introduction*. Heidelberg: Springer-Verlag.

Arnold, B. C. (1990). The Lorenz order and the effects of taxation policies. *Bulletin of Economic Research* **42**, 249–64.

Arnold, B. C., P. L. Brockett, C. A. Robertson, and B. Y. Shu (1987). Generating Ordered families of Lorenz curves by strongly unimodal distributions. *Journal of Business and Economic Statistics* **5**, 305–08.

Arrow, K. J. (1973). Some ordinalist-utilitarian notes on Rawls' theory of justice. *Journal of Philosophy* **70**, 245–63.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory* **2**, 244–263.

Atkinson, A. B. (1974). Poverty and income inequality in Britain. In D. Wedderburn (Ed.), *Poverty, Inequality and The Class Structure*. London: Cambridge University Press.

Atkinson, A. B. (1975). The distribution of wealth in Britain in the 1960s—the estate duty method re-examined. In J. D. Smith (Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Atkinson, A. B. (1983). *The Economics of Inequality* (Second edn.). Oxford: Clarendon Press.

Atkinson, A. B. (2007a). The long run earnings distribution in five countries: "remarkable stability," u, v or w? *Review of Income and Wealth* **53**, 1–24.

Atkinson, A. B. (2007b). Measuring top incomes: methodological issues. In A. B. Atkinson and T. Piketty (Eds.), *Top Incomes Over the 20th Century: A Contrast Between Continental European and English-speaking Countries*, Chapter 2, pp. 18–42. Oxford: Oxford University Press.

Atkinson, A. B. (2008). More on the measurement of inequality. *Journal of Economic Inequality* **6**, 277–83.

Atkinson, A. B. and F. Bourguignon (1982). The comparison of multi-dimensional distributions of economic status. *Review of Economic Studies* **49**, 183–201.

Atkinson, A. B. and F. Bourguignon (1987). Income distribution and differences in needs. In G. R. Feiwel (Ed.), *Arrow and the Foundations of the Theory of Economic Policy*, Chapter 12, pp. 350–70. New York: Macmillan.

Atkinson, A. B. and A. Brandolini (2001). Promise and pitfalls in the use of secondary data-sets: income inequality in OECD countries as a case study. *Journal of Economic Literature* **39**, 771–99.

Atkinson, A. B. and A. Brandolini (2009a, January). On analysing the world distribution of income. Bank of Italy Temi di Discussione (Working Paper) 701, Bank of Italy, Bank of Italy Economic Research Department via Nazionale 91-00184 Rome, Italy.

Atkinson, A. B. and A. Brandolini (2009b). On data: a case study of the evolution of income inequality across time and across countries. *Cambridge Journal of Economics* **33**, 381–404.

Atkinson, A. B., J. P. F. Gordon, and A. J. Harrison (1989). Trends in the shares of top wealth-holders in Britain 1923–1981. *Oxford Bulletin of Economics and Statistics* **51**, 315–32.

Atkinson, A. B. and A. J. Harrison (1978). *Distribution of Personal Wealth in Britain*. Cambridge: Cambridge University Press.

Atkinson, A. B., A. J. Harrison, and T. Stark (1978). *Wealth and Personal Incomes*. London: Pergamon Press Ltd on behalf of The Royal Statistical Society and the Social Science Research Council.

Atkinson, A. B. and J. Micklewright (1992). *Economic Transformation in Eastern Europe and the Distribution of Income*. Cambridge: Cambridge University Press.

Atkinson, A. B. and T. Piketty (Eds.) (2007). *Top Incomes over the 20th Century: A Contrast Between Continental European and English-speaking countries*. Oxford: Oxford University Press.

Atkinson, A. B., L. Rainwater, and T. M. Smeeding (1995). *Income Distribution in OECD Countries: The Evidence from the Luxembourg Income Study*. Number 18 in Social Policy Studies. OECD, Paris.

Atkinson, A. B. and J. E. Stiglitz (1980). *Lectures on Public Economics*. Basingstoke: McGraw Hill.

Barrett, R. and M. Salles (1995). On a generalisation of the Gini coefficient. *Mathematical Social Sciences* **30**, 235–44.

Barrett, R. and M. Salles (1998). On three classes of differentiable inequality measures. *International Economic Review* **39**, 611–21.

Basmann, R. L., K. J. Hayes, J. D. Johnson, and D. J. Slottje (1990). A general functional form for approximating the Lorenz curve. *Journal of Econometrics* **43**, 77–90.

Basmann, R. L., K. J. Hayes, and D. J. Slottje (1991). The Lorenz curve and the mobility function. *Economics Letters* **35**, 105–111.

Basmann, R. L. and D. J. Slottje (1987). A new index of income inequality—the B measure. *Economics Letters* **24**, 385–89.

Basu, K. (1987). Axioms for a fuzzy measure of inequality. *Mathematical Social Sciences* **14**(12), 275–88.

Batchelder, A. B. (1971). *The Economics of Poverty*. New York: Wiley.

Battistin, E., R. Blundell, and A. Lewbel (2009). Why is consumption more log normal than income? Gibrat's Law revisited. *Journal of Political Economy* **117**(6), 1140–54.

Bauer, P. T. and A. R. Prest (1973). Income differences and inequalities. *Moorgate and Wall Street* **Autumn**, 22–43.

Baxter, M. A. (1980). Minimum-variance unbiased estimation of the parameters of the Pareto distribution. *Metrika* **27**, 133–38.

Beach, C. M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies* **50**, 723–35.

Beach, C. M. and S. F. Kaliski (1986). Lorenz curve inference with sample weights: an application to the distribution of unemployment experience. *Applied Statistics* **35**(1), 38–45.

Beach, C. M. and J. Richmond (1985). Joint confidence intervals for income shares and Lorenz curves. *International Economic Review* **26**(6), 439–50.

Beblo, M. and T. Knaus (2001). Measuring income inequality in Euroland. *Review of Income and Wealth* **47**, 301–20.

Becker, G. S., T. J. Philipson, and R. R. Soares (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review* **95**, 277–91.

Beckerman, W. and R. Bacon (1970). The international distribution of incomes. In P. Streeten (Ed.), *Unfashionable Economics. Essays in Honour of Lord Balogh*. London: Weidenfeld and Nicolson.

Bénabou, R. (2000). Unequal societies: income distribution and the social contract. *American Economic Review* **90**, 96–129.

Bentzel, R. (1970). The social significance of income distribution statistics. *Review of Income and Wealth* **16**, 253–64.

Benus, J. and J. N. Morgan (1975). Time period, unit of analysis and income concept in the analysis of income distribution. In J. D. Smith (Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Bernadelli, H. (1944). The stability of the income distribution. *Sankhya* **6**, 351–62.

Berrebi, Z. M. and J. Silber (1987). Dispersion, asymmetry and the Gini index of inequality. *International Economic Review* **28**(6), 331–38.

Berry, A., F. Bourguignon, and C. Morrisson (1983a). Changes in the world distributions of income between 1950 and 1977. *The Economic Journal* **93**, 331–50.

Berry, A., F. Bourguignon, and C. Morrisson (1983b). The level of world inequality: how much can one say? *Review of Income and Wealth* **29**, 217–43.

Berry, D. A. and B. W. Lindgren (1996). *Statistics, Theory and Methods* (Second edn.). Belmont, CA: Duxbury Press.

Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics* **137**, 674–707.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics* **108**, 317–42.

Biewen, M. and S. P. Jenkins (2006). Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. *Oxford Bulletin of Economics and Statistics* **68**, 371–83.

Binder, D. A. and M. S. Kovacevic (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology* **21**(2) 137–45.

Bishop, J. A., S. Chakraborti, and P. D. Thistle (1989). Asymptotically distribution-free statistical inference for generalized Lorenz curves. *Review of Economics and Statistics* **71**(11), 725–27.

Bishop, J. A., S. Chakraborti, and P. D. Thistle (1991). Relative deprivation and economic welfare: a statistical investigation with Gini-based welfare indices. *Scandinavian Journal of Economics* **93**, 421–37.

Bishop, J. A., J.-R. Chiou, and J. P. Formby (1994). Truncation bias and the ordinal evaluation of income inequality. *Journal of Business and Economic Statistics* **12**, 123–27.

Bishop, J. A., J. P. Formby, and W. J. Smith (1991). International comparisons of income inequality: tests for Lorenz dominance across nine countries. *Economica* **58**, 461–77.

Bishop, J. A., J. P. Formby, and P. D. Thistle (1989). Statistical inference, income distributions and social welfare. In D. J. Slottje (Ed.), *Research on Economic Inequality I*. JAI Press.

Bjerke, K. (1970). Income and wage distributions part I: a survey of the literature. *Review of Income and Wealth* **16**, 235–52.

Blackorby, C. and D. Donaldson (1978). Measures of relative equality and their meaning in terms of social welfare. *Journal of Economic Theory* **18**, 59–80.

Blackorby, C. and D. Donaldson (1980). A theoretical treatment of indices of absolute inequality. *International Economic Review* **21**, 107–36.

Blitz, R. C. and J. A. Brittain (1964). An extension of the Lorenz diagram to the correlation of two variables. *Metron* **23**, 37–143.

Boadway, R. W. and N. Bruce (1984). *Welfare Economics*. Oxford: Basil Blackwell.

Bonferroni, C. (1930). *Elemente di Statistica Generale*. Firenze: Libreria Seber.

Bordley, R. F., J. B. McDonald, and A. Mantrala (1996). Something new, something old: parametric models for the size distribution of income. *Journal of Income Distribution* **6**, 91–103.

Borooah, V. K., P. P. L. McGregor, and P. M. McKee (1991). *Regional Income Inequality and Poverty in the United Kingdom*. Aldershot: Dartmouth Publishing Co.

Bosmans, K. (2007a). Comparing degrees of inequality aversion. *Social Choice and Welfare* **29**, 405–28.

Bosmans, K. (2007b). Extreme inequality aversion without separability. *Economic Theory* **32**(3), 589–94.

Bosmans, K. and F. A. Cowell (2010). The class of absolute decomposable inequality measures. *Economics Letters* **109**, 154–56.

Bosmans, K. and E. Schokkaert (2004). Social welfare, the veil of ignorance and purely individual risk: an empirical examination. *Research on Economic Inequality* **11**, 85–114.

Bossert, W. (1990). An axiomatization of the single series Ginis. *Journal of Economic Theory* **50**, 89–92.

Bossert, W. and A. Pfingsten (1990). Intermediate inequality: concepts, indices and welfare implications. *Mathematical Social Science* **19**, 117–34.

Boulding, K. E. (1975). The pursuit of equality. In J. D. Smith (Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica* **47**, 901–20.

Bourguignon, F., F. H. G. Ferreira, and P. G. Leite (2008). Beyond Oaxaca-Blinder: accounting for differences in household income distributions. *Journal of Economic Inequality* **6**, 117–48.

Bourguignon, F. and C. Morrisson (2002). Inequality among world citizens: 1820-1992. *American Economic Review* **92**, 727–44.

Bowen, I. (1970). *Acceptable Inequalities*. London: Allen and Unwin.

Bowman, M. J. (1945). A graphical analysis of personal income distribution in the United States. *American Economic Review* **35**, 607–28.

Braulke, M. (1983). An approximation to the Gini coefficient for population based on sparse information for sub-groups. *Journal of Development Economics* **2**(12), 75–81.

Brewer, M., A. Muriel, D. Phillips, and L. Sibieta (2008). Poverty and inequality in the UK: 2008. IFS Commentary 105, The Institute for Fiscal Studies.

Bronfenbrenner, M. (1971). *Income Distribution Theory*. London: Macmillan.

Broome, J. (1988). What's the good of equality? In J. Hey (Ed.), *Current Issues in Microeconomics*. Basingstoke, Hampshire: Macmillan.

Brown, A. J. (1972). *The Framework of Regional Economics in The United Kingdom*. London: Cambridge University Press.

Brown, J. A. C. (1976). The mathematical and statistical theory of income distribution. In A. B. Atkinson (Ed.), *The Personal Distribution of Income*. Allen and Unwin, London.

Bryan, K. A. and L. Martinez (2008). On the evolution of income inequality in the United States. *Economic Quarterly* **94**, 97–120.

Budd, E. C. (1970). Postwar changes in the size distribution of income in the US. *American Economic Review, Papers and Proceedings* **60**, 247–60.

Budd, E. C. and D. B. Radner (1975). The Bureau of Economic Analysis and current population survey size distributions: some comparisons for 1964. In J. D. Smith

(Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Buhmann, B., L. Rainwater, G. Schmaus, and T. Smeeding (1988). Equivalence scales, well-being, inequality and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database. *Review of Income and Wealth* **34**, 115–42.

Burkhauser, R. V., J. Butler, S. Feng, and A. J. Houtenville (2004). Long term trends in earnings inequality: what the CPS can tell us. *Economic Letters* **82**, 295–99.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2008, August). Estimating trends in US income inequality using the current population survey: the importance of controlling for censoring. Working Paper 14247, National Bureau of Economic Research.

Butler, R. J. and J. B. McDonald (1989). Using incomplete moments to measure inequality. *Journal of Econometrics* **42**, 109–19.

Campano, F. (1987). A fresh look at Champernowne's five-parameter formula. *Économie Appliquée* **40**, 161–75.

Canberra Group (2001). Expert group on household income statistics: final report and recommendations. Technical report, Ottawa.

Carlsson, F., D. Daruvala, and O. Johansson-Stenman (2005). Are people inequality averse or just risk averse? *Economica* **72**, 375–96.

Casella, G. and R. L. Berger (2002). *Statistical inference*. Thomson Press.

Chakravarty, S. R. (1988). Extended Gini indices of inequality. *International Economic Review* **29**(2), 147–56.

Chakravarty, S. R. (2001). The variance as a subgroup decomposable measure of inequality. *Social Indicators Research* **53**(1), 79–95.

Chakravarty, S. R. (2007). A deprivation-based axiomatic characterization of the absolute Bonferroni index of inequality. *Journal of Economic Inequality* **5**, 339–51.

Chakravarty, S. R. and A. B. Chakraborty (1984). On indices of relative deprivation. *Economics Letters* **14**, 283–87.

Chakravarty, S. R. and S. Tyagarupananda (1998). The subgroup decomposable absolute indices of inequality. In S. R. Chakravarty, D. Coondoo, and R. Mukherjee (Eds.), *Quantitative Economics: Theory and Practice*, Chapter 11, pp. 247–57. New Delhi: Allied Publishers Limited.

Chakravarty, S. R. and S. Tyagarupananda (2000). The subgroup decomposable absolute and intermediate indices of inequality. Mimeo, Indian Statistical Institute.

Champernowne, D. G. (1952). The graduation of income distribution. *Econometrica* **20**, 591–615.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal* **63**, 318–51.

Champernowne, D. G. (1956). Comment on the paper by P. E. Hart and S. J. Prais. *Journal of The Royal Statistical Society* A **119**, 181–83.

Champernowne, D. G. (1973). *The Distribution of Income Between Persons*. Cambridge: Cambridge University Press.

Champernowne, D. G. (1974). A comparison of measures of income distribution. *The Economic Journal* **84**, 787–816.

Champernowne, D. G. and F. A. Cowell (1998). *Economic Inequality and Income Distribution*. Cambridge: Cambridge University Press.

Chesher, A. and C. Schluter (2002). Measurement error and inequality measurement. *Review of Economic Studies* **69**, 357–78.

Chew, S.-H. (1983). A generalization of the quasi-linear mean with application to the measurement of income inequality. *Econometrica* **51**, 1065–92.

Ching, P. (1991). Size distribution of income in the Phillipines. In T. Mizoguchi (Ed.), *Making Economies More Efficient and More Equitable: Factors Determining Income Distribution*. Tokyo: Kinokuniya.

Chipman, J. S. (1974). The welfare ranking of Pareto distributions. *Journal of Economic Theory* **9**, 275–82.

Chotikapanich, D. and W. Griffiths (2005). Averaging Lorenz curves. *Journal of Economic Inequality* **3**, 1–19.

Chotikapanich, D., D. S. Prasada Rao, and K. K. Tang (2007). Estimating income inequality in China using grouped data and the generalized beta distribution. *Review of Income and Wealth* **53**, 127–47.

Clementi, F. and M. Gallegati (2005). Pareto's law of income distribution: evidence for Germany, the United Kingdom, and the United States. In A. Chatterjee, S. Yarlagadda, and B. K. Chakrabarti (Eds.), *Econophysics of Wealth Distributions*. Berlin: Springer.

Cortese, C. F., R. F. Falk, and J. K. Cohen (1976). Further consideration on the methodological analysis of segregation indices. *American Sociological Review* **41**, 630–37.

Coulter, F. A. E., F. A. Cowell, and S. P. Jenkins (1992a). Differences in needs and assessement of income distributions. *Bulletin of Economic Research* **44**, 77–124.

Coulter, F. A. E., F. A. Cowell, and S. P. Jenkins (1992b). Equivalence scale relativities and the extent of inequality and poverty. *The Economic Journal* **102**, 1067–82.

Cowell, F. A. (1975). Income tax incidence in an ageing population. *European Economic Review* **6**, 343–67.

Cowell, F. A. (1977). *Measuring Inequality* (First edn.). Oxford: Phillip Allan.

Cowell, F. A. (1980). On the structure of additive inequality measures. *Review of Economic Studies* **47**, 521–31.

Cowell, F. A. (1984). The structure of American income inequality. *Review of Income and Wealth* **30**, 351–75.

Cowell, F. A. (1985a). 'A fair suck of the sauce bottle'—or what do you mean by inequality? *Economic Record* **6**, 567–79.

Cowell, F. A. (1985b). Multilevel decomposition of Theil's index of inequality. *Review of Income and Wealth* **31**, 201–05.

Cowell, F. A. (1988a). Inequality decomposition—three bad measures. *Bulletin of Economic Research* **40**, 309–12.

Cowell, F. A. (1988b). Poverty measures, inequality and decomposability. In D. Bös, M. Rose, and C. Seidl (Eds.), *Welfare and Efficiency in Public Economics*, pp. 149–66. Berlin, Heidelberg: Springer-Verlag.

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics* **42**, 27–41.

Cowell, F. A. (1991). Grouping bounds for inequality measures under alternative informational assumptions. *Journal of Econometrics* **48**, 1–14.

Cowell, F. A. (1999). Estimation of inequality indices. In J. Silber (Ed.), *Handbook on Income Inequality Measurement*. Dewenter: Kluwer.

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Chapter 2, pp. 87–166. Amsterdam: North Holland.

Cowell, F. A. (2006). Theil, inequality indices and decomposition. *Reseach on Economic Inequality* **13**, 345–60.

Cowell, F. A. (2008a). Gini, deprivation and complaints. In G. Betti and A. Lemmi (Eds.), *Advances on Income Inequality and Concentration Measures*. London: Routledge.

Cowell, F. A. (2008b). Income distribution and inequality. In J. B. Davis and W. Dolfsma (Eds.), *The Elgar Companion To Social Economics*, Chapter 13. Cheltenham: Edward Elgar.

Cowell, F. A. (2008c). Inequality: Measurement. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd edn.). Basingstoke: Palgrave Macmillan.

Cowell, F. A. and U. Ebert (2004). Complaints and inequality. *Social Choice and Welfare* **23**, 71–89.

Cowell, F. A. and C. Fiorio (2009). Inequality decompositions—a reconciliation. Distributional Analysis Discussion Paper 99, STICERD, London School of Economics, London WC2A 2AE.

Cowell, F. A. and E. Flachaire (2007). Income distribution and inequality measurement: the problem of extreme values. *Journal of Econometrics* **141**, 1044–72.

Cowell, F. A. and K. A. Gardiner (2000). Welfare weights. OFT Economic Research Paper 202, Office of Fair Trading, Salisbury Square, London.

Cowell, F. A. and S. P. Jenkins (1995). How much inequality can we explain? A methodology and an application to the USA. *The Economic Journal* **105**, 421–30.

Cowell, F. A. and S. P. Jenkins (2003). Estimating welfare indices: household sample design. *Research on Economic Inequality* **9**, 147–72.

Cowell, F. A. and K. Kuga (1981). Inequality measurement: an axiomatic approach. *European Economic Review* **15**, 287–305.

Cowell, F. A. and F. Mehta (1982). The estimation and interpolation of inequality measures. *Review of Economic Studies* **49**, 273–90.

Cowell, F. A. and E. Schokkaert (2001). Risk perceptions and distributional judgments. *European Economic Review* **42**, 941–52.

Cowell, F. A. and M.-P. Victoria-Feser (1996). Robustness properties of inequality measures. *Econometrica* **64**, 77–101.

Cowell, F. A. and M.-P. Victoria-Feser (2002). Welfare rankings in the presence of contaminated data. *Econometrica* **70**, 1221–33.

Cowell, F. A. and M.-P. Victoria-Feser (2003). Distribution-free inference for welfare indices under complete and incomplete information. *Journal of Economic Inequality* **1**, 191–219.

Cowell, F. A. and M.-P. Victoria-Feser (2006). Distributional dominance with trimmed data. *Journal of Business and Economics Statistics* **24**, 291–300.

Cowell, F. A. and M.-P. Victoria-Feser (2007). Robust stochastic dominance: a semi-parametric approach. *Journal of Economic Inequality* **5**, 21–37.

Cowell, F. A. and M.-P. Victoria-Feser (2008). Modelling Lorenz curves: robust and semi-parametric issues. In D. Chotikapanich (Ed.), *Modeling Income Distributions and Lorenz Curves*, Chapter 13, pp. 241–55. New York: Springer.

Cramer, J. S. (1978). A function for the size distribution of income: comment. *Econometrica* **46**, 459–60.

Crawford, I. and Z. Smith (2002). Distributional aspects of inflation. Commentary 90, Institute for Fiscal Studies.

Creedy, J. (1977). The principle of transfers and the variance of logarithms. *Oxford Bulletin of Economics and Statistics* **39**, 153–58.

Crew, E. L. (1982). Double cumulative and Lorenz curves in weather modification. *Journal of Applied Meteorology* **21**, 1063–70.

Cronin, D. C. (1979). A function for the size distribution of income: a further comment. *Econometrica* **47**, 773–74.

Dagum, C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquée* **30**, 413–36.

Dagum, C. (1990). On the relationship between income inequality measures and social welfare functions. *Journal of Econometrics* **43**, 91–102.

Dahlby, B. G. (1987). Interpreting inequality measures in a Harsanyi framework. *Theory and Decision* **22**, 187–202.

Dalton, H. (1920). Measurement of the inequality of incomes. *The Economic Journal* **30**, 348–61.

Damjanovic, T. (2005). Lorenz dominance for transformed income distributions: a simple proof. *Mathematical Social Sciences* **50**, 234–37.

Dardanoni, V. and P. J. Lambert (1988). Welfare rankings of income distributions: a role for the variance and some insights for tax reform. *Social Choice and Welfare* **5**(5), 1–17.

Das, T. and A. Parikh (1981). Decompositions of Atkinson's measures of inequality. *Australian Economic Papers* **6**, 171–78.

Das, T. and A. Parikh (1982). Decomposition of inequality measures and a comparative analysis. *Empirical Economics* **7**, 23–48.

Dasgupta, P. S., A. K. Sen, and D. A. Starrett (1973). Notes on the measurement of inequality. *Journal of Economic Theory* **6**, 180–87.

David, H. A. (1968). Gini's mean difference rediscovered. *Biometrika* **55**, 573–75.

David, H. A. (1981). *Order Statistics* (2nd edn.). New York: John Wiley.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics* **150**(1), 30–40.

Davidson, R. and J.-Y. Duclos (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* **68**, 1435–64.

Davidson, R. and E. Flachaire (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics* **141**, 141–66.

Davies, J. B. and M. Hoy (1995). Making inequality comparisons when Lorenz curves intersect. *American Economic Review* **85**, 980–86.

Davies, J. B. and A. F. Shorrocks (1989). Optimal grouping of income and wealth data. *Journal of Econometrics* **42**, 97–108.

Davis, H. T. (1941). *The Analysis of Economic Time Series*. Bloomington, Indiana: Principia Press.

Davis, H. T. (1954). *Political Statistics*. Evanston, Illinois: Principia Press.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods*. Cambridge: Cambridge University Press.

Deaton, A. S. (1997). *The Analysis of Household Surveys*. Baltimore, Maryland: Johns Hopkins Press for the World Bank.

Deininger, K. and L. Squire (1997). A new data set measuring income inequality. *World Bank Economic Review* **10**, 565–91.

Deltas, G. (2003). The small-sample bias of the Gini coefficient: results and implications for empirical research. *The Review of Economics and Statistics* **85**, 226–34.

DeNavas-Walt, C., B. D. Proctor, and J. C. Smith (2008). Income, poverty, and health insurance coverage in the United States: 2007. Current Population Reports P60–235, U.S. Census Bureau, U.S. Government Printing Office, Washington, DC.

Department of Work and Pensions (2008). *Households Below Average Income: An Analysis of the Income Distribution 1994/95–2006/07*. London: Department for Work and Pensions.

Department of Work and Pensions (2009). *Households Below Average Income: An Analysis of the Income Distribution 1994/95–2007/08*. London: Department for Work and Pensions.

Devooght, K. (2003). Measuring inequality by counting 'complaints': theory and empirics. *Economics and Philosophy* **19**, 241–63.

Diez, H., M. C. Lasso de la Vega, A. de Sarachu, and A. M. Urrutia (2007). A consistent multidimensional generalization of the Pigou-Dalton transfer principle: an analysis. *The B.E. Journal of Theoretical Economics* **7**, Article 45.

Donaldson, D. and J. A. Weymark (1980). A single parameter generalization of the Gini indices of inequality. *Journal of Economic Theory* **22**, 67–68.

Dorfman, P. (1979). A formula for the Gini coefficient. *Review of Economics and Statistics* **61**, 146–49.

Duclos, J. and A. Araar (2006). *Poverty and Equity: Measurement, Policy, and Estimation with DAD*. New York: Springer.

Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indices. *American Sociological Review* **20**, 210–17.

Dutta, B. and J. Esteban (1992). Social welfare and equality. *Social Choice and Welfare* **9**, 267–76.

Ebert, U. (1987). Size and distribution of incomes as determinants of social welfare. *Journal of Economic Theory* **41**, 25–33.

Ebert, U. (1988). A family of aggregative compromise inequality measure. *International Economic Review* **29**(5), 363–76.

Ebert, U. (1995). Income inequality and differences in household size. *Mathematical Social Sciences* **30**, 37–55.

Ebert, U. (1999). Dual decomposable inequality measures. *Canadian Journal of Economics-Revue Canadienne D'Economique* **32**(1), 234–46.

Ebert, U. (2004). Social welfare, inequality, and poverty when needs differ. *Social Choice and Welfare* **23**, 415–44.

Ebert, U. (2007). Ethical inequality measures and the redistribution of income when needs differ. *Journal of Economic Inequality* **5**, 263–78.

Ebert, U. (2009). Taking empirical studies seriously: the principle of concentration and the measurement of welfare and inequality. *Social Choice and Welfare* **32**, 555–74.

Ebert, U. and H. Welsch (2009). How do Europeans evaluate income distributions? An assessment based on happiness surveys. *Review of Income Wealth* **55**(3), 803–19.

Eichhorn, W., H. Funke, and W. F. Richter (1984). Tax progression and inequality of income distribution. *Journal of Mathematical Economics* **13**(10), 127–31.

Eichhorn, W. and W. Gehrig (1982). Measurement of inequality in economics. In B. Korte (Ed.), *Modern Applied Mathematics—Optimization and Operations Research*, pp. 657–93. Amsterdam: North Holland.

Elbers, C., P. Lanjouw, J. A. Mistiaen, and B. Özler (2008). Reinterpreting between-group inequality. *Journal of Economic Inequality* **6**, 231–45.

Elliott, S. (2009). Why measure inequality? A discussion of the concept of equality. *Oxonomics* **4**(1), 32–41.

Eltetö, O. and E. Frigyes (1968). New income inequality measures as efficient tools for causal analysis and planning. *Econometrica* **36**, 383–96.

Esberger, S. E. and S. Malmquist (1972). *En Statisk Studie av Inkomstutveklingen*. Stockholm: Statisk Centralbyrån och Bostadssyrelsen.

Esteban, J. (1986). Income share elasticity and the size distribution of income. *International Economic Review* **27**, 439–44.

Evans, M., N. Hastings, and B. Peacock (1993). *Statistical Distributions*. New York: John Wiley.

Fase, M. M. G. (1970). *An Econometric Model of Age Income Profiles, A Statistical Analysis of Dutch Income Data*. Rotterdam: Rotterdam University Press.

Feldstein, M. (1998, October). Income inequality and poverty. Working Paper 6770, NBER, 1050 Massachusetts Avenue Cambridge, Massachusetts 02138-5398, USA.

Fellman, J. (1976). The effect of transformation on Lorenz curves. *Econometrica* **44**, 823–24.

Fellman, J. (2001). Mathematical properties of classes of income redistributive policies. *European Journal of Political Economy* **17**, 179–92.

Ferber, R., J. Forsythe, H. W. Guthrie, and E. S. Maynes (1969). Validation of a national survey of consumer financial characteristics. *Review of Economics and Statistics* **51**, 436–44.

Ferreira, F. H. G. and M. Ravallion (2009). Global poverty and inequality: a review of the evidence. In B. Nolan, F. Salverda, and T. Smeeding (Eds.), *Oxford Handbook on Economic Inequality*. Oxford, UK: Oxford University Press.

Fichtenbaum, R. and H. Shahidi (1988). Truncation bias and the measurement of income inequality. *Journal of Business and Economic Statistics* **6**, 335–37.

Fields, G. S. (2003). Accounting for income inequality and its change: a new method with application to distribution of earnings in the United States. *Research in Labor Economics* **22**, 1–38.

Fields, G. S. (2007). How much should we care about changing income inequality in the course of economic growth? *Journal of Policy Modelling* **29**, 577–85.

Fisk, P. R. (1961). The graduation of income distribution. *Econometrica* **29**, 171–85.

Fleurbaey, M. and G. Gaulier (2009). International comparisons of living standards by equivalent incomes. *Scandinavian Journal of Economics* **111**(3), 597–624.

Foster, J. E. (1983). An axiomatic characterization of the Theil measure of income inequality. *Journal of Economic Theory* **31**, 105–21.

Foster, J. E. (1984). On economic poverty: a survey of aggregate measures. *Advances in Econometrics* **3**, 215–51.

Foster, J. E. (1985). Inequality measurement. In H. P. Young (Ed.), *Fair Allocation*, pp. 38–61. Providence, R. I.: American Mathematical Society.

Foster, J. E. and E. A. Ok (1999). Lorenz dominance and the variance of logarithms. *Econometrica* **67**, 901–07.

Foster, J. E. and A. A. Shneyerov (1999). A general class of additively decomposable inequality measures. *Economic Theory* **14**, 89–111.

Francis, W. L. (1972). *Formal Models of American Politics: An Introduction*. New York: Harper and Row.

Freund, J. E. (2003). *Mathematical Statistics with Applications* (Seventh edn.). Harlow, Essex: Pearson Education.

Freund, J. E. and B. M. Perles (2007). *Modern Elementary Statistics* (Twelfth edn.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Frosini, B. V. (1985). Comparing inequality measures. *Statistica* **45**, 299–317.

Frosini, B. V. (1989). Aggregate units, within group inequality and decomposition of inequality measures. *Statistica* **49**, 349–69.

Frosztega, M. (2000). Comparisons of income data between the Family Expenditure Survey and the Family Resources Survey. Government Statistical Service Methodology Series 18, Analytical Services Division, Department of Social Security.

Gabaix, X. (2008). Power laws in economics and finance. Working Paper 14299, National Bureau of Economic Research.

Gaertner, W. (2006). *A Primer in Social Choice Theory*. Oxford: Oxford University Press.

Gail, M. H. and J. L. Gastwirth (1978a). A scale-free goodness-of-fit test for exponential distribution based on the Lorenz curve. *Journal of the American Statistical Association* **73**, 787–93.

Gail, M. H. and J. L. Gastwirth (1978b). A scale-free goodness of fit test for the exponential distribution based on the Gini statistic. *Journal of the Royal Statistical Society, Series B* **40**, 350–57.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica* **39**, 1037–39.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics* **54**, 306–16.

Gastwirth, J. L. (1974a). Large-sample theory of some measures of inequality. *Econometrica* **42**, 191–96.

Gastwirth, J. L. (1974b). A new index of income inequality. *International Statistical Institute Bulletin* **45**(1), 437–41.

Gastwirth, J. L. (1975). The estimation of a family of measures of economic inequality. *Journal of Econometrics* **3**, 61–70.

Gastwirth, J. L. and M. Glauberman (1976). The interpolation of the Lorenz curve and Gini index from grouped data. *Econometrica* **44**, 479–83.

Gastwirth, J. L., T. K. Nayak, and A. N. Krieger (1986). Large sample theory for the bounds on the Gini and related indices from grouped data. *Journal of Business and Economic Statistics* **4**, 269–73.

Gastwirth, J. L. and J. T. Smith (1972). A new goodness-of-fit test. *Proceedings of The American Statistical Association*, 320–22.

Gehrig, W. (1988). On the Shannon-Theil concentration measure and its characterizations. In W. Eichhorn (Ed.), *Measurement in Economics*. Heidelberg: Physica Verlag.

Giles, D. E. A. (2004). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics* **66**, 425–33.

Gini, C. (1912). Variabilità e mutabilità. *Studi Economico-Giuridici dell'Università di Cagliari* **3**, 1–158.

Glasser, G. J. (1962). Variance formulas for the mean difference and the coefficient of concentration. *Journal of the American Statistical Association* **57**, 648–54.

Glewwe, P. (1986). The distribution of income in Sri Lanka in 1969-70 and 1980-81. *Journal of Development Economics* **24**, 255–74.

Glewwe, P. (1991). Household equivalence scales and the measurement of inequality: transfers from the poor to the rich could decrease inequality. *Journal of Public Economics* **44**, 211–16.

Goodman, A. and Z. Oldfield (2004). Permanent differences? Income and expenditure inequality in the 1990s and 2000s. IFS Report 66, The Institute for Fiscal Studies.

Gottschalk, P. and T. M. Smeeding (1997). Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature* **35**, 633–87.

Graaff, J. (1957). *Theoretical Welfare Economics*. London: Cambridge University Press.

Grosfeld, I. and C. Senik (2010). The emerging aversion to inequality: evidence from subjective data. *Economics of Transition* **18**(1), 1–26.

Gupta, M. R. (1984). Functional forms for estimating the Lorenz curve. *Econometrica* **52**, 1313–14.

Hagenaars, A. J. M. (1986). *The Perception of Poverty*. Amsterdam: North-Holland.

Hagerbaumer, J. B. (1977). The Gini concentration ratio and the minor concentration ratio: a two-parameter index of inequality. *Review of Economics and Statistics* **59**, 377–79.

Hainsworth, G. B. (1964). The Lorenz curve as a general tool of economic analysis. *Economic Record* **40**, 426–41.

Hammond, P. J. (1975). A note on extreme inequality aversion. *Journal of Economic Theory* **11**, 465–67.

Hanna, F. A., J. A. Pechman, and S. M. Lerner (1948). *Analysis of Wisconsin Income*. New York: National Bureau of Economic Research.

Hannah, L. and J. A. Kay (1977). *Concentration in British Industry: Theory, Measurement and the UK Experience*. London: MacMillan.

Hardy, G., J. Littlewood, and G. Pólya (1934). *Inequalities*. London: Cambridge University Press.

Hardy, G., J. Littlewood, and G. Pólya (1952). *Inequalities* (Second edn.). London: Cambridge University Press.

Harrison, A. J. (1974). Inequality of income and the Champernowne distribution. Discussion paper 54, University of Essex, Department of Economics.

Harrison, A. J. (1979). The upper tail of the earnings distribution: Pareto or lognormal? *Economics Letters* **2**, 191–95.

Harrison, A. J. (1981). Earnings by size: a tale of two distributions. *Review of Economic Studies* **48**, 621–31.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* **61**, 434–35.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. *Journal of Political Economy* **63**, 309–21.

Hart, P. E. and S. J. Prais (1956). An analysis of business concentration. *Journal of The Royal Statistical Society* A **119**, 150–81.

Harvey, A. and J. Bernstein (2003). Measurement and testing of inequality from time series of deciles with an application to U.S. wages. *The Review of Economics and Statistics* **85**, 141–52.

Hasegawa, H. and H. Kozumi (2003). Estimation of Lorenz curves: a Bayesian nonparametric approach. *Journal of Econometrics* **115**, 277–91.

Hayakawa, M. (1951). The application of Pareto's law of income to Japanese data. *Econometrica* **19**, 174–83.

Helmert, F. R. (1876). Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus den ersten Potenzen der Differenzen gleichgenauer direkter Beobachtungen. *Astronomische Nachrichten* **88**, 127–32.

Hempenius, A. L. (1984). Relative income position, individual and social income satisfaction, and income inequality. *De Economist* **32**, 468–78.

Herfindahl, O. C. (1950). Concentration in the Steel Industry. Ph. D. thesis, Columbia University.

Heston, A. W., R. Summers, and B. Aten (2001). Price structures, the quality factor, and chaining. *Statistical Journal of the United Nations* ECE **18**, 77–101.

Hill, M. S. (1992). *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park, CA: Sage Publications.

Hill, T. P. (1959). An analysis of the distribution of wages and salaries in Great Britain. *Econometrica* **27**, 355–81.

Hills, J. R. (2004). *Inequality and the State*. Oxford: Oxford University Press.

HM Treasury (2003). *The Green Book: Appraisal and Evaluation in Central Government (and Technical Annex)*. London: (Third edn.). TSO.

Hobijn, B. and D. Lagakos (2005). Inflation inequality in the United States. *Review of Income and Wealth* **51**, 581–606.

Hochman, H. and J. D. Rodgers (1969). Pareto-optimal redistribution. *American Economic Review* **59**, 542–57.

Horrace, W., J. Marchand, and T. Smeeding (2008). Ranking inequality: applications of multivariate subset selection. *Journal of Economic Inequality* **6**, 5–32.

Howes, S. P. and P. Lanjouw (1994). Regional variations in urban living standards in urban China. In Q. Fan and P. Nolan (Eds.), *China's Economic Reforms: The Costs and Benefits of Incrementalism*. Basingstoke: Macmillan.

Hu, B. (1995). A note on calculating the Gini index. *Mathematics and Computers in Simulation* **39**, 353–58.

Hussain, A., P. Lanjouw, and N. H. Stern (1994). Income inequalities in China: evidence from household survey data. *World Development* **22**, 1947–57.

Iritani, J. and K. Kuga (1983). Duality between the Lorenz curves and the income distribution functions. *Economic Studies Quarterly* **34**(4), 9–21.

Jakobsson, U. (1976). On the measurement of the degree of progression. *Journal of Public Economics* **5**, 161–68.

Jasso, G. (1979). On Gini's mean difference and Gini's index of concentration. *American Sociological Review* **44**, 867–70.

Jasso, G. (1980). A new theory of distributive justice. *American Sociological Review* **45**, 3–32.

Jasso, G. and S. Kotz (2007). A new continuous distribution and two new families of distributions based on the exponential. *Statistica Neerlandica* **61**(3), 305–28.

Jencks, C. (1973). *Inequality*. London: Allen Lane.

Jenkins, S. P. (1991). The measurement of economic inequality. In L. Osberg (Ed.), *Readings on Economic Inequality*. Armonk, NY: M.E. Sharpe.

Jenkins, S. P. (1995). Accounting for inequality trends: decomposition analyses for the UK. *Economica* **62**, 29–64.

Jenkins, S. P. (2009). Distributionally-sensitive inequality indices and the GB2 income distribution. *Review of Income and Wealth* **55**, 392–98.

Jenkins, S. P. and F. A. Cowell (1994a). Dwarfs and giants in the 1980s: the UK income distribution and how it changed. *Fiscal Studies* **15**(1), 99–118.

Jenkins, S. P. and F. A. Cowell (1994b). Parametric equivalence scales and scale relativities. *The Economic Journal* **104**, 891–900.

Jenkins, S. P. and M. O'Higgins (1989). Inequality measurement using norm incomes—were Garvy and Paglin onto something after all? *Review of Income and Wealth* **35**, 245–82.

Jenkins, S. P. and P. Van Kerm (2005). Accounting for income distribution trends: a density function decomposition approach. *Journal of Economic Inequality* **3**, 43–61.

Jenkins, S. P. and P. Van Kerm (2008). The measurement of economic inequality. In W. Salverda, N. Nolan, and T. M. Smeeding (Eds.), *Oxford Handbook on Economic Inequality*, Chapter 3. Oxford: Oxford University Press.

Johnson, N. O. (1937). The Pareto law. *Review of Economic Statistics* **19**, 20–26.

Jones, F. (2008). The effects of taxes and benefits on household income, 2006/07. *Economic and Labour Market Review* **2**, 37–47.

Jorgenson, D. W. and D. T. Slesnick (1990). Inequality and the standard of living. *Journal of Econometrics* **43**, 103–20.

Kakamu, K. and M. Fukushige (2009). Multilevel decomposition methods for income inequality measures. *Japanese Economic Review* **60**(3), 333–44.

Kakwani, N. C. and N. Podder (1973). On the estimation of the Lorenz curve from grouped observations. *International Economic Review* **14**, 278–92.

Kampelmann, S. (2009). Inequality measures as conventions: new interpretations of a classic operationalization problem. *Socio-Economic Review* **7**, 669–94.

Kanbur, S. M. N. (2006). The policy significance of decompositions. *Journal of Economic Inequality* **4**, 367–74.

Kanbur, S. M. N. and Z. Zhang (1999). Which regional inequality? The evolution of rural–urban and inland–coastal inequality in China from 1983 to 1995. *Journal of Comparative Economics* **27**, 686–701.

Kanbur, S. M. N. and Z. Zhang (2005). Fifty years of regional inequality in China: a journey through central planning, reform, and openness. *Review of Development Economics* **8**, 87–106.

Kaplow, L. (2005). Why measure inequality? *Journal of Economic Inequality* **3**, 65–79.

Kendall, M. G., A. Stuart, and J. K. Ord (1994). *Kendall's Advanced Theory of Statistics* (Sixth edn.), Volume 1, Distribution theory. London: Edward Arnold.

Kendall, M. G., A. Stuart, J. K. Ord, and S. Arnold (1999). *Kendall's Advanced Theory of Statistics* (Sixth edn.), Volume 2A, Classical inference and the linear model. London: Edward Arnold.

Klass, O. S., O. Biham, M. Levy, O. Malcai, and S. Solomon (2006). The Forbes 400 and the Pareto wealth distribution. *Economics Letters* **90**, 290–95.

Kleiber, C. and S. Kotz (2002). A characterization of income distributions in terms of generalized Gini coefficients. *Social Choice and Welfare* **19**, 789–94.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken. NJ: John Wiley.

Kloek, T. and H. K. Van Dijk (1978). Efficient estimation of income distribution parameters. *Journal of Econometrics* **8**, 61–74.

Klonner, S. (2000). The first-order stochastic dominance ordering of the Singh-Maddala distribution. *Economics Letters* **69**(2), 123–128.

Kmietowicz, Z. W. (1984). The bivariate lognormal model for the distribution of household size and income. *The Manchester School of Economic and Social Studies 52*, 196–210.

Kmietowicz, Z. W. and H. Ding (1993). Statistical analysis of income distribution in the Jiangsu province ofChina. *The Statistician* **42**, 107–21.

Kmietowicz, Z. W. and P. Webley (1975). Statistical analysis of income distribution in the central province of Kenya. *Eastern Africa Economic Review* **17**, 1–25.

Kolm, S.-C. (1969). The optimal production of social justice. In J. Margolis and H. Guitton (Eds.), *Public Economics*, pp. 145–200. London: Macmillan.

Kolm, S.-C. (1976a). Unequal inequalities I. *Journal of Economic Theory* **12**, 416–42.

Kolm, S.-C. (1976b). Unequal inequalities II. *Journal of Economic Theory* **13**, 82–111.

Kondor, Y. (1971). An old-new measure of income inequality. *Econometrica* **39**, 1041–42.

Kondor, Y. (1975). Value judgement implied by the use of various measures of income inequality. *Review of Income and Wealth* **21**, 309–21.

Koo, A. Y. C., N. T. Quan, and R. Rasche (1981). Identification of the Lorenz curve by Lorenz coefficients. *Weltwirtschaftliches Archiv* **117**, 125–35.

Kopczuk, W. and E. Saez (2004). Top wealth shares in the United States, 1916–2000: evidence from estate tax returns. *National Tax Journal* **57**, 445–87.

Korinek, A., J. A. Mistiaen, and M. Ravallion (2006). Survey nonresponse and the distribution of income. *Journal of Economic Inequality* **4**, 33–55.

Kovacevic, M. S. and D. A. Binder (1997). Variance estimation for measures of income inequality and polarization. *Journal of Official Statistics* **13**, 41–58.

Kravis, I. B., A. W. Heston, and R. Summers (1978a). *International Comparisons of Real Product and Purchasing Power*. Baltimore: John Hopkins University Press.

Kravis, I. B., A. W. Heston, and R. Summers (1978b). Real GDP per capita for more than one hundred countries. *The Economic Journal* **88**, 215–42.

Krishnan, P. (1981). Measures of inequality for qualitative variables and concentration curves. *American Sociological Review* **46**, 368–71. Comment on Allison, *American Sociological Review* December 1978.

Kroll, Y. and L. Davidovitz (2003). Inequality aversion versus risk aversion. *Economica* **70**, 19–29.

Kuga, K. (1973). Measures of income inequality: an axiomatic approach. Discussion Paper 76, Institute of Social and Economic Research, Osaka University.

Kuga, K. (1980). The Gini index and the generalised entropy class: further results and a vindication. *Economic Studies Quarterly* **31**, 217–28.

Kuznets, S. (1959). *Six Lectures on Economic Growth*. Illinois: Free Press of Glencoe.

Kuznets, S. (1963). Quantitative aspects of the economic growth of nations: part VIII, distribution of income by size. *Economic Development and Cultural Change* **11**, 1–80.

Kuznets, S. (1966). *Modern Economic Growth*. New Haven, Connecticut: Yale University Press.

Lam, D. (1986). The dynamics of population growth, differential fertility and inequality. *American Economic Review* **76**, 1103–16.

Lambert, P. J. (1980). Inequality and social choice. *Theory and Decision* **12**, 395–98.

Lambert, P. J. (2001). *The Distribution and Redistribution of Income* (Third edn.). Manchester: Manchester University Press.

Lambert, P. J. and J. R. Aronson (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal* **103**(9), 1221–27.

Lambert, P. J. and G. Lanza (2006). The effect on inequality of changing one or two incomes. *Journal of Economic Inequality* **4**, 253–77.

Lambert, P. J., D. L. Millimet, and D. J. Slottje (2003). Inequality aversion and the natural rate of subjective inequality. *Journal of Public Economics* **87**, 1061–90.

Layard, A. (1994). *Cost Benefit Analysis* (Second edn.). Cambridge: Cambridge University Press.

Layard, R., S. Nickell, and G. Mayraz (2008, August). The marginal utility of income. *Journal of Public Economics* **92**(8–9), 1846–57.

Lebergott, S. (1959). The shape of the income distribution. *American Economic Review* **49**, 328–47.

Lefranc, A., N. Pistolesi, and A. Trannoy (2008). Inequality and opportunities vs. inequality of outcome: are western societies all alike? *Review of Income and Wealth* **54**, 513–46.

Lerman, R. I. and S. Yitzhaki (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters* **15**, 363–68.

Lerman, R. I. and S. Yitzhaki (1989). Improving the accuracy of estimates of the Gini coefficient. *Journal of Econometrics* **42**, 43–47.

Levine, D. and N. M. Singer (1970). The mathematical relation between the income density function and the measurement of income inequality. *Econometrica* **38**, 324–30.

Likes, J. (1969). Minimum variance unbiased estimates of the parameters of power function and Pareto's distribution. *Statistische Hefte* **10**, 104–10.

Lin, T. (1990). Relation between the Gini coefficient and the Kuznets ratio of MEP. *Jahrbuch von Nationalökonomie und Statistik* **207**(2), 36–46.

Lin, T., J. Zhuang, D. Yarcia, and F. Lin (2008). Income inequality in the People's Republic of China and its decomposition: 1990-2004. *Asian Development Review* **25**, 119–36.

Lindley, D. V. and J. C. P. Miller (1966). *Cambridge Elementary Statistical Tables*. London: Cambridge University Press.

Little, I. M. D. and J. A. Mirrlees (1974). *Project Appraisal and Planning in Developing Countries*. London: Heinemann.

Lomnicki, Z. A. (1952). The standard error of Gini's mean difference. *Annals of Mathematical Statistics* **23**, 635–37.

Lorenz, M. O. (1905). Methods for measuring concentration of wealth. *Journal of the American Statistical Association* **9**, 209–19.

Love, R. and M. C. Wolfson (1976). Income inequality: statistical methodology and Canadian illustrations. Occasional Paper 13-559, Statistics Canada.

Lydall, H. F. (1959). The long-term trend in the size distribution of income. *Journal of the Royal Statistical Society* A **122**, 1–36.

Lydall, H. F. (1968). *The Structure of Earnings*. Oxford: Clarendon Press.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica* **54**, 991–97.

Maasoumi, E. (1989). Continuously distributed attributes and measures of multivariate inequality. *Journal of Econometrics* **42**, 131–44.

Maccabelli, T. (2009). Measuring inequality: Pareto's ambiguous contribution. *History of Political Economy* **41**, 183–208.

Maddala, G. S. and S. K. Singh (1977). Estimation problems in size distribution of incomes. *Economie Appliquée* **30**, 461–80.

Majumder, A. and S. R. Chakravarty (1990). Distribution of personal income: development of a new model and its application to U.S. income data. *Journal of Applied Econometrics* **5**, 189–96.

Malik, H. J. (1970). Estimation of the parameters of the Pareto distribution. *Metrika* **15**, 126–32.

Mandelbrot, B. (1960). The Pareto-Lévy law and the distribution of income. *International Economic Review* **1**(2), 79–106.

Marfels, C. (1971). Einige neuere Entwicklungen in der Messung der industriellen Konzentration (some new developments in the measurement of industrial concentration). *Metrika* **17**, 753–66.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory and Majorization*. New York: Academic Press.

McClements, L. (1977). Equivalence scales for children. *Journal of Public Economics* **8**, 191–210.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica* **52**, 647–64.

McDonald, J. B. and B. Jensen (1979). An analysis of some properties of alternative measures of income inequality based on the gamma distribution function. *Journal of the American Statistical Association* **74**, 856–60.

McDonald, J. B. and A. Mantrala (1995). The distribution of personal income revisited. *Journal of Applied Econometrics* **10**, 201–04.

McDonald, J. B. and M. R. Ransom (1979). Functional forms, estimation techniques and the distribution of income. *Econometrica* **47**, 1513–25.

McDonald, J. B. and M. R. Ransom (1981). An analysis of the bounds for the Gini coefficient. *Journal of Econometrics* **17**, 177–218.

Meade, J. E. (1976). *The Just Economy*. London: Allen and Unwin.

Mehran, F. (1975). Bounds on the Gini index based on observed points of the Lorenz curve. *Journal of the American Statistical Association* **70**, 64–66.

Mera, K. (1969). Experimental determination of relative marginal utilities. *Quarterly Journal of Economics* **83**, 464–77.

Metcalf, C. E. (1969). The size distribution of income during the business cycle. *American Economic Review* **59**, 657–68.

Milanovic, B. (2007). Why we all care about inequality (but some of us are loathe to admit it). *Challenge* **50**, 109–120.

Milanovic, B. and S. Yitzhaki (2002). Decomposing world income distribution: does the world have a middle class? *Review of Income and Wealth* **48**, 155–78.

Modarres, R. and J. L. Gastwirth (2006). A cautionary note on estimating the standard error of the Gini index of inequality. *Oxford Bulletin of Economics And Statistics* **68**, 385–90.

Mookherjee, D. and A. F. Shorrocks (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal* **92**, 886–902.

Morduch, J. and T. Sicular (2002). Rethinking inequality decomposition, with evidence from rural China. *The Economic Journal* **112**, 93–106.

Morgan, J. N. (1962). The anatomy of income distribution. *Review of Economics and Statistics* **44**, 270–83.

Morgan, J. N., M. H. David, W. J. Cohen, and A. E. Brazer (1962). *Income and Welfare in The United States*. New York: McGraw-Hill.

Moyes, P. (1987). A new concept of Lorenz domination. *Economics Letters* **23**, 203–07.

Moyes, P. (1989). Equiproportionate growth of incomes and after-tax inequality. *Bulletin of Economic Research* **41**, 287–93.

Moyes, P. (2007). An extended Gini approach to inequality measurement. *Journal of Economic Inequality* **5**, 279–303.

Muellbauer, J. (1974). Inequality measures, prices and household composition. *Review of Economic Studies* **41**, 493–504.

Musgrave, R. A. and T. Thin (1948). Income tax progression, 1929–48. *Journal of Political Economy* **56**, 498–514.

Nair, U. S. (1936). The standard error of Gini's mean difference. *Biometrika* **28**, 428–36.

Nicholson, R. J. (1969). *Economic Statistics and Economic Problems*. London: McGraw-Hill.

Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics* **57**, 86–100.

Novotný, J. (2007). On the measurement of regional inequality: does spatial dimension of income inequality matter? *The Annals of Regional Science* **41**, 563–80.

Nygård, F. and A. Sandström (1981). *Measuring Income Inequality*. Stockholm, Sweden: Almquist Wicksell International.

Nygård, F. and A. Sandström (1985). Estimating Gini and entropy inequality parameters. *Journal of Official Statistics* **1**, 399–412.

Nygård, F. and A. Sandström (1989). Income inequality measures based on sample surveys. *Journal of Econometrics* **42**, 81–95.

OECD (2008). *Growing Unequal? Income Distribution And Poverty In OECD Countries*. Organisation For Economic Co-operation and Development.

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics* **62**, 123–29.

Ogwang, T. (2003). Bounds of the Gini index using sparse information on mean incomes. *Review of Income and Wealth* **49**, 415–23.

Ogwang, T. (2004). A convenient method of computing the Gini index and its standard error: some further results: reply. *Oxford Bulletin of Economics and Statistics* **66**, 435–37.

Ogwang, T. and U. L. G. Rao (2000). Hybrid models of the Lorenz curve. *Economics Letters* **69**, 39–44.

Ok, E. A. (1995). Fuzzy measurement of income inequality: a class of fuzzy inequality measures. *Social Choice and Welfare* **12**(2), 111–36.

Okner, B. A. (1972). Constructing a new data base from existing microdata sets: the 1966 MERGE file. *Annals of Economic and Social Measurement* **1**, 325–42.

Okner, B. A. (1975). Individual taxes and the distribution of income. In J. D. Smith (Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Okun, A. M. (1975). *Equality and Efficiency: The Big Trade-off*. Washington: Brookings Institution.

Olken, B. (2005). Revealed community equivalence scales. *Journal of Public Economics* **89**, 545–66.

Ortega, P., G. Martin, A. Fernandez, M. Ladoux, and A. Garcia (1991). A new functional form for estimating Lorenz curves. *Review of Income and Wealth* **37**(12), 447–52.

Paglin, M. (1975). The measurement and trend of inequality: a basic revision. *American Economic Review* **65**, 598–609.

Pareto, V. (1896). La courbe de la répartition de la richesse. In C. Viret-Genton (Ed.), *Recueil Publié par la Faculté de Droit à l'Occasion de l'Exposition Nationale Suisse, Geneva 1896*, pp. 373–87. Lausanne: Université de Lausanne.

Pareto, V. (1965). *Écrits sur La Courbe de la Repartition de la Richesse*, Volume 3 of *Oeuvres Complètes*. Geneva: Librairie Droz. Edited by Busino, G.

Pareto, V. (1972). *Manual of Political Economy*. London: Macmillan. Edited by Schwier, A. S. and Page, A. N.

Pareto, V. (2001). On the distribution of wealth and income. In M. Baldassarri and P. Ciocca (Eds.), *Roots of the Italian School of Economics and Finance: From Ferrara (1857) to Einaudi (1944)*, Volume 2, pp. 231–76. Houndmills: Palgrave.

Parker, S. C. (1999). The generalised beta as a model for the distribution of earnings. *Economics Letters* **62**(2), 197–200.

Paul, S. (1999). The population sub-group income effects on inequality: analytical framework and an empirical illustration. *Economic Record* **75**(229), 149–55.

Paul, S. (2004). Income sources effects on inequality. *Journal of Development Economics* **73**, 435–51.

Pen, J. (1971). *Income Distribution*. London: Allen Lane, The Penguin Press.

Pen, J. (1974). *Income Distribution* (Second edn.). London: Allen Lane, The Penguin Press.

Persky, J. (1992). Retrospectives: Pareto's law. *The Journal of Economic Perspectives* **6**(2), 181–92.

Phelps, E. S. (1973). *Economic Justice*. Harmondsworth: Penguin.

Pigou, A. C. (1952). *The Economics of Welfare* (Fourth edn.). London: Macmillan.

Pirttilä, J. and R. Uusitalo (2010). A 'leaky bucket' in the real world: Estimating inequality aversion using survey data. *Economica* **77**, 60–76.

Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities, 1968–2001. *Journal of Economic Inequality* **7**, 411–33.

Podder, N. and S. Chatterjee (2002). Sharing the national cake in post reform New Zealand: income inequality trends in terms of income sources. *Journal of Public Economics* **86**, 1–27.

Polanyi, G. and J. B. Wood (1974). How much inequality? Research monograph, Institute of Economic Affairs, London.

Prest, A. R. and T. Stark (1967). Some aspects of income distribution in the UK since World War II. *Manchester School* **35**, 217–43.

Preston, I. (2007). Inequality and income gaps. *Research on Economic Inequality* **15**, 33–56.

Pyatt, G. (1976). On the interpretation and disaggregation of Gini coefficients. *The Economic Journal* **86**, 243–55.

Quandt, R. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika* **10**, 55–82.

Rajaraman, I. (1975). Poverty, inequality and economic growth: Rural Punjab, 1960/1–1970/1. *Journal of Development Studies* **11**, 278–90.

Ram, R. (1979). International income inequality: 1970 and 1978. *Economics Letters* **4**, 187–90.

Ram, R. (1984). Another perspective on changes in international inequality from 1950 to 1980. *Economics Letters* **16**, 187–90.

Ram, R. (1987). Intercountry inequalities in income and 'index of net social progress': evidence from recent data. *Economics Letters* **24**, 295–98.

Ram, R. (1992). International inequalities in human development and real income. *Economics Letters* **38**, 351–54.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (Second edn.). New York: John Wiley.

Rao, U. L. P. and A. Y. P. Tam (1987). An empirical study of selection and estimation of alternative models of the Lorenz curve. *Journal of Applied Statistics* **14**, 275–80.

Rasche, R. H., J. Gaffney, A. Y. C. Koo, and N. Obst (1980). Functional forms for estimating the Lorenz curve. *Econometrica* **48**, 1061–62.

Ravallion, M. (1994). Poverty comparisons using noisy data on living standards. *Economics Letters* **45**, 481–85.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.

Rees, J. (1971). *Equality*. London: Pall Mall.

Rein, M. and S. M. Miller (1974, July/August). Standards of income redistribution. *Challenge* **July/August**, 20–26.

Reynolds, A. (2006). *Income and Wealth*. Westport, Connecticut: Greenwood Press.

Riese, M. (1987). An extension of the Lorenz diagram with special reference to survival analysis. *Oxford Bulletin of Economics and Statistics* **49**(5), 245–50.

Rietveld, P. (1990). Multidimensional inequality comparisons. *Economics Letters* **32**, 187–92.

Rohde, N. (2008). Lorenz curves and generalised entropy inequality measures. In D. Chotikapanich (Ed.), *Modeling Income Distributions and Lorenz Curves*, Chapter 15, pp. 271–84. New York: Springer.

Rosenbluth, G. (1955). Measures of concentration. In G. J. Stigler (Ed.), *Business Concentration and Price Policy*. Princeton: National Bureau of Economic Research.

Royal Commission on the Distribution of Income and Wealth (1975). *Initial Report on the Standing Reference, Cmnd 6171*. London: HMSO.

Royal Commission on the Distribution of Income and Wealth (1976). *Second Report on the Standing Reference, Cmnd 6626*. London: HMSO.

Royal Commission on the Taxation of Profits and Income (1955). *Final Report on the Standing Reference, Cmnd 9474*. London: HMSO.

Russet, B. M. (1964). Inequality and instability: the relation of land tenure to politics. *World Politics* **16**, 442–54.

Ryscavage, P. (1999). *Income Inequality in America: An Analysis of Trends*. New York: M.E. Sharpe.

Sala-i-Martin, X. (2006). The world distribution of income: falling poverty and . . . convergence, period. *Quarterly Journal of Economics* **121**, 351–97.

Salanié, B. (2000). *Microeconomics of Market Failures*. Cambridge, Massachusetts: MIT Press.

Salanié, B. (2003). *The Economics of Taxation*. Cambridge Massachusetts: MIT Press.

Salas, R. (1998). Welfare-consistent inequality indices in changing populations: the marginal population replication axiom. A note. *Journal of Public Economics* **67**, 145–50.

Salem, A. B. Z. and T. D. Mount (1974). A convenient descriptive model of income distribution: the Gamma density. *Econometrica* **42**, 1115–27.

Sandström, A., J. H. Wretman, and B. Walden (1985). Variance estimators of the Gini coefficient: simple random sampling. *Metron* **43**, 41–70.

Sandström, A., J. H. Wretman, and B. Walden (1988). Variance estimators of the Gini coefficient: probability sampling. *Journal of Business and Economic Statistics* **6**, 113–20.

Saposnik, R. (1981). Rank-dominance in income distribution. *Public Choice* **36**, 147–51.

Saposnik, R. (1983). On evaluating income distributions: rank dominance, the Suppes-Sen grading principle of justice and Pareto optimality. *Public Choice* **40**, 329–36.

Sarabia, J. M., E. Castillo, and D. J. Slottje (1999). An ordered family of Lorenz curves. *Journal of Econometrics* **91**, 43–60.

Sarabia, J. M., E. Castillo, and D. J. Slottje (2002). Lorenz ordering between McDonald's generalized functions of the income size distribution. *Economics Letters* **75**, 265–70.

Sastre, M. and A. Trannoy (2002). Shapley inequality decomposition by factor components: some methodological issues. *Journal of Economics* **Supp 9**, 51–90.

Sastry, D. V. S. and U. R. Kelkar (1994). Note on the decomposition of Gini inequality. *The Review of Economics and Statistics* **76**, 584–86.

Satchell, S. E. (1978). Source and subgroup decomposition inequalities for the Lorenz curve. *International Economic Review* **28**(6), 321–29.

Saunders, T. J. (1970). *Plato: The Laws*. Harmondsworth: Penguin.

Savaglio, E. (2006). Multidimensional inequality with variable population size. *Economic Theory* **28**, 85–94.

Schechtman, E. and S. Yitzhaki (1999). On the proper bounds of Gini correlation. *Economic Letters* **63**, 133–38.

Schluter, C. and M. Trede (2002). Tails of Lorenz curves. *Journal of Econometrics* **109**, 151–66.

Schutz, R. R. (1951). On the measurement of income inequality. *American Economic Review* **41**, 107–22.

Schwartz, J. E. and C. Winship (1980). The welfare approach to measuring inequality. *Sociological Methodology* **11**, 1–36.

Seidl, C. (1988). Poverty measurement: a survey. In D. Bös, M. Rose, and C. Seidl (Eds.), *Welfare and Efficiency in Public Economics*, pp. 71–147. Berlin, Heidelberg: Springer-Verlag.

Sen, A. K. (1970). *Collective Choice and Social Welfare*. Edinburgh: Oliver and Boyd.

Sen, A. K. (1973). *On Economic Inequality*. Oxford: Clarendon Press.

Sen, A. K. (1974). Informational bases of alternative welfare approaches. *Journal of Public Economics* **3**, 387–403.

Sen, A. K. (1976a). Poverty: an ordinal approach to measurement. *Econometrica* **44**, 219–31.

Sen, A. K. (1976b). Real national income. *Review of Economic Studies* **43**, 19–39.

Sen, A. K. (1977). On weights and measures: informational constraints in social welfare analysis. *Econometrica* **45**, 1539–72.

Sen, A. K. (1980). Equality of what? In S. McMurrin (Ed.), *Tanner Lectures on Human Values*, Volume 1. Cambridge: Cambridge University Press.

Sen, A. K. (1992). *Inequality Reexamined*. Cambridge, Mass: Harvard University Press.

Sen, A. K. (1997). From income inequality to economic inequality. *Southern Economic Journal* **64**, 383–401.

Sen, A. K. and J. E. Foster (1997). *On Economic Inequality* (Second edn.). Oxford: Clarendon Press.

Shorrocks, A. F. (1978). Income inequality and income mobility. *Journal of Economic Theory* **19**, 376–93.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica* **48**, 613–25.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica* **50**(1), 193–211.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica* **50**, 3–17.

Shorrocks, A. F. (1984). Inequality decomposition by population subgroups. *Econometrica* **52**, 1369–85.

Shorrocks, A. F. (1988). Aggregation issues in inequality measurement. In W. Eichhorn (Ed.), *Measurement in Economics*. Heidelberg: Physica Verlag.

Shorrocks, A. F. (2004). Inequality and welfare evaluation of heterogeneous income distributions. *Journal of Economic Inequality* **2**, 193–218.

Shorrocks, A. F. and J. E. Foster (1987). Transfer-sensitive inequality measures. *Review of Economic Studies* **54**, 485–98.

Shorrocks, A. F. and D. J. Slottje (2002). Approximating unanimity orderings: an application to Lorenz dominance. *Journal of Economics* **Supp 9**, 91–117.

Shorrocks, A. F. and G. Wan (2005). Spatial decomposition of inequality. *Journal of Economic Geography* **5**, 59–81.

Sicular, T., Y. Ximing, B. Gustaffson, and L. Shi (2007). The urban-rural gap and inequality in China. *Review of Income and Wealth* **53**, 93–126.

Sierminska, E., A. Brandolini, and T. M. Smeeding (2006). The Luxembourg Wealth Study—a cross-country comparable database for household wealth research. *Journal of Economic Inequality* **4**, 375–83.

Silber, J. (1989). Factor components, population subgroups and the computation of the Gini index of inequality. *Review of Economics and Statistics* **71**, 107–15.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika* **52**, 425–40.

Simon, H. A. (1957). The compensation of executives. *Sociometry* **20**, 32–35.

Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *American Economic Review* **48**, 607–17.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. New York: Springer.

Singh, S. K. and G. S. Maddala (1976). A function for the size distribution of income. *Econometrica* **44**, 963–70.

Slottje, D. J. (1984). A measure of income inequality based upon the beta distribution of the second kind. *Economics Letters* **15**, 369–75.

Slottje, D. J. (1987). Relative price changes and inequality in the size distribution of various components of income: a multidimensional approach. *Journal of Business and Economic Statistics* **5**, 19–26.

Smeeding, T. M., M. O'Higgins, and L. Rainwater (1990). *Poverty, Inequality and Income Distribution in Comparative Perspective*. Hemel Hempstead: Harvester Wheatsheaf.

Soltow, L. (1975). The wealth, income and social class of men in large northern cities of the United States in 1860. In J. D. Smith (Ed.), *The Personal Distribution of Income and Wealth*. New York: National Bureau of Economic Research.

Stark, T. (1972). *The Distribution of Personal Income in The United Kingdom 1949–1963*. London: Cambridge University Press.

Steindl, J. (1965). *Random Processes and the Growth of Firms: A Study of the Pareto Law*. New York: Hafner Press.

Stevenson, B. and J. Wolfers (2008, August). Happiness inequality in the United States. Working Paper 14220, National Bureau of Economic Research.

Summers, R. and A. W. Heston (1988). A new set of international comparisons of real product and price levels: estimates for 130 countries, 1950–1985. *Review of Income and Wealth* **34**, 1–25.

Summers, R. and A. W. Heston (1991). The Penn world table (mark 5): an expanded set of international comparisons 1950–1988. *Quarterly Journal of Economics* **106**, 327–68.

Taguchi, T. (1968). Concentration curve methods and structures of skew populations. *Annals of the Institute of Statistical Mathematics* **20**, 107–41.

Taille, C. (1981). Lorenz ordering within the generalized gamma family of income distributions. In P. Taille, G. P. Patil, and B. Baldessari (Eds.), *Statistical Distributions in Scientific Work*. Boston: Reidel.

Takahashi, C. (1959). *Dynamic Changes of Income and its Distribution in Japan*. Tokyo: Kinokuniya Bookstore Co.

Tarozzi, A. and A. S. Deaton (2009). Using census and survey data to estimate poverty and inequality for small areas. *The Review of Economics and Statistics* **91**, 773–92.

Tawney, H. R. (1964). *Equality*. London: Allen and Unwin.

Temkin, L. S. (1986). Inequality. *Philosophy and Public Affairs* **15**, 99–121.

Temkin, L. S. (1993). *Inequality*. Oxford: Oxford University Press.

Thatcher, A. R. (1968). The distribution of earnings of employees in Great Britain. *Journal of The Royal Statistical Society* A **131**, 133–70.

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North Holland.

Theil, H. (1979a). The measurement of inequality by components of income. *Economics Letters* **2**, 197–99.

Theil, H. (1979b). World income inequality and its components. *Economics Letters* **2**, 99–102.

Theil, H. (1989). The development of international inequality 1960–1985. *Journal of Econometrics* **42**, 145–55.

Thistle, P. D. (1989a). Duality between generalized Lorenz curves and distribution functions. *Economic Studies Quarterly* **40**(6), 183–87.

Thistle, P. D. (1989b). Ranking distributions with generalized Lorenz curves. *Southern Economic Journal* **56**, 1–12.

Thistle, P. D. (1990). Large sample properties of two inequality indices. *Econometrica* **58**, 725–28.

Thon, D. (1979). On measuring poverty. *Review of Income and Wealth* **25**, 429–39.

Thon, D. (1981). Income inequality and poverty; some problems. *Review of Income and Wealth* **27**, 207–10.

Thon, D. (1982). An axiomatization of the Gini coefficient. *Mathematical Social Science* **2**, 131–43.

Thon, D. (1983a). Lorenz curves and Lorenz coefficient: a sceptical note. *Weltwirtschaftliches Archiv* **119**, 364–67.

Thon, D. (1983b). A note on a troublesome axiom for poverty indices. *The Economic Journal* **93**, 199–200.

Thurow, L. C. (1970). Analysing the American income distribution. *American Economic Review* **60**, 261–69.

Thurow, L. C. (1975). *Generating Inequality*. New York: Basic Books.

Tobin, J. (1970). On limiting the domain of inequality. *Journal of Law and Economics* **13**, 263–77. Reprinted in Phelps (1973).

Toyoda, T. (1980). Decomposability of inequality measures. *Economic Studies Quarterly* **31**(12), 207–46.

Tsui, K.-Y. (1995). Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. *Journal of Economic Theory* **67**, 251–65.

Tuomala, M. (1990). *Optimal Income Tax and Redistribution*. London: Oxford University Press.

United Nations Economic Commission for Europe (1957). *Economic Survey of Europe in 1956*. Geneva: United Nations.

UNU-WIDER (2005). World income inequality database, version 2.0a. Technical report, UNU-WIDER.

van der Wijk, J. (1939). *Inkomens—En Vermogensverdeling*. Number 26 in Nederlandsch Economisch Instituut. Haarlem: De Erven F. Bohn, N. V.

Van Kerm, P. (2002). Inference on inequality measures: a Monte Carlo experiment. *Journal of Economics* **Supp 9**, 283–06.

Victoria-Feser, M.-P. (1999). The sampling properties of inequality indices: comment. In J. Silber (Ed.), *Handbook on Income Inequality Measurement*, pp. 260–67. Dewenter: Kluwer.

Weisbrod, B. A. and W. L. Hansen (1968). An income-net worth approach to measuring economic welfare. *American Economic Review* **58**, 1315–29.

Weiss, Y. (1972). The risk element in occupational and educational choices. *Journal of Political Economy* **80**, 1203–13.

Welniak, E. J. (2003). Measuring household income inequality using the CPS. In J. Dalton and B. Kilss (Eds.), *Special Studies in Federal Tax Statistics 2003*. Washington, DC: Statistics of Income Directorate, Inland Revenue Service.

Weymark, J. A. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences* **1**, 409–30.

Weymark, J. A. (2006). The normative approach to the measurement of multidimensional inequality. In *Inequality and Economic Integration*, F. Farina and E. Sarvaglio (Eds). London: Routledge.

Wiles, P. J. D. (1974). *Income Distribution, East and West*. Amsterdam: North Holland.

Wiles, P. J. D. and S. Markowski (1971). Income distribution under communism and capitalism. *Soviet Studies* **22**, 344–69, 485–511.

Wilfling, B. (1996). The Lorenz-ordering of generalized beta-II income distributions. *Journal of Econometrics* **71**, 381–88.

Wilfling, B. and W. Krämer (1993). The Lorenz-ordering of Singh-Maddala income distributions. *Economics Letters* **43**, 53–57.

Williams, R. and D. P. Doessel (2006). Measuring inequality: tools and an illustration. *International Journal for Equity in Health* **5**, 1–8.

Wilson, J. (1966). *Equality*. London: Hutchinson.

Wold, H. O. A. and P. Whittle (1957). A model explaining the Pareto distribution of wealth. *Econometrica* **25**, 591–95.

Wolff, E. N. and A. Zacharias (2007). The distributional consequences of government spending and taxation in the U.S., 1989 and 2000. *Review of Income and Wealth* **53**, 692–715.

World Bank (2004). *A Better Investment Climate for Everyone: World Development Report 2005*. New York: The World Bank and Oxford University Press.

World Bank (2005). *Equity and Development: World Development Report 2006*. New York: The World Bank and Oxford University Press.

Yaari, M. E. (1988). A controversial proposal concerning inequality measurement. *Journal of Economic Theory* **44**(4), 381–97.

Yitzhaki, S. (1979). Relative deprivation and the Gini coefficient. *Quarterly Journal of Economics* **93**, 321–24.

Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review* **24**(10), 617–28.

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality* **8**, 13–30.

Yitzhaki, S. and R. Lerman (1991). Income stratification and income inequality. *Review of Income and Wealth* **37**(9), 313–29.

Yoshida, T. (1991). Social welfare rankings on income distributions: alternative views of efficiency preference. Discussion Paper 10, Osaka University.

Young, A. A. (1917). Do the statistics of the concentration of wealth in the United States mean what they are commonly supposed to mean? *Journal of The American Statistical Association* **15**, 471–84.

Yu, L., R. Luo, and L. Zhan (2007). Decomposing income inequality and policy implications in rural China. *China and World Economy* **15**, 44–58.

Yule, G. U. and M. G. Kendall (1950). *An Introduction to The Theory of Statistics*. London: Griffin.

Zheng, B. (1997). Aggregate poverty measures. *Journal of Economic Surveys* **11**(2), 123–62.

Zheng, B. (2007). Unit-consistent decomposable inequality measures. *Economica* **74**(293), 97–111.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Reading, Mass: Addison-Wesley.

Index