Measuring Inequality

Frank A. Cowell

May 2000

Third Edition

http://sticerd.lse.ac.uk/research/frankweb/MeasuringInequality/index.html

Abstract

- Part of the *LSE Economics Series*, published by Oxford University Press

- This book is dedicated to the memory of my parents.

Contents

Preface to the First Edition **xiii**

Preface to the Second Edition **xv**

1 FIRST PRINCIPLES **1**
1.1 A PREVIEW OF THE BOOK 3
1.2 INEQUALITY OF WHAT? 4
1.3 INEQUALITY MEASUREMENT, JUSTICE AND POVERTY . 7
1.3.1 Scalar Inequality 7
1.3.2 Numerical Representation 9
1.3.3 Income Differences 10
1.3.4 Given Population. 11
1.4 INEQUALITY AND THE SOCIAL STRUCTURE 12
1.5 QUESTIONS 13

2 CHARTING INEQUALITY **15**
2.1 DIAGRAMS 16
2.2 INEQUALITY MEASURES 21
2.3 RANKINGS 27
2.4 FROM CHARTS TO ANALYSIS 31
2.5 QUESTIONS 33

3 ANALYSING INEQUALITY **35**
3.1 SOCIAL-WELFARE FUNCTIONS 36
3.2 SWF-BASED INEQUALITY MEASURES 44
3.3 INEQUALITY AND INFORMATION THEORY 47
3.4 BUILDING AN INEQUALITY MEASURE 54
3.5 CHOOSING AN INEQUALITY MEASURE 60
3.6 SUMMARY 65
3.7 QUESTIONS 67

4 MODELLING INEQUALITY **69**
4.1 THE IDEA OF A MODEL 70
4.2 THE LOGNORMAL DISTRIBUTION 71
4.3 THE PARETO DISTRIBUTION 78

4.4 HOW GOOD ARE THE FORMULAS? 85
4.5 QUESTIONS 88

5 FROM THEORY TO PRACTICE 91
5.1 THE DATA 92
5.1.1 *Method 1: Asking people.* 92
5.1.2 *Method 2: Compulsion.* 93
5.1.3 *What is included?* 95
5.1.4 *Which heads are counted?* 96
5.1.5 *What time period?* 98
5.1.6 *What valuation procedure has been used?* 98
5.1.7 *Which economic assumptions have been made?* 99
5.2 COMPUTATION OF THE INEQUALITY MEASURES 99
5.3 APPRAISING THE CALCULATIONS 115
5.4 SHORTCUTS: FITTING FUNCTIONAL FORM[1] 120
5.5 INTERPRETING THE ANSWERS 126
5.5.1 What cardinal representation has been used? 127
5.5.2 Has the cake shrunk? 128
5.5.3 Is the drop in inequality an optical illusion? 129
5.5.4 How do we cope with problems of non-comparability? . . 130
5.5.5 Is the trend toward equality large enough to matter? . . . 131
5.6 A SORT OF CONCLUSION 132
5.7 QUESTIONS 133

A TECHNICAL APPENDIX 135
A.1 OVERVIEW 135
A.2 MEASURES AND THEIR PROPERTIES 135
A.2.1 Discrete Distributions 136
A.2.2 Continuous distributions 136
A.3 FUNCTIONAL FORMS OF DISTRIBUTION 138
A.3.1 The lognormal distribution 138
A.3.2 The Pareto distribution (type I) 140
A.3.3 Other Functional Forms 141
A.4 INTERRELATIONSHIPS BETWEEN INEQUALITY MEASURES144
A.4.1 Atkinson (A_ε) and Dalton (D_ε) Measures 145
A.4.2 The Logarithmic Variance (v) and the Variance of Logarithms (v_1) 146
A.5 DECOMPOSITION OF INEQUALITY MEASURES 147
A.5.1 By subgroups 147
A.5.2 By income components 152
A.6 NEGATIVE INCOMES 152
A.7 ESTIMATION PROBLEMS 153
A.7.1 Micro Data 153
A.7.2 Grouped Data 156

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

A.8 USING THE WEBSITE 159

B NOTES ON SOURCES AND LITERATURE 161

B.1 CHAPTER 1 161
B.2 CHAPTER 2 162
B.3 CHAPTER 3 165
B.4 CHAPTER 4 167
B.5 CHAPTER 5 169
B.6 TECHNICAL APPENDIX 172

List of Tables

1.1 Four inequality scales 9

3.1 How much should R give up to finance a č1 bonus for P? 39
3.2 Is P further from Q than Q is from R? 54
3.3 The break-down of inequality: poor East, rich West 57
3.4 The break-down of inequality: the East catches up 58
3.5 Which measure does what? 66

4.1 Pareto's α and "average/base" Inequality B 82
4.2 Pareto's α for income distribution in the UK and the USA 86

5.1 The McClements Equivalence Scale 97
5.2 Distribution of Income Before Tax. USA 1995. Source: Internal Revenue Service 105
5.3 Values of Inequality indices under a variety of assumptions about the data. US 1995 110
5.4 Approximation Formulas for Standard Errors of Inequality Measures 117
5.5 Individual distribution of household net per capita annual income. Czechoslovakia 1988. 119
5.6 Average income, taxes and benefits by decile groups of all households. UK 1998-9 133
5.7 Observed and expected frequencies of household income per head. Jiangsu, China. 134

A.1 Inequality measures for discrete distributions 137
A.2 Inequality measures for continuous distributions 139
A.3 Decomposition of inequality in Chinese provinces, Rural and Urban subpopulations 151
A.4 Source files for tables and figures 160

List of Figures

1.1 Two Types of Inequality 8
1.2 An Inequality Ranking 9

2.1 The Parade of Dwarfs. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987 17
2.2 Frequency Distribution of Income. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987 17
2.3 Cumulative Frequency Distribution. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987 18
2.4 Lorenz Curve of Income. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987 19
2.5 Frequency Distribution of Income (Logarithmic Scale). UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987 20
2.6 The Parade with Partial Equalisation 23
2.7 The Stark Approach 26
2.8 The Parade and the Quantile Ranking 28
2.9 Quantile ratios of earnings of adult males, UK 1968-1998. Source: New Earnings Survey, 1998, Part A Table 28 29
2.10 Ranking by Shares. UK 1984/5 Incomes before and after tax. Source: as for Figure 2.1 30
2.11 Lorenz Curves Crossing 31
2.12 Change at the bottom of the income distribution 32
2.13 Change at the top of the income distribution 32

3.1 Social utility and relative income 40
3.2 The relationship between welfare weight and income. 41
3.3 The Generalised Lorenz Curve Comparison 43
3.4 Distribution of Income and Distribution of Social Utility 44
3.5 The Atkinson and Dalton Indices 45
3.6 The Theil Curve 49
3.7 Theil's Entropy Index 51
3.8 A variety of distance concepts 53
3.9 Lorenz Curves for Equivalised Disposable Income per Person. Switzerland and USA. 64

3.10 Inequality Aversion and Inequality Rankings, Switzerland and USA. Source: as for Figure 3.9 64

4.1 The Normal Distribution 72
4.2 The Lognormal Distribution 73
4.3 The Lorenz curve for the Lognormal distribution 75
4.4 Inequality and the Lognormal parameter σ^2 76
4.5 The Pareto Diagram. UK Income Before Tax, 1984/5. Source: *Economic Trends*, November 1987. 79
4.6 The Pareto Distribution in the Pareto Diagram 80
4.7 Paretian frequency distribution 81
4.8 The Lorenz curve for the Pareto distribution 83
4.9 Inequality and Pareto's α 84
4.10 The Distribution of Earnings. UK Male Manual Workers on Full-Time Adult Rates. Source: New Earnings Survey, 1998 85
4.11 Pareto Diagram. UK Wealth Distribution 1995. Source: Inland Revenue Statistics 1998, Table 13.3 88

5.1 Frequency Distribution of Income, UK 1996/97, Before and After Tax. Source: Inland Revenue Statistics, 1998, Table 3.3 93
5.2 Disposable Income (Before Housing Costs). UK 1996/7. Source: Households Below Average Income, 1999 94
5.3 Disposable Income (After Housing Costs). UK 1996/7. Source: Households Below Average Income, 1999 96
5.4 Income Observations Arranged on a Line 102
5.5 Frequency Distribution of Disposable Income, UK 1996/7 (After Housing Costs), Unsmoothed. Source: as for Figure 5.3 103
5.6 Estimates of Distribution Function. Disposable Income, UK 1996/7. (After Housing Costs), Moderate Smoothing. Source: as for Figure 5.3 104
5.7 Estimates of Distribution Function. Disposable Income, UK 1996/7. (After Housing Costs), High Smoothing. Source: as for Figure 5.3 104
5.8 Frequency distribution of income before tax. US 1995 107
5.9 Lower Bound Inequality, Distribution of Income Before Tax. US 1995 108
5.10 Upper Bound Inequality, Distribution of Income Before Tax. US 1995 108
5.11 The coefficient of variation and the upper bound of the top interval. 111
5.12 Lorenz Co-ordinates for Table 5.2 112
5.13 Upper and Lower Bound Lorenz Curves 113
5.14 The "split histogram" compromise. 115
5.15 The Atkinson index for grouped data, US 1995. Source, as for Table 5.2. 118
5.16 The Atkinson Index for Grouped Data: First interval deleted. Czechoslovakia 1988 120

5.17 The Atkinson Index for Grouped Data: All data included. Czechoslovakia 1988 121
5.18 Fitting the Pareto diagram for the data in Table 5.2 126
5.19 The minimum income growth to offset a 1% growth in inequality 129

A.1 Relationships Between Functional Forms 145
A.2 Density Estimation with a Normal Kernel 155

Preface to the First Edition

> "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 (1945), *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 weed-killer. 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 161 tff. 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 full bibliography follows the notes. References to the bibliography (in either the main text or the notes) are made by citing the author, or the author and date. If more than one work in the same year is cited, these are distinguished by appending 'a', 'b' as appropriate; thus: von Obertrauser (1976a).

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

$$\sum_{i=l}^{n} x_i$$

representing the sum of n numbers indexed by the subscript i, thus: $x_1 + x_2 + x_3 + ... + x_n$. 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 the appendix in which formal definitions are listed, and where proofs of some of the assertions in the text are given. The appendix also serves as a glossary of symbols used for inequality measures and other expressions.

I would like to thank Professor M. Bronfenbrenner for the use of the table on page 86, and Dr T. Stark and Professor A. B. Atkinson for each allowing me to see in advance copies of forthcoming work. 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.

Preface to the Second Edition

A lot has happened to the pattern of income distribution in the UK and the USA (the two countries from which the worked examples were principally drawn in the first edition), but since this book is principally about measurement rather than about economic history I have not altered the structure of the text much to accommodate these changes.

However, in some respects, recent changes have had to be accommodated in this new edition. A wealth of material has appeared in the literature on the axiomatic foundation of inequality analysis, on the formal relationships between inequality and social welfare, and related topics. This material is now covered in chapters 2 and 3. Some of the developments in techniques of estimation and computation are covered in the new chapter 5. But perhaps the most important changes concern data availability. The data series on which several of the examples in the early chapters in the first edition were based – the CSO "Blue Book" series published in *Economic Trends* – is now missing, presumed dead. For the sake of continuity I have included the last known representative of this series in the present edition (see page 17), but for practical-minded readers to make realistic progress I have also included the early results a more recent UK data source: *Households Below Average Income*. These data and the other data sources are described in more detail in the expanded "Notes on Sources" section – see page 161.

This new material is one of the reasons for a further innovation in the second edition. Packaged with this book you will find a diskette; this contains - in both ASCII files and Lotus 1-2-3TM worksheet WK3 files [this is now handled on the Website, http://sticerd.lse.ac.uk/research/frankweb/MeasuringInequality/index.html] - all the material used for the examples provided in the second edition, some of which is not easily available elsewhere. Although you should be able to read the text without having to use it, 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 this edition now comes with a number of questions and problems at the end of chapter: you may well find that the easiest course is to pick up the data for these straight from the diskette [website] rather than doing them by hand or keying in the numbers into a computer yourself. This is described further in the Technical Appendix (page 159), but to get going with this diskette you only need a standard IBM PC-compatible: put the diskette in drive **A**: and look at the README.TXT file for guidance [go to the welcome page of the website].

In preparing the second edition I have 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 2 on page 134. Christian Schlüter helped greatly with the 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.

STICERD, LSE

Chapter 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. Indeed, a few 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.
- *Equalisation 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 crystallisation:* 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 you might want to know about the properties of inequality measures could try turning to 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. The Appendix 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 definition of definitions and results that are described verbally or graphically in the main text. The Notes section has been used mainly to tidy away 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. The Notes section 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 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. Firstly, 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. Secondly, 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 of 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. Firstly, 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. Secondly, 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 recognised:

- 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 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 measured inequality and total income in Chapter 3 (particularly onwards), and examine the practical implications of a growing – or dwindling – cake in Chapter 5 (see page 128.

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 summarises 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.

1.3.1 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 straightjacket 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.

Firstly, 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. In fact we shall see some practical examples (in Chapters 3 and 5) where we do exactly that.

Secondly, 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 Figure 1.1, which reveals that there is an amount I_1 of type-1 inequality, and I_2 of type-2

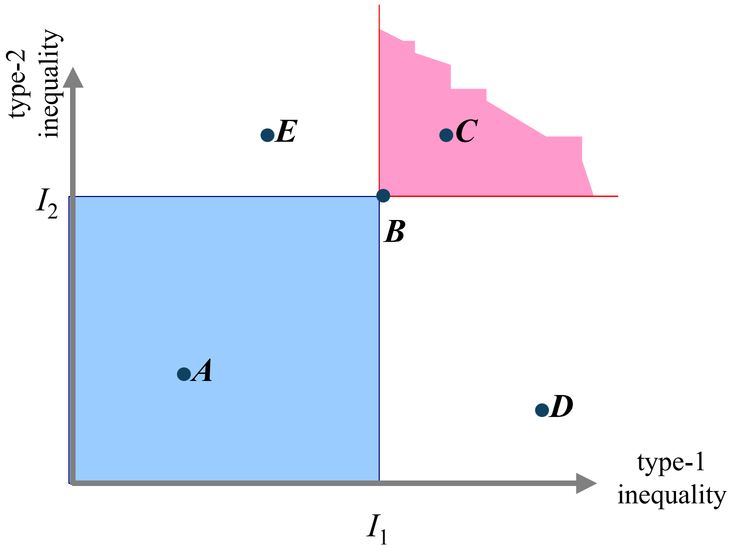

Figure 1.1: Two Types of Inequality

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 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 Figure 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.

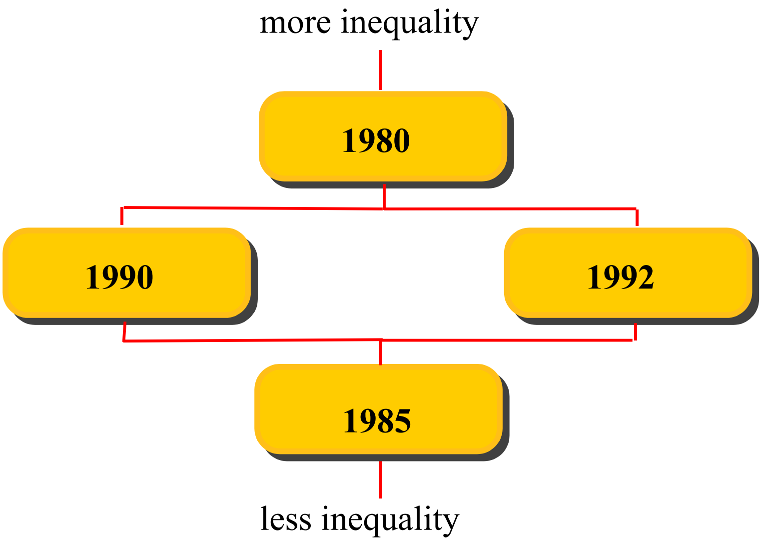

Figure 1.2: An Inequality Ranking

	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

Table 1.1: Four inequality scales

1.3.2 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, realised 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 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): inequality now needs to have the same kind of "numerical representation" as temperature on a thermometer.

1.3.3 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 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 5.1.4).

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 would be 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.

[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, i.e.$I_1 = f(I_4)$, where $dI_1/dI_4>0$. An example of such a function is $\log(I)$.

1.3.4 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. Firstly, 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. 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.

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 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 are 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 advantage – and perhaps concern about the inordinately rich - that transcends the personal twinges of altruism and envy. 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 Figure 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 4 on page 13.) 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, the nuclear family or the 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 characterised 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 heterogenous 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: where $\alpha > 0$ and $0 < \beta < 1$.

2. Because of this, they argue, inequality of consumption must be less than inequality of income. Provide an intuitive argument for this. Ruritanian society consists of three groups of people: Artists, Bureaucrats and Chocolatiers. Each Artist has high income (15 000 Ruritanian Marks) with a 50% probability, and low income (5 000 RM) with 50% probability. Each Bureaucrat starts working life on a salary of 5 000 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?

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

4. In 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].

Chapter 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 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% to wages, 35% 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. In fact 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 them up in order of height and let them march past in some given time interval – let us say one hour. Then the sight that meets our eyes is represented by the curve in Figure 2.1. 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 income 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 OD crossing the base line towards its left-hand end. And in order to keep the diagram on the page, we have plotted the point 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 the notes to this chapter. An example is found in Figure 2.2. Suppose you are looking down on a field. On one side - the axis $0y$ - there is a long straight fence marked off 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

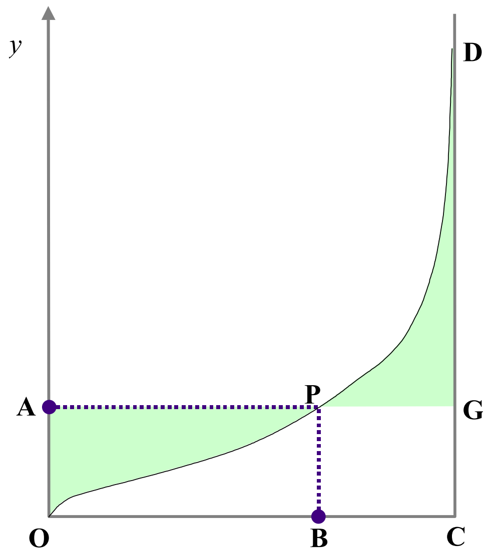

Figure 2.1: The Parade of Dwarfs. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

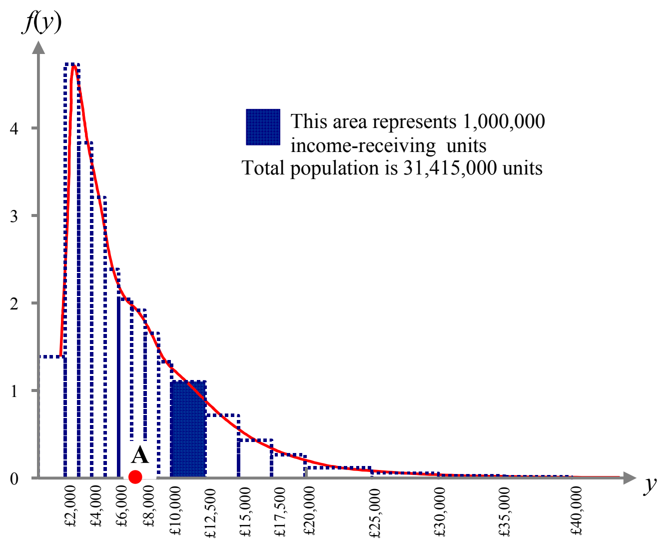

Figure 2.2: Frequency Distribution of Income. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

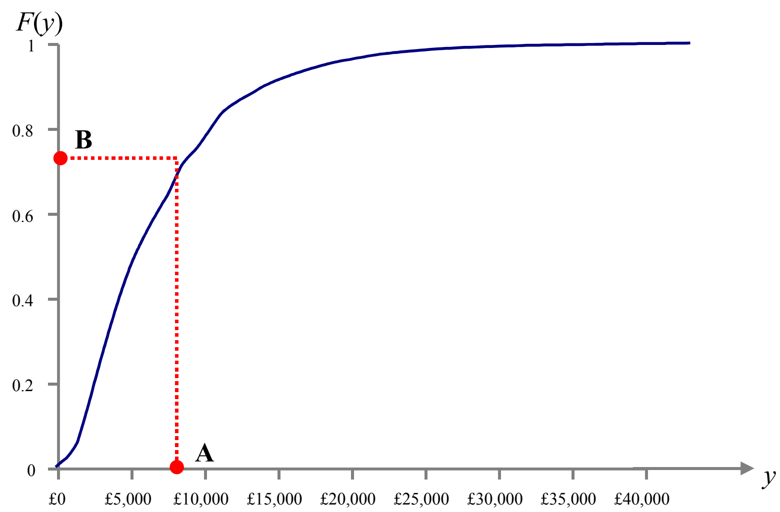

Figure 2.3: Cumulative Frequency Distribution. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

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 Figure 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 Figure 2.2. The relationship $f(y)$ charted by this curve is sometimes known as a *density function*, where the scale is chosen such that the area under the curve and above the line $0y$ is standardised at unity.

The frequency distribution shows what is happening in the middle income ranges more clearly. But perhaps it is not so readily apparent what is happening 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 Figure 2.3 from Figure 2.2. The horizontal scale of each figure is identical. On the vertical scale of Figure 2.3 we plot "cumulative frequency". For any income y this cumulative frequency, written $F(y)$, is proportional to the area under the curve and to the left of y in Figure 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 $y = \pounds 30\ 000$, we plot a point in Figure 2.3 that corresponds to the proportion of the population with $\pounds 30\ 000$ or less. And we can repeat this operation for every point on either the empirical

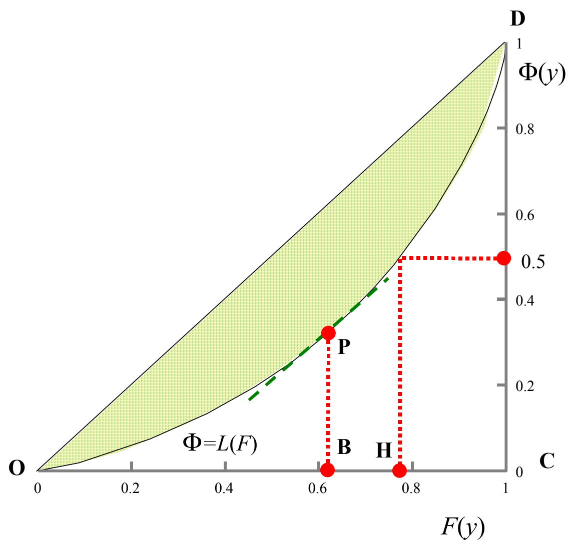

Figure 2.4: Lorenz Curve of Income. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

curve or on the smooth theoretical curve.

The visual relationship between Figures 2.1 and 2.3 is now obvious. As a further point of reference, the position of mean income has been drawn 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 the 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 Figure 2.4. Once point C is reached everyone has gone by, so $F(y) = 1.0$. Now as each person passes, hand him his share of the "cake" - i.e. 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 Figure 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% have filed by $(F(y_1) = .10)$and you have handed out 4% of the cake $(\Phi(y_1) = .04)$; then by the time the next 10% of the people go by $(F(y_2) = .20)$, you must have handed out at least 8% of the cake $(\Phi(y_2) = .08)$. Why? - because we arranged the parade in ascending order of cake-receivers. Notice too that if the Lorenz curve lay along OD we

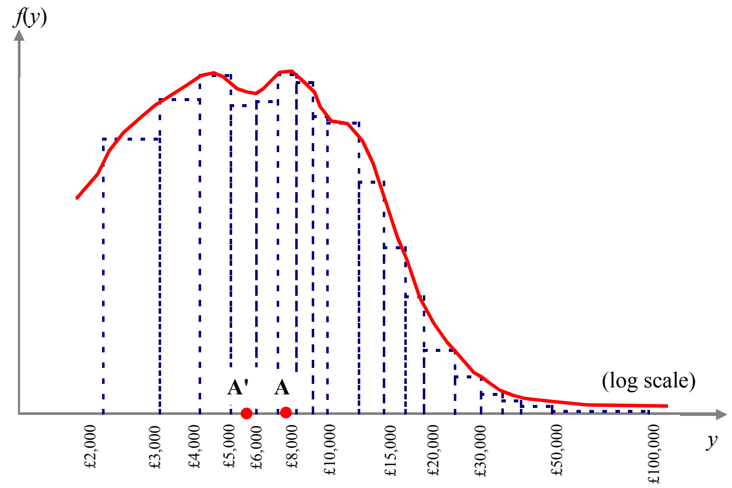

Figure 2.5: Frequency Distribution of Income (Logarithmic Scale). UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

would have a state of perfect equality, for along that line the first 5% get 5% of the cake, the first 10% get 10% ... 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 30) and also in Chapter 3 (pages 42 and 55). And again there is a nice relationship with Figure 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 Figure 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 Figure 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. If incomes are non-negative, then 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 emphasise 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 in emphasis is partly reflected in the inequality measures derived from the diagrams.

2.2 INEQUALITY MEASURES

We can use Figures 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 it we will briefly consider their principal properties: (a proper job must wait until page 60, 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. Firstly, we have the range, which we define simply as the distance CD in Figure 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 standardise 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% and the top 5% 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 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 Figure 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 is this area, the greater is inequality. (Try drawing the Parade with more giants and more dwarfs.) It is conventional to standardise the inequality measure in unit-free terms, so let us divide by the total income (which equals area OCGA). In terms of the diagram then the relative mean deviation is then:[1]

$$M = \frac{\text{area OAP} + \text{area PGD}}{\text{area OCGA}}$$

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 Figure 2.6.

[1]You are invited to check the technical appendix (pp. 135 ff) for formal definitions of this and other inequality measures.

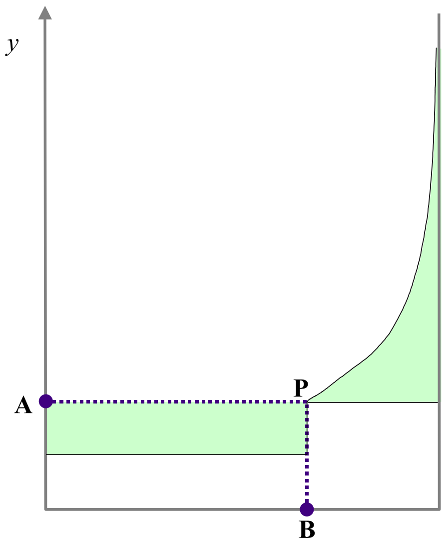

Figure 2.6: The Parade with Partial Equalisation

But notice that the two shaded regions in Figure 2.6 are exactly the same area as in Figure 2.1: so the value of M has not changed. Whatever reallocation you arrange among people to the left of C only, or among people to the right of C only, inequality according to the relative mean deviation stays the same.

The relative mean deviation can be easily derived from the Lorenz curve (Figure 2.4). From the Technical Appendix, page 139. 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 Figure 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 page 137 and 139 for a formal definition. The main disadvantage of G is that its places a rather curious implicit relative value on changes that may occur in different parts of the distribution. An income transfer from a richer person to a poorer person has a much greater effect on G if the two persons are near the middle rather than at either end of the parade. 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 Figure

2.4. Two have been suggested in connection with the problem of measuring inequality in the distribution of power, as reflected in voting strength. Firstly, consider the income level y_0 at which half the national cake has been distributed to the parade; i.e. $\Phi(y_0) = \frac{1}{2}$. 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. Secondly, 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 (the equal shares coefficient): in effect they measure changes in inequality by only recording transfers between two broadly based groups.

Now let us consider Figures 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*:

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

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 Figure 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 criticised more generally on grounds similar to those on which G was criticised. 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 have 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 1 to 5 where we have = £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.[2]

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 curve) 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 Figure 2.2 as Figure 2.7. Stark's study concentrated specifically on UK incomes, but the ideas it embodies could be applied more generally. The distance OP in Figure 2.7 we

[2] 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 146.

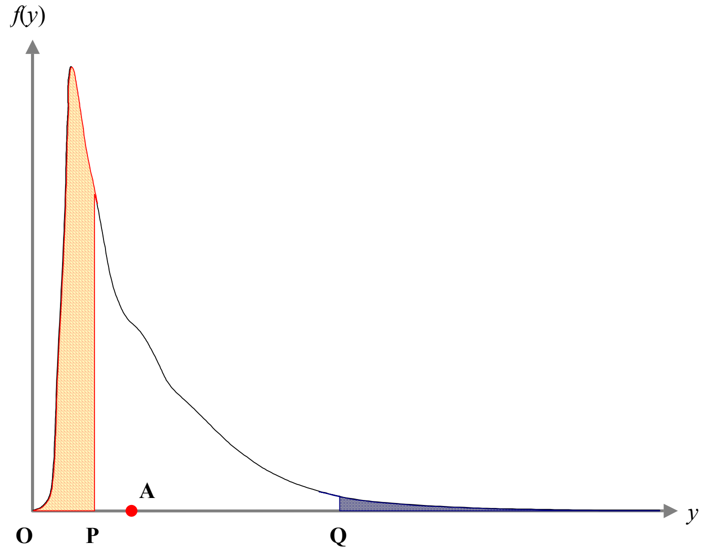

Figure 2.7: The Stark Approach

will call the range of "low incomes": in the case of the UK, P would have been fixed from the basic national assistance (supplementary benefit) scale plus a percentage to allow for underestimation of income and income disregarded in applying for assistance (benefit) – this is very similar to the specification of a "poverty line" that has been attempted by some researchers . The point R would have been determined by the level at which one becomes liable to any special taxation levied on the rich (at that time in the UK it was surtax), adjusted for need.[3] Stark's *high/low* index is then total shaded area between the curve and the horizontal axis.

The high/low index is imaginative and practical, but it suffers from three important weaknesses. Firstly, 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 among the "middle income receivers". Secondly, 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 increases only at the same rate as mean 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 increases with year-to-year increases in the social standard of reference - i.e. the national assistance scales (with a similar argument affecting the movement of point R. Then if the

[3]Note that in a practical application the positions of both P and ... a point which we are deferring until later. Figure 2.7 illustrates one type

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 the definitions of "poor" people. Some compromise between the two courses must be chosen and the results derived for a particular application treated with caution.[4] Thirdly, there is the point that in practice the contribution of the shaded area in the upper tail to the inequality measure would be negligible. Then 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 (Figures 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.

[4] There is a further complication in the specific UK application .... "been" rather than "is". 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;

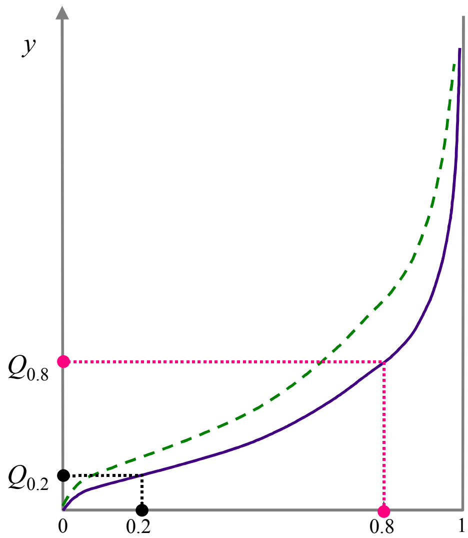

Figure 2.8: The Parade and the Quantile Ranking

We have already encountered quantiles when we were discussing the incomes of the 3rd and 57th minute people as an alternative to R (page 22). Quantiles are best interpreted using either the Parade diagram or its equivalent the cumulative frequency distribution (Figure 2.3). Take the Parade diagram and enlarge the vertical a scale a little so that it is easier to see what is going on: this gives us Figure 2.8. Mark the point 0.2 on the horizontal axis, and read off the corresponding income on the vertical axis: this gives the 20- percent *quantile* point (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-percent 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.[5]

How might we use a set of quantiles to compare income distributions? We could produce something like Figure 2.9, which shows the proportionate movements of the quantiles of the frequency distribution of earnings in the UK in

---

but the basic problem of comparisons over time that I have highlighted would remain. So too, of course, would problems of comparisons between countries.

[5] 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 you patience and your knowledge of Latin prefixes permits.

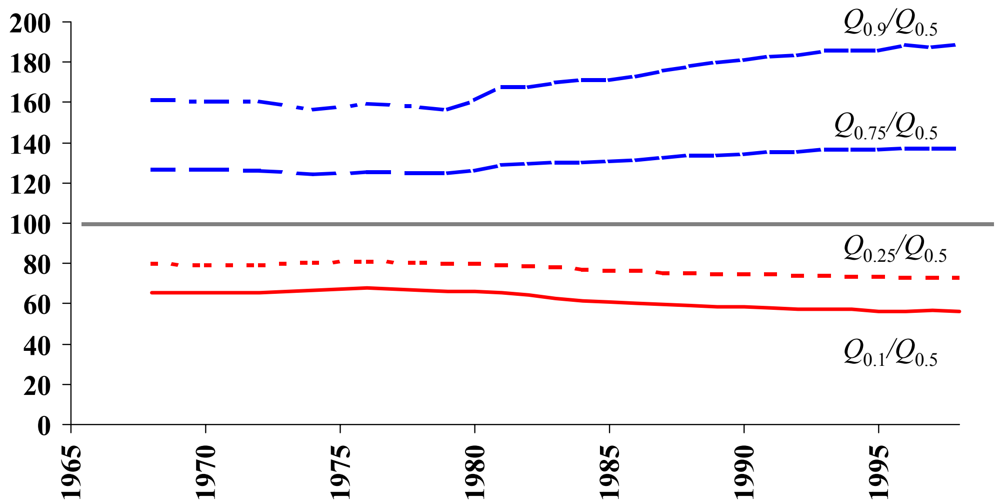

Figure 2.9: Quantile ratios of earnings of adult males, UK 1968-1998. Source: New Earnings Survey, 1998, Part A Table 28

recent years (the diagram has been produced by standardising 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 Figure 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.[6]

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 move up as shown by the broken curve in Figure 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 of a clear quantile ranking is telling us something interesting about the income distribution which we will discuss more

[6] 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 has 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% has actually fallen (from 5/36 to 4/36) and the share of the top 25% has risen (from 13/36 to 14/36).

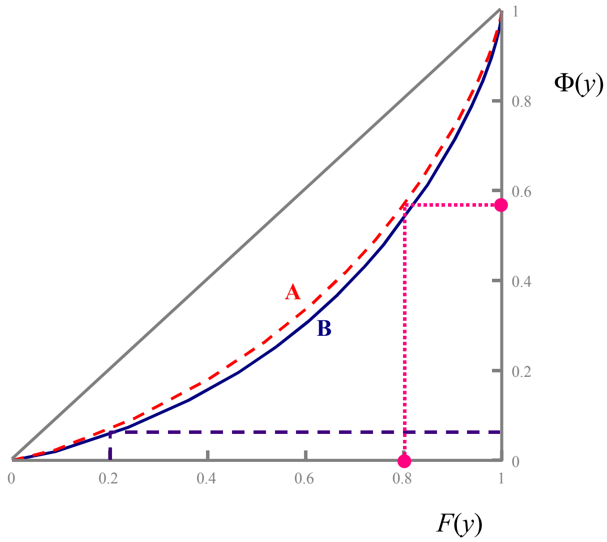

Figure 2.10: Ranking by Shares. UK 1984/5 Incomes before and after tax. Source: as for Figure 2.1

in the next chapter.

Shares by contrast are most easily interpreted in terms of Figure 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 Figure 2.10, and again do an exercise similar to that which we carried out for the quantiles in Figure 2.8: for reference let us mark in the share that would accrue to the bottom 20 percent and to the bottom 80 percent in distribution B (which is the distribution Before tax – the same as the Lorenz curve that we had in Figure 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 income. As we might have expected, Figure 2.10 shows that people in the bottom 20 percent 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 percent 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 percent 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

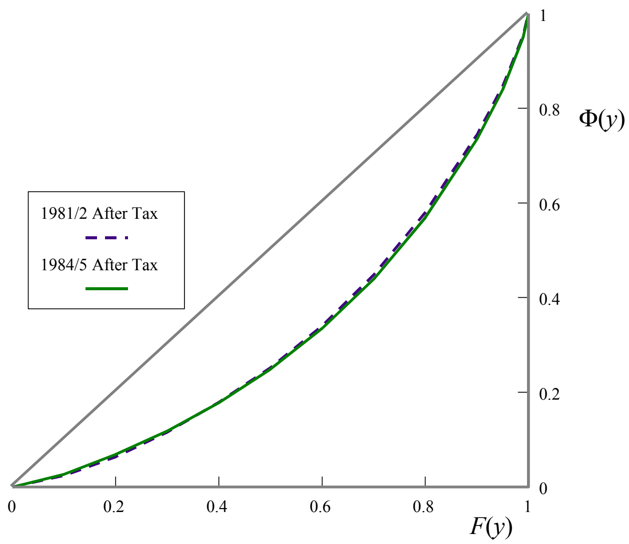

Figure 2.11: Lorenz Curves Crossing

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 just 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 Figure2.11, which depicts the way in which the distribution of income after tax changed from 1981/2 to 1984/5. Notice that the bottom 20 percent of the population did proportionately better under 1984/5 than in 1981/2 (see also the close-up in Figure 2.12), whilst the bottom 80% did better in 1981/2 than in 1984/5 (see also Figure 2.12). 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.

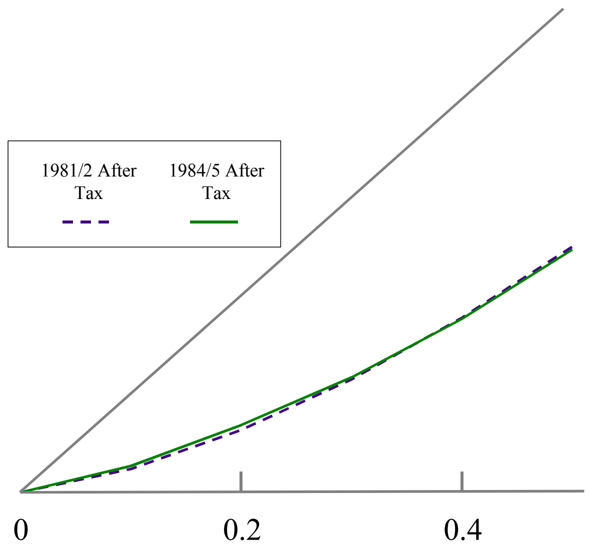

Figure 2.12: Change at the bottom of the income distribution

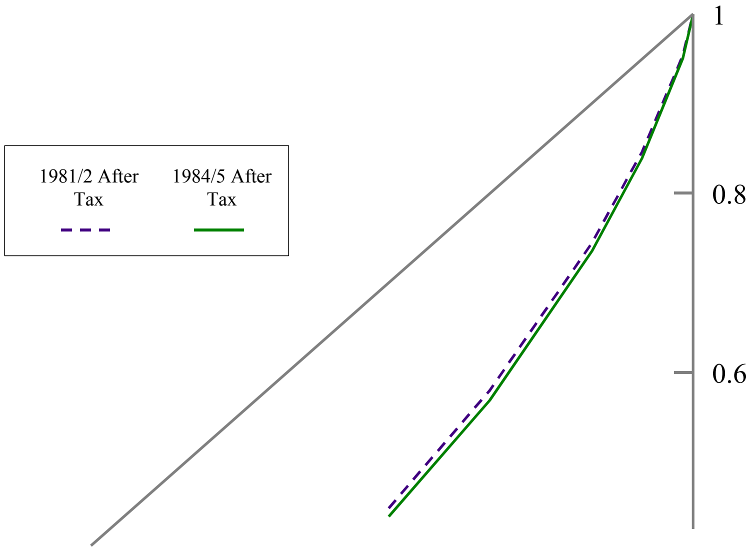

Figure 2.13: Change at the top of the income distribution

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 Pen's Parade would look if there were some individuals with negative incomes.

2. Explain how the Lorenz curve would look if (a) there were some individuals with negative incomes but mean income were still positive, (b) there were so many individuals with negative incomes that mean income itself were negative. [See the Technical Appendix, page 152. for more on this]

3. Reconstruct the histogram for the UK 1984/5, before-tax income, using the worksheet "ET84-5" on the website (see the Technical Appendix page 159 for guidance on how to use the data 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.

4. Using the same data source for the UK 1984/5, before-tax income, construct the distribution function corresponding to the histogram drawn in question 3. 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?.

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

$$T = t[y - y_0]$$

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 $y_1, y_2, ..., y_n$, use the formulas given in the Technical Appendix (page 137) to write down the coefficient of variation and the Gini coefficient for after-tax income. Comment on your results.

6. Suppose the income distribution before tax is represented by a set of numbers $\{y_{[1]}, y_{[2]}, ..., y_{[n]}\}$, where $y_{[1]} \leq y_{[2]} \leq y_{[3]}....$ Write down an expression for the Lorenz curve. If the tax system were to be of the form given in question 5, 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)].

7.

(a) Ruritania consists of six districts that are approximately of equal size in terms of population. In 1992 per-capita incomes in the six districts were:

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

What is 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 worksheet "EastWest" in the data file, and you should ignore any income differences within any one district).

(b) By 1993 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 1993 data. Did inequality rise or fall from 1992 to 1993? [See the discussion on page 59 below for an explanation of this phenomenon].

Chapter 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 judgments from 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

An obvious 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 criticised on their own merits.

In its simplest form, a SWF 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 a 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:

  $$W_{\mathrm{A}} = W(y_{1\mathrm{A}}, y_{2\mathrm{A}}, ..., y_{n\mathrm{A}})$$

  and if $y_{i\mathrm{B}} \geq y_{i\mathrm{A}}$ implies, ceteris paribus, that $W_{\mathrm{B}} \geq W_{\mathrm{A}}$, 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,

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

  That is, 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

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

  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:

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

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 $U(y_i)$ 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

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

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

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

[1] Notice that I have used a slightly different cardinalisation 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 discussed in either edition.

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 standardisation 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 standardised incomes y_i as a measure that puts everyone in the population on an equal footing as regards needs and deserts, 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 $W_{\rm B} - W_{\rm A}$, 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, $W_{\rm B} - W_{\rm A}$ 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, 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". 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: $\triangle y_1, \triangle y_2, ..., \triangle y_n$. What is the change in social welfare? It is simply

$$dW = U'(y_1)\triangle y_1 + U'(y_2)\triangle y_2 + ... + U'(y)\triangle y_n$$

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.

[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 "equivalised" incomes: see page 96 for more on this.

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
...	...

Table 3.1: How much should R give up to finance a č1 bonus for P?

- ☐ **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 have 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 income for all.

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

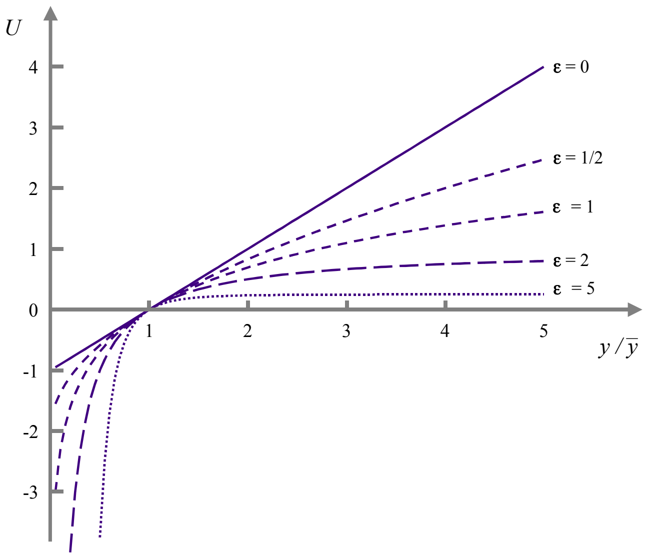

Figure 3.1: Social utility and relative income

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 Figure 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 $\varepsilon = 2$ 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.

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.

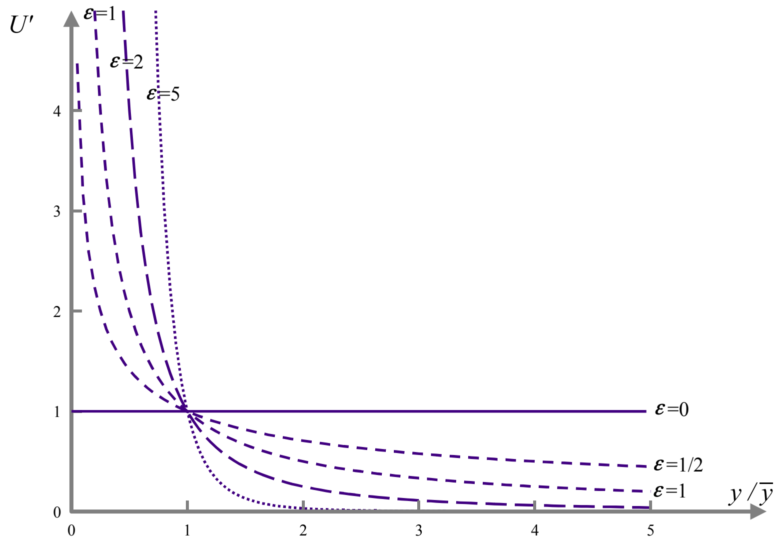

Figure 3.2: The relationship between welfare weight and income.

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 Figure 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% of the mean in order to be assigned a welfare weight 5 times as great as the weight of the person at mean income.

Let us now put the concept of the SWF to work. First consider the ranking by quantiles that we discussed in connection with Figure 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* $W_{\mathrm{A}} > W_{\mathrm{B}}$ *for any individualistic, additive and symmetric social-welfare function* W.

7So if the Parade of distribution A lies everywhere above the Parade of distribution B (as in the example of Figure 2.8 on page 28), social welfare must be higher for this class of SWFs. In fact 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* $(y_{1A}, y_{2A}, ..., y_{nA})$ *and social state B have income distribution* $(y_{1B}, y_{2B}, ..., y_{nB})$*, 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* $W_B > W_A$ *for any individualistic, increasing, symmetric and strictly concave social-welfare function* W*.*

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 Figure 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 crystallise 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 *generalised Lorenz curve* by multiplying the vertical co-ordinate of the Lorenz curve by mean income (so now the vertical axis runs from 0 to rather than 0 to 1).

Theorem 3 *The Generalised Lorenz curve for state A lies wholly above the generalised 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 Figures 2.11-2.13). Now let us consider the generalised Lorenz curves for the same two datasets, which are depicted in Figure 3.3. Notice that the vertical axes is measured in monetary units, by contrast with that for Figures 2.4 and

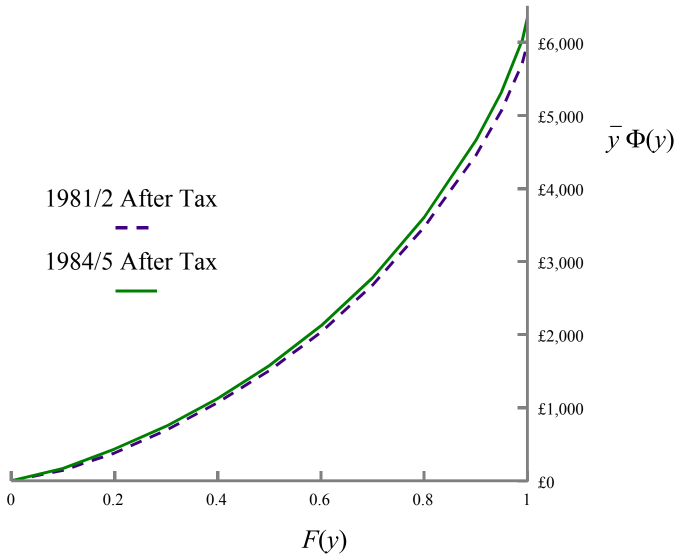

Figure 3.3: The Generalised Lorenz Curve Comparison

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 Figure 3.3 if the mean of distribution A is higher than the distribution B, then the generalised 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 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 generalised Lorenz curves. Moreover we often desire a unique numerical value for inequality in order to make comparisons of different changes in inequality. This is an issue that we shall tackle right away: we use the social-welfare function to find measures of inequality.

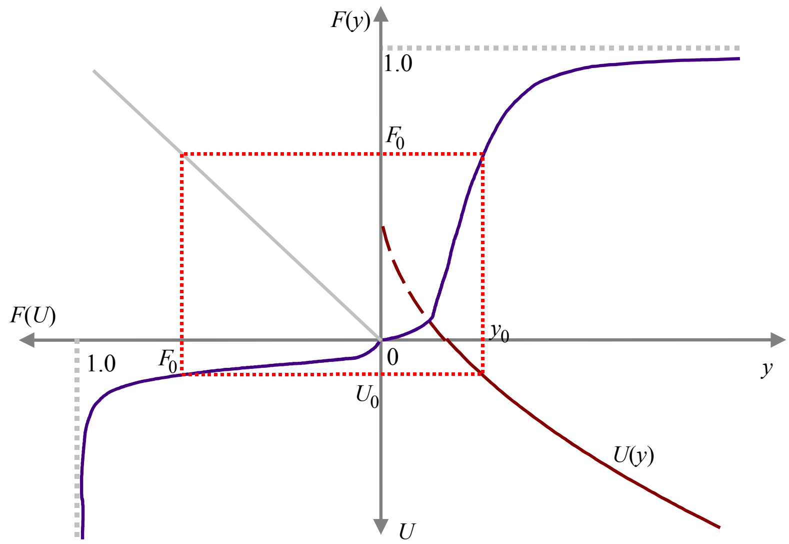

Figure 3.4: Distribution of Income and Distribution of Social Utility

3.2 SWF-BASED INEQUALITY MEASURES

In fact from 1 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 Figure 2.2 – and the frequency distribution of social utility U.

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

In the top right-hand quadrant you will recognise 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

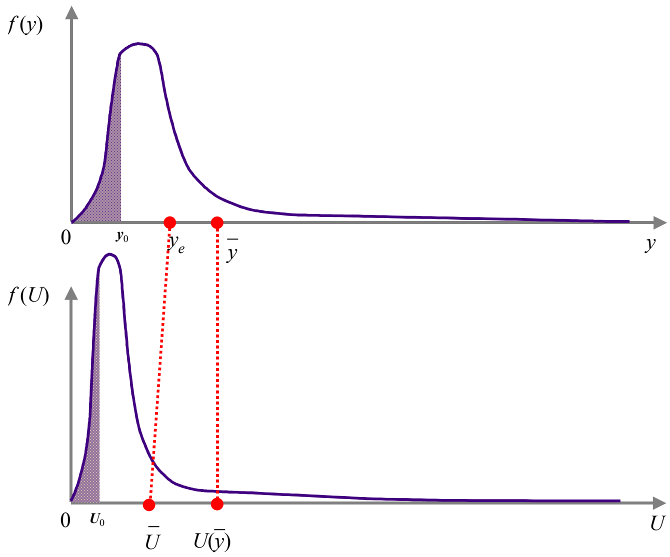

Figure 3.5: The Atkinson and Dalton Indices

plot the F and U-values in a new diagram (bottom left-hand corner) - this is done by using 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 Figure 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 Figure 3.5: notice that the points y_0 and U_0 correspond to the points y_0 and U_0 in Figure 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 Figure 3.5) mark the position of the mean, $\bar{y}$, on $0y$. Do the same for the distribution of social utility – point $\bar{U}$ on $0U$. 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 Figure 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 Figure 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 Figure 3.4; if ε were changed, then so would the frequency distribution in the bottom half of Figure 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 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

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

We may note that this is zero for perfectly equally distributed incomes (where we would have $\bar{U} = U(\bar{y})$. (Atkinson 1970) criticises 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, 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 $\bar{y}$. So define *Atkinson's Inequality Index* (for inequality aversion ε) as

$$A_\varepsilon = 1 - \frac{1}{\bar{y}}U^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\right)U(y_i)$$

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

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

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

which means that

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

Clearly the choice between 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.[3] Of much greater significance is the choice of the value of ε, especially where Lorenz curves intersect, as in Figure 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. If a low value of ε is used we are particularly sensitive to changes in distribution at the top end of the parade; if a high value is employed, then it is the bottom end of the parade which concerns us most – 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

[3] 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 as A_ε. However, even apart from the fact that $D_\varepsilon$depends on the cardinalisation 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.

attach probabilities p_1, p_2, p_3,... Each p_i is not less than zero (which represents total impossibility of 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, 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

$$h(p_1 p_2) = h(p_1) + h(p_2) \tag{3.3}$$

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 "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 characterise 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 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, where of course if is mean income, and y_i is the income of person i:

$$s_i = \frac{y_i}{n\overline{y}}$$

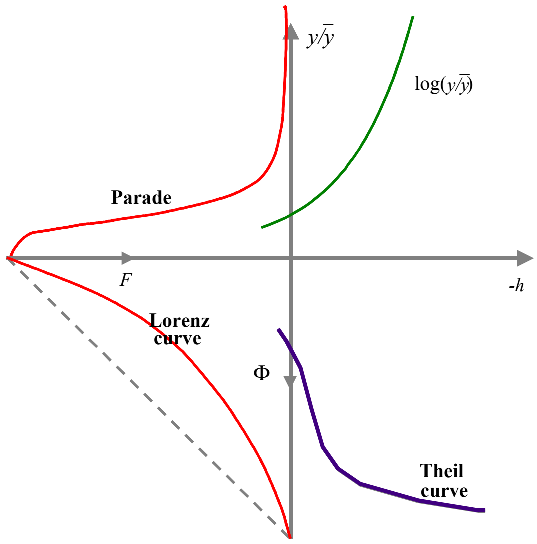

Figure 3.6: The Theil Curve

so that, of course:

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

Then subtracting the actual entropy of the income distribution (just replace all the $p_i s$ with $s_i s$ 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.

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

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

A diagrammatic representation of T can be found in Figures 3.6 and 3.7. In the top right-hand corner of Figure 3.6, the function $\log(\frac{y}{\bar{y}})$ is plotted (along the horizontal axis) against $\frac{y}{y}$ (along the vertical axis). In the top left-hand corner we have the Parade, slightly modified from Figure2.1, whiles 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 Figure 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/μ - 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 $\frac{y}{y}$ 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 Figure 3.6 since it crosses the 0Φ axis ; to remedy this I have drawn a fuller picture of the curve in Figure 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:

$$\Delta T = \Delta s \left[h\left(s_2\right) - h\left(s_1\right) \right]$$

$$= -\Delta s \log \left(\frac{s_2}{s_1} \right)$$

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.

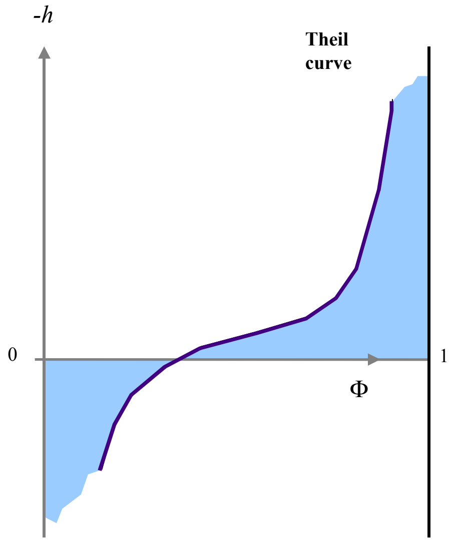

Figure 3.7: Theil's Entropy Index

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 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 aids 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,[4] although $h(s) = -\log(s)$ gives us a reasonably sensible inequality measure, T. The function, $-\log(s)$ can be seen as a member of a much wider class of functions, illustrated in Figure 3.8. This figure charts members of the

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

family of curves given by[5]

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

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:

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

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

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

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

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

$$= \left[h\left(s_2\right) - h\left(s_1\right)\right]\Delta s$$

You get the same effect by transferring s 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" $h(s_2) - h(s_1)$.

The special case where $\beta = 0$ simply yields the measure T once again. As we noted, this implies a relative concept of distance: 1 and 2 are the same distance apart as 3 and 4 if the ratios s_1/s_2 and s_3/s_4 are equal. Another interesting special case is found by setting $\beta = 1$. 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* indexis simply

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

[5] 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.

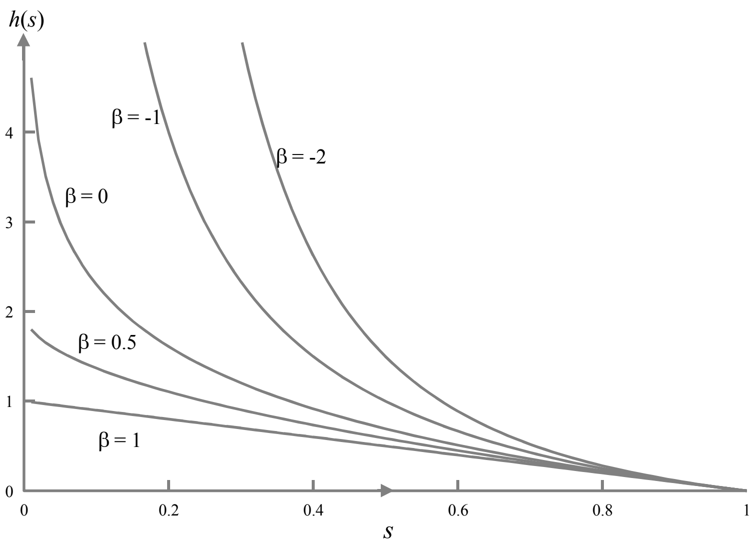

Figure 3.8: A variety of distance concepts

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

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

Thus the distance between a person with a 1% share and one with a 2% share is considered to be the same as the distance between a person with a 4% share and one with a 5% 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 Figure 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 Figure 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 the accompanying panel: the top part of this gives the income for poor P, rich R and quite-well-off Q and their respective shares in total income (which is £1 000 000); the bottom part gives the 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 β=-ε

	income	*share*
person P	£2 000	0. 2%
person Q	£10 000	1%
person R	£50 000	5%
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	$\log(\frac{s_i}{s_j})$	5	5
1	$s_j - s_i$	0.008	0.04

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

the (P,Q)-gap is ranked as greater than the (Q,R)-gap; for β=1 the reverse is true; and for β=0, the two gaps are regarded as equivalent.

Notice the obvious formal similarity between choosing one of the curves in Figure 3.8 and choosing a social utility function or welfare index in Figure 3.1. If we write β = - ε , 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 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 Figure 2.10 on page 30 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 emphasises the importance of this principle for choosing between inequality measures. As we have seen $V, c, G, T, H,\ A_\varepsilon,\ D_\varepsilon\ (\varepsilon > 0)$ 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.[6]

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.[7] This is immediately obvious 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 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

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

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

East
A:(6,7,8)
B:(6,6,9)

	A	B
$\bar{y}$	7.00	7.00
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 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

Table 3.3: The break-down of inequality: poor East, rich West

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 sub-group. 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 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

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	.126	0.149

Table 3.4: The break-down of inequality: the East catches up

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, then 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. Table3.3 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.

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.[8]

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 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.

In fact 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 *generalised entropy measures.* A quick comparison of this formula with that of the modified information-theoretic measures defined on page 52, shows that the two are very closely related: in fact the generalised entropy measures are just the modified-information theoretic family again, now normalised so that they satisfy the population principle, and with the parameter θ set equal to $\beta - 1$.[9] In view of this "family connection"

[8] In fact there is a bit more to the decomposability story and the Gini coefficient, which is explained in the technical appendix – see page 147.

[9] 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 normalisation of the Generalised 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.

it is clear that the generalised entropy measures has other connections too: inspection of the generalised 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.

As with our discussion of welfare-based and information- theory based measures we have 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 what 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 to use in the pattern, of we might consider first the broader choice between families of curves or shapes - squares, circles, triangles, ellipses,... – 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 generalised 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 51 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 Figure 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 Figure 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 generalised entropy family which, as we have just seen is its 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 generalised 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 formalise this. For, given a particular d, one may derive a particular inequality measure by using the strong principle as a fundamental axiom.[10] 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.[11] 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 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

[10] For the other axioms required see Cowell and Kuga (1981) and the discussion on page 167 which give an overview of the development of this literature.

[11] This is clear from the second in the three ways in which the information-theoretic measure was written down on page 52.

if they are on opposite sides. This property can be deduced from the effect of the particular redistribution illustrated in Figure 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 location 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 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 normalised 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 normalised without altering its cardinal properties simply by dividing by that maximum value.[12]

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. In fact, the family E_θ (all 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 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.

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

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 interest will be ordinally equivalent to the generalised 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 E_θ, for $\theta = 1 - \varepsilon$ where $\varepsilon > 0$). Let us then take the generalised entropy family of measures[13] – 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* of the inequality index to information about different parts of the distribution. We have already seen this issue in our discussion on page 54 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 generalised 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 39 how specification of the parameter ε implies a particular willingness to trade income loss from the leaky bucket against further equalisation 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 Figure 3.9[14]. 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 percent in Switzerland is higher than the USA, but so too is the share of the top ten percent. 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 Figure 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.

[13] 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.

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

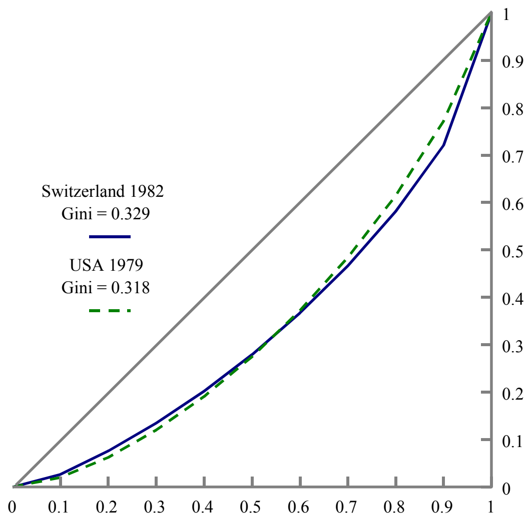

Figure 3.9: Lorenz Curves for Equivalised Disposable Income per Person. Switzerland and USA.

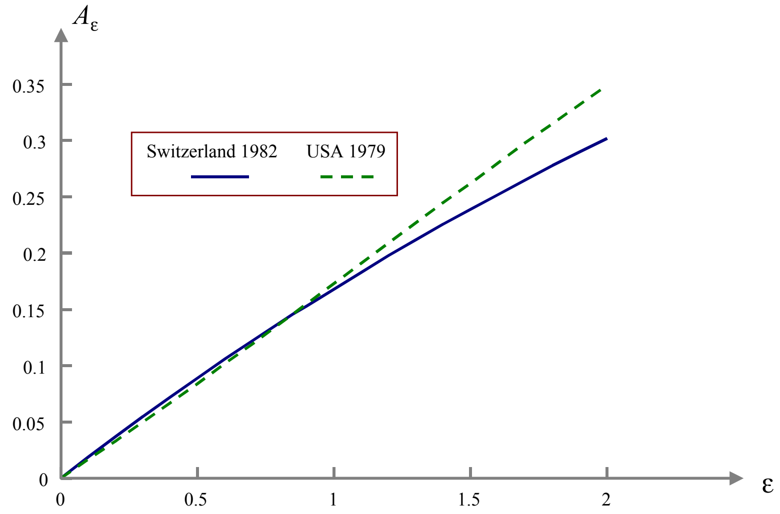

Figure 3.10: Inequality Aversion and Inequality Rankings, Switzerland and USA. Source: as for Figure 3.9

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, then, 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 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. Standardisation of the measure in a given range (such as 0 to 1) has only a superficial attractiveness to recommend it: it may be well worth while 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 summarised 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.

	Principle of Transfers	Distance Concept
Variance, V	strong	Absolute differences
Coeff. of variation, c	weak	As for variance
Relative mean deviation, M	just fails	0, if incomes on same side of $\bar{y}$, or 1 otherwise
Logarithmic variance, v	fails	Differences in (log-income)
Variance of logarithms, v_1	fails	As for logarithmic variance
Equal shares coefficient	just fails	As for relative mean deviation
Minimal majority	just fails	Similar to M (critical income is y_0, not $\bar{y}$)
Gini, G	weak	Depends on rank ordering
Atkinson's index, A_ε	weak	Difference in marginal social utilities
Dalton's index, D_ε	weak	As for Atkinson's index
Theil's entropy index, T	strong	Proportional differences
Herfindahl's index, H	strong	As for variance
Generalised entropy, E_θ	strong	Power function

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

Table 3.5: Which measure does what?

Decomposable?	Independent of income scale & population size?	Range in interval [0,1] ?	
Yes	No: increases with income	No	
Yes	Yes	No	
No	Yes	No: in [0,2]	
No	Yes	No	
No	Yes	No	
No	Yes	Yes	1.1
No	Yes	Yes	
No	Yes	Yes	
Yes	Yes	Yes	
Yes	No	No	
Yes	Yes	No	
Yes	No: decreases with population	Yes: but min > 0	
Yes	Yes	No	

3.7 QUESTIONS

1. Show that the inequality aversion parameter ε is the elasticity of social marginal utility defined on page 37.

2.

   (a) Use the UK 1984/5 data (see file ET84-5 on the Website) to compute Atkinson's inequality index with $\varepsilon = 2$, making the same assumptions as in question 4 of Chapter 2.

   (b) Recompute the index in part (a) after dropping the first income class from the data set. 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 data set.

3. Suppose that the assumption of constant relative inequality aversion (page 37) 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 A:(1,4,7,10,13) B:(1,5,6,10,13) 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) , 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 generalised Lorenz curve for A is steeper than the generalised Lorenz curve of B. [see Thistle (1989b)].

Chapter 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 which will enable us to get an easier grip upon particular features of the income distribution. This entails meeting some more specialised jargon, and so it is probably a good idea at the outset to consider in general terms why it is worth while 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 data set 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 characterise 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 areuseful in the analysis of inequality, however, we are not of course immediately interested in straight lines, but rather in curves which will look like Figures 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 (Figure 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 summarise 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 characterises 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 realised. 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 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

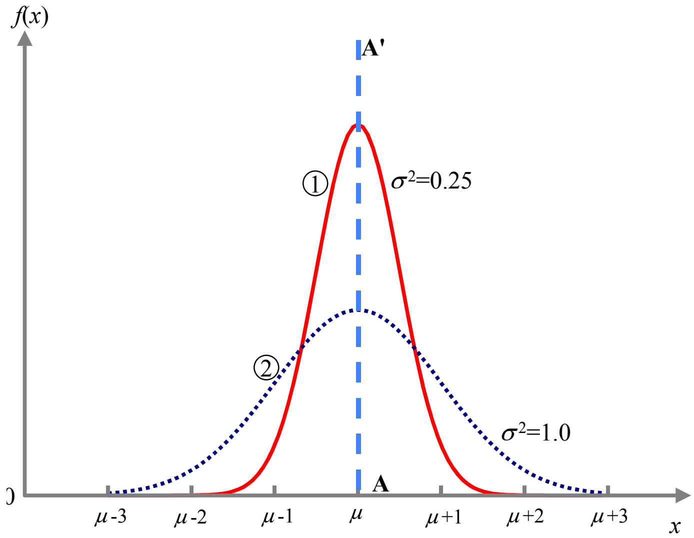

Figure 4.1: The Normal Distribution

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 Figure 4.1. As you can see, the curves are symmetrical about the line $A'A$; A marks the value ì 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 does not fit the normal curve (although there are some latter day Ptolemaians

[1] It has now been long recognised 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.

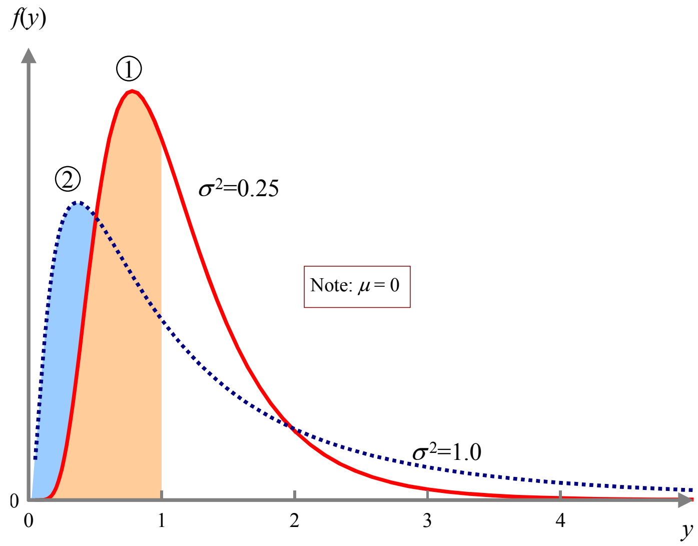

Figure 4.2: The Lognormal Distribution

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 long right-hand tail - this is even 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 Figure 3.1), we will find that it has the normal distribution like the curves in Figure 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 Figure 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 himself, why choose a logarithmic transformation to produce a distribution of the "right" shape? There are four reasons. Firstly, the lognormal distribution has a lot of convenient properties, some of which are explained below. Secondly, 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 the 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] Thirdly, 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" does. 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".[3] Fourthly, the lognormal provides a reasonable sort of fit to many actual sets of data. This I shall consider later.

- □ **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 look up 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.
- "standardise" this number using the two parameters to calculate $z = \frac{x-\mu}{\sigma}$

[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 very useful in analysing the inequality in the size distribution of firms.

[3] Incidentally, Champernowne's (1953, 1973) use of the term "income power" to describe $\log(y)$ is blameless on this score. This is simply a matter of terminological convenience so that he can look at income proportions rather than incomes themselves.

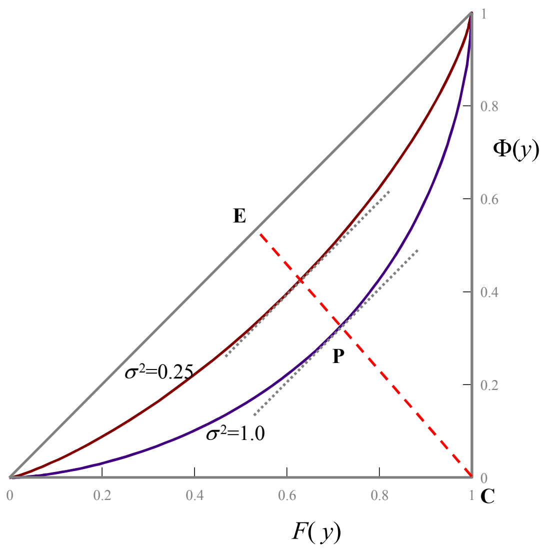

Figure 4.3: The Lorenz curve for the Lognormal distribution

- look up$F(z)$ in the 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 Figure 4.3: the Lorenz curves are symmetric about the line CP, where P 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 Figure 4.3.[4] The important conclusion to be derived from this observation is this: *given any two members of the lognormal family of distributions, one will unambiguously 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

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

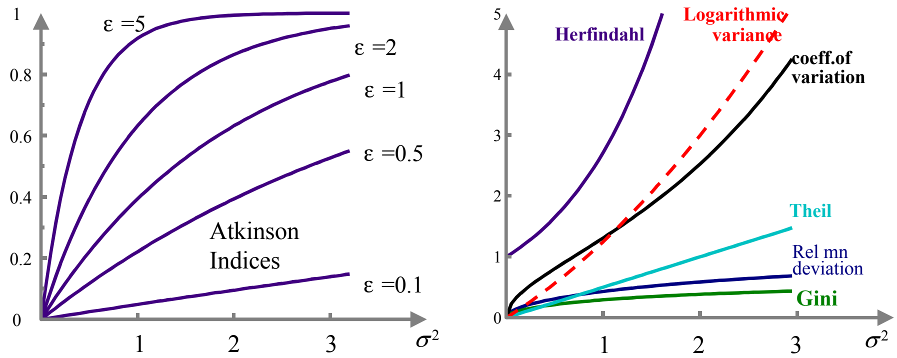

Figure 4.4: Inequality and the Lognormal parameter σ^2

lognormal distribution, written $\wedge(y;\mu,\sigma^2)$, and the normal distribution. It is evident by definition that ì 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% of the distribution lies to the left of the value $y = e^\mu$ - see the shaded area in Figure 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.[5] 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 25.

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 appendix. Some of those which were discussed in the last two chapters are sketched against the corresponding values of σ^2 in Figure 4.4. Thus to find, say, the value of the Gini coefficient in a population with the lognormal distribution, locate the relevant

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

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.

The final point may seem a little mystifying, though it can be useful. It follows in fact 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 many country's personal tax schedules are approximated reasonably by the formula

$$t = y - Ay^h$$

where t is individual tax liability and y is income.[6] 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.

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

[6] 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.

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 17. If you cover up the left-hand end of Figure 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 "Pareto's 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.[7] 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. Secondly, the distribution may once again be shown to be related to a

[7] Curious readers are invited to check the notes to this chapter for details.

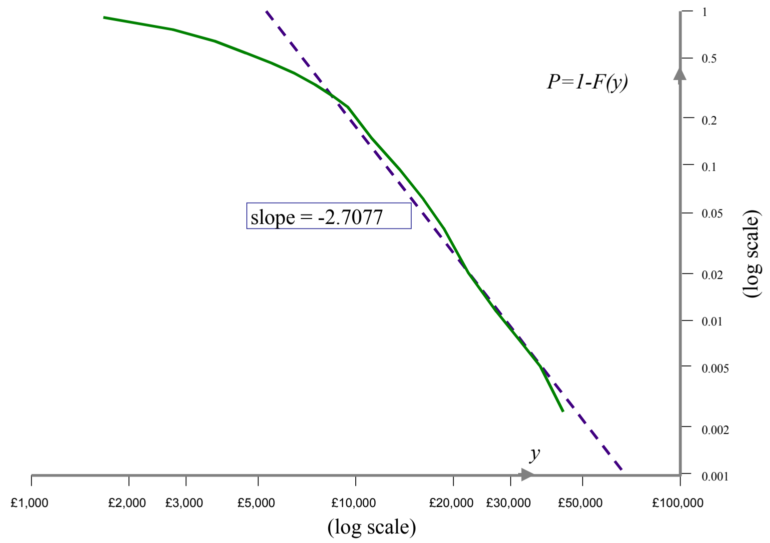

Figure 4.5: The Pareto Diagram. UK Income Before Tax, 1984/5. Source: *Economic Trends*, November 1987.

simple "random process" theory of individual income development.[8] The principle is very similar to the process referred to on page 74, the main difference being that a device is introduced to prevent an indefinite increase in dispersion over time, which as the effect of erecting a "lower barrier" income y below which no one can fall. Thirdly, the Paretian form can be shown to result from simple hypotheses about the formation of individual remuneration within bureaucratic organisations. 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. Fourthly, 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 Figures 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 [9]
- For any income level y transform the cumulative income proportions $F(y)$

[8] The details of this are set out in full in Champernowne (1953, 1973) and the non-technical reader will find a simple summary in Pen (1971 1974).

[9] This is a scale similar to that used in Figure 2.5.

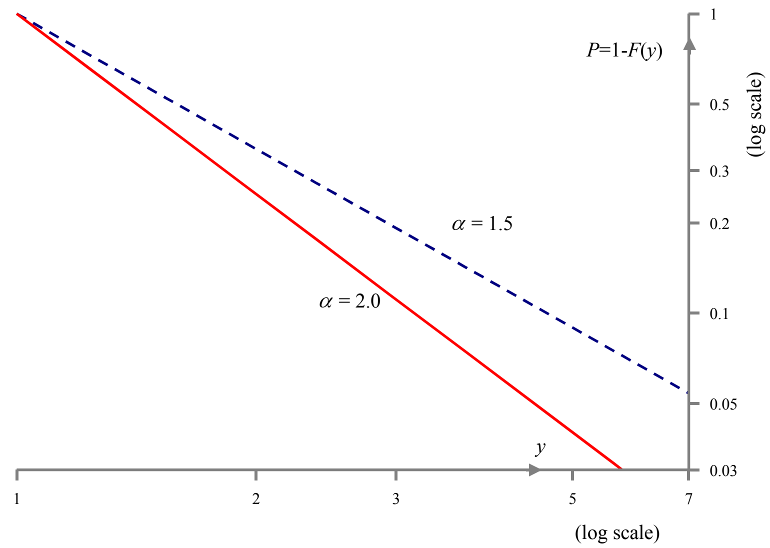

Figure 4.6: The Pareto Distribution in the Pareto Diagram

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 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 of 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 Figure 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 inter-

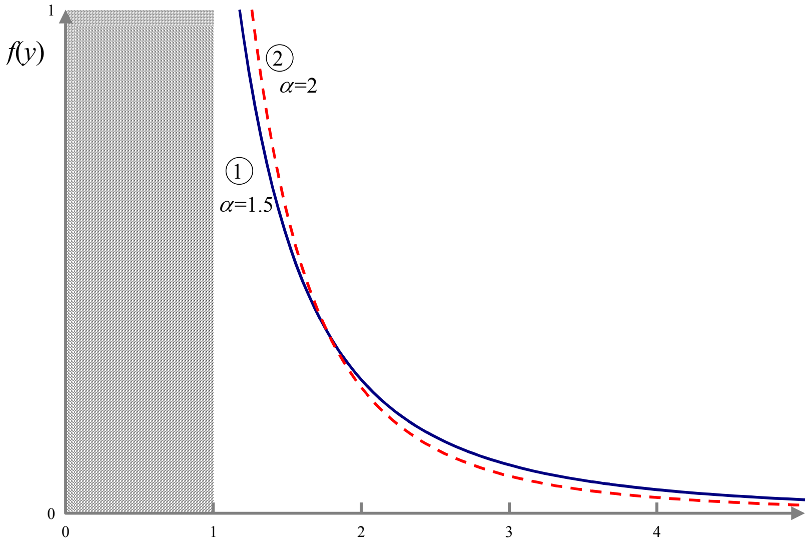

Figure 4.7: Paretian frequency distribution

cept 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 Figure 4.6. The corresponding frequency distributions are drawn in Figure 4.7. It is apparent from a superficial comparison of this picture with Figure 2.2 or other frequency distributions based on different data sets that, for income distributions at 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

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

Table 4.1: Pareto's α and "average/base" Inequality B

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.[10] 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 Figure 4.8, and can be readily inferred from the explicit formula for the Lorenz curve of the Pareto distribution given in the technical appendix. From that formula it may be seen that if we choose any value of F in Figure 4.8 (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.

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

[10] 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 143.

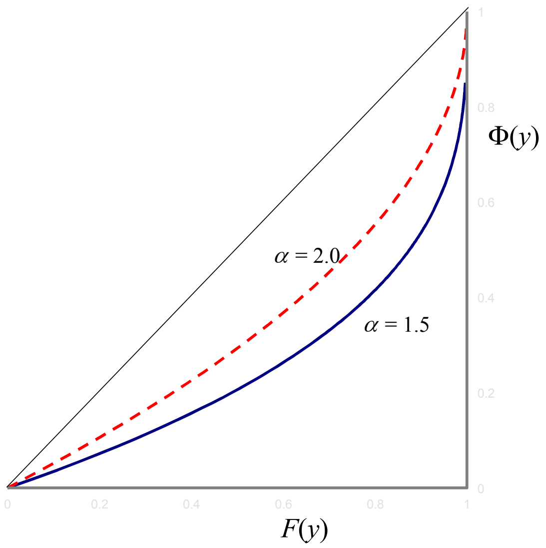

Figure 4.8: The Lorenz curve for the Pareto distribution

for all inequality measures satisfying the weak principle of transfers.[11]

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 appendix, or using Figure 4.9. This figure is to be interpreted in a manner very similar to that of Figure 4.4 in the case of the lognormal distribution. The interpretation of the parameter $\underline{y}$ can easily be seen from Figure 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.[12]

[11] 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.

[12] 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

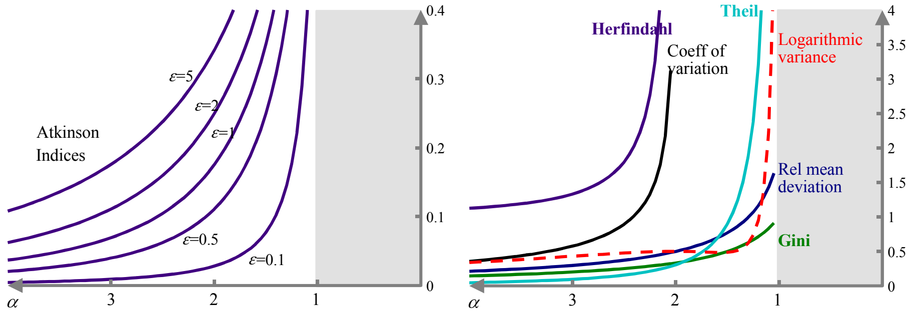

Figure 4.9: Inequality and Pareto's α

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 from the appendix we find that the proportion of the population with incomes 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 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.

income (for example the two curves in Figure 4.7). We find that social welfare *decreases* with α although, as we ave seen, inequality also decreases for any "sensible" mean-independent inequality measure. Why does this occur? It is simply that as α is increased (held constant) mean income, (which equals $\alpha/[\alpha - 1]$) 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 constant, social welfare is increased for any individualistic, additive and concave social welfare function.

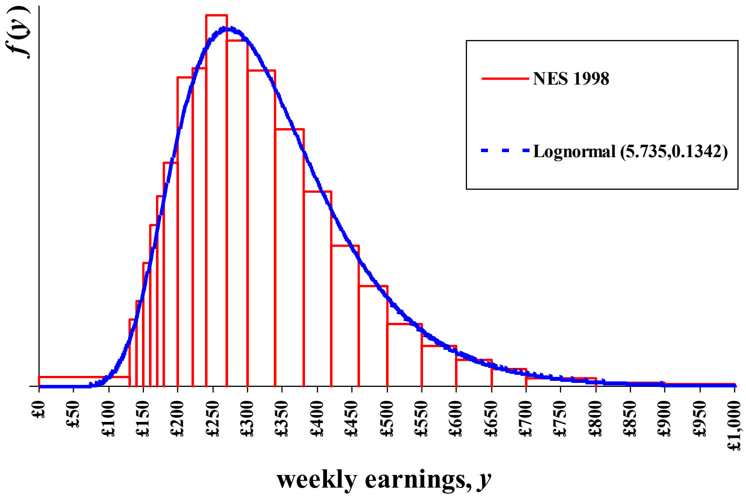

Figure 4.10: The Distribution of Earnings. UK Male Manual Workers on Full-Time Adult Rates. Source: New Earnings Survey, 1998

4.4 HOW GOOD ARE THE FORMULAS?

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 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 sectors of the labour market. 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.

When we look at more comprehensive populations a difficulty arises in that

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
1949-50	1.95	1945	1.95
1954-55	2.10		
1959-60	2.32		
1964-65	2.33		
1969-70	2.55		

Sources: For UK 1949 onwards, Board of Inland Revenue (1972),p.2.
Other data taken from Bronfenbrenner (1971), p.46
Figures refer to income before tax.

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

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 Figure 4.10 illustrates that the lognormal distribution is not 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% 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 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 Figure 2.5 with the "ideal" shape of Figure 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 Figure 4.5 bears out. In the case of incomes, the values of α tend to be in the range 1.5 to 2.5 and generally reveal a secularly increasing trend - see Table 4.2. 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 works 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 usually compiled with any accuracy only for the moderately wealthy and above. Hence – excluding those whose wealth is unrecorded – we typically find 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 Figure 4.11; notice that the straight line approximation is particularly good if we drop the first three observations (see the broken line) rather than trying to fit a line to all the observations. 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. Once again, we usually find an increase in α over time indicating, for that part of the distribution where the Paretian approximation is suitable, a trend toward greater equality.

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 (for the UK) and in Steindl (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 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 preliminary glimpse of 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 has been increasing nearly everywhere until recently.

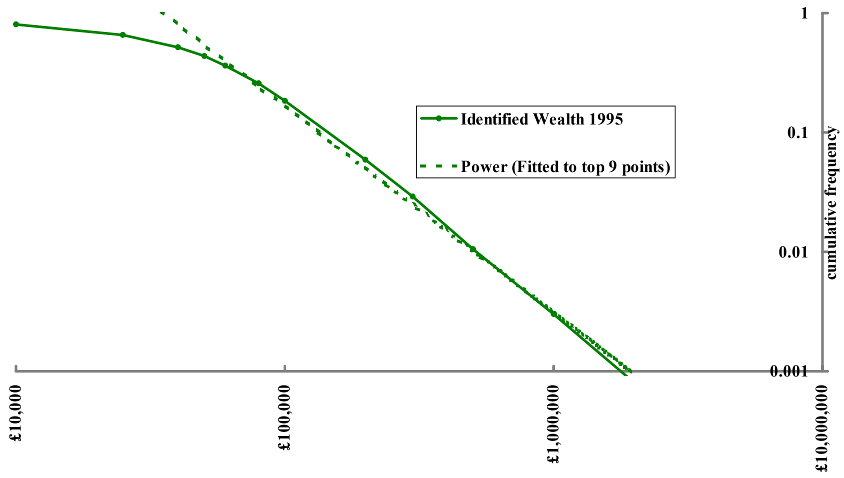

Figure 4.11: Pareto Diagram. UK Wealth Distribution 1995. Source: Inland Revenue Statistics 1998, Table 13.3

- 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,[13] 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 kronor.
- 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.

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 $\{u_1, u_1, , .., u_t, ...\}$ 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:

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

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

or $t = 1, , ...,$ 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 $\{u_1, u_1, , .., u_t, ...\}$. 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 $t = 1, 2, 3, ...$ 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 earnings distribution ("NES 1998" on the Website) compute the mean and the coefficient of variation on the assumptions (i) from the raw data and (ii) using the fitted lognormal distribution illustrated in 10 (use the relevant formula in Table A.2 on page 139).

4. Show that the "first guess" at the Pareto distribution given by the formula for the frequency distribution on page 4.3 really does correspond to the formula for the distribution function F on page 140of the Technical Appendix. What is the relationship of the constants k_1 and k_2 to the parameters y and α?

5. Use the formula given in the Table A.2 to sketch the relationship between coefficient of variation c and α in Figure 4.9. Why is c undefined for $\alpha \leq 2$?

6. Use the formula on page 140 of the Technical Appendix. to derive the generalised Lorenz curve for the Pareto distribution.

7.
   (a) Using the data for the UK wealth distribution 1995 (see the file "IR wealth 94,95" 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?

8.

(a) Using the same source on the Website as in question 7 for the lower bound of each wealth interval y compute P (as defined on page 80) and then use ordinary least squares to fit the equation

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

then the estimate of Pareto's α? Use this estimate to compute the Gini coefficient on 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 7.

(d) Suppose the Pareto-type density given on page 78 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 140.

(e) 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 114 in Chapter 5 and page A.7.2 in the Technical Appendix].

Chapter 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 issues 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.

5.1.1 *Method 1: Asking people.*

This approach is commonly used by those organisations 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 Expenditure Survey, and in the Current Population Surveys conducted by the US Bureau of the Census. Obviously the principal difficulty is, as I mentioned, that of non-response or misinformation by those approached in the survey. The 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% to

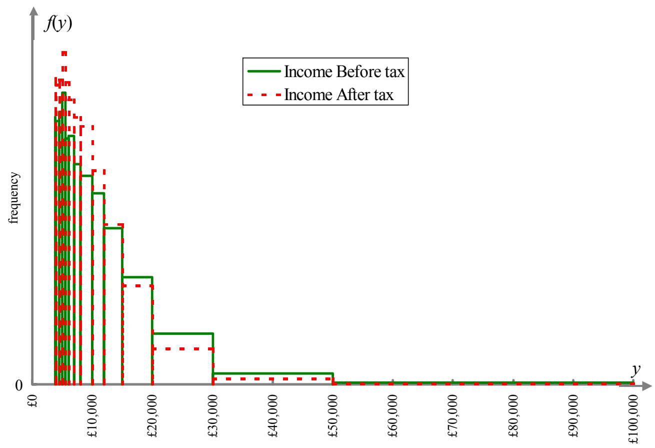

Figure 5.1: Frequency Distribution of Income, UK 1996/97, Before and After Tax. Source: Inland Revenue Statistics, 1998, Table 3.3

80%), 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 volun teers 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.

5.1.2 *Method 2: Compulsion.*

Useful information on income and wealth is often obtained as a by-product to 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 in 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 econ omic 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

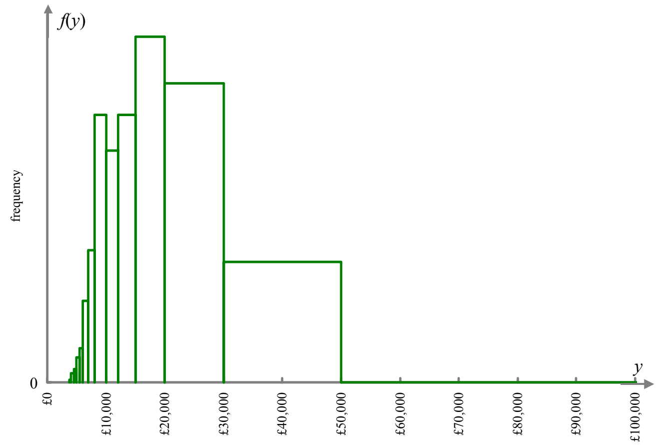

Figure 5.2: Disposable Income (Before Housing Costs). UK 1996/7. Source: Households Below Average Income, 1999

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 Figure 5.1, which illustrates the UK distribution of income before and after tax in 1996/97, 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 Figure 5.1 here with Figure 2.2 on page 17 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 four 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 alternative data source. xv*Households Below Average Income* (HBAI) provides estimates of disposable income in based on the UK's Family Resources Survey, the results of which are summarised in Figure 5.2, using the same income groupings as in Figure 5.1. In comparing this figure with the *Inland Revenue Statistics* distribution of income after tax (the solid line in Figure 5.1) we immediately notice the interesting shape of the lower tail in Figure 5.2 by contrast to the manifestly incomplete picture of the lower tail in Figure 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 checklist in the accom panying panel. Let us briefly examine each of these five items in turn.

5.1.3 *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 suggested on page 93 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 Figure 5.2. Some have argued that, because of the way in which housing is pro vided in the

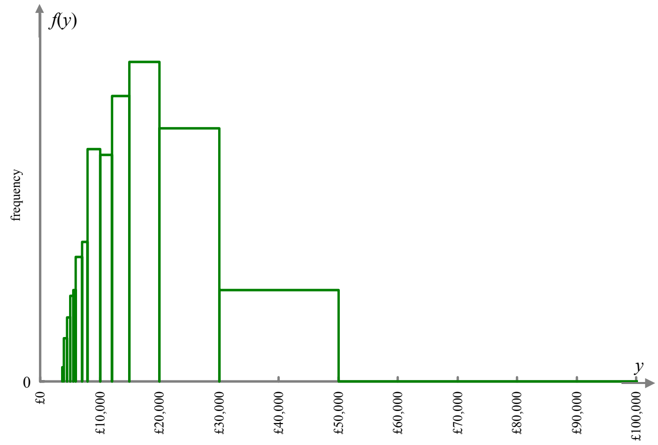

Figure 5.3: Disposable Income (After Housing Costs). UK 1996/7. Source: Households Below Average Income, 1999

UK, the costs of housing 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 Figure 5.3 which presents the after-housing-cost distribution using the same income groupings as for Figures 5.1 and 5.2.

5.1.4 *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 enfranchised 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:

- Adjusting each family's income to allow for differences in needs between different types of families. The process - known in the jargon as "equivalising" the incomes – involves dividing the income by an index such as

	before	*after*
	housing costs	
Two adults	1.00	1.00
Single adult	0.61	0.55
Second adult	0.42	0.45
Third adult	0.36	0.45
Fourth adult	0.36	0.40
Child aged 0-1 years	0.09	0.07
2-4	0.18	0.18
5-7	0.21	0.21
8-10	0.23	0.23
11-12	0.25	0.26
13-15	0.27	0.28
16-18	0.36	0.38
	Examples	
Married couple	1.00	1.00
couple + child aged 4	1.18	1.18
couple + children aged 2,4	1.36	1.36
couple + children aged 2,4,13	1.63	1.64

Source: (Coulter et al. 1992a)

Table 5.1: The McClements Equivalence Scale

that displayed in the accompanying box (for example on this scale a married couple with two children aged 2 and 4 and a nominal income of £10 000 would have an equivalised income of £10 000/1.36 =£7 352.94). 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.

5.1.5 *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, sufficiently detailed data are just not available, and fairly sophisticated techniques must be used to draw inferences about the interpersonal distribution of this quantity.

5.1.6 *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 recognised 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 recognise 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 signifi cantly 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.

5.1.7 *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 distribu tion 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 proceedings from a table of raw data to a number giving the realised value of an inequality measure. We proceed by describing a number of inequality measures that were intro duced 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 so that if it is

[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

summed over the entire income range the result is exactly one; formally:

$$\int_0^\infty f(y)\mathrm{dy} = 1$$

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. 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 TablesA.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

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

or Theil's inequality measure for which we find

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

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

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

and then consider the transformation $1 - J1/^{1-\varepsilon}$ we find that we have A_ε, Atkinson's inequality index with inequality aversion parameter ε. It is worth re-emphasising 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.

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 Figure 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 caluclated 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".

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. However, 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 $y_1, y_2, y_3, ..., y_n$ 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 $y_{[1]}, y_{[2]}, ..., y_{[n]}$, (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 Figure 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:

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

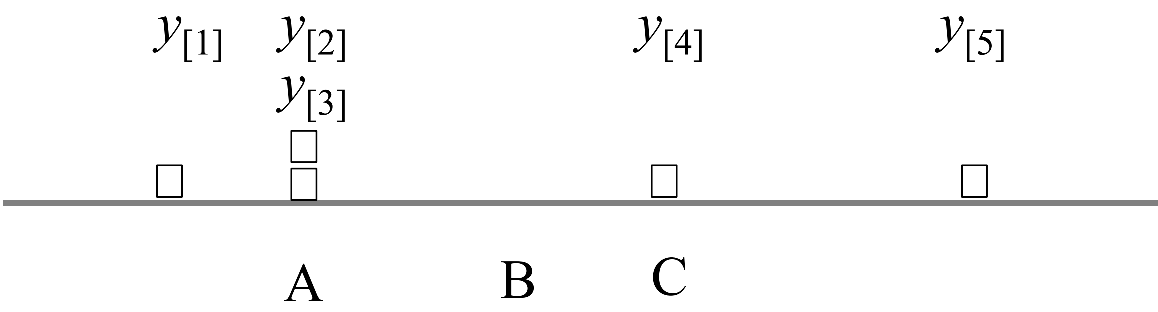

Figure 5.4: Income Observations Arranged on a Line

In fact 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 setting out the n observations in order along the income line as represented by the little blocks in Figure 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 Figure 5.4. This is in fact how the published HBAI data are presented – see Figure 5.5. Of course the picture that emerges will be sensitive to the arbitrary width that is used in this exercise (compare Figure 5.5 with the deliberately coarse groupings used for the same data in Figure 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 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 Figure 5.5 is provided in Figures 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 154ff 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

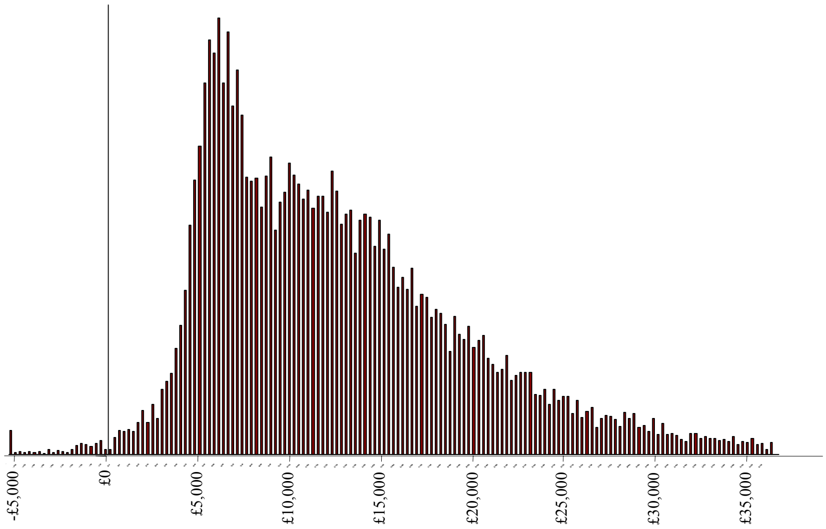

Figure 5.5: Frequency Distribution of Disposable Income, UK 1996/7 (After Housing Costs), Unsmoothed. Source: as for Figure 5.3

this is that much 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. Many 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 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 computerised 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 data sets on to storage media that will make it relatively easy to analyse on a micro-computer. 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 $\$a_3$

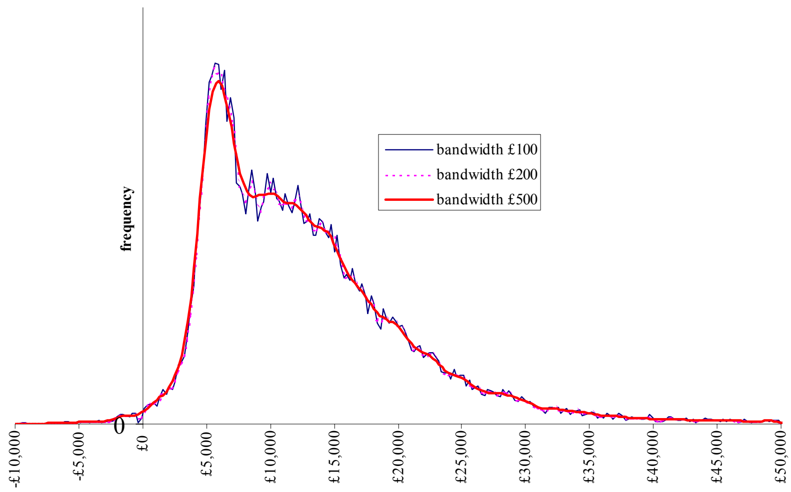

Figure 5.6: Estimates of Distribution Function. Disposable Income, UK 1996/7. (After Housing Costs), Moderate Smoothing. Source: as for Figure 5.3

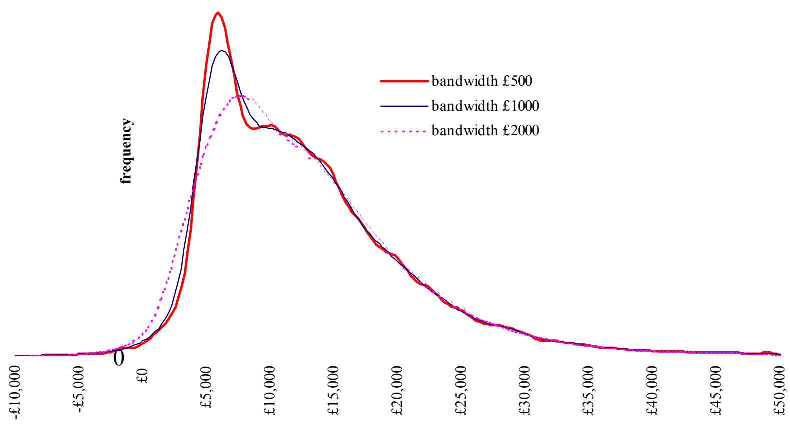

Figure 5.7: Estimates of Distribution Function. Disposable Income, UK 1996/7. (After Housing Costs), High Smoothing. Source: as for Figure 5.3

lower boundary of of income range	*number in groups ('000)*	*group mean income*	*relative freq pop*	*relative freq inc*	*cumulative freq pop*	*cumulative freq inc*
(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
(<$1 000)	3 204	-$16 823		-0.013	0.000	0.000
$1 000	6 526	$1 992	0.057	0.003	0.057	0.003
$3 000	5 860	$3 959	0.051	0.005	0.108	0.009
$5 000	5 680	$5 986	0.049	0.008	0.157	0.017
$7 000	5 593	$8 028	0.049	0.011	0.206	0.027
$9 000	5 372	$9 996	0.047	0.013	0.252	0.040
$11 000	5 555	$11 989	0.048	0.016	0.301	0.055
$13 000	5 344	$13 997	0.046	0.018	0.347	0.073
$15 000	4 837	$15 981	0.042	0.018	0.389	0.091
$17 000	4 402	$18 015	0.038	0.019	0.428	0.110
$19 000	6 507	$20 501	0.057	0.031	0.484	0.141
$22 000	5 610	$23 494	0.049	0.031	0.533	0.173
$25 000	7 848	$27 421	0.068	0.051	0.601	0.223
$30 000	12 380	$34 774	0.108	0.101	0.709	0.325
$40 000	9 099	$44 686	0.079	0.096	0.788	0.421
$50 000	13 679	$60 553	0.119	0.195	0.907	0.616
$75 000	5 374	$85 318	0.047	0.108	0.954	0.724
$100 000	4 075	$130 552	0.035	0.125	0.989	0.849
$200 000	1 007	$290 070	0.009	0.069	0.998	0.918
$500 000	178	$675 843	0.002	0.028	0.999	0.946
$1 000 000	87	$2 616 092	0.001	0.054	1.000	1.000
all ranges	118 217					
positive inc	115 013	$36 892				

Table 5.2: Distribution of Income Before Tax. USA 1995. Source: Internal Revenue Service

and less than \$$a_4$,...". In addition we may be told that the average income of people in the first group (\$$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 5.2 are an example of this kind of presentation. Notice the difference between having the luxury of knowing the individual incomes $y_1, y_2, y_3, ..., y_n$ and of having to make do with knowing the numbers of people falling between the arbitrary class limits $a_1, a_2, a_3, ...$ which have been set by the compilers of the official statistics.

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

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

If we assume for the moment that $a_1 = 0$ and $a_k + 1 = 4$, 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:

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

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 Figure 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 Figure 5.1 we would get a picture like Figure 5.8. However this is not in fact 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 ni by the number in the total population n to give the *relative frequency* in class i (columns 2 and 4 in Table 5.2), 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 upper and lower limits on the unknown theoretical value J. That

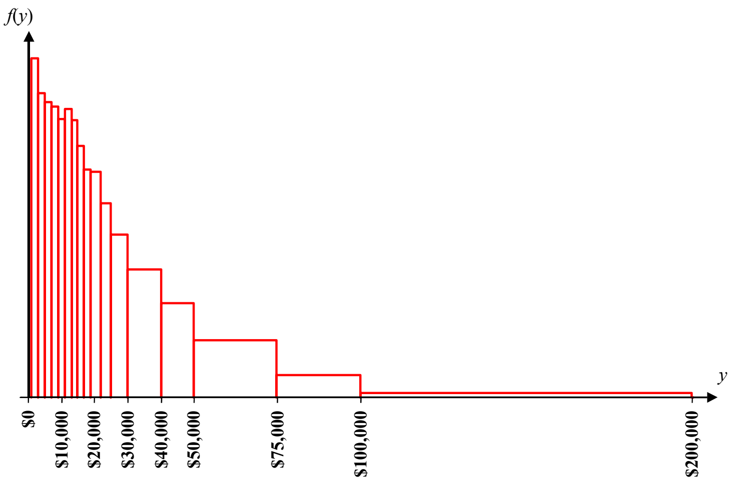

Figure 5.8: Frequency distribution of income before tax. US 1995

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 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, $\$\mu_2$, ... and so on. So, to compute J_L one imagines that there is no inequality within classes (a_i, a_{i+1}) for every $i = 1, 2, ..., k$, as depicted in Figure 5.9. Given that the population relative frequency in income class i is n_i/n (column 4 in Table 5.2) 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, $\mu_1, \mu_2, \mu_3, ..., \mu_k$, we do not need to know the class limits $a_1, a_2, a_3, ..., a_{k+1}$, 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

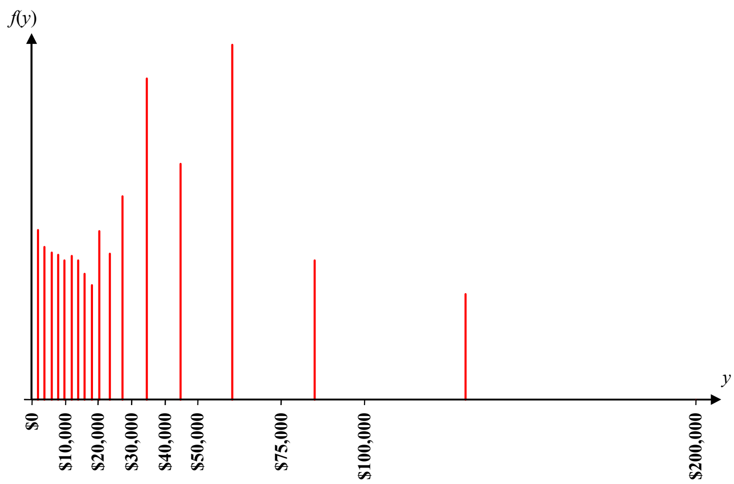

Figure 5.9: Lower Bound Inequality, Distribution of Income Before Tax. US 1995

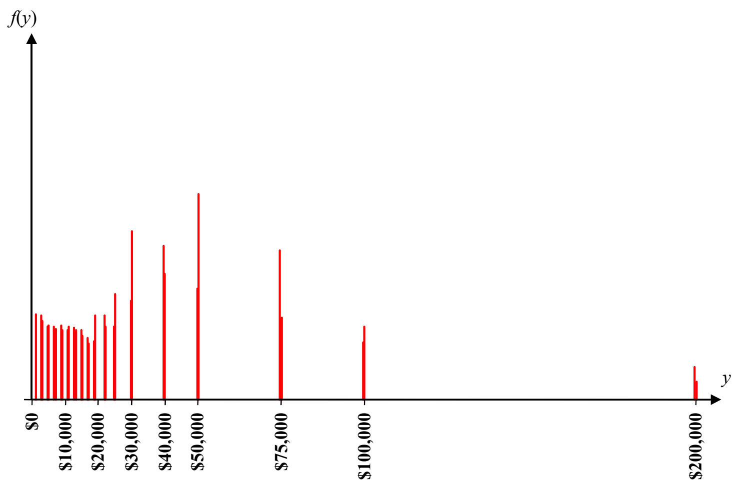

Figure 5.10: Upper Bound Inequality, Distribution of Income Before Tax. US 1995

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 Figure 5.10.

We now have our two numbers J_L, J_U which will meet our requirements for upper and lower 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 an ordinary pocket calculator.

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 5.3: this has been constructed from the data of Table 5.2, on the basis of a variety of alternative assumptions about the underlying distribution of income. Because of the negative mean in the first interval the coimputations 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 Figures 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

	Lower Bound	Comp-romise	Upper Bound		Lower romise	Comp-romise	Upper Bound	
		c				$A_{0.5}$		
(1)	2.281	***	***	***	0.249	0.253	0.260	(0.313)
(2)	**2.281**	2.481	**2.838**	(0.358)	**0.249**	0.252	**0.257**	(0.334)
(3)	2.210	2.406	2.756	(0.359)	0.226	0.228	0.233	(0.336)
		G				A_1		
(1)	0.532	0.535	0.537	(0.667)	0.431	0.435	0.443	(0.330)
(2)	**0.532**	0.535	**0.537**	(0.667)	**0.431**	0.435	**0.442**	(0.328)
(3)	0.508	0.512	0.514	(0.667)	0.380	0.383	0.388	(0.336)
		T				A_2		
(1)	0.632	0.675	***	***	0.690	0.701	0.725	(0.327)
(2)	**0.632**	0.646	**0.673**	(0.335)	**0.690**	0.701	**0.725**	(0.327)
(3)	0.587	0.601	0.629	(0.335)	0.590	0.593	0.600	(0.335)

(1) Top interval is a Pareto tail, bottom interval included
(2) Top interval closed at \$5.5mn, bottom interval included
(3) Top interval closed at \$5.5mn, bottom interval excluded

Table 5.3: Values of Inequality indices under a variety of assumptions about the data. US 1995

and that the highest possible income a_{k+1}, was \$5 500 000. 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 1.0% for the Gini coefficient, 5.0% for Atkinson (A_2) and 6.0% for Theil, it is as much as 22.5% for the coefficient of variation![2]

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}, (\$5 500 000). 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 5.3. 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 imact 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 47). Here the proportionate gap is dramatically cut to 1.8%. This suggests that the practical usefulness of a measure such as this will depend crucially on

[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 46.4%.

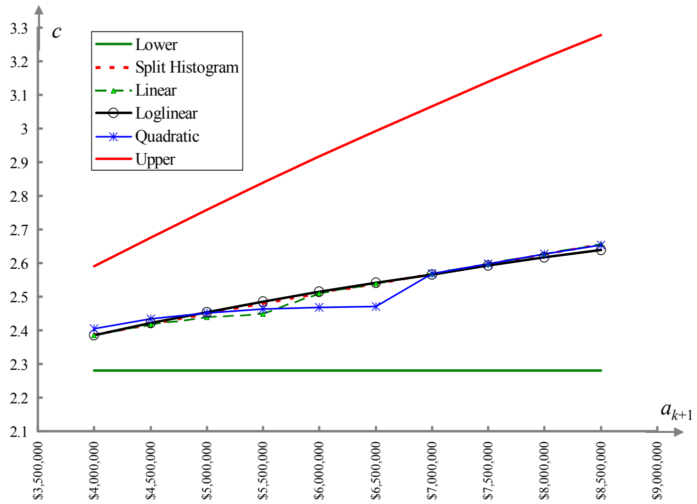

Figure 5.11: The coefficient of variation and the upper bound of the top interval.

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 5.3. 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 in the top interval k were Paretian: this would then yield the results in row (1) of Table 5.3. 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 47 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 1.72, and because this less than 2, the coefficient of variation is effectively infinite hence the asterisks in Table 5.3. All this confirms that estimates of c –

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

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

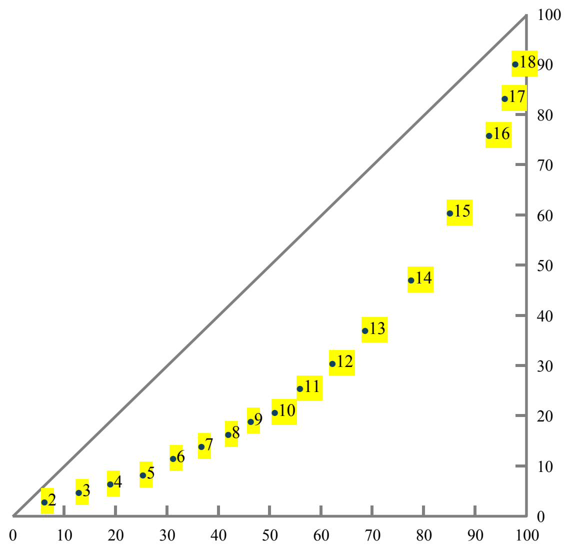

Figure 5.12: Lorenz Co-ordinates for Table 5.2

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 5.3 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 Figure 5.11; the two outer curves represent the lower- and upper-bound assumptions, and the four curves in the middle represent four possible compromise assumptions 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 5.2 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 cumulat ive 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 1995. Some of these points[5] are plotted in Figure 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,

[5] Three of the upper observations have been left out of the diagram for reasons of clarity.

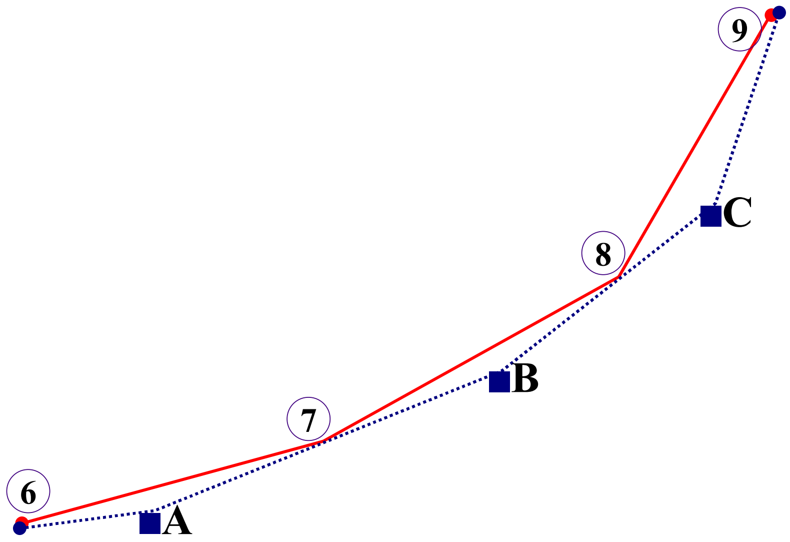

Figure 5.13: Upper and Lower Bound Lorenz Curves

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 inequal ity within income classes – in other words, this corresponds to the use of J_L, the lower bound on the calculated inequality measure, (also illustrated by the distribution in Figure 5.9). This method is shown in detail by the solid lines connecting vertices (6),(7),(8),(9) in Figure 5.13 which is an enlargement of the central portion of Figure 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 Figure 5.13 (in turn this corresponds to J_U and Figure 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 (100,100) in Figure 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 these reasons for being particularly interested in fitting a curve satisfying these requirements is that the observed points on the Lorenz curve in Table 5.2 (columns 6 and 7) only give us the income shares of the bottom 5.7%, the bottom 10.8%,... and so on, whereas we would be more interested in the shares of, say, the bottom 10%, the bottom 20%, 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 (given on page 140 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 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 Figure 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 Figure 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[6] produce inequality estimates that are remarkably similar. This is illustrated in Figure 5.11 where I have plotted the results of four alternative interpolation methods. These three methods (which are more easily explained using the associated density function) are:

- a "split histogram" density function density function in each interval. This is illustrated in Figure 5.14: contrasting this with Figure 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 5.3
- a separate straight line density function fitted to each interval[7]
- 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 nterpolation 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 straightforward enough to use any of these three methods to compute an compromise value for an inequality measure. But in fact if you do not need moon-shot 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 5.3: this column gives, for each inequality

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

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

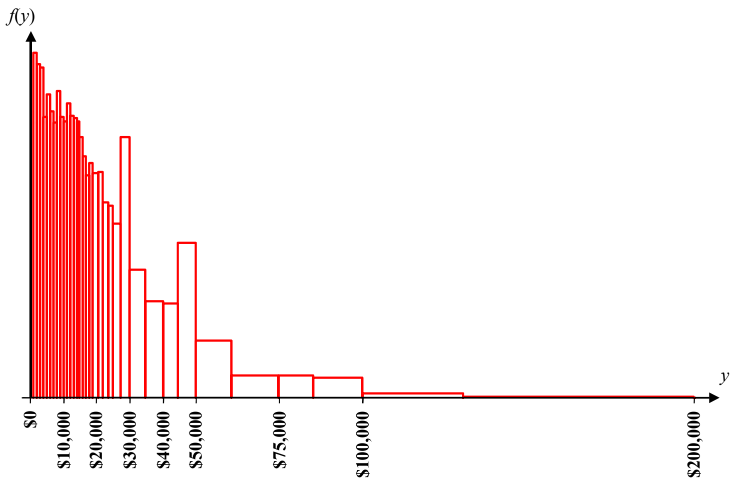

Figure 5.14: The "split histogram" compromise.

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 5.3. 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 data set, 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 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. In fact, 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 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 appendix and the notes to this chapter.

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 $y_1, y_2, y_3, ..., y_n$ 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[8] 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 $y_1, ..., y_n$. 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 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

[8] 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.

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 and Stuart (1977)

Table 5.4: Approximation Formulas for Standard Errors of Inequality Measures

inequality is not due to sampling variability alone; if this is at least three times the standard error, then it is almost certain that the change in c is not a result of sampling variability and 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 5.4, but I should emphasise 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 5.2, 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 $c = 2.481$ given in Table 5.3 were the actual value computed from ungrouped data. What would the standard error be? Noting that the sample size is about 116 million, the standard error is about

$$2.481 \times \sqrt{\frac{1 + 2 \times 2.481^2}{116 \times 10^6}} = 8.404\,2 \times 10^{-4}$$

Hence we can be virtually certain that sampling variability introduces an error of no more than three times this, or 0.001384 on the ungrouped value of c. Contrast this with the gap between the upper bound and lower bound estimates found from Table 5.3 as $2.838 - 2.281 = 0.557$. Hence for this kind of distribution, the grouping error may be of the order of six 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 case. To see this, consider Atkinson’s measure A_ε for the data of Table 5.2. Instead of tabulating

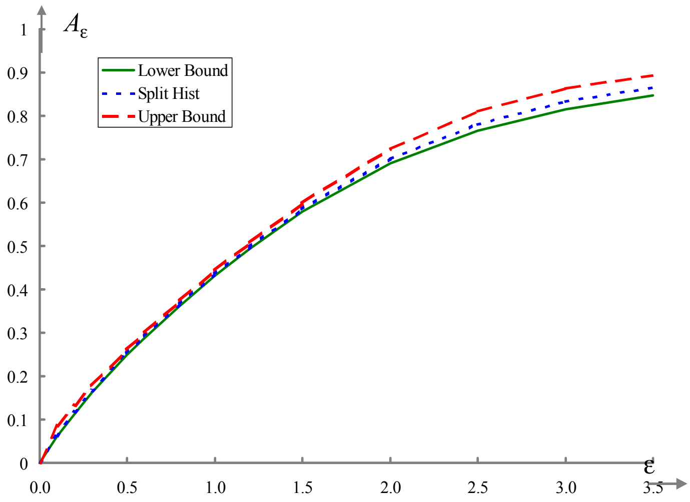

Figure 5.15: The Atkinson index for grouped data, US 1995. Source, as for Table 5.2.

the calculations of the lower and upper bounds and compromise value as in Table 5.3, let us represent them pictorially as in Figure 5.15. This gives a glimpse of the problem that arises when lower incomes are included. For values of $\varepsilon \geq 1$, the range of possible values of the measure is extremely large. In fact for $\varepsilon = 2$, the range of possible values amounts to some 42% of the compromise value of the inequality measure itself. Obviously, then, it will be hard to say unequivocally whether A_2 is greater or less than it was in some other year simply because the grouping bounds are so wide.

However, Figure 5.15 in some respects under-represents the problem: the principal reason for this is that in analysing the inequality represented by the data in Table 5.2 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 Figure 5.16 – 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 Figure 5.17 emerges. Notice that not only is the upper-bound estimate of inequality seriously affected for $\varepsilon > 1$ (which 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

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

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

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

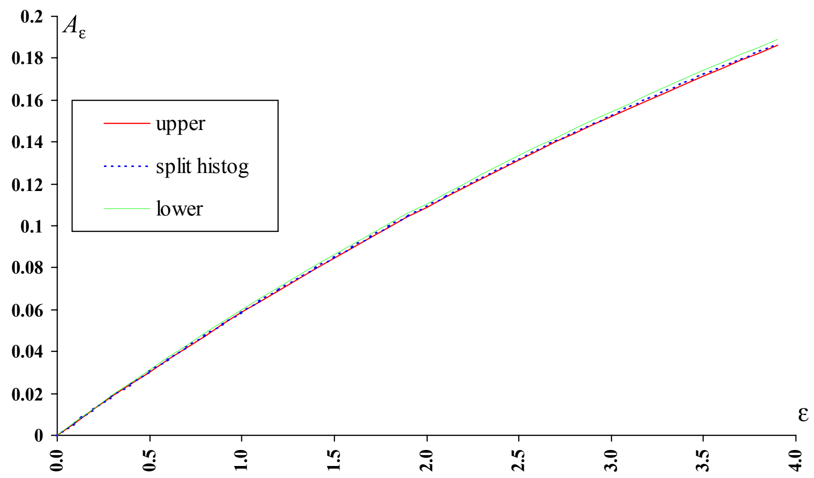

Figure 5.16: The Atkinson Index for Grouped Data: First interval deleted. Czechoslovakia 1988

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 FORM[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 characterise the particular family member appropriate to the data.

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

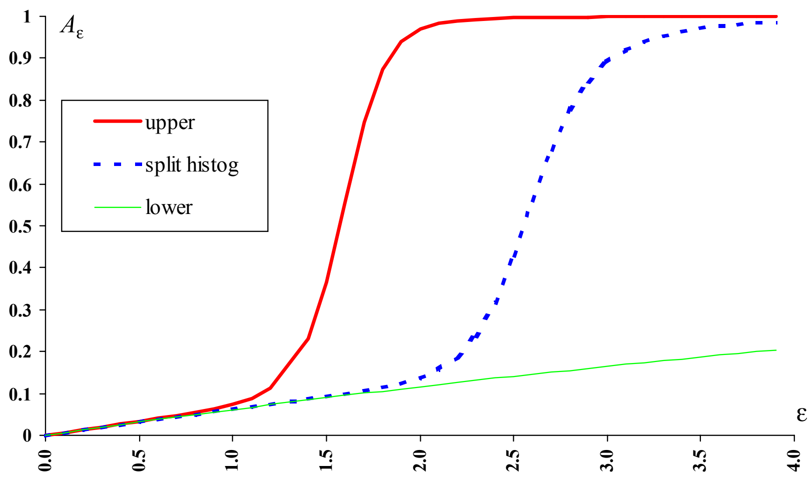

Figure 5.17: The Atkinson Index for Grouped Data: All data included. Czechoslovakia 1988

- Given the formula for a particular inequality measure in terms of the family parameters,[11] 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 139).

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.

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.

[11] See the technical appendix for these formulae.

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, $y_1, y_2, ..., y_n$ - 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 , 2 for the lognormal distribution. Calculate the mean, and the Herfindahl index (the sum of the squares of the shares – see page52) for these n incomes:

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

Then we find:

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

$$\widetilde{\mu} = \log(\overline{y}) - \frac{1}{2}\widetilde{\sigma}^2$$

While this is very easy, it is not as efficient[13] 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 $y_1, y_2, ..., y_n$ to their logarithms $x_1, x_2, ..., x_n$. 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 can be applied similarly to the way it was done in the ungrouped case, provided that in the computation of H an appropriate correction is made to allow for the grouping of observation.

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 $y_1, y_2, ..., y_n$ in Parade order $y_{[1]}, y_{[2]}, ..., y_{[n]}$, (as on page 101). It can be shown that the

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

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 $n\underline{y}/[n-1]$. Work out the observed mean income $\overline{y}$. We already know (from page 83) 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:

$$\widetilde{\alpha} = \frac{\overline{y} - \frac{y_{[1]}}{n}}{\overline{y} - y_{[1]}}$$

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

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. In fact 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, $\log(\underline{y})$.

Given a set of ungrouped observations $y_1, y_2, ..., y_n$ 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.[14] Given the values of the independent variable $x_1, x_2, ..., x_n$ (calculated from the y-values) we may then write down the following set of regression equations:

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

where $e_1, e_2, ..., e_n$ are error terms. One then proceeds to obtain least squares estimates of α and z in the usual way by minimising 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 has been used, as in the construction of Figure 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 n 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

[14] 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 $i = 1, 2, 3, ...k$. The p_i's are then found by cumulating the f_i's upwards from interval i and taking logarithms, thus:

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

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 index on the supposition that the data actually follow the Pareto law. If

$$G_L \leq G_\Pi \leq G_U$$

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 IRS data of Table 5.2 and examine the Pareto law. Since we expect only higher incomes to follow this Law, we shall truncate incomes below \$15 000. First of all we work out from column 5 or 6 of Table 5.2 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 ai given in column 1 of Table 5.2, in order to set up the regression equations. Using ordinary least squares on these last 13 intervals we find our estimate of α as 1.688 with a standard error of 0.0106, and $R^2 = 0.985$. Using the formula for the Gini coefficient on the hypothesis of the Pareto distribution (see page 121 above) we find

$$G = \frac{1}{2 \times 1.688 - 1} = 0.421.$$

Now, noting that the upper and lower bounds on the Gini, for incomes over \$17 000 are 0.413 and 0.422 respectively, we find that the Pareto certainly seems to be an acceptable fit for the last 13 income classes.

Had we relied on the R^2 criterion alone, however, we might have been seriously misled, for if we reworked the calculations for all incomes above \$1 000 we would still have a high R^2 (0.831) but a much lower value of α (1.143); the implied value of $G_\Pi = 0.777$ lies well above the upper bound $G_U = 0.537$ recorded for this group of the population in Table 5.3, thus indicating that the Pareto curve is in fact a rather poor fit for all incomes above \$1 000. It is not hard to see what is going on once we draw the Pareto diagram for these data in Figure 5.18: as we can see from the broken regression line, the straight-line Pareto assumption is quite reasonable for \$15 000 and over; if we were to fit a straight line to all the data (the solid line), we might 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

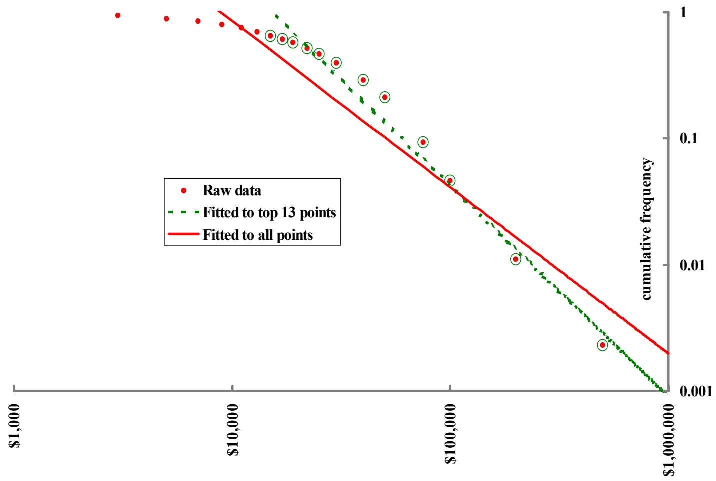

Figure 5.18: Fitting the Pareto diagram for the data in Table 5.2

characteristic of income distribution data: Cf the results for the UK given in Figure 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 must 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 import ant 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 CHECKLIST

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 asked often enough.

5.5.1 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 5.2 where we noted that the minimum and maximum values of c were 2.281 and 2.838, which means that there is a difference in measured inequality of about 22.5% 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 46.4%. In fact 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 $x\%$ 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 results is not substantially different.[17]

5.5.2 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,[18] 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 $y_1, y_2, ..., y_n$ we are specifying both an inequality ranking and a tradeoff between average income and an inequality index consistent with this ranking.[19] Atkinson's measure A_ε and the social-welfare function specified on page 37 form a good example of this approach: by 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

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

[17] A technical note. It is not sufficient to normalise 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" where m is any positive number, and of course, J and J" are ordinally, but not cardinally, equivalent.

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

[19] 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.

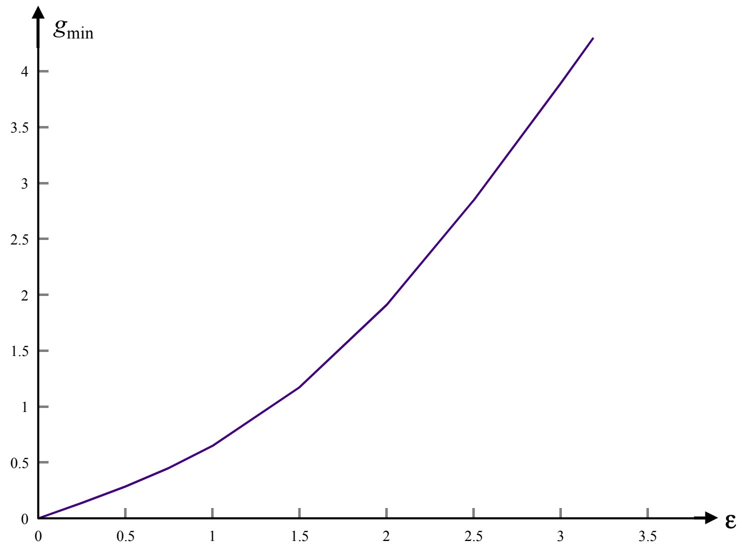

Figure 5.19: The minimum income growth to offset a 1% growth in inequality

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 Figure 5.15 that for $\varepsilon = \frac{1}{2}$, $A_\varepsilon = 0.25$, we find that on this criterion $g_{\min} = 0.33$: a one percent reduction in inequality would be exactly wiped out by a 0.33% 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 percent 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 judgment; then the choice of ε along the horizontal axis in Figure 5.19 is obviously a matter of social values too.

5.5.3 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 treat ment of others is less clear. Let us follow through two examples.

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 fact). 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. In fact, if your chances of physical survival are closely linked to your income, the appearance that inequality is decreasing can be quite misleading, sine the death rate may have been substantially boosted by the greater inequality among the old.

There are obviously several social and economic factors which 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.

5.5.4 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 both by 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 demo graphic 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:

$$I_{total} = w_1 I_1 + w_2 I_2 + ... + w_k I_k + I_{between}$$

where I_{total} is the value of inequality in the aggregate, $I_1, I_2, ..., I_k$ is the value of inequality within subgroup $1, 2, ..., k$ respectively, $w_1, w_2, ..., w_k$ form a set of weights, and Ibetween 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 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.

Firstly, suppose that each group corresponds to a particular family-size class, with the family taken as the funda mental 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 of other reasons listed on page 98, 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 Ibetween 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.

5.5.5 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 reduc tion in inequality in the population as a whole during a number of years with the change in inequality over the life cycle as observed for the age groups in any one year. Alternatively, we might consider the secular change in inequality alongside the apparent[20] redistribution 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 pers pective.

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% 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%, 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 in equality 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."

[20]The qualification "apparent" is included because, as we noted on page 99, the observed distribution of income before tax is not equivalent to the theoretical distribution of income "without the tax".

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

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

- Piet Hein.

5.7 QUESTIONS

1. The data in Table 5.6 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 Generalised Lorenz curves. What effect on income inequality does each tax or benefit component appear to have? Does the distribution of final income welfare-dominate the distribution of original income according to the principles in Theorem 3 on page 42 . [See file "ET TaxesAndBenefits" on the Website for a copy of the data and a hint at the answers].

2. Consider an income distribution in which there are two families. Family 1 contains 1 person with an income of $\$10,000$; family 2 contains 2 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 generalised 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 reweighting affect

		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
			1000	5	3.3	*800*	346	330.4
						1000	237	248.3
						1500	40	38.4
						2000	13	8.8
						5000		
all ranges	*985*	*985*		*1498*	*1498*		*3400*	*3400*

y: lower limit of income interval (yuan pa)

Source: Statistical office, Jiangsu Province,, Rural household budget survey.

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

the results?[See the file "Equiv. and Weight" on the Website for the answers. See also Coulter et al. (1992b) for further discussion.]

3 Suppose you have income data which has been grouped into three intervals: (\$0,\$2000),(\$2000,\$4000),(\$4000,\$6000). There are 1000 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 112.

4 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 5.2 (see file "IRS" on the Website) find the numbers f_1 and f_2 for each interval.

5 For the same data set as in question 2 verify the lower bound and the upper bound estimates of the Atkinson index $A_{0.5}$ given in Table 5.3.

6 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 as well as covering 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 (continuous and discrete) 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.

A.2.1 Discrete Distributions

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

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

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

From this we may define the share of person i in total income to be $s_i = y_i / [n\bar{y}]$.

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 value of A_ε and D_ε is 1, but not otherwise. Note that the minimum value of each measure is zero with the exception of the Herfindahl index for hich 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 where it is assumed that $y_j > y_i$.

A.2.2 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 range of y; i.e. over $[0, \infty)$ or, equivalently for F, over the range $[0, 1]$ The arithmetic mean and the geometric mean are defined as

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

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

Name	**Definition**	Maximum	**Transfer effect**
Variance	$V = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2$	$\bar{y}^2\left[n-1\right]$	$\frac{2}{n}\left[y_j - y_i\right]$
Coefficient of variation	$c = \frac{\sqrt{V}}{\bar{y}}$	$\sqrt{n-1}$	$\frac{y_j - y_i}{n\bar{y}\sqrt{V}}$
Range	$R = y_{\max} - y_{\min}$	$n\bar{y}$	1 if $y_i = y_{\min}$ or $y_j = y_{\max}$, 2 if $y_i = y_{\min}$ and $y_j = y_{\max}$, 0 otherwise
Rel.mean deviation	$M = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i}{\bar{y}} - 1\right\rvert$	$2 - \frac{2}{n}$	$\frac{2}{n\bar{y}}$ if $y_i \leq \bar{y} \leq y_j$ 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	$\frac{1}{2n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert y_i - y_j\rvert$	$\frac{n-1}{n}$	$\frac{[j]-[i]}{n^2\bar{y}}$
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^*	$\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{n\bar{y}^{1-\varepsilon}\left[1-A_\varepsilon\right]^{-\varepsilon}}$
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^*	$\frac{1-\varepsilon}{n}\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{\bar{y}^{1-\varepsilon}-1}$
Generalised entropy	$E_\theta = \frac{1}{\theta^2-\theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{\theta} - 1\right]$	$\frac{n^{\theta-1}-1}{\theta^2-\theta}$ or ∞^{**}	$\frac{y_i^{\theta-1} - y_j^{-\varepsilon}}{n\bar{y}^{\theta}}$
Herfindahl	$H = \frac{1}{n}\left[c^2 + 1\right]$	∞	$\frac{2}{n^2\bar{y}^2}\left[y_j - y_i\right]$
Theil	$T = \sum_{i=1}^{n} s_i \log\left(ns_i\right)$	$\log n$	$\frac{1}{n\bar{y}}\log\frac{y_j}{y_i}$

Notes: * 1 if $\varepsilon \geq 0$; ** ∞ if $\theta < 1$

Table A.1: Inequality measures for discrete distributions

From this 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.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

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-2u} du,$$

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.

A.3.1 The lognormal distribution

The basic specification is:

$$\begin{aligned}
F(y) &= \Lambda\left(y;\ \mu, \sigma^2\right) = \int_0^y \frac{1}{\sqrt{2\pi}\sigma\, x} \exp\left(-\frac{1}{2\sigma^2}\left[\log x - \mu\right]^2\right) dx \\
\Phi(y) &= \Lambda\left(y;\ \mu + \sigma^2, \sigma^2\right) \\
\bar{y} &= e^{\mu + \frac{1}{2}\sigma^2} \\
y^* &= e^{\mu}
\end{aligned}$$

and the Lorenz curve is given by:

$$\Phi = N\left(N^{-1}\left(F\right) - \sigma\right)$$

Name	**Definition**	$\Lambda(y;\mu,\sigma^2)$	$\Pi(y;\underline{y},\alpha)$
Variance	$V=\int\left[y_i-\bar{y}\right]^2dF$	$e^{2\mu+\sigma^2}\left[e^{\sigma^2}-1\right]$	$\frac{\alpha\underline{y}^2}{[\alpha-1]^2[\alpha-2]}$
Coefficient of variation	$c=\frac{\sqrt{V}}{\bar{y}}$	$\sqrt{e^{\sigma^2}-1}$	$\sqrt{\frac{1}{\alpha[\alpha-2]}}$
Rel.mean deviation	$M=\int\left\lvert\frac{y}{\bar{y}}-1\right\rvert dF$	$2\left[2N\left(\frac{\sigma}{2}\right)-1\right]$	$2\frac{[\alpha-1]^{\alpha-1}}{\alpha^\alpha}$
logarithmic variance	$v=\int\left[\log\frac{y}{\bar{y}}\right]^2dF$	$\sigma^2+\frac{1}{4}\sigma^4$	$\log\frac{\alpha-1}{\alpha}+\frac{1}{\alpha}+\frac{1}{\alpha^2}$
variance of logarithms	$v_1=\int\left[\log\frac{y}{y^*}\right]^2dF$	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	$G=1-2\int\Phi dF$	$2N\left(\frac{\sigma}{\sqrt{2}}-1\right)$	$\frac{1}{2\alpha-1}$
Atkinson	$A_\varepsilon=1-\left[\int\left[\frac{y}{\bar{y}}\right]^{1-\varepsilon}dF\right]^{\frac{1}{1-\varepsilon}}$	$1-e^{-\frac{1}{2}\varepsilon\sigma^2}$	$1-\left[\frac{\alpha-1}{\alpha}\right]\left[\frac{\alpha}{\alpha+\varepsilon-1}\right]^{\frac{1}{1-\varepsilon}}$
Generalised entropy	$E_\theta=\frac{1}{\theta^2-\theta}\left[\int\left[\frac{y_i}{\bar{y}}\right]^\theta dF-1\right]$	$\frac{e^{\left[\theta^2-\theta\right]\frac{\sigma^2}{2}}-1}{\theta^2-\theta}$	$\frac{1}{\theta^2-\theta}\left[\left[\frac{\alpha-1}{\alpha}\right]^\theta\frac{\alpha}{\alpha-\theta}-1\right]$
Theil	$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{\bar{x}}\log\left(\frac{x_i}{\bar{x}}\right)$		

Table A.2: Inequality measures for continuous distributions

A.3.2 The Pareto distribution (type I)

The basic specification is:

$$\begin{aligned}
F(y) &= \Pi(y;\ \underline{y},\alpha) = 1 - \left[\frac{\underline{y}}{y}\right]^{\alpha} \\
\Phi(y) &= \Pi(y;\ \underline{y},\alpha-1) \\
\bar{y} &= \frac{\alpha}{\alpha-1}\underline{y} \\
y^{*} &= e^{1/\alpha}\underline{y}
\end{aligned}$$

Clearly 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:

$$\Phi(y) = \exp\left(\frac{\log\frac{1-F(y)}{1-F_0}\log\frac{1-\Phi_1}{1-\Phi_0}}{\log\frac{1-F_1}{1-\Phi_0}}\right)$$

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 $\{a_1, a_2, a_3, ..., a_k, a_{k+1}\}$, the proportions of the population in each group $\{f_1, f_2, f_3, ..., f_k\}$, and the average income of each group $\{\mu_1, \mu_2, \mu_3, ..., \mu_k\}$, are all known. Then, as described on page 112, a "maximum inequality" Lorenz curve may be drawn through the observed points using this information. However 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 157.

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

$$z(y) = \frac{\int_y^{\infty} uf(u)du}{\int_y^{\infty} f(u)du} = \bar{y}\frac{1-\Phi(y)}{1-F(y)}$$

From the above, for the Pareto distribution (Type I) we have

$$
\begin{aligned}
z(y) &= \frac{\alpha}{\alpha-1}\underline{y}\left[\frac{\underline{y}}{y}\right]^{\alpha-1}\left[\frac{\underline{y}}{y}\right]^{-\alpha} \\
&= \frac{\alpha}{\alpha-1}y.
\end{aligned}
$$

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. Letting $z(y) = \gamma y$ where γ is a constant and rearranging, we have

$$\int_y^\infty uf(u)du = \gamma y \int_y^\infty f(u)du$$

where $f(\cdot)$ is unknown. Differentiate this with respect to y:

$$-yf(y) = -\gamma y f(y)\gamma + [1 - F(y)]$$

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

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

A.3.3 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

$$F(y) = 1 - e^{-ky}\left[\gamma + \delta y\right]^{-\alpha}$$

where $k, \gamma \geq 0$, and $\alpha, \delta > 0$. By putting $k = 0$ in the above equation we obtain the Pareto type II distribution. By putting $\gamma = 0$ and $\delta = 1/\underline{y}$ in the type II 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 proved by Rao and Tam (1987)

Singh and Maddala (1976) suggested as a useful functional form the following:

$$F(y) = 1 - \left[\gamma + \delta y^\beta\right]^{-\alpha}$$

where α, β, γ, δ are parameters such that $F(0) = 0$, $F(\infty) = 1$, and $F'(y) = f(y) \geq 0$. From this we can derive the following special cases.

- If $\beta = 1$ we have the Pareto type II distribution.
- If $\gamma = 1$, $\delta = [1/\alpha]\, k^\beta$ 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 $\Gamma(x) = \int uxe^{-u}du$.
- 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 $\alpha = \gamma = 1$ and $\delta = \underline{y}^{-\beta}$, 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

  $$\mu'_r = r\underline{y}^r \frac{\pi}{\beta \sin\left(\frac{r\pi}{\beta}\right)},\ -\beta < r < \beta.$$

  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:

$$\mu_r' = \underline{y}^r \frac{\pi}{\theta}\frac{\sin\left(\frac{r\theta}{\beta}\right)}{\sin\left(\frac{r\pi}{\beta}\right)}, \ -\beta < r < \beta.$$

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

$$\mu_r' = \sum_{i=1}^{n} \frac{\rho n!}{\rho - n}\Delta_{n,r}\ ,\ \rho > r$$

where

$$\Delta_{n,r} = \begin{cases} [-1]^{r-n}\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=j}^{n}...[ijk...]}_{r-n \text{ terms}} & \text{if } n < r \\ 1 & \text{if } n = r \end{cases}$$

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:

$$f(y) = \frac{y^\xi\,[1-y]^\eta}{B(\xi, \eta)}$$

[1] The Beta and Gamma functions are extensively tabulated. Their analytical properties are discussed in many texts on statistics, for example Keeping (1962), Weatherburn (1949).

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 $[\xi+\eta]^{-r}\Gamma(\xi+r)/\Gamma(\xi)$.

The relationships mentioned in the previous paragraphs are set out in Figure A.1. Solid arrows indicate that one distribution is a special case of another. Dotted 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

[2] This restriction means that y must be normalised by dividing it by its assumed maximum value.

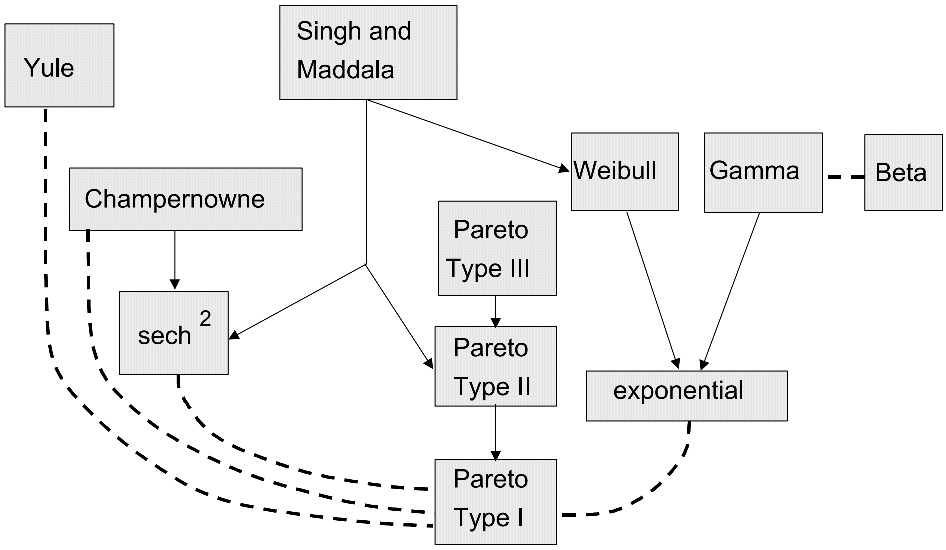

Figure A.1: Relationships Between Functional Forms

distinguishing features.

A.4.1 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:

$$\log\left(1 - A_\varepsilon\right) + \frac{1-\varepsilon}{1-A_\varepsilon}\frac{\partial A_\varepsilon}{\partial \varepsilon} = \frac{\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}\log\left(\frac{y_i}{\bar{y}}\right)}{\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}}$$

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:

$$\frac{\partial A_\varepsilon}{\partial \varepsilon} = \frac{1 - A_\varepsilon}{\bar{x}\left[1-\varepsilon\right]^2}\left[\frac{1}{n}\sum_{i=1}^{n} x_i \log\left(x_i\right) - \bar{x}\log\bar{x}\right]$$

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 cardinalisation 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 cardinalisation for the function U – a problem that does not arise with the Atkinson index). We find that if $\varepsilon \neq 1$:

$$D_\varepsilon = 1 - \frac{\bar{y}^{1-\varepsilon}\left[1 - A_\varepsilon\right]^{1-\varepsilon} - 1}{\bar{y}^{1-\varepsilon} - 1}$$

and in the limiting case $\varepsilon = 1$:

$$D_\varepsilon = 1 - \frac{\log\left(\bar{y}\left[1 - A_1\right]\right)}{\log\left(\bar{y}\right)}$$

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.

A.4.2 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 $x \lesseqgtr e = 2.71828....$ 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 or y^* according as the measure under consideration is v or v_1.

A.5 DECOMPOSITION OF INEQUALITY MEASURES

A.5.1 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}.

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

$$I_{\text{total}} = \Xi \left(I_1, I_2, ..., I_k; \bar{y}_1, \bar{y}_2, ..., \bar{y}_k; n_1, n_2, ..., n_k \right)$$

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.

- 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 characterises each measure. The prime example of this is the generalised entropy class E_θ introduced on page 59 for which the scale independence

[3] This is equivalent to the term "relative frequency" used in Chapter 5.

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 generalised entropy family E_θ and the associated Atkinson indices. For the generalised 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 Generalised 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 Generalised 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 subgroup, although the weights w_j do not necessarily sum to one.
- the within-group component weights will only sum to one if $\theta = 0$ or if $\theta = 1$ in which case of course we have Theil's index T.

[4] Notice that this is the same as the expression given for the generalised 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.

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

$$A_\varepsilon = 1 - \left[\left[\theta^2 - \theta\right] E_\theta + 1\right]^{1/\theta}$$

for $\theta \neq 0$.[5] 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 Generalised Entropy measure above. If we do this then – for the case where I is the Atkinson index with parameter ε - we get the following:

$$I_{\text{between}} = 1 - \left[\sum_{j=1}^{k} f_j \left[\frac{\bar{y}_j}{\bar{y}}\right]^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$$

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

$$\left[1 - I_{\text{total}}\right]^{1-\varepsilon} = \left[\left[1 - I_{\text{between}}\right]^{1-\varepsilon} + \left[1 - I_{\text{within}}\right]^{1-\varepsilon} - 1\right]$$

and from these we can deduce

$$I_{\text{within}} = 1 - \left[1 - \sum_{j=1}^{k} f_j^{\varepsilon} g_j^{1-\varepsilon} \left[\left[1 - I_j\right]^{1-\varepsilon} - 1\right]\right]^{\frac{1}{1-\varepsilon}}$$

To restate the point: the decomposition formula given here for the Atkinson index is different from that given on page 148 for the generalised 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. In fact we have

$$A_2 = 1 - \frac{1}{2E_{-1} + 1}$$

pplying this formula and using a self-explanatory adaptation of our earlier notation the allocation of the components of inequality is as follows:

$$\begin{aligned} E_{-1[\text{within}]} &= \sum_{j=1}^{k} \frac{f_j^2}{g_j} E_{-1[j]} \\ E_{-1[\text{between}]} &= \frac{1}{2}\left[\sum_{j=1}^{k} \frac{f_j^2}{g_j} - 1\right] \\ E_{-1[\text{total}]} &= E_{-1[\text{between}]} + E_{-1[\text{within}]} \end{aligned}$$

[5] For $\theta = 0$ the relationship becomes:$A_1 = 1 - e^{-E_0}$.

whreas

$$A_{2[\text{total}]} = \frac{A_{2[\text{between}]} + A_{2[\text{within}]} - A_{2[\text{between}]}A_{2[\text{within}]}}{1 - A_{2[\text{between}]}A_{2[\text{within}]}}$$

Now let us consider the situation in China represented in Table A.3. Te top part of Table A.3 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 Generalised Entropy measure or its ordinally equivalent Atkinson measure.
- the between-group and within-group components sum exactly to total inequality in the case of the Generalised 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 equalisation 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.

	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

Table A.3: Decomposition of inequality in Chinese provinces, Rural and Urban subpopulations

A.5.2 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; in fact 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 probably look much the same as that depicted in figure 1 of Chapter 2, but the axes will have been shifted vertically.

To see how the shape of the Lorenz curve and the Generalised 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 Generalised 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 Generalised 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 to 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 Generalised Lorenz curve of the distribution $\left(\frac{y_1}{\bar{y}}, \frac{y_2}{\bar{y}}, ..., \frac{y_n}{\bar{y}}\right)$ nd the absolute Lorenz curve is the generalised Lorenz curve of the distribution $(y_1 - \bar{y}, y_2 - \bar{y}, ..., y_n - \bar{y})$

The reason that many conventional inequality tools will not work in the presence of negative incomes can be seen from "evaluation function" $h(\cdot)$ introduced on page 100. Recall that many inequality measures can be defined in terms of the evaluation function. Consider for example the generalised 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 α 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 generalised 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

A.7.1 Micro Data

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. 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 which case these weights will need to be used both in the point estimates and in the standard errors - see the notes for further discussion of this point.

Now consider the standard errors of inequality estimates. As we noted on page 141 inequality measures can be expressed in terms of standard statistical moments. Correspondingly in situations where we are working with a sample $\{y_1, y_2, ...y_n\}$ 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:

$$\mathcal{E}\left(m'_r\right) = \mu'_r$$

$$\text{var}\left(m'_r\right) = \frac{1}{n}\left[\mu'_{2r} - \left[\mu'_r\right]^2\right]$$

and an unbiased estimate of the sample variance of m is

$$\widehat{\text{var}}\left(m'_r\right) = \frac{1}{n-1}\left[m'_{2r} - \left[m'_r\right]^2\right]$$

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 generalised 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 $n\left[\psi\left(m'_r, m'_1\right) - \psi\left(\mu'_r, \mu'_1\right)\right]$ is asymptotically normally distributed thus:

$$N\left(0, \frac{\partial^2\psi}{\partial m'^2_r}\text{var}\left(m'_r\right) + \frac{\partial^2\psi}{\partial m'_1 \partial m'_r}\text{cov}m'_r, m'_1 + \frac{\partial^2\psi}{\partial m'^2_1}\text{var}\left(m'_1\right)\right)$$

Finally let us consider the problem of estimating the density function from a set of n sample observations. As explained on page 101 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

$$\hat{f}(y) = \frac{1}{w}\sum_{i=1}^{n} K\left(\frac{y - y_i}{w}\right)$$

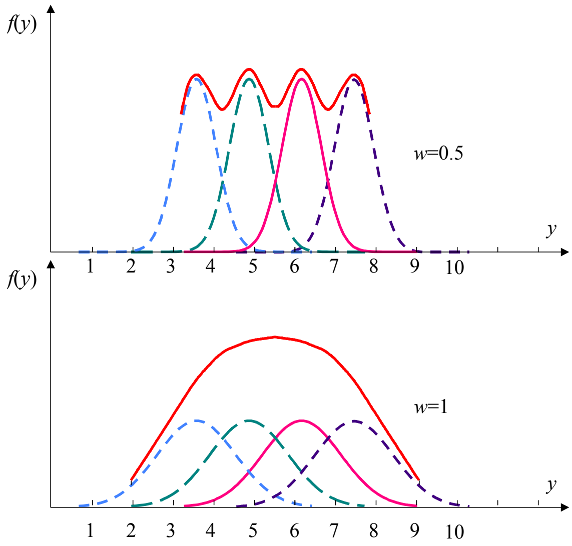

Figure A.2: Density Estimation with a Normal Kernel

– 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 figure 5 of Chapter 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

$$K\left(\frac{y - y_i}{w}\right) = \begin{array}{l} 1 \text{ if } \begin{array}{c} a_j \leq y < a_{j+1} \text{ and} \\ a_j \leq y_i < a_{j+1} \end{array} \\ 0 \text{ otherwise} \end{array}$$

However, this histogram rule is crude: each observation makes an "all or nothing" contribution to the densty estimate. So it may be more useful to take a 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 Figure A.2 for the case where there are just four income observations. The upper part of Figure 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 $y_1 ... y_4$ 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) .

A.7.2 Grouped Data

Now let us suppose that you do not have micro-data 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 upper and lower 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, 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 $\phi_i$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 101 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

$$\begin{aligned} F(y) &= F_i + \int_{a_i}^{y} \phi_i(x)dx \\ \Phi(y) &= \Phi_i + \int_{a_i}^{y} x\phi_i(x)dx \end{aligned}$$

So, to find the share of the bottom 20%, 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.

$$\phi_i(y) = \frac{f_i}{a_{i+1} - a_i}\ ,\ a_i \leq y < a_{i+1}.$$

Using the formulas given on page 140 above, we can see that the *Paretian density* in any closed interval is given by

$$\begin{aligned}\phi_i(y) &= \frac{\alpha a_i^{\alpha}}{y^{\alpha+1}}\ ,\ a_i \leq y < a_{i+1}.\\ \alpha &= \frac{\log(1 - f_i)}{\log \frac{a_i}{a_{i+1}}}\end{aligned}$$

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 107. 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.

[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}.

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

$$\phi_i(y) = \begin{cases} \frac{f_i}{a_{i+1}-a_i}\frac{a_{i+1}-\mu_i}{\mu_i-a_i} & , a_i \leq y < \mu_i, \\ \frac{f_i}{a_{i+1}-a_i}\frac{\mu_i-a_i}{a_{i+1}-\mu_i} & , \mu_i \leq y < a_{i+1}. \end{cases}$$

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

$$\phi_i(y) = \frac{c}{y^{\alpha+1}}\ , \ a_i \leq y < a_{i+1}$$

where

$$c = \frac{\alpha f_i}{a_{i+1}^{-\alpha} - a_i^{-\alpha}}$$

and α is the root of the following equation:

$$\frac{\alpha}{\alpha - 1}\frac{a_{i+1}^{1-\alpha} - a_i^{1-\alpha}}{a_{i+1}^{-\alpha} - a_i^{-\alpha}} = \mu_i$$

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. Letting $i = k$ and $a_{k+1} \to \infty$ we have $a_{k+1}^{-\alpha} \to 0$ if $\alpha < 0$ and so 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

$$\phi_i(y) = b + cy,\ a_i \leq y < a_{i+1}$$

where

$$
\begin{aligned}
b &= \frac{12\mu_i - 6\left[a_{i+1} - a_i\right]}{\left[a_{i+1} - a_i\right]^3} f_i \\
c &= \frac{f_i}{a_{i+1} - a_i} - \frac{1}{2}\left[a_{i+1} + a_i\right] b.
\end{aligned}
$$

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 – which can be used on any personal computer that is compatible with IBM file format – makes it straightforward to do that. The files are accessed from the Website at `http://sticerd.lse.ac.uk/research/frankweb/MeasuringInequality/index.html`. Ideally you should be able to read Excel 97 files, but, for those without this facility, the contents have also been saved the comma-separated-variable (csv) versions.

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 summarised in Table A.4. Individual files and their provenance are cited in detail in Appendix B.

Table		**Figure**		**Figure**	
3.3	East-West	**2.1**	ET84-5	**5.1**	IR income
3.4	East-West	**2.2**	ET84-5	**5.2**	HBAI
5.2	IRS	**2.3**	ET84-5	**5.3**	HBAI
5.3	IRS ineq	**2.4**	ET84-5	**5.5**	HBAI
5.5	Czechoslovakia	**2.5**	ET84-5	**5.6**	HBAI
5.6	Czechoslovakia	**2.9**	Earnings Quantiles	**5.7**	HBAI
5.7	Jiangsu	**2.10**	81-2,84-5 compared	**5.8**	IRS
A.3	Decomp	**2.11**	81-2,84-5 compared	**5.9**	IRS
		2.12	81-2,84-5 compared	**5.10**	IRS
		2.13	81-2,84-5 compared	**5.11**	IRS ineq
		3.1	Atkinson SWF	**5.12**	IRS
		3.2	Atkinson SWF	**5.14**	IRS
		3.9	LIS comparison	**5.15**	IRS ineq
		3.10	LIS comparison	**5.16**	Czechoslovakia
		4.5	ET84-5	**5.17**	Czechoslovakia
		4.10	NES		
		4.11	IR wealth		

Table A.4: Source files for tables and figures

Appendix B

NOTES ON SOURCES AND LITERATURE

These notes describe the data sets which have used for particular examples in each chapter; cite the sources which have been used for the discussion in the text, and provide 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.

B.1 CHAPTER 1

A useful general discussion of conventional terminology and approach is found in Prest and Bauer (1973); reference may also be made to Chapter 1 of Atkinson (1983) and Chapter 2 of Thurow (1975).

The question "inequality of what" 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 (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 recognised (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) and then if you can manipulate the mathematics, go on to the other papers in Phelps. 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 "equivalisation" – and its impact upon measured inequality is discussed in Coulter et al. (1992b, 1992a) . Alternative approaches to measuring inequality in the presence of household heterogeneity are discussed in Cowell (1980) , Glewwe (1991) Jenkins and O' Higgins (1989), Jorgenson and Slesnick (1990).

On some of the classical principles of justice and equality, see Rees (1971), Chapter 7, Wilson (1966). On the desirability of equality *per se* see Broome (1988). 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) concern about equality as a consumption externality see Hochman and Rodgers (1969). A notable landmark in modern though 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.

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, middle incomes, or to high incomes. 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 page 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) . The link between a measure that captures the depth of poverty and the Gini index 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) and Seidl (1988). The relationship between inequality and poverty measures is discussed in some particularly useful papers by Thon (1981, 1983a).

B.2 CHAPTER 2

The main examples here are from the tables in *Economic Trends*, November 1987, which are reproduced on the Webpage in the file "ET84-85": the income intervals used are those that were specified in the original tables. If you load this file you will also see exactly how to construct the histogram for yourself; experience suggests that it is well worth running through this as an exercise. The example in Figure 2.9 is taken from the 1998 New Earnings Survey data –

see file "Earnings Quantiles" on the Webpage. The reference to the Plato as an early precursor of inequality measurement is to be found in Saunders (1970) pp 214-215.

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 popularised. 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 any good introductory statistics text such as Allen (1949), Connor and Morrell (1964) or Keeping (1962); for an extensive empirical application of Pen's parade see Jenkins and Cowell (1994) . 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 19) needs to be qualified in one very special case: where the mean of the thing that you are charting is itself negative – see page 152 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); see Fellman (1976) for a neat result 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).

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 to 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 137. A convenient alternative form of the standard definition is given in Dorfman (1979):

$$G = 1 - \frac{1}{\bar{y}} \int_0^\infty P(y)^2 dy \text{ where } P(y) = 1 - F(y).$$

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 equalisation 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 generalisations of the Gini are discussed by Bossert (1990), Donaldson and Weymark (1980), Weymark (1981), and Yaari (1988); see also Lin (1990). The Gini coefficient has also been used as the basis for constructing indices of relative deprivation – see Bishop, Chakraborti, and Thistle (1991), Chakravarty and Chakraborty (1984), Yitzhaki (1979).

The properties of the more common ad hoc inequality measures are discussed at length in Atkinson (1970, pp.252- 257; 1973), pp.53-58), Champernowne(1974 page 805), Foster (1985), Jenkins (1991) and Sen (1973, 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 skew to the right. Creedy (1977) discusses the properties of the variance of logarithms. The use of the skewness statistic was proposed by Young (1917) , and 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 101) Further details on the use of moments may be found in texts such as Keeping. 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.138-143). Some of the criticisms of Stark's high/low measure are raised in Polanyi and Wood (1974). Another such practical measure with a similar flavour is Wiles (1974) semi-decile ratio: (Minimum income of top 5%)/(maximum income of bottom 5%). 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 thinks that the semi-decile ratio focuses on the essential feature of income inequality to the exclusion of others.

Wiles and Markowski (1971) argue 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 (1973, Chapter 3) who suggests 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. The method of percentiles is used extensively by Lydall (1959) and Polanyi and Wood (1974). The formalisation of this approach as a "comparative function" was suggested by Esberger and Malmquist (1972).

[1] But only in this respect, since they reject the Lorenz curve as an 'inept choice', preferring to use histograms instead.

B.3 CHAPTER 3

The data set used for the example on page 64 is given in the file "LIS comparison" on the Website.

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), Mayston (1974) 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 (cf. Dasgupta and Pearce(1972, Chapter 2), Little and Mirrlees (1974, Chapter 3). Applications of SWF analysis include taxation design (Atkinson and Stiglitz (1980), 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).

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), Dagum (1990), Dahlby (1987), and Sen (1973, 1992). The formal relationships between inequality and social welfare are discussed in Ebert (1987) and Dutta and Esteban (1992). The association of Rawls (1972) concept of justice (where society gives priority to improving the position of the least advantaged person) with a SWF exhibiting extreme inequality aversion is discussed in Sen (1974, pp.395-398), Arrow (1973) and Hammond (1975). Inequality measures of the type first suggested by (Dalton 1920)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, so that as we increase a person's income y by one unit (pound, dollar, etc.) his welfare weight U' drops by $\alpha\%$ where α 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 generalising both this and Atkinson's measure. See also Bossert and Pfingsten (1990), Yoshida (1991). 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.

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 fairly sophisticated mathematics are found in (Dasgupta et al. 1973), Kolm (1969) and Sen (1973, pp.49-58). In fact a lot of this work was anticipated by Hardy, Littlewood, and Pólya (1934, 1952); Marshall and Olkin (1979) develop this approach and cover in detail relationships involving Lorenz curves, generalised Lorenz curves and concave functions (see also Arnold (1987)): readers who are happy with an undiluted mathematical presentation may find this the most useful single reference on this part of the subject. Shorrocks (1983) introduced the concept of Generalised 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 TaxBenQ) then the generalised Lorenz ordering of distributions is preserved if and only if ϕ is concave. Iritani and Kuga (1983) and Thistle (1989b, 1989a, and the Generalised Lorenz curve and the distribution function. A further discussion and overview of these topics is to be found in Lambert (1993). A novel 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 inequality are to be found in Maasoumi (1986, 1989), Rietveld (1990).

Types of permissible "distance" function, and their relationship with inequality is 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 Allisson (1981) discusses the use of the Theil index as a measure of inequality interpreted in terms of average distance.

Herfindahl (1950)'s measure (which is obviously 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 I_β 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).

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:

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

The special case $\beta = 0$ which becomes $\sum_{i=1}^{n}\log(\bar{y}/y_i)/n$ is discussed in Theil (1967, Chapter 4 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 generalised entropy measure with parameter $\theta = 0$ (equivalently Atkinson index with $\varepsilon = 1$).

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 Frosini (1985), Eichhorn and Gehrig (1982) and Kolm (1976a,1976b). Shorrocks and Foster (1987) examine the issue of transfer sensitivity. The Atkinson and Generalised Entropy families are examples of the application of the concept of the quasi-linear mean, which is discussed in Hardy, Littlewood, and Pólya (1934, 1952) and Chew (1983)

The axiomatic approach to inequality measurement discussed on page 59 is not of course restricted to the generalised 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) axiomatisation 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). Using a simulation study Kuga (1979, 1980) and shows that distributional rankings of the Theil index are often similar to those of the Gini coefficient. The problematic cases highlighted in the examples on page 34 and 59 are based on Cowell (1988a). Ebert (1988) discusses the principles on which a generalised type of the relative meand deviation may be based.

B.4 CHAPTER 4

Most texts on introductory statistical theory give an introduction to the normal distribution – for example Keeping (1962) or (Mood et al. 1974). For the less mathematically inclined reader a more tender treatment is given in Reichman (1961) or Statistics (1966). 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).

Pareto's original work can be consulted in Pareto (1896,1965) or in Pareto (1972), which deals in passing with some of Pareto's late views on the law of income distribution. 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.). Unfortunately, oversimplified interpretations of the Law persist – Adams (1976) suggests 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). Several of the other results in the text are formally proved in Chipman (1974). Nicholson (1969, pp.286-292) and Bowman (1945) give a simple account of the use of the Pareto diagram.

The example of earnings displayed on page 85 can be reproduced from file "NES" on the Website; the income example of 80 is taken from the Website file "ET84-85" again, and the wealth example on page 88 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 (1975a, Appendix C; 1976b, Appendix E). Bjerke (1970) dealing with the structure of wages in Copenhagen in 1953 shows that the more homogenous the occupation, the more likely it would be that the distribution of earnings within it was lognormal. Hill (1959) shows that merging normal distributions with different variances leads to a 'leptokurtosis' (more of the population in the 'tails' than expected from a normal distribution) – typical feature of the distribution of the logarithm of income. Other useful references on the lognormal distribution in practice are Fase (1970), Takahashi (1959), Thatcher (1968). Evidence for lognormality is discussed in the case of India (Rajaraman 1975), Kenya (Kmietowicz and P. 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.

Atkinson (1975) and Soltow (1975) produce evidence on the Pareto distribution and the distribution of wealth in the UK and the USA of the 1860's respectively. For further evidence on the variability of Pareto's α in the USA, see Johnson (1937), a cautious supporter of Pareto. Some of the less orthodox applications of the Pareto curve are in Zipf (1949). 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 (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 civilisation 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.

Finally, a brief consideration of other functional forms that have been claimed to fit observed distributions more or less satisfactorily (see the technical appendix page 141). Some of these are generalisations of the lognormal or Pareto forms, such as the three-parameter lognormal – Metcalf (1969)) – or the generalised Pareto-Levy law, which attempts to take account of the lower tail – see Mandelbrot (1960), Arnold (1983). Indeed, note that 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 – see Quandt (1966), Hayakawa (1951). 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). Other suggestions are the 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 Champernowne (1952) distribution Fisk (1961), and the Yule distribution (Simon 1955,1957; Simon and Bonini 1958); see also Campano (1987). Evans et al. (1993) 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); Cf also the closely related model by Dagum (1977). A generalised 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 generalised beta distribution; on this distribution see alsoMajumder and Chakravarty (1990). Alternative approaches to parameterising the Lorenz curve are discussed in Basmann, Hayes, Johnson, and Slottje (1990, 1991), Kakwani and Podder (1973).

B.5 CHAPTER 5

The UK data used for Figure 5.1 are from *Inland Revenue Statistics* (see file "IR income" on the Website),and the US data in Table 5.2 from Internal Revenue Service, *Statistics of Income: Individual Tax Returns* (see file "IRS"). The UK data used for Figures 5.2-5.7 are taken from the fairly recent Households Below Average Income (HBAI); summary charts and results are published in Department of Social Security (1998) Appendix 9 on FRS and FES; extracts from the micro-data of the HBAI are provided on the Website in annualised form in the Website file "HBAI".

If you want a fuller introduction to the problem of specifying an income or wealth variable, see Atkinson Atkinson (1983). The quality of the data, of course, depends crucially on the type of tax administration and official 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, such as the UK's Central Statistical Office, 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, Royal Commission on the Distribution of Income and Wealth (1975a, Appendices F and H). For detail on income data in the USA, and the quality of sample surveys in particular see Budd and Radner (1975) and the references therein, and (Ferber et al.) . Wealth data in the UK are considered in detail in Atkinson and Harrison (1978). Since publication of the first edition of this book large comprehensive datasets of individual incomes have become much more readily available; it is impossible to do justice to them. Two particular cases from the USA that deserve 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), and the Panel Study of Income Dynamics described in Hill (1992).

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. In this latter reference and in Morgan (1962) the effect of family grouping on measured inequality is considered; Prest and Stark (1967) do this for the UK. 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 standard approach to equivalence scales in the UK is that of McClements (1977). The relationship between equivalence scales and measured inequality is examined in Coulter et al. (1992b): 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): see also Benus and Morgan (1975). The key reference on the theoretical and empirical importance of price changes on measured inequality is Muellbauer (1974). 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.

International comparisons of inequality within countries are found in Paukert (1973) and Jain (1975), though neither discusses fully the problems of international comparison of data; an early treatment of this is found in Kuznets (1963, 1966). The topic is treated exhaustively in 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". Atkinson and Micklewright (1992) compare the income distributions in Eastern European economies in the process of transition. 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. For a systematic analysis of this issue using decomposition techniques see also Berry et al.(1983, 1981), Theil (1979b, 1989).

A comprehensive overview of many of the statistical issues is to be found in 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 much 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), McDonald and Ransom (1981), 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 $a_1, a_2, ..., a_{k+1}$ and Hagerbaumer (1977) suggests the upper bound. of the Gini index as an inequality measure in its own right. In the two Gastwirth references 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. 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 $\{a_1, a_2, ..., \}$ which will minimise the Gini index; Davies and Shorrocks (1989) refine the technique for larger data sets.

For general information on the concept of the standard error see Kendall and Stuart (1977) or (fairly easy) Reichman (1961); formulas for standard errors of inequality measures for can be found in the following references: Weatherburn (1949, p.144)[coefficient of variation], Kendall and Stuart (1977), 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 Gastwirth et al. (1986), Glasser (1962), Lomnicki (1952), Nygård and Sandström (1989), and Sandström et al. (1985, 1988). Allison (1978) discusses issues of estimation and testing based on microdata using the Theil index, coefficient of variation and Theil index. The statistical properties of the Generalised 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) and Beach and Richmond (1985); for the generalised Lorenz curve estimation refer to Bishop, Chakraborti, and Thistle (1989), and Bishop, Formby, and Thistle (1989).

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 curve a fit is made 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. 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 and Stuart (1977).

B.6 TECHNICAL APPENDIX

Functional forms for distributions are discussed in Evans et al. (1993). The formulas in the appendix for the decomposition of inequality measures are standard – see Bourguignon (1979), Cowell (1980) , Das and Parikh (1981, 1982) and Shorrocks (1980). For applications of the decomposition technique see Anand

(1983), Borooah et al. (1991), Ching (1991), Cowell (1984, 1985b), Frosini (1989), Glewwe (1986), Mookherjee and Shorrocks (1982). For a characterisation of some general results in decomposition see Shorrocks (1984, 1988). Decomposition by income components is discussed by Satchell (1978), Shorrocks (1982) and Theil (1979a). The relationship between decomposition of inequality and the measurement of poverty is examined in Cowell (1988b). An obvious omission from the decomposition formulae is the Gini coefficient, the decomposition of which 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) and Lerman and Yitzhaki (1984, 1989), Yitzhaki and Lerman (1991) . Braulke (1983) examines the Gini decomposition on the assumption that within-group distributions are Paretian. Silber (1989) discusses the decomposition of the Gini index by subgroups of the population (for the case of non-overlapping partitions) and by income components. The empirical example from China is based on the work of Howes and Lanjouw (1991) and Hussain et al. (1991)

Bibliography

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.

Adams, K. (1976). Pareto's answer to inflation. *New Scientist' 71*, 534–537.

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*.

Aigner, D. J. and A. S. Goldberger (1970). Estimation of Pareto's law from grouped observations. *Journal of the American Statistical Association 65*, 712–723.

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*.

Aitchison, J. and J. A. C. Brown (1957). *The Lognormal Distribution*. London: Cambridge University Press.

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*, 207–218.

Allen, R. G. D. (1949). *Statistics for Economists*. London: Hutchinson.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review 43*, 865– 880.

Allisson, P. D. (1981). Inequality measures for nominal data (reply to krishnan). *American Sociological Review 46*, 371–372.

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, and A. Polovin (1996). Inequality amongst the kibbutzim. *Economica 63*, S63–S85.

Anand, S. (1983). *Inequality and poverty in Malaysia*. London: Oxford University Press.

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.

Arrow, K. J. (1973). Some ordinalist-utilitarian notes on Rawls' theory of justice. *Journal of Philosophy 70*, 245–263.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory 2*, 244–263.

Atkinson, A. B. (1973). More on the measurement of inequality. mimeo, University of Essex.

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 ed.). Oxford: Clarendon Press.

Atkinson, A. B. and F. Bourguignon (1982). The comparison of multidimensional 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*. New York: Macmillan.

Atkinson, A. B. and A. J. Harrison (1978). *Distribution of Personal Wealth in Britain*. 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 J. E. Stiglitz (1980). *Lectures on Public Economics*. Basingstoke: McGraw Hill.

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*, 385–389.

Basu, K. (1987). Axioms for a fuzzy measure of inequality. *Mathematical Social Sciences 14*(12), 275–288.

Batchelder, A. B. (1971). *The Economics of Poverty*. New York: Wiley.

Baxter, M. A. (1980). Minimum-variance unbiased estimation of the parameters of the Pareto distribution. *Metrika 27*.

Beach, C. M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies 50*, 723–735.

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), 439–450.

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.

Bentzel, R. (1970). The social significance of income distribution statistics. *Review of Income and Wealth*, 253–264.

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*, 351–362.

Berrebi, Z. M. and J. Silber (1987). Dispersion, asymmetry and the Gini index of inequality. *International Economic Review 28*(6), 331–338.

Berry, A., F. Bourguignon, and C. Morrisson (1981). The level of world inequality: how much can one say? *Review of Income and Wealth 29*, 217–243.

Berry, A., F. Bourguignon, and C. Morrisson (1983). Changes in the world distributions of income between 1950 and 1977. *Economic Journal 93*, 331–350.

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), 725–727.

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*, 421–437.

Bishop, J. A., J. P. Formby, and W. P. Smith (1991). International comparisons of income inequality: Tests for Lorenz dominance across nine countries. *Economica 58*, 461–477.

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*.

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*, 107–136.

Blitz, R. C. and J. A. Brittain (1964). An extension of the Lorenz diagram to the correlation of two variables. *Metron 23*.

Boadway, R. W. and N. Bruce (1984). *Welfare Economics*. Oxford: Basil Blackwell.

Board of Inland Revenue (1972). *Survey of Personal Incomes*. London: HMSO.

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.

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*, 117–134.

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*, 901–920.

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*, 607–628.

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.

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.

Budd, E. C. (1970). Postwar changes in the size distribution of income in the US. *American Economic Review, Papers and Proceedings 60*, 247–260.

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.

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*, 161–175.

Chakravarty, S. R. (1988). Extended Gini indices of inequality. *International Economic Review 29*(2), 147–156.

Chakravarty, S. R. and A. B. Chakraborty (1984). On indices of relative deprivation. *Economics Letters 14*, 283–287.

Champernowne, D. G. (1952). The graduation of income distribution. *Econometrica 20*, 591–615.

Champernowne, D. G. (1953). A model of income distribution. *Economic Journal 63*, 318–351.

Champernowne, D. G. (1956). Comment on the paper by P.E. Hart and S.J. Prais. *Journal of The Royal Statistical Society A 119*, 181–183.

Champernowne, D. G. (1973). *The Distribution of Income Between Persons.* Cambridge University Press.

Champernowne, D. G. (1974). A comparison of measures of income distribution. *Economic Journal 84*, 787–816.

Chew, S.-H. (1983). A generalization of the quasi-linear mean with application to the measurement of income inequality. *Econometrica 51*, 1065–1092.

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*, 275–282.

Connor, L. R. and A. J. H. Morrell (1964). *Statistics in Theory and Practice* (5th ed.). London: Pitman.

Cortese, C. F., R. F. Falk, and J. K. Cohen (1976). Further consideration on the methodological analysis of segregation indices. *American Sociological Review 41*, 630–637.

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. *Economic Journal 102*, 1067–1082.

Cowell, F. A. (1975). Income tax incidence in an ageing population. *European Economic Review 6*, 343–367.

Cowell, F. A. (1977). *Measuring Inequality* (First ed.). Oxford: Phillip Allan.

Cowell, F. A. (1980). On the structure of additive inequality measures. *Review of Economic Studies 47*, 521–531.

Cowell, F. A. (1984). The structure of American income inequality. *Review of Income and Wealth 30*, 351–375.

Cowell, F. A. (1985a). 'A fair suck of the sauce bottle'-or what do you mean by inequality? *Economic Record 6*, 567–579.

Cowell, F. A. (1985b). Multilevel decomposition of Theil's index of inequality. *Review of Income and Wealth 31*, 201–205.

Cowell, F. A. (1988a). Inequality decomposition - three bad measures. *Bulletin of Economic Research 40*, 309–312.

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. 149–166. 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. 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*, 273–290.

Cramer, J. S. (1978). A function for the size distribution of >income: Comment. *Econometrica 46*, 459–460.

Creedy, J. (1977). The principle of transfers and the variance of logarithms. *Oxford Bulletin of Economics and Statistics 39*, 153–8.

Crew, E. L. (1982). Double cumulative and Lorenz curves in weather modification. *Journal of Applied Meteorology 21*, 1063–1070.

Cronin, D. C. (1979). A function for the size distribution of income: A further comment. *Econometrica 47*, 773–774.

Dagum, C. (1977). A new model of personal income distribution: Specification and estimation. *Economie Appliquée 30*, 413–436.

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. *Economic Journal 30*(9), 348–361.

Das, T. and A. Parikh (1981). Decompositions of Atkinson's measures of inequality. *Australian Economic Papers 6*, 171–178.

Das, T. and A. Parikh (1982). Decomposition of inequality measures and a comparative analysis. *Empirical Economics 7*, 23–48.

Dasgupta, A. K. and D. W. Pearce (1972). *Cost-Benefit Analysis*. London: Macmillan.

Dasgupta, P. S., A. K. Sen, and D. A. Starrett (1973). Notes on the measurement of inequality. *Journal of Economic Theory 6*, 180–187.

David, H. A. (1968). Gini's mean difference rediscovered. *Biometrika 55*, 573–575.

David, H. A. (1981). *Order Statistics* (2nd ed.). New York: John Wiley.

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: the Principia Press.

Davis, H. T. (1954). *Political Statistics*. Evanston, Illinois: Principia Press.

Department of Social Security (1998). *Households Below Average Income: A Statistical Analysis, 1979-1996/7*. London: The Stationery Office.

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*, 146–149.

Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indices. *American Sociological Review 20*, 210–217.

Dutta, B. and J. Esteban (1992). Social welfare and equality. *Social Choice and Welfare 9*, 267–276.

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), 363–376.

Eichhorn, W., H. Funke, and W. F. Richter (1984). Tax progression and inequality of income distribution. *Journal of Mathematical Economics 13*(10), 127–131.

Eichhorn, W. and W. Gehrig (1982). Measurement of inequality in economics. In B. Korte (Ed.), *Modern Applied Mathematics - optimization and operations research*, pp. 657–693. Amsterdam: North Holland.

Eltetö, O. and E. Frigyes (1968). New income inequality measures as efficient tools for causal analysis and planning. *Econometrica 36*, 383–396.

Esberger, S. E. and S. Malmquist (1972). *En Statisk Studie av Inkomstutveklingen.* Stockholm: Statisk Centralbyr å n och Bostadssyrelsen.

Esteban, J. (1986). Income share elasticity and the size distribution of income. *International Economic Review 27*, 439–444.

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.

Fellman, J. (1976). The effect of transformation on Lorenz curves. *Econometrica 44*, 823–824.

Ferber, R., J. Forsythe, H. W. Guthrie, and E. S. Maynes. Validation of a national survey of consumer financial characteristics. *Review of Economics and Statistics 51*, 436–444.

Fisk, P. R. (1961). The graduation of income distribution. *Econometrica 29*, 171–185.

Foster, J. E. (1983). An axiomatic characterization of the Theil measure of income inequality. *Journal of Economic Theory 31*, 105–121.

Foster, J. E. (1984). On economic poverty: a survey of aggregate measures. *Advances in Econometrics 3*, 215–251.

Foster, J. E. (1985). Inequality measurement. In H. P. Young (Ed.), *Fair Allocation*, pp. 38–61. Providence, R. I.: American Mathematical Society.

Francis, W. L. (1972). *Formal Models of American Politics: An Introduction,.* New York: Harper and Row.

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*.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica 39*, 1037–1039.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics 54*, 306–316.

Gastwirth, J. L. (1974a). Large-sample theory of some measures of inequality. *Econometrica 42*, 191–196.

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*, 479–483.

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*, 269–273.

Gastwirth, J. L. and J. T. Smith (1972). A new goodness-of-fit test. *Proceedings of The American Statistical Association*, 320–322.

Gehrig, W. (1988). On the Shannon - Theil concentration measure and its characterizations. In W. Eichhorn (Ed.), *Measurement in Economics*. Heidelberg: Physica Verlag.

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*, 648–654.

Glewwe, P. (1986). The distribution of income in Sri Lanka in 1969-70 and 1980-81. *Journal of Development Economics 24*, 255–274.

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*, 211–216.

Graaff, J. (1957). *Theoretical Welfare Economics*. London.

Gupta, M. R. (1984). Functional forms for estimating the Lorenz curve. *Econometrica 52*, 1313–1314.

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*, 377–379.

Hainsworth, G. B. (1964). The Lorenz curve as a general tool of economic analysis. *Economic Record*.

Hammond, P. J. (1975). A note on extreme inequality aversion. *Journal of Economic Theory 11*, 465–467.

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 ed.). London: Cambridge University Press.

Harrison, A. (1981). Earnings by size: A tale of two distributions. *Review of Economic Studies 48*, 621–631.

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*, 191–195.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy 61*, 434–435.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. *Journal of Political Economy 63*, 309–321.

Hart, P. E. and S. J. Prais (1956). An analysis of business concentration. *Journal of The Royal Statistical Society A 119*, 150–181.

Hayakawa, M. (1951). The application of Pareto's law of income to Japanese data. *Econometrica 19*, 174–183.

Helmert, F. R. (1876). Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus den ersten Potenzen der Differenzen gleichgenauer direkter Beobachtungen. *Astronomische Nachrichten 88*, 127–132.

Herfindahl, O. C. (1950). *Concentration in the Steel Industry*. Ph. D. thesis, Columbia University.

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*, 355–381.

Hochman, H. and J. D. Rodgers (1969). Pareto-optimal redistribution. *American Economic Review 59*, 542–557.

Howes, S. P. and P. Lanjouw (1991). Regional variations in urban living standards in urban China. Development Economics Research Programme Discussion Paper CP17, STICERD,London School of Economics, Houghton St, London WC2A 2AE.

Hussain, A., P. Lanjouw, and N. H. Stern (1991). Income inequalities in China: Evidence from household survey data. Development Economics Research Programme Discussion Paper CP18, STICERD,London School of Economics, Houghton St, London WC2A 2AE.

Iritani, J. and K. Kuga (1983). Duality between the Lorenz curves and the income distribution functions. *Economic Studies Quarterly 34*(4), 9–21.

Jain, S. (1975). *Size Distribution of Income. A Compilation of Data*. Washington: World Bank.

Jakobsson, U. (1976). On the measurement of the degree of progression. *Journal of Public Economics 5*, 161–168.

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.

Jencks, C. (1973). *Inequality*. London: Allen Lane.

Jenkins, S. P. (1991). The measurement of economic inequality. In L. Osberg (Ed.), *Readings on Economic Inequality*. M.E. Sharpe, Armonk, N.Y.

Jenkins, S. P. and F. A. Cowell (1994). Dwarfs and giants in the 1980s: The UK income distribution and how it changed. *Fiscal Studies 15*(1), 99–118.

Jenkins, S. P. and M. O' Higgins (1989). Inequality measurement using norm incomes. *Review of Income and Wealth 35*(9), 245–282.

Johnson, N. O. (1937). The Pareto law. *Review of Economic Statistics 19*, 20–26.

Jorgenson, D. W. and D. T. Slesnick (1990). Inequality and the standard of living. *Journal of Econometrics 43*, 103–120.

Kakwani, N. C. and N. Podder (1973). On the estimation of the Lorenz curve from grouped observations. *International Economic Review 14*, 278–292.

Keeping, E. . S. (1962). *Introduction to Statistical Inference*. New York: Van Nostrand Reinhold.

Kendall, M. and A. Stuart (1977). *The Advanced Theory of Statistics*. London: Griffin.

Kloek, T. and H. K. Van Dijk (1978). Efficient estimation of income distribution parameters. *Journal of Econometrics 8*, 61–74.

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 of China. *The Statistician 42*, 107–121.

Kmietowicz, Z. W. and W. P. (1975). Srarisrical aalysis 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*, 416–442.

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*, 1041–1042.

Kondor, Y. (1975). Value judgement implied by the use of various measures of income inequality. *Review of Income and Wealth*, 309–321.

Koo, A. Y. C., N. T. Quan, and R. Rasche (1981). Identification of the Lorenz curve by Lorenz coefficients. *Weltwirtschaftliches Archiv 117*, 125–135.

Kravis, I. B., A. W. Heston, and R. Summers (1978a). *International Comparisons of Real Product and Purchasing Power*. Baltimore: Johns Hopkins University Press.

Kravis, I. B., A. W. Heston, and R. Summers (1978b). Real GDP per capita for more than one hundred countries. *Economic Journal 88*, 215–242.

Krishnan, P. (1981). Measures of inequality for qualitative variables and concentration curves. *American Sociological Review 46*, 368–371. comment on Allison, American Sociological Review December 1978.

Kuga, K. (1973). Measures of income inequality: An axiomatic approach. Discussion Paper 76, Institute of Social and Economic Research, Osaka University.

Kuga, K. (1979). Comparison of inequality measures: a Monte Carlo study. *Economic Studies Quarterly 30*, 219–235.

Kuga, K. (1980). The Gini index and the generalised entropy class: further results and a vindication. *Economic Studies Quarterly 31*, 217–228.

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*.

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*(12), 1103–1116.

Lambert, P. J. (1993). *The Distribution and Redistribution of Income* (second ed.). Manchester: Manchester University Press.

Lambert, P. J. and J. R. Aronson (1993). Inequality decomposition analysis and the Gini coefficient revisited. *Economic Journal 103*(9), 1221–1227.

Lebergott, S. (1959). The shape of the income distribution. *American Economic Review 49*, 328–347.

Lerman, R. I. and S. Yitzhaki (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters 15*, 363–368.

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*, 324–330.

Likes, J. (1969). Minimum variance unbiased estimates of the parameters of power function and Pareto's distribution. *Statistische Hefte 10*, 104–110.

Lin, T. (1990). Relation between the Gini coefficient and the Kuznets ratio of MEP. *Jahrbuch von Nationalökonomie und Statistik 207*(2), 36–46.

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*.

Lorenz, M. O. (1905). Methods for measuring concentration of wealth. *Journal of the American Statistical Association 9*, 209–219.

Love, R. and M. C. Wolfson (1976). Income inequality: statistical methodology and Canadian illustrations. *Statistics Canada, Catalogue 13-559 Occasional*.

Lydall, H. F. (1959). The long-term trend in the size distribution of income. *Journal of the Royal Statistical Society A122*, 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*(7), 991–997.

Maasoumi, E. (1989). Composite indices of income and other developmental indicators: a general approach. *Research on Economic Inequality 1*.

Majumder, A. and S. R. Chakravarty (1990). Distribution of personal income: Development of a new model and its application to US income data. *Journal of Applied Econometrics 5*, 189–196.

Malik, H. J. (1970). Estimation of the parameters of the Pareto distribution. *Metrika 15*, 126–132.

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*, 753–766.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory and Majorization.* New York: Academic Press.

Mayston, D. J. (1974). *The Idea of Social Choice*. London.: Macmillan.

McClements, L. (1977). Equivalence scales for children. *Journal of Public Economics 8*.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica 52*, 647–664.

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 M. R. Ransom (1979). Functional forms, estimation techniques and the distribution of income. *Econometrica 47*, 1513–1525.

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*, 464–477.

Metcalf, C. E. (1969). The size distribution of income during the business cycle. *American Economic Review 59*, 657–668.

Mood, A. M., F. Graybill, and D. Boes (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.

Mookherjee, D. and A. F. Shorrocks (1982). A decomposition analysis of the trend in UK income inequality. *Economic Journal 92*, 886–902.

Morgan, J. N. (1962). The anatomy of income distribution. *Review of Economics and Statistics 44,*, 270–283.

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*, 203–207.

Moyes, P. (1989). Equiproprortionate growth of incomes and after-tax inequality. *Bulletin of Economic Research 41*, 287–293.

Muellbauer, J. (1974). Inequality measures, prices and household composition. *Review of Economic Studies 41* (10), 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*.

Nicholson, R. J. (1969). *Economic Statistics and Economic Problems.* London: McGraw-Hill.

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.*

Nygård, F. and A. Sandström (1989). Income inequality measures based on sample surveys. *Journal of Econometrics 42*, 81–95.

Okner, B. A. (1972). Constructing a new data base from existing microdata sets: The 1966 MERGE file. *Annals of Economic and Social Measurement 1*, 325–342.

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 Institute.

Paglin, M. (1975). The measurement and trend of inequality: a basic revision. *American Economic Review 65*(9), 598–609.

Pareto, V. (1896). Ecrits sur la courbe de la répartition de la richesse. In *Oeuvres complètes de Vilfredo Pareto.* Giovanni Busino. Librairie Droz, Genève, 1965.

Pareto, V. (1965). *Écrits sur La Courbe de la Repartition de la Richesse*, Volume 3 of *Oeuvres Complètes.* Geneva: Librairie Droz.

Pareto, V. (1972). *Manual of Political Economy.* London: Macmillan.

Paukert, F. (1973). Income distribution at different levels of development: A survey of the evidence. *International Labour Review 108*, 97–125.

Pen, J. (1971). *Income Distribution.* London: Allen Lane, The Penguin Press.

Pen, J. (1974). *Income Distribution* (second ed.). London: Allen Lane, The Penguin Press.

Phelps, E. S. (1973). *Economic Justice.* Harmondsworth: Penguin.

Pigou, A. C. (1952). *The Economics of Welfare* (4th ed.). London: Macmillan.

Polanyi, G. and J. B. Wood (1974). How much inequality? Research monograph, Institute of Economic Affairs, London.

Prest, A. R. and P. T. Bauer (1973). Income differences and inequalities. *Moorgate and Wall Street*, 22–43.

Prest, A. R. and T. Stark (1967). Some aspects of income distribution in the UK since World War II. *Manchester School 35*, 217–243.

Pyatt, G. (1976). On the interpretation and disaggregation of Gini coefficients. *Economic Journal 86*, 243–255.

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*, 278–290.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 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*, 275–280.

Rasche, R. H., J. Gaffney, A. Y. C. Koo, and N. Obst (1980). Functional forms for estimating the Lorenz curve. *Econometrica 48*, 1061–1062.

Ravallion, M. (1994). Poverty comparisons using noisy data on living standards. *Econoomics Letters 45*, 481–485.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Pres.

Rawls, J. (1972). *A Theory of Justice*. Oxford: Oxford University Press.

Rees, J. (1971). *Equality*. Pall Mall, London.

Reichman, W. J. (1961). *Use and Abuse of Statistics*. London: Methuen.

Rein, M. and S. M. Miller (1974, July/August). Standards of income redistribution. *Challenge*, 20–26.

Riese, M. (1987). An extension of the Lorenz diagram with special reference to survival analysis. *Oxford Bulletin of Economics and Statistics 49*(5), 245–250.

Rietveld, P. (1990). Multidimensional inequality comparisons. *Economics Letters 32*, 187–192.

Rosenbluth, G. (1955). Measures of concentration. In G. J. Stigler (Ed.), *Business Concentration and Price Policy*.

Royal Commission on the Distribution of Income and Wealth (1975a). *Initial Report on the Standing Reference, Cmnd 6171*. London: HMSO.

Royal Commission on the Distribution of Income and Wealth (1976b). *Initial 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*, 442–454.

Salem, A. B. Z. and T. D. Mount (1974). A convenient descriptive model of income distribution: The Gamma density. *Econometrica 42*, 1115–1127.

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*, 113–120.

Saposnik, R. (1981). Rank-dominance in income distribution. *Public Choice 36*, 147–151.

Saposnik, R. (1983). On evaluating income distributions: Rank dominance, the Suppes-Sen grading principle of justice, and Pareto optimality. *Public Choice 40*, 329–336.

Satchell, S. E. (1978). Source and subgroup decomposition inequalities for the Lorenz curve. *International Economic Review 28*(6), 321–329.

Saunders, T. J. (1970). *Plato: The Laws*. Harmondsworth: Penguin.

Schutz, R. R. (1951). On the measurement of income inequality. *American Economic Review 41*, 107–122.

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*, 219–231.

Sen, A. K. (1976b). Real national income. *Review of Economic Studies 43*.

Sen, A. K. (1977). On weights and measures: informational constraints in social welfare analysis. *Econometrica 45*, 1539–1572.

Sen, A. K. (1980). Equality of what? *The Tanner Lectures on Human Values 1*.

Sen, A. K. (1992). *Inequality Reexamined*. Harvard University Press.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica 48*, 613–625.

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*, 1369–1385.

Shorrocks, A. F. (1988). Aggregation issues in inequality measurement. In W. Eichhorn (Ed.), *Measurement in Economics*. Physica Verlag Heidelberg.

Shorrocks, A. F. and J. E. Foster (1987). Transfer-sensitive inequality measures. *Review of Economic Studies 54*, 485–498.

Silber, J. (1989). Factor components, population subgroups and the computation of the Gini index of inequality. *Review of Economics and Statistics 71*, 107–115.

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*, 425–440.

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*, 607–617.

Singh, S. K. and G. S. Maddala (1976). A function for the size distribution of income. *Econometrica 44*, 963–970.

Slottje, D. J. (1984). A measure of income inequality based upon the beta distribution of the second kind. *Economics Letters 15*, 369–375.

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.

Statistics, E. (1966). New York: John Wiley.

Steindl, J. (1965). *Random Processes and the Growth of Firms: A Study of the Pareto Law.* New York: Hafner Press.

Summers, R. and A. 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. Heston (1991). The Penn world table (mark 5): an expanded set of international comparisons 1950-1988. *Quarterly Journal of Economics 106*, 327–368.

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). *Dtnamic Changes of Income and its Distribution in Japan.* Tokyo.

Tawney, H. R. (1964). *Equality.* London: Allen and Unwin.

Temkin, L. S. (1986). Inequality. *Philosophy and Public Affairs 15*, 99–121.

Thatcher, A. R. (1968). The distribution of earnings of employees in Great Britain. *Journal of The Royal Statistical Society A131*, 133–170.

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North Holland.

Theil, H. (1979a). The measurement of inequality by components of income. *Economics Letters 2*, 197–9.

Theil, H. (1979b). World income inequality and its components. *Economics Letters 2*, 99–102.

Theil, H. (1989). The development of international inequality. *Journal of Econometrics 42*, 145–155.

Thistle, P. D. (1989a). Duality between generalized Lorenz curves and distribution functions. *Economic Studies Quarterly. 404* (6), 183–187.

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*, 725–728.

Thon, D. (1979). On measuring poverty. *Review of Income and Wealth 25*, 429–439.

Thon, D. (1981). Income inequality and poverty; some problems. *Review of Income and Wealth 27*, 207–210.

Thon, D. (1982). An axiomatization of the Gini coefficient. *Mathematical Social Science 2*, 131–143.

Thon, D. (1983a). Lorenz curves and Lorenz coefficient: a sceptical note. *Weltwirtschaftliches Archiv*, 364–367. Band 119.

Thon, D. (1983b). A note on a troublesome axiom for poverty indices. *Economic Journal 93*, 199–200.

Thurow, L. C. (1970). Analysing the American income distribution. *American Economic Review 60*, 261–269.

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*, 263–277. Reprinted in Phelps (1973).

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.

van der Wijk, J. (1939). *Inkomens- En Vermogensverdeling.* Number 26. Haarlem: Nederlandsch Economisch Instituut.

Weatherburn, C. E. (1949). *A First Course in Mathematical Statistics* (2nd ed.). London: Cambridge University Press.

Weisbrod, B. A. and W. L. Hansen (1968). An income-net worth approach to measuring economic welfare. *American Economic Review 58*, 1315–1329.

Weiss, Y. (1972). The risk element in occupational and educational choices. *Journal of Political Economy 80*, 1203–1213.

Weymark, J. A. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences 1*, 409–430.

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*, 344–369,485–511.

Wilson, J. (1966). *Equality*. London: Hutchinson.

Yaari, M. E. (1988). A controversial proposal concerning inequality measurement. *Journal of Economic Theory 44*(4), 381–397.

Yitzhaki, S. (1979). Relative deprivation and the Gini coefficient. *Quarterly Journal of Economics 93*, 321–324.

Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review 24*(10), 617–628.

Yitzhaki, S. and R. Lerman (1991). Income stratification and income inequality. *Review of Income and Wealth 37*(9), 313–329.

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*, 471–484.

Yule, G. U. and M. G. Kendall (1950). *An Introduction to The Theory of Statistics*. London: Griffin.

Zipf, G. K. (1949). *Human Behavior and the Principles of Least Effort*. Reading, Mass: Addison-Wesley.

Index

Abel-Smith, B., 170
Adams, K., 168
Additivity, 38, 128
Addo, H., 161, 164
Aghevli, B.B., 172
Aigner, D.J., 165, 172
Aitchison, J., 76, 85, 144, 167, 172
Alker, H.R.J., 161, 163, 164
Allen, R.G.D., 163
Allisson, P.D., 163, 166, 172
Amiel, Y., 68, 162, 163, 167
Anand, S., 173
Arnold, B.C., 166, 169
Aronson,J.R., 173
Arrow, K.J., 165
Atkinson index, 46, 47, 54, 59, 63, 65, 67, 100, 115, 118, 134, 150, 165, 167
    and Dalton index, 47, 145, 146
    and generalised entropy index, 60, 149
    and information-theoretic measure, 54
    and SWF, 47, 128, 129
    decomposition, 149, 150
    sensitivity of, 111, 117, 118
Atkinson, A.B., xii, xiv, 46, 54, 118, 161, 164–166, 168, 170, 171

Bacon, R., 171
Bagley, C., 170
Basmann, R. L., 164
Basmann, R.L., 169
Basu, K., 164
Batchelder, A.B., 162
Bauer, P.T., 161
Baxter, M.A., 172
Bayeux
    bishop of, 168
Beach, C.M., 172
Beckerman, W., 171
Bentzel, R., 165
Benus, J., 170
Bernadelli, H., 169
Berrebi, Z.M., 164
Berry, A., 171
Beta distribution, 143, 144, 169
Beta function, 143
Bishop, J.A., 63, 164, 172
Bjerke. K., 168
Blackorby, C., 165
Blitz, R. C., 163
Boadway, R., 165
Boes, D., 167
Bonini, C.P., 169
Borooah, V.K., 173
Bossert, W., 164, 165
Boulding, K.E., 161
Bourguignon, F., 166, 171, 172
Bowen, I., 162
Bowman, M.J., 168
Braulke, M., 173
Brazer, A.E., 170
Brittain, J. A., 163
Bronfenbrenner, M., xii, 86, 167
Broome, J., 162
Brown, A.J., 167
Brown, J.A.C., 76, 85, 144, 167, 172
Bruce, N., 165
Budd, E.C., 163, 170
Butler, R.J., 164

Campano, F., 169

Cardinal equivalence, 9, 10, 37, 46, 53, 54, 127, *see* Ordinal equivalence
Cardinal representation, 127, 148
Chakraborti, S., 164, 172
Chakraborty, A.B., 164
Chakravarty, S.R., 164, 169
Champernowne distribution, 143
Champernowne, D.G., 74, 79, 143, 162–164, 167, 169, 172
Ching, P., 173
Chipman, J.S., 168
Coefficient of variation, 24, 25, 33, 59, 61, 89, 111, 112, 127, 152, 153
  and decomposition, 152
  standard error of, 117, 172
Cohen, J.K., 163
Cohen, W.J., 170
Comparative function, 164
Concavity, 37, 38, 40–43, 166
Concentration ratio, 163
Connor, L.R., 163
Constant elasticity, 37, 39, 41
Constant relative inequality aversion, 37, 39, 67
Constant residual progression, 84
Continuous distribution, 101, 136, 138
Cortese, C.F., 163
Coulter, F.A.E., 134, 162, 170
Cowell, F.A., 37, 59, 68, 134, 162, 163, 166, 167, 170–173
Cramer, J.S., 169
Creedy, J., 164
Crew, E.L., 163
Cronin, D.C., 169
Cumulative frequency, 18, 28, 44, 45
  and Parade, 19, 163
Current Population Survey, 92

Dagum, C., 165, 169
Dahlby, B.G., 165
Dalton index, 54, 56, 146
  and Atkinson index, 47, 145, 146
Dalton, H., 54, 163
Das, T., 172
Dasgupta, A.K., 165
Dasgupta, P., 165
Data collection, 92
David, H.A., 163, 172
David, M.H., 170
Davidson, R., 172
Davies, J.B., 172
Davis, H.T., 164, 168
Decomposition of inequality, 11, 59, 60, 62, 65, 130, 173
  by income components, 144, 152, 173
  by population subgroups, 147, 148, 150, 171
Density function, 18, 70, 71, 114, 140, 143
  and interpolation, 114, 156, 171
  estimation of, 102, 154
Ding, H., 168
Distance, 38, 54, 63, 65, 147, 166
  and income differences, 20–22
  and income shares, 51–53, 61, 65, 125
  and principle of transfers, 62
Distance function, 54, 166
Donaldson, D., 164, 165
Dorfman, P., 163
Duncan, B., 163
Duncan, O. D., 163
Dutta, B., 165

Earnings, 5, 28, 85, 86, 88, 89, 95, 99, 129, 130, 168, 170
  lognormal distribution, 85, 86, 168
Ebert, U., 165, 167
Eichhorn, W., 33, 167
Eltetö, O., 164
Entropy, 48, 49
Esberger, S.E., 164
Esteban, J., 165, 169
Evaluation function, 100, 101, 153
Evans, M., 169, 172

Falk, R.F., 163
Family Expenditure Survey, 92, 170
Family Resources Survey, 95, 170
Fase, M.M.G, 168
Fellman, J., 163
Ferber, R., 170
Firms, 87, 166
Fisk, P.R., 142, 169
Formby, J.P., 63, 164, 172
Forsythe, J., 170
Foster, J., 162, 164, 167
Francis, W.L., 164, 166
Frequency distribution, 16, 18, 20, 24, 25, 28, 44–46, 78, 79, 114, 156, 163, *see* Density function
  estimation, 102
  Pareto, 81
Frigyes, E., 164
Frosini, B.V., 167, 173
Functional form, 70, 71, 78–82, 172
  empirical justification, 85–87
  fitting, 120–126, 138, 140–144, 169, 172
  particular distributions, *see* separate entries
Funke, H., 33

Gamma distribution, 144, 169
Gamma function, 142, 144
Gastwirth, J., 163
Gastwirth, J.L., 171, 172
Gehrig, W., 166, 167
Generalised entropy index, 59–61, 63, 133, 137, 139, 147–150, 153, 167, 172
  and moments, 154
  and quasi-linear mean, 167
Generalised Lorenz curve, 42, 43, 68, 152, 166
  and negative income, 152
  estimation, 172
  for Pareto distribution, 89
Geometric mean, 21, 76, 122, 136, 150
Gini coefficient, 23, 33, 34, 58, 63, 65, 68, 76, 89, 90, 100, 101, 110, 115, 121, 125, 136, 153, 162, 163, 171
  and criterion of fit, 124
  and decomposition, 59, 147, 173
  axiomatisation, 167
  generalisations of, 164
  grouped data, 109, 115, 172
  standard error of, 117, 172
Gini, C., 163
Glasser, G.J., 172
Glauberman, M., 171
Glewwe, P., 162, 173
Goldberger, J.S., 172
Graaf, J. de V., 165
Graybill, F., 167
Guthrie, H.W., 170

Hagerbaumer, J.B, 171
Hagerbaumer, J.B., 162
Hainsworth, G.B., 163
Hammond, P.J., 165
Hanna, F.A., 170
Hannah, L., 166
Hansen, W.L., 170
Hardy, G., 165, 167
Harrison, A.J., 169, 170
Harsanyi, J., 162
Hart, P.E., 87
Hastings, N., 169, 172
Hayakawa, M., 169
Hayes, K.J., 169
Heins, A.J., 165
Helmert, F.R., 163
Herfindahl index, 52, 60, 61, 110, 122, 127, 166
Herfindahl, O.C., 57, 166
Heston, A.W., 171
Hill, M.S., 170
Hill, T.P., 168
Histogram
  split, 158
Hochman, H., 162
Hoel, P.G., 167
Howes, S.R., 173

Hussain, A., 173

Income
comparability, 5, 6, 98, 161, 171
differences, 10, 11, 16, 20, 34, 58, 61
equivalised, 38, 96, 162
growth, 129, 132, 162
lifetime, 4, 5, 13, 98
measurability, 6, 161
negative, 33, 68, 152, 153, 163
specification, 95
specification of, 4, 86, 94, 95, 127, 170
time period, 98
Income distribution, 11, 15, 18, 93, 96
analogy with probability, 47, 49
and Lorenz ranking, 42
and Pareto diagram, 79
and Pen' parade, 16
comparison, 6, 8, 21, 27, 29, 43, 58, 60, 61, 65, 75, 164, 171
components of, 130
distance, 3, 53
functional form, 70, 87, 138, 140, 143, 144, 168, 169
grouped data, 106, 109, 115
Laws, 167
sample, 97
truncation, 120
typical shape, 73, 126
Income power, 74
Income share, 2, 29, 30, 38, 50, 53, 56, 60, 65, 112, 125, 147, 166
Inequality
and distance, 51–54, 60–63, 65, 68, 125, 147, 166
and justice, 7
and poverty, 14, 162
aversion, 37, 39, 46, 54, 63, 65, 67, 120, 132, 146, 147, 165
concern for, 11
decomposition, 11, 59, 60, 62, 65, 130, 144, 147, 148, 150, 152, 171, 173
Inequality index, *see* Inequality measures
Inequality measures, *see* see individual measures
and Lorenz curve, 23, 47, 63
approaches to, 35, 36, 60
computation, 99
construction of, 54
definitions, 135–138
for continuous distributions, 138
for discrete distributions, 136
grouped data, 107
how to choose, 60
interrelationships, 144
meaning of, 7
properties, 65, 135–138
sensitivity of, 47, 63
Information function, 51
Information theory, 47, 48, 52–54, 166
Interpolation, 114, 135, 156–158, 171
log-linear, 158
Pareto, 140
Pareto distribution, 171
split histogram, 158
straight line, 158
Iritani, J., 166

Jain, S., 171
Jakobsson, U., 33
Jasso, G., 163, 166
Jencks, C., 161
Jenkins, S.P., 134, 162–164, 170
Jensen, B, 169
Johnson, J.D., 169
Johnson, N.O., 168, 172
Jorgenson, D.W., 162
Justice, 2, 3, 10, 11, 35, 162, 165, 166

Kakwani, N., 169
Kaliski, S.F., 172
Kay, J.A., 166
Keeping, E.S., 143, 163, 164, 167
Kendall, M. G., 72, 172

Kernel function, 154–156
Kloek, T., 169
Kmietowicz Z.M., xiv
Kmietowicz, Z.M., 168
Kolm index, 148
Kolm, S., 165, 167
Kondor, Y., 164, 167
Koo, A.Y.C., 163
Kravis, I.B., 171
Krieger, A.N., 172
Krishnan, P., 166
Kuga, K., xii, 166, 167
Kurtosis, 168, 172
Kuznets, S., 163, 171

Lam, D., 163
Lambert, P.J., xiv, 166, 173
Lanjouw, P., 173
Least squares, 90, 123–125, 172
Lebergott, S., 73
Lerman, R., 173
Lerner, S.M., 170
Levine, D.B., 163
Likes, J., 172
Lin, T., 164
Lindley, D.V., 138
Little, I.M.D., 165
Littlewood, J., 165, 167
Log variance, 25, 34, 65
  and principle of transfers, 146
  non-decomposability, 147
  standard error of, 117
Logarithmic transformation, 20, 21, 73, 74, 87, 150
Logisitic function, 142
Lognormal distribution, 25, 71–73, 76, 77, 83, 84, 89, 121, 122, 138, 168, 172
  and aggregation, 86
  and Lorenz curve, 75, 138
  estimation of, 122
  three-parameter, 76, 144
Lomnicki, Z.A., 172
Lorenz curve, 19, 20, 23, 33, 43, 133, 134, 149, 163, 166, *see* Generalised Lorenz curve
  absolute, 153
  and hypothesis testing, 172
  and incomplete moments, 164
  and inequality measures, 23, 47, 63
  and interpolation, 114, 156
  and negative income, 152, 163
  and principle of transfers, 55
  and shares ranking, 30, 31
  and SWFs, 42
  and Theil curve, 50
  and variance of logarithms, 76
  bounds on Gini, 171
  computation, 101, 113
  convexity, 19, 113, 163
  definition, 138
  empirical, 112
  for lognormal distribution, 75, 138, 167
  for Pareto distribution, 82
  parameterisation, 142, 169
  symmetric, 75
Lorenz, M.O., 163
Love, R., 166
Lydall, H.F., 86, 164

Maasoumi, E., 166
Maddala, G.S., 142, 169
Majumder, A., 169
Malik, H.J., 172
Malmquist, S., 164
Mandelbrot, B., 169
Marfels, C., 166
Markowski, S., 164, 170
Marshall, A.W., 165
Maximum likelihood estimate, 122
Maynes, E.S., 170
Mayston, D.J., 165
McClements L., 170
McDonald, J.B., 164, 169, 171
McGregor, P.P.L., 173
McKee, P.M., 173
Mead, J.E., 165
Mehran, F., 171, 172
Mehta, F., 171
Mera, K., 165

Metcalf, C.E., 169, 171
Method of moments, 122
Method of percentiles, 164
Micklewright, J., xiv, 118, 171
Miller, J.C.P., 138
Miller, S.M., 161
Minimal majority measure, 24, 26, 164
Mirrlees, J.A., 165
Mobility, 1, 11, 169
Moments, 141–144, 153, 164
 method of, 122, 123
Mood, A.M., 167
Mookherjee, D., 173
Morgan, J.N., 170
Morrell, A.J.H., 163
Morrisson, C., 171
Mount, T.D., 169
Moyes, P., 153, 166
Muellbauer, J., 171
Mumbo Jumbo, 168
Musgrave, R. A., 77

Nair, U.S., 172
Nayak, T.K., 172
Nicholson, R.J., 168
Non-comparability, 6, 130, 131
Normal distribution, 71, 73–77, 138, 167, 172
Nygård, F., 171, 172

O'Higgins, M., 162, 171
Okner, B.A., 170
Okun, A.M., 162
Olkin, I., 165
Ordinal equivalence, 10, 13, 47, 59–62, 112, 121, 127, 149, 150, 152, 163, 166, *see* Cardinal equivalence

Paglin, M., 162
Parade of Dwarfs, 16, 18, 19, 21–25, 28, 33, 35, 68
Parade of incomes, 41, 43, 47, 50, 62, 101, 130
Parade ranking, *see* Quantile ranking
Pareto distribution, 78–80, 82–84, 86, 89, 111, 141, 169
 and inequality, 82
 and interpolation, 114
 criteria of fit, 125
 estimation, 121–123
 evidence, 168
 Generalised Lorenz curve of, 89
 interpolation form, 140
 Lorenz curve of, 82
 properties, 138, 140
 type III, 141
Pareto's α, 78, 80–84, 87, 89, 111, 114, 168
 and average/base index, 82
 and French revolution, 168
 and inequality, 83
 and van der Wijk's law, 141
 estimation of, 121, 123–125, 157, 168, 172
 in practice, 86, 87, 168
Pareto, V., 167, 168
Pareto-Levy law, 169
Parikh, A., 172
Paukert, F., 171
Peacock, B., 169, 172
Pearce, D.W., 165
Pechman, J.A., 170
Pen, J., 16, 33, 79, 163, 168
Pfingsten, A., 165
Phelps, E.S., 162
Pigou, A.C., 168
Plato, 21, 163
Podder, N., 169
Polanyi, G., 164
Polovin, A., 163
Polya G., 165, 167
Poverty, 12, 14, 15, 25, 27
 and inequality, 14, 162
 line, 12, 26
 measure, 56
Prais, S.J., 87
Prest, A.R., 161, 170
Principle of population, 56, 57
Principle of transfers
 and log variance, 146

and Lorenz curve, 55
strong, 60–62, 166
weak, 55, 56, 59, 65, 75, 83, 146, 157
Principle of transfwers
weak, 128
Ptolemaic system, 71
Pyatt, G., 173

Quan, N.T., 163
Quandt, R., 169, 172
Quantile ranking, 29, 30
and SWF, 41
Quantiles, 28, 29, 50
and dispersion, 29
Quasi-linear mean, 167

R-squared, 124–126
Radner, D.B., 170
Rainwater, L., 171
Rajaraman, I., 168
Random process, 79, 167
Ranking, 27, 28, 30, 36, 61, 76, 128, 152, 164
and decomposition, 57, 59, 60
ordinal equivalence, 47
quantiles, 29, 41
shares, 30, 31, 42
Ransom, M.R., 169, 171
Rao, C.R., 154
Rasche, R., 163
Ravallion, M., 162
Rawls, J., 162, 165
Rees, J., 162
Reichmann, W.J., 167, 172
Rein, M., 161
Relative mean deviation, 22–24, 34, 35, 65, 100, 111, 115, 153
distance concept, 61
non-decomposability, 59, 147
relation to Lorenz curve, 23
standard error of, 117, 172
Richmond, J., 172
Richter, W.F., 33
Riese, M., 163
Rietveld, P., 166
Rodgers, J.D., 162
Rosenbluth, G., 166
Russet, B.M., 161, 164

Salem, A.B.Z., 169
Sandström, A., 171, 172
Saposnik, R., 165
Satchell S.E., 173
Schutz, R.R., 163
Seidl, C., 162
Semi-decile ratio, 164
Sen, A.K., 161, 162, 164, 165
Shorrocks, A.F., 166, 167, 172, 173
Silber, J., 164, 173
Simon, H.A., 169
Singer, N.M., 163
Singh, S.K., 142, 169
Skewness, 25, 164, 172
Slesnick, D.T., 162
Slottje, D.J., 164, 169
Smeeding, T.M., 171
Smith, J.T., 172
Smith, W.P., 63
Social utility, 37, 38, 40, 44, 45, 54, 67, 146
Social wage, 5
Social-welfare function, 36, 39, 41–44, 47, 61, 68, 74, 128, 129, 132, 165
Soltow, L., 168
Standard error, 116, 117, 122, 125, 153, 172
particular indices, *see* separate entries
Stark, T., 25, 162, 170
high/low index, 26, 164
Starrett, D.A., 165
Steindl, J., 87
Stern, N.H., 173
Stiglitz, J.E., 165
Stuart, A, 172
Summers, R., 171
SWF, *see* Social-welfare function

Taille, C., 169
Takahashi, C., 168

Tawney, R.H., 167
Tax returns, 93, 94, 96, 103, 170
Thatcher, A.R., 168, 169
Theil Curve, 50
Theil index, 50, 57–59, 61, 66, 100, 115, 136, 148, 166
  and distance concept, 166
  and transfers, 51
  axiomatisation, 167
  decomposition, 171
  estimation of, 172
Theil, H., 48, 166, 171, 173
Thin, T., 77
Thistle, P.D., 68, 166, 172
Thon, D., 14, 162, 167
Thurow, L.C., 161, 169
Tobin, J., 161
Tuomala, M., 165

Utility function, 38, 46, 74, 165

Van der Wijk's law, 81–83, 140, 143, 168
Van der Wijk, J., 168
Van Dijk, H.K., 169
Variance, 24, 56, 59, 61, 65, 70, 72, 77, 90, 134
  and Normal distribution, 77
  decomposability, 150
  of log income, 24
  sample, 154
Variance of logarithms, 59, 76, 117, 122, 146, 147, 150, 164, 172
  non-decomposability, 150
Villeins, 168
Voting, 2, 24, 92, 96, 161

Walden, B., 172
Wealth, 4, 5, 12, 36, 37, 44, 71, 87, 91, 92, 99, 103, 128, 168
  and income, 170
  and Pareto distribution, 78, 87, 90, 168
  and slaves, 168
  data on, 94–96, 98, 103
  definition of, 170
  distribution of, 19, 24, 71, 73, 87
  valuation of, 5, 98
Weatherburn, C.E., 143, 172
Webley, P., 168
Website, 159
Weibull distribution, 142
Weisbrod, B.A., 170
Weiss, Y., 85
Welfare index, 6, 37, 38, 40, 44, 54
Weymark, J.A., 164
Wiles, P.J. de la F., 164, 170
Wilson, J., 162
Wolfson, M.C., 166
Wood, J.B., 164
Wretman, J.H., 172

Yaari, M., 164
Yitzhaki, S., 164, 173
Yoshida, T., 165
Young, A.A., 164
Yule distribution, 144, 169
Yule, G. U., 72

Zipf, G.K., 168

@@

REFERENCES

Hempenius, A.L. (1984), “Relative income position individual and social income satisfaction, and income inequality”, De Economist, 32, 468-478.

Lambert, P.J. (1980), “Inequality and social choice”, Theory and Decision, 12, 395-398.

Shorrocks, A.F. (1978), “Income inequality and income mobility”, Journal of Economic Theory, 19, 376-393.