The Economics of Poverty and Inequality: Introduction

Frank Cowell
London School of Economics

Revised August 2002
15 Dynamics 25

16 Functional Forms of Income and Wealth Distribution 25
  16.1 The choice of functional form .......................... 26
  16.2 Inequality in parametric models ..................... 27
  16.3 Estimation ........................................... 27

17 Statistical Issues 27

List of Figures

1 $G$ first-order dominates $F$ ................................. 7
2 $G$ second-order dominates $F$ ............................. 8
3 $G$ (relative) Lorenz-dominates $F$ ...................... 9
4 The equally-distributed equivalent is less than the mean ... 11
1 Introduction

These two volumes focus on the economics of distributional equity and the way general criteria for comparisons of income distributions can be used to inform the analysis of inequality and poverty. The issues addressed by the literature covered in them include:

² The nature of general ranking rules for comparing economic states based on simple ethical principles.

² The close relationship between the analysis of poverty and that of inequality or social welfare.

² The structure and properties of inequality and poverty indices.

The following synopsis will provide an overview of the modern approach to these issues and an indication of how some of the classic papers fit into this literature. References to the brief combined bibliography at the end of this introduction are given by author and year of publication; references to papers in the collection are marked with an asterisk.

1.1 Income distributions

The elements of the modern approach to inequality and poverty measurement involve the definition of an income concept, an ethical or other basis for distributional comparisons and a set of assumptions or axioms that give meaning to an ordering or ranking principle.

"Income" in practice may mean wealth or expenditure. For much of the modern literature "income" plays the role of a personal welfare index or utility; this is sometimes articulated as "equivalised income" – nominal income normalised by an index of needs (see Section 12). Formally we assume that an individual's income $x$ belongs to an interval $\mathcal{X} := [\underline{x}, \overline{x}] \cap \mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers and $\mathcal{X}$ is a proper interval: $\mathcal{X}$ incorporates an implicit assumption about the logically possible values that $x$ could adopt and will be determined by the precise economic definition of "income". So an income distribution is a vector

$$x = (x_1, x_2, \ldots)$$

and the set of valid income distributions is given by $\mathcal{X} := f(\mathcal{X}^n : n = 2, 3, \ldots)$. A welfare ordering is a relationship $< \text{ defined on } \mathcal{X}$ such that for every $x, x^0 \in \mathcal{X}$ either $x < x^0$ is true or $x^0 < x$ is true or both are true. The
statement \( x < x^0 \) means "the welfare represented by the income vector \( x \) is at least as great as that represented by \( x^0 \)." We also write \( x \succ x^0 \) where both \( x < x^0 \) and \( x^0 < x \) are true and we use the terminology \( x \preceq x^0 \) where it is true that \( x < x^0 \) but it is not true that \( x^0 < x \). A welfare ranking means a quasi-ordering \( \prec \) on \( X \) that permits also cases where neither \( x < x^0 \) nor \( x^0 < x \) is true.

1.2 Axioms

In the modern literature on social welfare, inequality and poverty the basic properties of a welfare ranking or ordering are usually "programmed in" by introducing an axiom system: this is essentially just a way of thinking about a subject and need not have any particular moral or ethical base, nor any rigorous empirical support – although an axiom system that is completely at variance with practical experience is likely to be of little relevance. Some of the standard axioms are

Axiom 1 Anonymity. \( \prec \) is invariant under a permutation of the components of \( x \). 2 \( X \).  

Axiom 2 The population principle. Let \( x[m] \) denote an \( m \)-fold replication of \( x \): if \( x \prec x^0 \), then \( x[m] \succ x \).  

Axiom 3 Monotonicity. For any \( x, x^0 \in X^n \): \( x > x^0 \) implies \( x \preceq x^0 \).  

Axiom 4 The transfer principle. If \( x^0 \in X^n \) can be obtained from \( x \in X^n \) by a poorer-to-richer income transfer then \( x \preceq x^0 \).  

Axiom 5 Scale invariance. 1 For any \( x, x^0 \in X^n \) and any positive scalar \( \lambda \) such that \( \lambda x, \lambda x^0 \in X^n \): \( x < x^0 \) \( \lambda x < \lambda x^0 \).

The implications of these will become evident in the discussion of the papers contained in sections 2 and 4.

1.3 Distributions: an alternative representation

Given axioms 1 and 2 the welfare theory of income distributions can be conveniently represented using the "F-form" approach (Cowell 2000). Let \( F \) be the space of all univariate probability distributions with support \( X \): then \( x \in X \) is a particular value of income and \( F \) is one possible distribution

\[ \text{Also known as homotheticity. An alternative is the translation invariance axiom: For any } x, x^0 \in X^n \text{ and a vector } a \in \mathbb{R}^n \text{ such that } x + a, x^0 + a \in X^n \text{ where } x < x^0 \text{ then } x + a < x^0 + a. \]
of income in the population; \( F(x) \) is the proportion of the population with income less than or equal to \( x \). Using the concept of the \( F \)-function one can conveniently capture a wide range of theoretical and empirical distributions, including some important special cases - see Section 16. In particular, for a ranking \( < \) defined on \( X \) one can define a corresponding ranking \( <^w \) defined on \( F \) : if \( F, G \) are the distribution functions representing vectors \( x, x^0 \) in \( X \) then \( x < x^0 \) implies \( F <^w G \) and vice versa.

In the case of welfare orderings, the terminology can be made more friendly and the underlying concept more transparent: if one introduces a continuity assumption, then standard results imply that \( <^w \) can represent a welfare ordering generally as a functional \( W : F \rightarrow \mathbb{R} \) (Fishburn 1970). In other words there is a real-valued functional \( W \) such that \( "F <^w G" \) is equivalent to \( "W(F) < W(G)". \) This is what is meant by a social-welfare function.

This formal apparatus is useful for understanding the modern theoretical literature on both poverty and inequality.

2 The Welfare Basis of Distributional Analysis

The papers in section 2 deal with the welfare-economic basis of the analysis of inequality and poverty. The axioms outlined above include two key ethical properties:

2 Axiom 3 means that if some of the mass of the distribution \( F \) is shifted to the “right” this must improve welfare.

2 Axiom 4 owes its origin to a reworking by Dalton (1920)* of an idea by Pigou (1912)* – a mean-preserving income transfer from any one person to someone richer must reduce social welfare and increase inequality.

The transfer principle is the criterion that is overwhelmingly used in the literature to introduce a concern for inequality into judgments about income distribution. But why should one accept this as appropriate? On what is concern for inequality supposed to be based? There are several possibilities:

2 Analogy with risk. One is to exploit a relationship with the analysis of risk (Harsanyi 1955)*. Suppose we impose some structure on the function \( W \) by making it additively separable:

\[
W(F) = \int_{x \geq x^0} u(x) dF(x)
\]  

(2)
where \( u : X \rightarrow \mathbb{R} \) is an evaluation function of individual incomes. Then (2) has a convenient interpretation: social welfare is found by assigning a social evaluation to each income \( -u(x) \) – weighting the result by the proportion of the population that have that income \( -dF(x) \) – and summing. It can be seen as “expected utility” where a representative person in the population regards it as equally likely that he should have any of the income-entitlements in the income distribution. Social aversion to inequality is then determined by individual aversion to risk.

² Individual tastes on distribution. The basis of inequality-aversion could be rooted in other forms of individual distributional judgments – for example, the individual valuation of an externality involving other people’s incomes or living standards (Hochman and Rodgers 1969)*. Inequality aversion is determined by the marginal utility of the externality.

² Pragmatic approaches. An alternative approach is to suppose that social values, including inequality aversion, will be revealed by public policy decisions. Amiel and Cowell (1992)* examine the extent to which the standard axioms applied in inequality analysis conform to the way in which lay people make distributional comparisons.

² Philosophy. Temkin (1986)* has suggested a formulation based on an individualistic (but not welfarist) notion of “complaints” and this has been used to formulate a class of inequality measures (Cowell and Ebert 2002). Vallentyne (2000)* suggests an approach based on the priority of the worst off.

3 Welfare and Inequality Rankings

A single social-welfare function \( W \) is arguably an over-restrictive concept with which to work. Much of the welfare-economics literature has focused on important classes of SWFs from which broad lessons can be drawn and the papers in section 3 provide a representative example of this approach.

The basic formal tools required are these: we let \( W \) denote the set of functions \( W \) of the form (2), \( W_1 \) be the subclass of \( W \) where \( u \) is increasing; and \( W_2 \) the subclass of \( W_1 \) where \( u \) is also concave. The SWF subclasses \( W_1 \) and \( W_2 \) play a crucial role in interpreting two fundamental ranking principles – first- and second-order distributional dominance (stochastic dominance). It should be noted that these concepts are not abstruse constructs. They have a close relationship with intuitive practical tools – quantiles and income shares.
Figure 1: \( G \) rst-order dominates \( F \)

In particular the second order dominance criterion is closely related to one of the most important tools of inequality analysis – the Lorenz curve (Lorenz 1905)*.

3.1 First-order dominance

First-order dominance criteria are based on the quantiles of the distribution which are yielded by the inverse of the distribution function \( F \). This concept is defined as:

Definition 1 For all \( F \in F \) and for all \( 0 \leq q \leq 1 \), the quantile functional is defined by:

\[
Q(F; q) := \inf \{ x \mid F(x) \geq q \} = x_q
\]

For example \( Q(F; 0.1) \) is the rst decile of the distribution \( F \). For any distribution of income \( F \), the graph of \( Q \) describes Pen’s Parade (Pen 1971)*. The quantiles contain important information about economic welfare: if every quantile in distribution \( G \) is greater than the corresponding quantile in distribution \( F \) – written \( G \succ Q F \) (see Figure 1) – then distribution \( G \) will be assigned a higher welfare level by every SWF in class \( W_1 \) (Saposnik 1981*, 1983):

Theorem 1 \( G \succ Q F \) if, and only if, \( W(G) \succ W(F) \in (W_2 W_1) \).
3.2 Second-order dominance

The first-order ranking-principle \( \pi_q \) does not employ all the standard principles of social welfare analysis: it neglects the principle of transfers. For this reason it is useful to introduce the second-order dominance criterion which requires the following concept

Definition 2
For all \( F \geq G \) and for all \( 0 \leq q \leq 1 \), the cumulative income functional is defined by:

\[
C(F; q) := \int \mu(F(x)) q(x) dF(x),
\]

(4)

Note that, by definition, \( C(F; 0) = 0 \), \( C(F; 1) = \mu(F) \), and that, for a given \( F \geq G \), the graph of \( C(F; q) \) against \( q \) describes the generalised Lorenz curve (GLC). If every income cumulation in distribution \( G \) is greater than the income cumulation in distribution \( F \) – written \( G \preceq C F \) (see Figure 2) – then distribution \( G \) will be assigned a higher welfare level by every SWF in class \( W \) (Marshall and Olkin 1979, Shorrocks 1983):

Theorem 2
\( 8F, G \geq F \): \( G \preceq C F \) if, and only if, \( W(G) \succeq W(F) \) \( 8(W \geq W) \).

The second-order dominance relationship is closely related to the Lorenz curve: normalise the cumulative income functional by the mean:

\[
L(F; q) := \frac{C(F; q)}{\mu(F)}
\]

(5)
The graph of $L(F; q)$ against $q$ – the Lorenz curve – yields a simple distributional-shares ranking: in Figure 3 it is evident that the income share of the bottom 100$q\%$ of the population must be higher in distribution $G$ than in $F$, whatever the value of $q$.

The essence of Theorem 2 was originally obtained for distributions with a given mean (Atkinson 1970*, Kolm 1969*). Let $F(\mu)$ denote the subset of $F$ that has a given mean $\mu$:

**Theorem 3** $8F, G \supset F(\mu): G \circ L, F$ if, and only if, $W(G) \geq W(F) 8(W 2 W 2)$.

In cases where Lorenz curves cross further restrictions can usefully be imposed upon the $W$-classes so as to generate third and higher-order dominance concepts (Fishburn and Willig 1984)*.

### 4 Inequality Measurement

An inequality index is a functional $I : F(\mu) \rightarrow R$ that is a counterpart to the concept of SWF introduced above. Some of the most persuasive inequality measures have been founded (at least in part) on intuitive appeal: the obvious example is the Gini coefficient (Gini 1921)*. Some have been developed using analogies to measurement in the context of other issues involving
distributions, such as information theory (Theil 1967)* or choice under un-
certainty (Yaari 1988)*. However the bulk of the recent literature has focused
in one of two main directions:

2 explicit links with social-welfare analysis – see section 5.
2 examination of issues of formal structure – see section 6.

5 Inequality: Welfare Approach

Theorem 3 is fundamental to the modern work on the welfare-economic ap-
proach to inequality. Although the original derivation used a restriction on
$W_2$ that limited the SWF to the form (2) Dasgupta et al. (1973)* and Roth-
schild and Stiglitz (1973)* showed that it applies to a broader class than
those of the form (2). This work also lays the basis for the derivation of spe-
cific welfare-based inequality measures. The papers in section 5 cover the
key contributions that develop the link between welfare theory and inequality
measurement.

A simple transformation of the SWF yields an important practical tool:
the equally-distributed equivalent can be defined as a money measure of so-
cial welfare. Let $H(x)$ denote a distribution that puts all the mass of a
distribution at the single point $x$; we can use this to provide an implicit
definition of a number $\xi$ such that

$$W(H(\xi)) = W(F)$$

This can be used to yield the equally-distributed equivalent as a functional
$F \forall R$; in other words, given a distribution $F$, $\xi(F)$ may be extracted from
equation (6). The expression $\xi(F)$ is that income which, if it were imputed to
every income-receiver in the population would yield the same level of social
welfare as the actual income distribution $F$ (Atkinson 1970*, Kolm 1969*).

Figure 4 illustrates the idea. Let point $F$ represent an income distribution
in a two-person economy; then mean income

$$\mu(F) = \int x dF(x)$$

can be found as the abscissa of the point $M$ where the 45° line through $F$
intersects the equality ray; the equally-distributed equivalent $\xi$ is the abscissa
of the point $E$ where the $W$-contour through $F$ intersects the equality ray.
The normalised gap between $\xi$ and $\mu$ provides a natural basis for an inequality
index:

$$I_A(F) := 1 \text{i} \frac{\xi(F)}{\mu(F)}$$
Expression (8) permits a general approach to social-welfare values interpreted as aversion to inequality: for any given income distribution the more sharply convex to the origin is the contour in Figure 4, the greater is the gap between $\xi$ and $\mu$; in an extreme case, given that $W 2 W$, one would get L-shaped “max-min” contours (Hammond 1975). In the absence of the restriction to $W$ other concepts of extreme inequality aversion such as Meade’s “superegalitarianism” could be introduced (Meade 1976*, page 49).

To obtain a specification inequality measure we need to impose more structure on $W$. If we also require that Axiom 5 hold (this means that inequality contours are homothetic) then $\xi$ in (8) becomes a kind of generalised mean (Atkinson 1970)*:

$$I^{\eta}(F) := \frac{1}{\mu(F)} \int Z x^{\eta} dF(x)$$

where $\eta > 0$ is a parameter defining (relative) inequality aversion. Alternative assumptions about structure and normalisation of $W$ and $\xi$ could be made which will induce alternative families of inequality measures. Suppose that instead of scale invariance, we require that inequality or welfare comparisons satisfy translation invariance — see note 1. Then, instead of the “relative”

$$I^{\eta}(F) := 1 \int_0^1 \exp \int \log(x) dF(x)$$

The limiting form of (9) as $\eta \to 1$ is $I^1(F) := 1 \int_0^1 \exp \int \log(x) dF(x)$. 

\[ \]
measures (9), we will instead the “absolute” indices

\[ I^\beta_K(F) := \frac{1}{\beta} \cdot \mathcal{Z} e^{\beta x_i \mu(F)} dF(x) \]  

(10)

where \( \beta > 0 \) is a sensitivity parameter (Kolm 1976a)*.

6 Inequality: Structure

An alternative approach to the characterisation of inequality measures is to derive them from considerations concerning the structure of distributional comparisons. The papers in section 6 touch on two major aspects of this.

6.1 Inequality and income transformations

The assumed relation between inequality and transformations of the income distribution can be a useful way of narrowing down the large range of mathematical tools that are available for use in inequality measurement. For example, requiring that inequality should remain unchanged under proportional increases or decreases in all incomes simultaneously imposes an important restriction on the class of inequality measures - this issue is discussed in Blackorby and Donaldson (1978)* and Kolm (1976b)*.

Furthermore indices that have an appealing intuitive interpretation can sometimes be completely characterised by reasonably plausible systems of axioms. A good example of this is the Gini index which has long played a central role in the inequality literature. This index can be expressed in a number of equivalent forms:

\[ I_{\text{Gini}}(F) := \frac{1}{2 \mu(F)} \mathcal{Z} \mathcal{Z} j x_i x_0 j dF(x) dF(x_0) \]  

(11)

\[ = \frac{1}{Z} \int_{\mathcal{Z} 0} L(F; q) dq \]  

(12)

\[ = \int_{\mathcal{Z} 0} x \kappa(x) dF(x) \]  

(13)

where \( x, x^0, x^2 \) and \( 8F \ g \), \( x \ g \), and \( \kappa(x) := [F(x^i) + F(x^+) \ g \mu(F) \ g 1] / \mu(F) \).

The Gini coefficient has a number of practical advantages: for example, it deals with negative incomes and it satisfies both the scale-invariance and translation-invariance principles. Such scale- and translation-invariant measures are sometimes known as compromise indices (Blackorby and Donaldson 1980b) - see Ebert (1988)* for a general characterisation.
6.2 Decomposition

Suppose people are distinguished by attributes that permit partitioning the population into distinct subgroups. This is essential in analysing the relationship between overall inequality and inequality within and between subgroups categorised by gender, ethnicity and the like, or in attempts to “account for” the level of, or trend in, inequality by components of the population. Suppose the population consists of a set of $N$ individuals: a partition of $N$ is a collection of $J$ subgroups

$$\pi = f N_1, N_2, ..., N_J g.$$  \hspace{1cm} (14)

such that everyone belongs to one and only one subgroup. An example of a simple principle of decomposability can be seen in the case of social welfare:

Axiom 6 Subgroup consistency: if welfare in a subgroup increases, overall welfare must increase.

Then, if Axiom 6 holds for all $\pi$, the SWF can be expressed in the form of (2).

To see how similar concepts can be applied in the case of inequality take a partition $\pi$ such that a proportion $p_j$ of the population belong to subgroup $j$, $j = 1, 2, ..., J$; let $F(j)$ be the income distribution in group $j$ and

$$s_j := \frac{p_j F(j)}{I(F)}$$

be the income share of group $j$. Then the definition of decomposability can be expressed in a number of equivalent forms. Applying Axiom 6 in the inequality context requires that inequality overall $I(F)$ can be expressed using the basic decomposition relation:

$$I(F) = \bigoplus I \left[ I \left[ F^{(1)} \right] \right] + I \left[ I \left[ F^{(2)} \right] \right] + \cdots + I \left[ I \left[ F^{(J)} \right] \right] \frac{p_1, p_2, ..., p_J, s_1, s_2, ..., s_J}{15}$$

where $\bigoplus$ is increasing in each of its first $J$ arguments (Shorrocks 1984)*.

This can be seen as a minimal requirement for decomposability, but one might wish for a more demanding interpretation of decomposability. Additive decomposability requires:

$$I(F) = \sum_{j=1}^{J} \omega_j I \left( F^{(j)} \right) \frac{p_1, p_2, ..., p_J, s_1, s_2, ..., s_J}{16}$$

where

$$\omega_j = w(p_j, s_j) \geq 0,$$  \hspace{1cm} (17)

and

$$\sum_{j=1}^{J} \omega_j = 1$$  \hspace{1cm} (18)

where the distribution $F_j$ could be the distribution that would arise if everyone in group $j$ had the mean income of group $j$. A more demanding interpretation of decomposability imposes the additional restriction

$$\sum_{j=1}^{J} \omega_j = 1$$

13
which is perhaps an “accountant’s approach” to decomposition: the weights in the within-group component sum exactly to 100 percent.

Theorem 4 Inequality measures that are scale-invariant and additively separable must take the form:

\[ I_{\alpha}^{GE}(F) := \frac{1}{\alpha^2} \int \frac{x^{\alpha}}{\mu(F)} \cdot 1 \cdot dF(x) \]  

(19)

The class (19) is known as the Generalised Entropy (GE) indices and \( \alpha \) is a parameter that captures the sensitivity of a particular GE index to particular parts of the distribution: for \( \alpha \) large and positive the index is sensitive to changes in the distribution that affect the upper tail; for \( \alpha \) negative the index is sensitive to changes in the distribution that affect the lower tail. In this case the weights in (16) take the form

\[ \omega_j = \omega(p_j, s_j) = p_j^{1 - \alpha} s_j^\alpha. \]  

(20)

If one further requires the property (18) then only two measures are available: the MLD index where the weights are population shares \( (\alpha = 0 \) in equation 20), and the Theil index where the weights are income shares \( (\alpha = 1) \) (Bourguignon 1979*, Cowell 1980*, Shorrocks 1980).

7 Multidimensional approaches

In many cases the problem of analysing income distributions is essentially one of multivariate rather than univariate analysis – the papers in section 7 deal with three aspects of this difficult area. The general problem is inherently complex because one has to take into account the interaction amongst variables, whether interpreted as the interrelation between income and non-income personal attributes, or as multiple components of income involved in a multidimensional generalisation of the Lorenz curve and related concepts (Atkinson and Bourguignon 1982, Cowell 1985*, Kolm 1977*).

One specific application of the multivariate case is of particular interest – the decomposition of inequality by income source. The idea can be appreciated in the simplified case where we take the variance as a measure of inequality and income \( x \) consists of two components; then, taking a bivariate distribution we have by definition an elementary variance decomposition:

\[ \text{var}(x_1 + x_2) = \text{var}(x_1) + \text{var}(x_2) + 2\text{cov}(x_1, x_2). \]  

(21)

Is it possible to apply the same sort of decomposition rule to determine the impact of the inequality of income component \( j \) upon the inequality of total
income for other inequality measures and an arbitrary number \( r \) of income components? For the Gini coefficient, using (13), we get

\[
I_{\text{Gini}}(F) = \int x \kappa(x) \, dF(x) = \sum_{j=1}^{r} x_j \kappa(x_j) \, dF(x),
\]

where \( F \) is the joint distribution function of the \( r \) variables \( x_1, x_2, \ldots, x_r \) and \( \hat{F} \) the distribution of \( x = \sum_{j=1}^{r} x_j \). The term inside the brackets in (22) is typically used as the basis for specifying the “contribution” to inequality of income component \( j \); but this term is not a true inequality index but a concentration coefficient. However, without further restriction on the decomposition rule, the assignment of these inequality-contributions is non-unique (Shorrocks 1982)*.

8 Polarisation

The concept of polarisation is distinct from that of inequality, although the reasons for concern with the two phenomena may be similar. Whereas inequality is mainly concerned with the convergence to the global mean, polarisation has more to do with clustering around local means. “Polarisation” has been employed to explain distributional processes common to some developed economies such as the disappearing middle class (Wolfson 1994)*.

The axioms used to construct a summary measure of polarisation are different from those used in inequality measurement. The classic reference Esteban and Ray (1994)* defines polarisation using the following basic principles:

1. The higher the degree of homogeneity within each group the larger is polarisation.
2. The higher the degree of heterogeneity across groups the larger is polarisation.
3. There must be a small number of significantly sized groups.
4. The smaller the number of relevant groups the larger is polarisation.

9 Horizontal Equity

Horizontal inequality represents, along with the topic of section 8, a further aspect concern for income distribution that is not adequately captured by
conventional concepts of inequality. The central issue concerns the comparison of two income distributions when the relative positions, or ranks, of individuals in the distributions change: comparisons of Lorenz curves or of inequality measures in the usual way reveal none of this position-changing since, by Axiom 1, inequality remains unchanged under an arbitrary permutation of the income receivers. The relationship between Lorenz comparisons, the notion of reranking and income mobility is examined in Atkinson (1980)*. Plotnick (1981)* introduces a specific index of inequity building on this use of Lorenz curves.

10 Poverty Concept and Poverty Line

Although there is a plethora of techniques and indices for poverty measurement (see Callan and Nolan 1991, Foster 1984, Hagenaars 1986, Jäntti and Danziger 2000, for surveys) much of the fundamentals of formal poverty analysis can be derived from the concepts and methods introduced in sections 2 and 4. Poverty can be seen as another aspect of “diswelfare” and there are several connections between the modern theory of inequality measurement and poverty analysis; the contributions in section 10 cover the basic issues involved.

There are three principal logical steps to poverty measurement:

1. The fundamental partition of the population into the poor and the non-poor. We need to be clear about the economic or social norms which separate the two groups.

2. The way in which persons are to be identified individually as poor or non-poor.

3. The way in which information about the income distribution of the poor and the non-poor is to be used, perhaps to produce some sort of aggregate index.

10.1 Partitioning the population

One of the principal threads connecting the poverty analysis to the work of previous sections is the structural analysis of the type considered in Section 6.2. An operational approach to poverty requires the specification of a poverty line $x^*$ that induces the fundamental partition of the population into poor and non-poor - a special case of partitioning discussed in Section 6.2.
In the case of inequality the anonymity axiom induces a symmetry of treat-
ment of the component subgroups. However, in the case of the fundamental 
poor/non-poor partition one specifically wants to treat the members of the 
two groups differently.

In the standard case one imagines that the population is arranged ac-
ording to one or more observable personal attributes: these attributes could 
include various indicators of resources and of needs. In principle we could 
partition the population according to any of those attributes and then adopt 
an approach similar to the kind of “accounting” framework to the analy-
sis of inequality (Section 6.2). However, more importantly, we could try to 
choose a fundamental partition of the population that – according to the 
specified resources- and needs- indicators – will then categorise every actual 
or hypothetical member of the population as “poor” or “non-poor”, accord-
ing to economic status and other relevant characteristics. Alternatively one 
sets up an economic model of the determinants of poverty and derives the 
implied incomes or expenditures that correspond to specified poverty criteria 
contribution by Ravallion (1996)* discusses many of these issues and the way 
that they interact.

10.2 Identifying the Poor

Once the criterion for dividing up the population is determined then any 
one person's poverty status can be determined – the identification problem 
in terms of Sen (1979)*. There is more than one way of characterising an 
individual's poverty status:

2 A simple “yes/no” as to whether the person's income \( x \) falls short of \( x^\text{m} \).

2 The person's actual income \( x \). This may be alternatively expressed as 
the poverty gap:

\[
g(x, x^\text{m}) := \max \{0, x^\text{m} - x\}
\]  
\( (23) \)

as an intuitive concept of individual poverty status.

2 The individual's income relative to others in the distribution.

10.3 Information aggregation

Each of the three types of information about the person's status can be incor-
porated into the final step of poverty measurement. This involves some type
of aggregation rule covering the poverty status indicators for each member of the population (the aggregation problem in terms of Sen 1979*). A standard approach is to construct an ordering of distributions of poverty gaps, a device which effectively filters out the (irrelevant) information about the non-poor (Jenkins and Lambert 1997). The distribution of poverty gaps \( F^\mu \) is a simple transform of \( F \), censored at the poverty line (Takayama 1979):

\[
F^\mu(g) = \begin{cases} 
1 & \text{if } x < x^\mu \\
0 & \text{otherwise}
\end{cases}
\]

where \( g \) is given by (23); many of the tools that are commonly applied to income distribution may be adapted to the problem of poverty measurement. Distributional dominance as discussed in Section 3 translate into criteria for poverty dominance – see section 11.4. Alternatively one can derive an explicit poverty index: for example, standard families of inequality indices translate into poverty indices (Blackorby and Donaldson 1980a*, Cowell 1988, Foster 1984, Foster et al. 1984*, Ravallion 1994, Sen 1976*) – see section 11. But in doing so difficulties arise at each of the three stages mentioned above.

2 There is room for debate as to what should determine the poverty line, in principle, and how it should be adjusted through time. The value \( x^\mu \) may be unique and exogenously given, some functional of the distribution \( F \), or one of a set of possible values:

\[
x^\mu_{\min} \cdot x^\mu \cdot x^\mu_{\max}
\]

(Atkinson 1987)*. The partition itself may be ambiguous in which case a “fuzzy-set” approach may be appropriate (Cerioli and Zani 1990)*.

2 It may not be clear what the appropriate choice should be from the information about each person that is available in practice: should we use income or expenditure, or perhaps some measure of personal wealth, for the categorisation of each individual?

2 Although there seems to be an obvious “solution” as to how to aggregate information about individuals – just to count the number of persons or families who fall into the poor group – this is not the only “obvious” method. A number of criticisms have been made of the “head-counting” method and other “obvious” approaches have been suggested on theoretical grounds – see section 11.2 below.
11 Poverty Measures

11.1 Additively separable poverty

By analogy with SWFs and inequality measures we can characterise poverty measures as a class of functionals $P : F \to \mathbb{R}$. Section 11 contains some of the key papers on specific poverty indices, on classes of poverty indices and on poverty rankings.

Consider first an important subclass of poverty indices $P$ - the Additively Separable Poverty (ASP) class $P$. Given the poverty line $x^a$ and an individual's income $x$, let an individual's poverty evaluation be

$$p(x, x^a) = \begin{cases} 8 & \text{if } x < x^a \\ 0 & \text{otherwise} \end{cases}$$

(25)

where the function $p$ is positive, non-increasing in $x$, non-decreasing in $x^a$. Then define aggregate poverty as

$$P_{ASP}(F, x^a) = \int_{x^a}^x p(x, x^a) dF(x)$$

(26)

aggregate poverty is simply a sum of individual poverty evaluations. By specifying $p$ in (25) one could read off a variety of poverty indices from (26) although there are important exceptions. Clearly (26) is the counterpart of (2) and $p$ is the counterpart of $u$. The poverty evaluation function is an approach to the issue of how one should count the poor.

11.2 Poverty evaluation

The simplest response to the question of how to count the poor is just to count the number of mouths to be fed or bodies to be housed. In this case $p$ in (25) takes the form

$$p(x, x^a) = 1$$

(27)

Poverty evaluation is done on a simple "yes/no" basis, so that if the person's poverty gap is positive the evaluation is 1, otherwise it is 0. But it is also reasonable to claim that in counting the poor due attention should be given to the depth of poverty of individual poor people, not just to the fact of whether they are poor or not.\(^3\) Then the poor could be counted by evaluating each

\(^3\)See for example the recommendations in Panel on Poverty and Public Assistance (1995).
person’s poverty in proportion to the poverty gap for example:

\[ p(x, x^m) = \frac{x^m \cdot x}{x^m} \quad (28) \]

However, (27) and (28) have the feature that a transfer within the poor set—from a less poor to a poorer person—has a nil impact on poverty. This may not be a desirable property and there may be a case for a more complicated method of counting the poor: one that takes into account the dispersion of incomes amongst the poor. This can be done evaluating the poverty status of each person in a way that is dependent on either the person’s income \( x \) or the poverty gap \( g \), but which is not proportional to the poverty gap.

Three examples from the literature are:

1. the proportionate shortfall of the individual below the poverty line:

\[ p(x, x^m) = \log \left( \frac{x^m}{x} \right) \quad (29) \]

2. a transformed version of (28):

\[ p(x, x^m) = \frac{x^m \cdot x^\alpha}{x^m} \quad (30) \]

3. a transformed version of (29):

\[ p(x, x^m) = \frac{1}{\beta} \left[ \frac{x}{x^m} \right]^\beta \quad (31) \]

where \( \alpha \) and \( \beta \) are sensitivity parameters and play a role akin to \( \alpha \) in (19) in the context of inequality—see, respectively, Watts (1968)*, Foster et al. (1984)* and Clark et al. (1981)*.

If we work out the overall poverty count in the population using (26) and the approach of (27), then this yields a well-known poverty index, the simple (normalised) head count, the number of the poor divided by the population. If instead we use (28) this gives the index known as the (normalised) poverty deicit. Poverty evaluation functions of the forms (29)-(31) yield respectively the Watts index, Foster-Greer-Thorbecke (FGT) class and the

\[ ^4 \text{Clearly (30) is strictly convex in income for } \alpha > 1 \text{ and becomes (27) if } \alpha = 0 \text{ and (28) if } \alpha = 1. \text{ If we write (30) in terms of the poverty gap } g \text{ the family resemblance between Foster et al. (1984)* indices and the inequality indices (9) and (19) is clear. For (31) to be strictly convex one must have } \beta < 1. \]
Clark-Hemming-Ulph second measure. The relationships amongst these and other types of poverty index are discussed by Hagenaars (1987)*.

However, Sen’s original contribution (Sen 1976)* did not conform to this pattern. He introduced into individual poverty evaluation the person’s position in the cumulative distribution $F$ alongside $x$ and the poverty line $x^u$. His suggested formulation has the quasi-additive form:

$$Z(p(x, x^u, F(x))dF(x))$$

(32)

which he specialised to:

$$Z \frac{x^u}{A + B} \left[ x^u - x \right] F(x)dF(x)$$

where $A$ and $B$ are constants of normalisation. In effect Sen suggested that we take the income gaps and weight them by the relative position of each poor person in the society.

11.3 Axiomatic approach

Why adopt an axiomatic approach? Some of the standard inequality indices provide only ad hoc answers to the question “what is poverty?”. The “father” of the axiomatic approach in this area is Sen (1976)* who took the view that the “head-count” approach to poverty measurement was misleading or misguided, and that what one really ought to do – as in the case of inequality – is to specify a list of properties a priori that should characterise poverty comparisons. Each of the axioms introduced in the context of welfare and inequality has its counterpart in poverty analysis. In addition we require:

**Axiom 7** Focus: $P$ is independent of the income distribution of those above the poverty line

By invoking the standard axioms one can generate many of the standard poverty indices. So, for example, the additive structure of (26) will ensure subgroup consistency; the monotonicity property requires that $p$ be strictly decreasing in its first argument for $x < x^u$; application of a scale invariance requirement generates the FGT class.

There are well-known problems with specifying an axiom system. One is that there are typically several similar, but not identical, ways of stating an individual poverty axiom, depending on whether it is defined for incomes strictly less than $x^u$ or for incomes less than or equal to $x^u$. A more serious problem, discovered early in the development of the literature, is that
axioms that are apparently appealing when taken individually may be mutually contradictory when assembled into a system. For example, there is a conflict between the monotonicity principle, the focus axiom and the transfer principle when an individual is close to the poverty line (Kundu and Smith 1983)*.

11.4 Poverty Rankings

As an alternative to picking just one or two individual poverty measures, what about the inferences that we may draw concerning the class of poverty measures that satisfy some of the key poverty axioms? What conclusions could be drawn about general forms of the function $p$ and its associated poverty measures?

Clearly one could apply standard ranking tools to $F^m$, the distribution of poverty gaps that has “coded out” information about the non-poor. However, more can be said about poverty comparisons in cases where the poverty line $x^m$ is not immutably fixed – a point that is important for comparisons over time or between countries. Atkinson (1987)* based results on the portfolio literature concerning “below-target returns” (Fishburn 1977):

Theorem 5 Given a range of poverty lines of the form (24), and poverty measures of the form (26), a necessary and sufficient condition for poverty to be lower in distribution $F$ than in distribution $G$ is that the poverty deficit be no greater in $F$ than in $G$ for all $x \cdot x^m_{\text{max}}$.

This is equivalent to requiring that the second-order dominance condition hold for all $x^m$. Foster and Shorrocks (1988a, 1988b*) have a similar approach to orderings by $P$, but one which concentrates on the FGT index’s particular functional form of the poverty measure (30):

Theorem 6 Poverty orderings are equivalent to (a) rst-order welfare dominance for the case $a = 0$, (b) second-degree welfare dominance for the case $a = 1$, (c) equivalent to third-order welfare dominance for $a = 2$.

12 Welfare, Inequality and Needs

A basic problem with applying the welfare criteria underlying inequality and poverty analysis is that persons, families or households differ dramatically in terms of needs and other attributes as well as differing in income: one
cannot just take income levels as proxy for standards of living when comparing different types of income-receiver. Section 12 deals with theoretical and practical aspects of this problem.

The standard approach is to introduce an equivalence scale which defines a “rate of exchange” between conventionally-de ned income $y$ and an adjusted concept of income $x$ - equivalised income. Imagine that a complete description of a family or household’s circumstances other than money income can be given by some list of attributes $a$ (age of each family member, health indicators, ...) then we suppose that there is some functional relationship $\chi$ such that

$$ x = \chi(a, y) \quad (33) $$

This relationship is usually expressed in the form

$$ x = \frac{y}{\nu(a)} \quad (34) $$

where $\nu(.)$ is a function determining the number of equivalent adults. There is a range of difficulties associated with a speci cation such as (34); for example it is not clear what the appropriate analytical basis for the function $\chi$ in (33) should be, nor even why there should be a proportional relationship between $x$ and $y$ (Coulter et al. 1992a). Deaton and Muellbauer (1986)* show the way in which the costs of children can be appropriately allowed for in an estimate of $\chi$. The effect on inequality and poverty of applying different equivalence scales is discussed in Coulter et al. (1992b)* and Lazear and Michael (1980)*.

An alternative approach would be to see how much can be said about distributional comparisons without precommitment to a particular equivalence scale. Instead of assuming an equivalising function $\chi$ suppose instead that the population can be unambiguously partitioned into $J$ different needs categories ordered in terms of decreasing need. Category $j$ is a set $N_j$ and there is a “categorical” social evaluation function – the income evaluation $u$ in the additive SWF (2) depends upon each income-receiver’s needs category. Then welfare is

$$ W(F) = \sum_{j \in N} u(j, y) dF(a, y) \quad (35) $$

If the marginal social-utility gap between needs levels become smaller at high levels of incomes; so for every category $j$ the expression

$$ \frac{\partial u(j, y)}{\partial y} \approx \frac{\partial u(j + 1, y)}{\partial y} \quad (36) $$
should be positive and decreasing in \( y \).

Let us denote by \( W_3 \) the subclass of \( W_2 \) such that this condition on (36) holds. Then we can introduce the concept of sequential generalised-Lorenz dominance. Let \( F^{(\cdot, j)} \) denote the distribution covering the subpopulation of the first \( j \) most needy groups. Notice the contrast here with subgroup-decomposition analysis: in (16) each observation in the whole population appears in one and only one subgroup distribution; here if the attributes of a particular person \( i \) belong to \( N_j \) then his income will appear in the distribution \( F^{(\cdot, j)} \) and also in \( F^{(\cdot, j+1)}, F^{(\cdot, j+2)}, \ldots \).

Then we have the following extension of the second-order-dominance criterion (see the definition of \( C_F \) in Section 3.2) to the heterogeneous household case (Atkinson and Bourguignon 1987)*:

**Theorem 7**

\[
C^{(\cdot, j)} F^{(\cdot, j)} \geq C \quad \text{if, and only if} \quad W(G) \geq W(F) \quad \text{for} \quad W_2 \quad \text{and} \quad W_3
\]

13 Relative Deprivation

Relative deprivation is a sociological concept whose social-welfare analytic counterpart has developed from the relationship between inequality measures and SWFs; it has structural similarity to the formal work on poverty measurement.

The very features of the Gini coefficient that make it awkward for some branches of the modern literature on inequality make it particularly attractive for embodying the relative deprivation concept of Runciman (1966). Relative deprivation lends itself to a natural expression in terms of income differences: the rôle of rank in defining the Gini coefficient can be reinterpreted in terms of social disadvantage. Suppose the relative deprivation experienced by a person with income \( x_0 \) is measured by

\[
Z \int_{x_0}^{x} [x \cdot \alpha] dF(x)
\]

then the aggregated value of this over the distribution \( F \) is

\[
Z \mu(F) \int_{0}^{2} C(F; q) dq
\]

which is simply \( \mu(F) I_{\text{Gini}}(F) \) - the “absolute Gini”. 

24
The form (38) shows the close relationship between this interpretation of deprivation and GLC rankings (Hey and Lambert 1980) and a number of straightforward generalisations of the concept have been proposed – see Chakravarty and Chakraborty (1984)*, Yitzhaki (1979, 1982*).

14 Progressivity

The ranking of income distributions induced by the Lorenz curve and related concepts outlined in section 3 can be applied to policy-oriented questions such as whether one tax system is more "progressive" than another in terms of its impact on the resulting distribution of after-tax income. Musgrave and Thin (1948)* were among the first to develop and discuss a number of specific concepts of tax progressivity. The relationship to Lorenz-rankings was later shown by Fellman (1976)* and Jakobsson (1976)*. For a general treatment see Lambert (1993).

15 Dynamics

The two papers in this section focus on key issues that have become standard in the applied poverty literature. The path-breaking paper of Lillard and Willis (1978)* showed how to use panel-data on family incomes in a model of dynamics and the ways in which to interpret the error structure of the fitted model in terms of fixed and time-varying unobserved components. From this type of model it is possible to construct a story of the frequency and duration of poverty. Bane and Ellwood (1986)* focus on the analysis of poverty spells as they affect households or families rather than individuals.

16 Functional Forms of Income and Wealth Distribution

The papers in section 16 represent a departure from the intellectual approach of most of this collection. In this introductory discussion it has generally been assumed that inequality or poverty analysis is to be undertaken for an unspecified income distribution: $F$ could have any shape. However, for a variety of reasons, it is sometimes convenient to impose a specific form on $F$. A specific functional form can be particularly useful for estimation of inequality indices or other statistics in cases where information is sparse. Furthermore
some standard functional forms claim attention, not only for their suitability in modelling some features of many empirical income distributions but also because of their role as equilibrium distributions in economic processes (Gibrat 1957)*.

16.1 The choice of functional form

Some of the most important functional forms used in the inequality literature include the following:

² the Pareto model (Pareto 1896*, Chipman 1974*),

\[ F_{\xi,\alpha}(x) = 1 - \frac{\xi^\alpha}{x^\alpha}, \quad (39) \]

where \( \xi > 0 \) is a location parameter and \( \alpha \) is a parameter that is inversely related to dispersion.

² the lognormal model (Aitchison and Brown 1954*, 1957)

\[ F_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^x \frac{1}{y} e^{-\frac{1}{2\sigma^2} [\log(y) - m]^2} dy, \quad (40) \]

where \( m \) and \( \sigma^2 \) are parameters specifying the mean and variance of log-income.

The functional form that is appropriate for modelling distributions depends on the definition of income and the particular part of the distribution in which we happen to be interested. For example, the Pareto model is appropriate for analysing upper incomes; in general lognormal models are appropriate for individual earnings in a homogeneous population (Harrison 1981).

The families of two-parameter models are limited in the variety of shapes of income distributions that they can describe. One way forward is to consider extensions to the basic forms to make them more flexible; for example, multi-parameter generalisations of the Pareto and of the lognormal have been suggested – see Singh and Maddala (1976)* on Pareto-type models.

Other families of distributions have merit in capturing some important features of the distribution; many of these functional forms are interrelated, in the sense that one is a special form of another, or one approximates another asymptotically (Champernowne 1952)*.
16.2 Inequality in parametric models

The use of a functional form induces a structure on $F$ that usually makes the distributional ranking problem very easy. For example, the Lorenz curves of Pareto distributions never intersect; the same is true for lognormal distributions (the Lorenz curve lies further from the line of equality the lower is the parameter $\alpha$ and the higher is the parameter $\sigma^2$ in (39) and (40) respectively; so adoption of either of these families as a paradigm for the admissible class of distributions means that first- and second- order dominance criteria are always clear.

If one does adopt a specific functional form, then inequality can be expressed in terms of its parameters: see Aitchison and Brown (1957) for the lognormal and Chipman (1974)* for the Pareto.

16.3 Estimation

It is first necessary to choose a functional form or model that is appropriate in an economic sense - i.e. a general family form $F_{\mu}$ that captures the general shape of the distribution, or part of the distribution, that one wishes to model, as described in 16.1 above. Then the model parameters can be estimated using an appropriate algorithm: this means an algorithm chosen according to criteria which implicitly define the term “appropriate” in the statistical sense.

For example, “appropriateness” is often interpreted in terms of efficiency of the estimator. For the maximum likelihood estimator for a variety of functional forms see McDonald and Ransom (1979)*.

17 Statistical Issues

The final section of papers deals with the formal statistical issues that arise when inequality and poverty measures are confronted with real data.

Historically data on income distribution were often provided only in grouped form - i.e. where one has information on the number of income-receivers (and possibly the amount of income) that fall into a specified set of income ranges. Although the provision of microdata has improved greatly in recent years it is still often convenient to work with such data, for example from summary tables provided by tax authorities. Cowell and Mehta (1982)* show how sensible estimates of inequality can be obtained from this type of data.
There is an extensive literature on large-sample (asymptotic) methods applied to inequality and poverty statistics. In many cases, for single inequality and poverty indices, one can adapt some of the standard statistical methods for estimating and performing inference on the moments of a distribution. However, where one wants to implement ranking criteria a set of such statistics is involved (for example the set of decile shares represented as ordinates on the Lorenz curve) and the issues of how to construct an appropriate test become more complex (Beach and Davidson 1983)*.

For small samples it may be appropriate to use a resampling scheme—the bootstrap (Mills and Zandvakili 1997)*. These methods may be appropriate even where the overall sample is reasonably large: focusing on the inequality or poverty within one population subgroup may drastically reduce the effective sample size.

Where the data may be contaminated by observations that do not really “belong,” special care is needed. Inequality measures can be remarkably sensitive to these rogue outliers and it may be appropriate to use robust methods of estimation (Cowell and Victoria-Feser 1996)*.

References


