9. Distributional Analysis: a Robust Approach

Frank A. Cowell and Maria-Pia Victoria-Feser

1. Introduction

This paper addresses the issue of how practical comparisons of income distributions can be founded on a sound statistical and economic base when there is good reason to believe that the data in at least one of the distributions are "dirty". Dirtiness can mean not only the possibility of obvious gross errors in the data (such as arise from coding or transcribing mistakes) but also of other more innocuous observations that in some sense do not really belong to the income-data set.

The economic base consists of ranking theorems that are fundamental to the analysis of income distributions: as abstract theoretical constructs they provide a connection between the philosophical basis of welfare judgments and elementary statistical tools for describing distributions. In practical applications they suggest useful ways in which simple computational procedures may be used to draw inferences from collections of empirical income distributions. However, formal welfare propositions can only be satisfactorily invoked for empirical constructs if the sample data can be taken as a reasonable representation of the underlying income distributions which we want to compare.

The statistical base for the approach consists of applications of recent work on robustness. Because the data may be contaminated by recording errors, measurement errors and the like, it is important to be clear about the way in which the possibility of contamination can affect the welfare conclusions which may be drawn from the data. The use of robust analysis can help us with insights on the properties of individual tools of income distribution, such as inequality indices. However they can be of more general use in a practical approach to the problem of comparing income distributions from contaminated data. Consider the two fundamental questions:

1. On what basis are judgments about income distribution to be made?

2. How is data contamination to be incorporated in a formal model of distributional comparisons?

For the first question, it is useful to distinguish three distinct approaches to welfare judgments about income distributions. The most obvious of these is to specify an explicit social-welfare function (SWF), whether it be an unsophisticated criterion such as national income or a more complicated specification from the Bergson-Samuelson class of SWFs. The second approach is more ambitious: we might specify a particular inequality measure, or class of inequality measures, and use the information about inequality and mean income jointly to draw conclusions about welfare. Thirdly we could attempt to make judgements about welfare comparisons that are valid for a class of SWFs: how wide the class is will depend on the set of properties that the member functions are required to fulfil.

The first of these three approaches has the attraction of simplicity but is rather restrictive. The second approach will run into a class of problems that arise in connection with the estimation of inequality measures. Here we pursue the third route: the use of general dominance results to draw inferences about welfare rankings for broad classes of SWFs. We combine consideration of sensible *ad hoc* techniques (section 2) with an investigation of the relationship between economic ranking principles and statistical tools (sections 3 and 4): the discussion will cover both first-order and second-order dominance criteria, and also associated concepts as the relative Lorenz curve (RLC) and the absolute Lorenz curve (ALC). In section 5 we introduce considerations of data contamination and their likely impact on the estimates of statistics associated with distributional dominance. We show that considerable caution may be required in applying some commonly used welfare-dominance criteria. Section 6 discusses a number of ways in which one can make allowance for the distorting effect of data contamination; finally section 7 illustrates one of these – a family of dominance comparisons based on the statistical concept of the trimmed mean – with an application to Lorenz comparisons over time and between countries using the data-base of the Luxembourg Income Study.

2. Informal approaches

Empirical studies of income distribution use informal ranking criteria as a matter of routine. There is a variety of good reasons for doing so: they are usually involve easy computations, and they have a direct intuitive appeal; more importantly, they are usually connected to deeper points that are particularly relevant to applied welfare economics, as we shall see. Some of the more prominent examples of the informal approach include:

- Pragmatic indices involving *quantiles* have been proposed. These include the semi-decile ratio (Wiles 1974), (Wiles and Markowski 1971) and the comparative function of Esberger and Malmquist (1972). An extreme example of the same type is the *range* – literally the maximum minus the minimum income, but sometimes implemented in practice as a difference between extreme quantiles.
- The "*parade of incomes*" introduced by Pen (1971). This provides a persuasive picture of snapshot inequality and of the implications of an income distribution that is changing through time – see for example Jenkins and Cowell (1994).
- The use of *distributive shares* (sometimes known as quantile shares).

The quantile method can be explicitly linked to formal welfare criteria. For example in Rawls' work on a theory of justice there is a discussion of how to implement his famous "difference principle" which focuses upon the least advantaged: to do this Rawls himself suggests that it might be interpreted relative to the median of the distribution.[1] So too can the distributive shares approach:

[1] See Rawls (1972) page 98.

changes in the relative income shares of, say, the richest and the poorest 10% slices of the distribution can be directly interpreted in terms of the principle of transfers enunciated by Pigou (1912) and Dalton (1920).

3. A formal framework

To put these ideas into a form suitable for rigorous economic and statistical analysis we introduce a simplified description of the problem of comparing income distributions.

3.1. Notation and definitions

Assume that the concept of income and of income receiver have been well defined. Let $\Re$ denote the real line and $\Re_+$ denote the strictly positive subset of $\Re$. An individual's income is a number $x \in \mathfrak{X}$, where $\mathfrak{X} \subseteq \Re$. Let $\mathfrak{F}$ be the set of all univariate probability distributions (distribution functions) with support $\mathfrak{X}$. An *income distribution* means just one particular member $F \in \mathfrak{F}$: the distribution function F is a fundamental concept for economic and statistical approaches to distribution analysis.

3.1.1 Statistics. In this approach we may express a *statistic* of any distribution $F \in \mathfrak{F}$ as a functional $T(F)$. A simple example of a standard summary statistics expressed in these terms is the mean; this is just the functional $\mu : \mathfrak{F} \mapsto \Re$ given by

$$\mu(F) := \int x dF(x). \tag{9.1}$$

The properties of any given functional T may play a number of roles in both economic and statistical interpretations. Of particular interest here is the case where the range of T is a profile of values rather than a single number as in the example of (9.1); T is then in effect a *family* of statistics. Individual family members may be of interest in their own right; the behaviour of the whole family when applied to a pair of distributions F and G will provide important information about distributional comparisons that is richer than that provided by a single real-valued functional.

3.1.2 Ranking Criteria. The basic distributional concept employed here is that of a *ranking criterion* which amounts to a partial ordering on the space of distributions $\mathfrak{F}$. Use the symbol $\succeq_T$ to denote the ranking induced on $\mathfrak{F}$ by a statistic T: the expression $F \succeq_T G$ is means that distribution F weakly dominates distribution G according to the statistic T. From this concept a number of other concepts are immediately derived:

Definition 1. *For all* $F, G \in \mathfrak{F}$*:*

(a) *(strict dominance)* $G \succ_T F \Rightarrow G \succeq_T F$ *and* $F \not\succeq_T G$

(b) *(equivalence)* $G \sim_T F \Rightarrow G \succeq_T F$ *and* $F \succeq_T G$

(c) *(non-comparability)* $G \perp_T F \Rightarrow G \not\succeq_T F$ *and* $F \not\succeq_T G$.

For example statement (a) reads in plain language "Distribution G strictly T-dominates distribution F if G weakly T-dominates F and F does not weakly T-dominate G; likewise (c) would read "Distribution G does not T-dominate distribution F, nor does distribution F T-dominate distribution G." We will use the T-ranking concept to motivate a discussion of welfare economic issues in distributional analysis and their statistical implementation.

3.2. Data contamination

To assume that data will automatically give a reasonable picture of the "true" picture of a distributional comparison would obviously be reckless in the extreme. A prudent applied researcher will anticipate that, because of miscoding and misreporting and other types of mistake, some of the observations will be incorrect, and this may have a serious impact upon distributional comparisons (Van Praag et al. 1983). Obviously if one had reason to suspect that this sort of error were extensive in the data sets under consideration the problem of distributional comparison might have to be abandoned because of unreliability. However, it is possible that there might be a fairly serious problem of comparison even if the amount of contamination were fairly small, so that the data might be considered "reasonably clean".

3.2.1 Models of contamination. Consider a standard model of this type of problem.[2] Suppose that the "true" distributions that we wish to compare are denoted by F and G; but because of the problem of data-contamination we cannot assume that the data we have to hand have really been generated by F and G. What we actually observe instead of F is a distribution that is in some neighbourhood of F. To represent this distribution that is "close to" F we need a specific model of data contamination. For example the distribution

$$dH(\mathbf{x}) = \begin{cases} \alpha_1 & \text{if } x = z_1 \\ \ldots & \\ \alpha_m & \text{if } x = z_m \end{cases}, \tag{9.2}$$

where $\forall i : \alpha_i \geq 0$, $\sum \alpha_i = 1$, $z_1, ..., z_m \in \mathfrak{X}$, consists of m discrete "spots". A simplified version of (9.2) is the elementary distribution $H^{(z)}$ that has a unit point mass at z and zero mass elsewhere:

$$H^{(z)}(x) = \iota(x \geq z) \tag{9.3}$$

where ι is the *indicator function* defined by

$$\iota(D) = \begin{cases} 1 & \text{if } D \text{ is true} \\ & \\ 0 & \text{if } D \text{ is false} \end{cases}.$$

Then the distribution that is actually observed will not be the true distribution F but the mixture distribution:

$$F_\varepsilon^{(z)}(x) := [1 - \varepsilon]F(x) + \varepsilon H^{(z)}(x). \tag{9.4}$$

where the parameter ε captures the importance of the contamination relative to the true data: an observation drawn from $F_\varepsilon^{(z)}$ has probability $1 - \varepsilon$ of being generated by F and probability ε of being equal to z. The issue can be illustrated by the elementary case depicted in Figure 9.1 depicting the mixture distribution

[2] This approach is based upon the work of Hampel (1968, 1974), Hampel et al. (1986), Huber (1986).

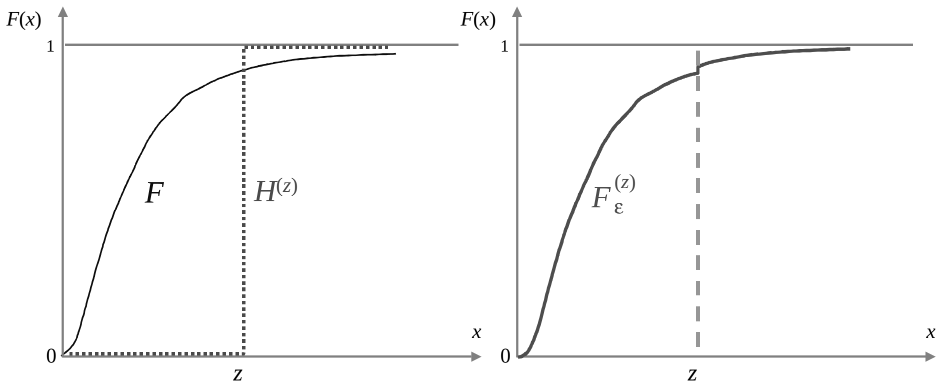

Figure 9.1: Contamination modelled as a mixture of distributions

(9.4).

The central issue with which we are concerned can then be stated as follows. Suppose we wish to rank two distributions F and $G \in \mathfrak{F}$ in terms of their welfare properties. Will the welfare-ranking criteria applied to the associated observables such as $F_\varepsilon^{(z)}$ or $G_\varepsilon^{(z)}$ give very misleading answers? If the amount of contamination were large relative to the true data then we might reasonably conclude that nothing much could be expected from the ranking criteria. However, if the amount of contamination were relatively small, we might reasonably expect that welfare rankings should be robust under contamination, and might be concerned were this not to be the case.

3.2.2 The Influence Function. For any statistic T this idea can be made more precise by introducing the *influence function* IF.[3] This is obtained by taking the derivative with respect to ε of the statistic at $F_\varepsilon^{(z)}$ when $\varepsilon \to 0$ thus:

$$\mathrm{IF}(z; T, F) := \lim_{\varepsilon \to 0} \left[\frac{T(F_\varepsilon^{(z)}) - T(F)}{\varepsilon} \right] = \frac{\partial}{\partial \varepsilon} T(F_\varepsilon^{(z)}) \bigg|_{\varepsilon \to 0}. \tag{9.5}$$

[3]The IF was first introduced by Hampel (1968, 1974) in the framework of *Robust Statistics*. It has been widely used since not only to study robustness properties of statistics but also to build robust estimators and robust test procedures (see e.g. Hampel et al. 1986, Heritier and Ronchetti 1994, Victoria-Feser and Ronchetti 1997, Victoria-Feser 1997).

The IF for the statistic T measures the impact upon the estimate of an infinitesimal amount of contamination at the point z. It is a function of z, the point at which the contamination occurs. Then under the given model of data-contamination (9.4) the statistic T is *robust* if IF in (9.5) is bounded for all $z \in \mathfrak{X}$. If the IF is unbounded for some value of z it means that the T-statistic may be catastrophically affected by data-contamination at income values close to z.

One might ask how the IF can be used to derive results on the robustness properties of stochastic ordering tools. For parameter estimation, an estimator is said to be non robust if its IF is unbounded. This implies that in principle its asymptotic bias can be infinite. However, most stochastic ordering statistics are actually bounded in that they can take values in a bounded interval. Their IF can nevertheless be unbounded. Therefore, saying that the bias on the statistic can be infinite is not appropriate here. On the other hand, if one interprets the IF as the slope of the function $\varepsilon \rightarrow \max \left| T(F_{\varepsilon}^{(z)}) - T(F) \right|$ when $\varepsilon \rightarrow 0$ (see Hampel et al. 1986), then the interpretation becomes clearer. Although the (asymptotic) bias of the stochastic ordering statistic cannot be unbounded, its value can drastically change with an infinitesimal amount of contamination introduced in the data if its IF is unbounded. The obvious implication is that the ordering between two distributions can be different with and without contamination. This point will be illustrated in section 5.2.

4. Welfare judgments

Quantiles and incomplete moments are often used as convenient tools for judgments about income distributions, as we mentioned in section 2. To give economic meaning to a class of distributional rankings it is appropriate to introduce standard welfare criteria expressed in terms of a *social-welfare function* (SWF).

4.1. Social-welfare functions

In economic terms the SWF embodies the ethical judgments of a normative analyst or policy maker; in statistical terms the SWF is just a statistic of the distribution. To get specific results it is useful to focus upon a particular *additively separable* class of SWF:

Definition 2.

$$\mathfrak{W} := \left\{ W : \mathfrak{F} \mapsto \Re | W(F) = \Psi\left(\int u(x)dF(x)\right)\right\}. \tag{9.6}$$

where $u : \mathfrak{X} \mapsto \Re$ *is an* evaluation function *of individual incomes, and* $\Psi : \Re \mapsto \Re$ *is monotonic.*

From $\mathfrak{W}$ we derive two important subclasses.

- First, let $\mathfrak{W}_1$ denote the subclass for which the evaluation function is everywhere increasing: this monotonicity criterion is consistent with the assumption of the *Pareto principle* and the absence of externalities in the SWF (Amiel and Cowell 1994).
- Second, denote by $\mathfrak{W}_2$ the subclass of $\mathfrak{W}_1$ for which the evaluation function is also concave: for additive social welfare functions concavity of the welfare function implies and is implied by the *principle of transfers*, that any mean-preserving richer-to-poorer transfer will (Atkinson 1970).

The SWF subclasses $\mathfrak{W}_1$ and $\mathfrak{W}_2$ will play a crucial role in interpreting two fundamental ranking principles – first- and second-order distributional dominance – and have a close relationship with the informal quantiles and shares criteria introduced in Section 2.[4]

4.2. First-order distributional dominance.

In order to connect the formal concepts just expounded with the intuitive approaches to distributional analysis, we need to introduce formal definitions of the ideas in section 2. Begin with the quantiles of the distribution:

[4]For further discussion of the role played by this type of SWF in distributional comparisons see the survey in Cowell (2000).

Definition 3. *For all* $F \in \mathfrak{F}$ *and for all* $0 \leq q \leq 1$, *the* quantile functional $Q : \mathfrak{F} \times [0,1] \mapsto \mathfrak{X}$ *is defined by:*[5]

$$Q(F; q) = \inf\{x|\ F(x) \geq q\} \tag{9.7}$$

For example $\{Q(F;0.1), Q(F;0.2), ..., Q(F;0.9)\}$ are the deciles of the distribution F.Where there is no ambiguity as to the distribution in question we will write $Q(F;q) = x_q$. For any distribution of income F, the graph of Q – the family of statistics $\{Q(\cdot;q) : q \in [0,1]\}$ in Definition 3 – is in effect a formal description of Pen's Parade mentioned in section 2, Associated with this family of statistics is the following principle and result:

Definition 4. *(Q-ranking) For any* $F, G \in \mathfrak{F}$, $G \succeq_Q F$ *if and only if* $\forall q \in [0,1]$: $Q(G;q) \geq Q(F;q)$

Theorem 1. *(Quirk and Saposnik 1962, Saposnik 1981, Saposnik 1983). For all* $F, G \in \mathfrak{F}$, $G \succeq_Q F$ *if and only if welfare in distribution G is at least as great as that in distribution F for all additively separable SWFs that are monotonic increasing in income.*

Theorem 1 provides an appealing criterion for the welfare-ranking of income distributions: if every quantile in distribution G is greater than the corresponding quantile in distribution F – if some persons "grow" (and nobody shrinks) as in the $F \to G$ transformation depicted in Figure 9.2 – then distribution G will be assigned a higher welfare level by every SWF in class $\mathfrak{W}_1$.

4.3. Second-order distributional dominance.

The first-order dominance criterion $\succeq_Q$ is sometimes considered to be less than ideal. One objection is a on practical grounds: in empirical applications it often happens that neither distribution first-order dominates the other. However,

[5] *See Gastwirth (1971). Alternative definitions are available but redefinition does not affect the results that follow: see Appendix 9.1.*

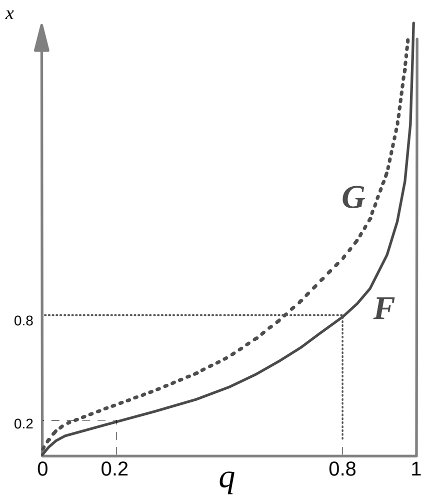

Figure 9.2: G first-order dominates F

it should be noted that Bishop et al. (1991) argue that in international comparisons the second-order criteria discussed below do not resolve many of the "incomparable cases" where $G \perp_Q F$. Second, there is a theoretical point: $\succeq_Q$ does not employ all the standard principles of social welfare analysis: in particular it does not incorporate the principle of transfers. So, whether or not the first-order ranking principle is decisive in practice, it is of interest to consider a criterion – second-order dominance – that takes into account the "transfer principle" mentioned in section 4.1. This introduces the second key distributional concept to be derived from F.

The application of the second-order dominance criterion requires the following:

Definition 5. *For all* $F \in \mathfrak{F}$ *and for all* $0 \leq q \leq 1$, *the* cumulative income functional $C : \mathfrak{F} \times [0, 1] \mapsto \mathfrak{X}$ *is defined by:*

$$C(F; q) := \int_{\underline{x}}^{Q(F;q)} x dF(x). \tag{9.8}$$

where $\underline{x} := \inf \mathfrak{X}$.

The importance of this concept in practical analysis of income distributions is considerable: note, for example, that the mean functional emerges as one

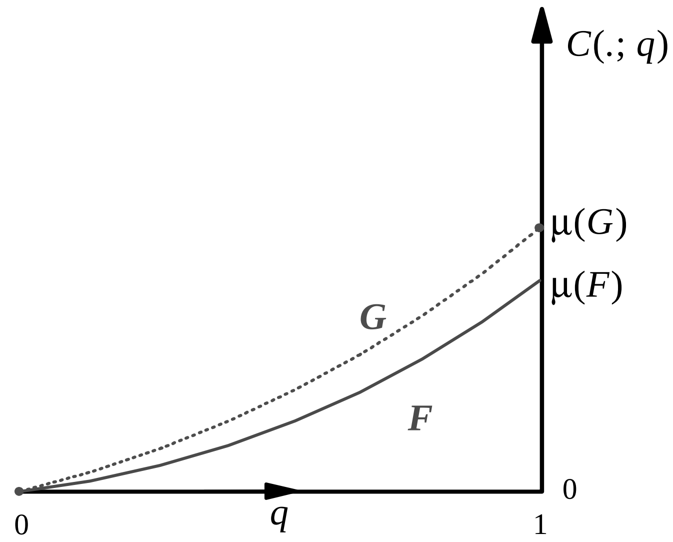

Figure 9.3: G second-order dominates F

particular case ($\mu(\cdot) = C(\cdot, 1)$) and the income share of the bottom q of the population is given by $C(\cdot, q)/C(\cdot, 1)$. For any distribution of income F, the graph of C – the family of statistics $\{C(\cdot; q) : q \in [0, 1]\}$ – characterizes the *generalized Lorenz curve* (GLC). Associated with this family of statistics is the following principle and result:

Definition 6. *For any* $F, G \in \mathfrak{F}$, $G \succeq_C F$ *if and only if* $\forall q \in [0, 1] : C(G; q) \geq C(F; q)$

Theorem 2. *(Kolm 1969, Marshall and Olkin 1979, Shorrocks 1983). For any* $F, G \in \mathfrak{F}$, $G \succeq_C F$ *if and only if welfare in distribution* G *is at least as great as that in distribution* F *for all additively separable SWFs that are monotonic increasing and concave in income.*

Theorem 2 means that if every income-cumulation in distribution G is greater than the corresponding income-cumulation in distribution F then distribution G will be assigned a higher welfare level by every SWF in class $\mathfrak{W}_2$ – see Figure 9.3.

From the fundamental concept of the cumulative income functional (9.8) two other important analytical tools distributional for drawing welfare-conclusions from income data.

Definition 7. *The* Relative Lorenz curve *(RLC) (Lorenz 1905) is the graph* $\{L(\cdot;q) : q \in [0,1]\}$ *with ordinates:*

$$L(F;q) := \frac{C(F;q)}{\mu(F)} \tag{9.9}$$

for any $F \in \mathfrak{F}$.

Definition 8. *The* Absolute Lorenz Curve *(ALC) (Moyes 1987) is the graph* $\{A(\cdot;q) : q \in [0,1]\}$ *with ordinates:*

$$A(F;q) := C(F;q) - q\mu(F) \tag{9.10}$$

The RLC, closely related to the *first moment function* [6], is just a standardized version of the GLC and encapsulates the intuitive principle of the distributional-shares ranking referred to in Section 2 We will examine the implementation of (9.9) and (9.10) in Section 7 below. The RLC and ALC are also second-order dominance criteria based on C that are widely used to provide empirical distributional rankings. The particular interpretations that they permit can be understood by restricting the admissible SWFs to a subset of $\mathfrak{W}_2$.

Take first, the subclass that have the additional property that proportional increases in all incomes yield welfare improvements:

$$\left\{W \,\middle|\, W \in \mathfrak{W}_2; \forall F \in \mathfrak{F}, k > 1 : W\left(F^{(\times k)}\right) > W(F)\right\}. \tag{9.11}$$

where

$$\forall k \in \Re_+,\ F^{(\times k)}(x) = F\left(\frac{x}{k}\right). \tag{9.12}$$

Then distribution G dominates F for SWFs in this restricted class if and only if $G \succeq_L F$ and $\mu(G) \geq \mu(F)$.[7] This is illustrated in Figure 9.4.

[6] This is a function $\Phi : \mathfrak{X} \mapsto [0,1]$ defined for any $F \in \mathfrak{F}$ as $\Phi(x) = L(F;F(x)) = \frac{1}{\mu(x)}\int^x y dF(y)$ - (Kendall and Stuart 1977).

[7] The basic insights of the income-cumulation function were originally obtained for $\mathfrak{F}(\mu)$ the

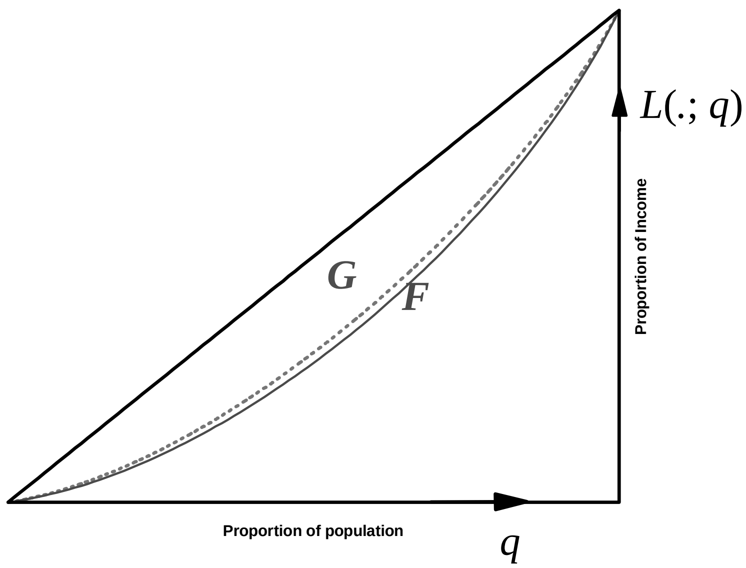

Figure 9.4: G (relative-) Lorenz-dominates F

Alternatively take the subclass for which uniform absolute increases in all incomes yield welfare improvements:

$$\left\{W \,\middle|\, W \in \mathfrak{W}_2; \forall F \in \mathfrak{F}, k > 0 : W\left(F^{(+k)}\right) > W(F)\right\}. \tag{9.13}$$

where

$$\forall k \in \Re,\ F^{(+k)}(x) := F(x-k).$$

Then $G \succeq_A F$ (see Figure 9.5) and $\mu(G) \geq \mu(F)$ if, and only if, $W(G) \geq W(F)$ for all W in $\mathfrak{W}_2$ that also satisfy (9.13).

subset of $\mathfrak{F}$ consisting of distributions with a given mean μ :

$$(\forall F, G \in \mathfrak{F}(\mu) : \ G \succeq_L F) \Leftrightarrow (\forall W \in \mathfrak{W}_2 : W(G) \geq W(F))$$

- see Atkinson (1970).

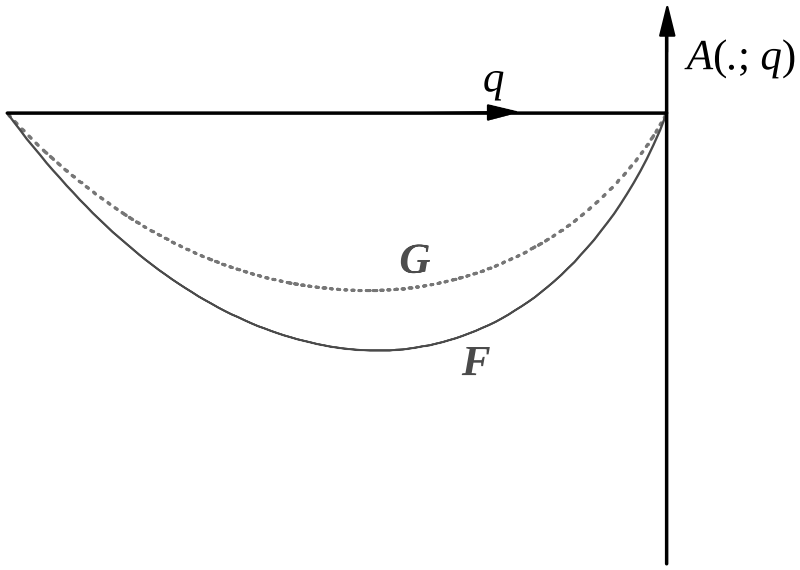

Figure 9.5: G absolute Lorenz-dominates F

4.4. Higher-order dominance

As with the first-order criterion, in practice one often finds that second-order criteria are indecisive. Where Lorenz curves intersect one possible way forward is to supplement the restrictions on the class of SWFs (9.6) by imposing a further restriction on the income-evaluation function. An additional restriction that could reasonably be imposed upon the $\mathfrak{W}$-classes is the "principle of diminishing transfers" (Kolm 1976a): that a small transfer from an individual with income x to one with income $x - k$ should have a greater impact on welfare or inequality the lower x is located in the distribution.[8] Denote by $\mathfrak{W}_3$ the subset of $\mathfrak{W}_2$ for which this principle holds. Then the following can be established:

Theorem 3. *(Atkinson 1973, Davies and Hoy 1994, Dardanoni and Lambert 1988, Muliere and Scarsini 1989) For all SWFs in* $\mathfrak{W}_3$ $W(F) > W(G)$ *if* $F, G \in \mathfrak{F}$ *satisfy: (i)* $\mu(F) = \mu(G)$, *(ii)* $\text{var}(F) \leq \text{var}(G)$, *and (iii)* $\exists q^* \in (0,1)$ *such that* $\forall q < q^* : L(F;q) > L(G;q)$ *and* $\forall q > q^* : L(F;q) < L(G;q)$. [9]

[8] The principle is implied by the principle of "transfer sensitivity" (Shorrocks and Foster 1987) or "aversion to downside inequality" (Davies and Hoy 1995).

[9] *Davies and Hoy (1995) extend the analysis to cases of multiple Lorenz intersections.*

This result is closely linked to a concept of "third-order" dominance (Shorrocks and Foster 1987); an extension of the idea of dominance to an arbitrary order is discussed in Fishburn and Willig (1984) and Kolm (1974, 1976b).

5. Dominance results and contamination

It might be thought that standard results on the structure of distributional comparisons in economics permit one to draw conclusions about the role of contamination in a mixture distribution. For example the concept of decomposability – related to, but weaker than, the additive separability used in definition 2 – is often invoked in standard approaches within the field of distributional analysis, including the measurement of inequality or social welfare and the measurement of risk. We may state a form of it as:

Definition 9. *A statistic T is* generally decomposable *if $\forall F, G, K \in \mathfrak{F}$ such that $\mu(F) = \mu(G)$, and $\forall \lambda \in [0, 1]$:*

$$(G \succeq_T F) \Leftrightarrow ([1-\lambda]\, G + \lambda K \succeq_T [1-\lambda]\, F + \lambda K)$$

This seems promising as a tool for disentangling the impact of contamination in comparing income distributions. An apparently neat conclusion can be drawn from Definition 9 and (9.4) if two "true" distributions F and G are subjected simultaneously to exactly the same contamination: for any generally decomposable statistic T if $G_\varepsilon^{(z)} \succeq_T F_\varepsilon^{(z)}$ we may safely conclude that $G \succeq_T F$. However, it runs into two serious problems. The first is that many of the ranking statistics in which one is interested are not generally decomposable; this is indeed the case with first- and second-order ranking criteria $\succeq_Q$ and $\succeq_C$ introduced in definitions 4 and 6 below.[10] The second difficulty with this superficially attractive result is that this story of contamination is a very special case. It assumes that, although contamination is not observable, it may nevertheless be taken to be

[10] More restricted notions of decomposability are available: in particular if the contamination distribution is "non-overlapping" with F and G (Cowell 1988) (Ebert 1988) then the ranking criteria will be decomposable (an example of this type of non-overlappingness is the case on page 227). However this modification of the definition is of no help with the second objection.

exactly the same for two empirical distributions; the conditions under which one might reasonably accept this seem rather contrived.

A general treatment of the impact of contamination on welfare judgments requires the detailed examination of the properties of the IF for the particular statistic T associated with a given ranking principle. This requires a two-stage approach: first we look at the impact of contamination upon individual statistics used in distributional comparisons (section 5.1); then we consider the implications of this for the behaviour of ranking tools that use *families* of these statistics (section 5.2).

5.1. Robustness properties of distributional statistics

5.1.1 First-order statistics. The functional $Q(\cdot, q)$ is an useful tool for distributional analysis in its own right – consider for example the widespread use of the median or the interquartile range as informative descriptive statistics – and it is interesting to see the effect of contamination on a typical quantile. This can be done by considering $F_{\varepsilon}^{(z)}$ instead of F and applying equation (9.5) to find the influence function $\mathrm{IF}(z; Q(\cdot, q), F)$. We may write

$$Q(F_{\varepsilon}^{(z)}; q) = Q\left(F; \frac{q - \iota(x_q \geq z)\varepsilon}{1 - \varepsilon}\right). \tag{9.14}$$

Therefore,

$$\mathrm{IF}(z; Q(\cdot, q), F) = \frac{q - \iota(x_q \geq z)}{f(Q(F; q))}. \tag{9.15}$$

where the density $f(\cdot)$ is defined. Note that for any $F \in \mathfrak{F}$ and for all z : $\iota(x_q \geq z) = \iota(q \geq F(z))$. To interpret (9.15) it is convenient to introduce the concept of the hazard rate $h(x) := \frac{f(x)}{1-F(x)}$.

Theorem 4. *$\forall z \in \mathfrak{X}$ and $\forall F \in \mathfrak{F}$:*

(a) $\forall q \neq 0, 1$: $\mathrm{IF}(z; Q(\cdot; q), F)$ is bounded if and only if $f(x_q) > 0$

(b) If the hazard rate is non-decreasing for large x then $\mathrm{IF}(z; Q(\cdot; q), F)$ is bounded as $q \to 1$.

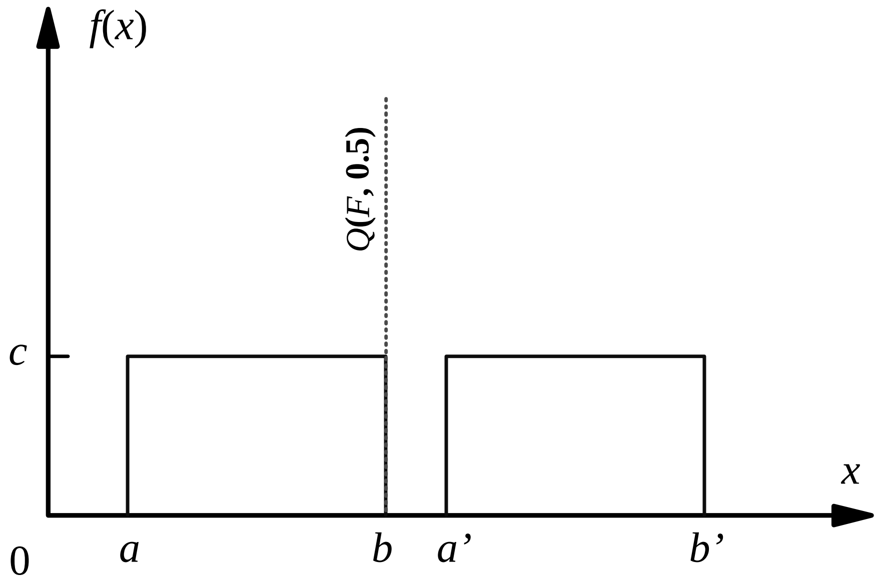

Figure 9.6: $Q(F, 0.5)$ in a case with a "dead" interval

(c) If $\lim_{x \to \inf \mathfrak{X}} f(x) > 0$ *or if* $f(x)$ *has positive slope as* $x \to \inf \mathfrak{X}$*, then* $\text{IF}(z; Q(\cdot; q), F)$ *is bounded as* $q \to 0$.

Proof: Part (a) is immediate from (9.15). (b) $\forall z < \sup \mathfrak{X}$, $\exists (\delta > 0) \mid (q = 1 - \delta)$ then $\iota(x_q \geq z) = 1$ and IF becomes $\lim_{x \to \sup \mathfrak{X}} \left(-\frac{1}{h(x)} \right)$ and so the second part follows.[11] (c) Likewise, $\forall z > \inf \mathfrak{X}$, $\exists (\delta > 0) \mid (q = \delta)$ then $\iota(x_q \geq z) = 0$ and IF becomes $\lim_{x \to \inf \mathfrak{X}} \left(\frac{F(x)}{f(x)} \right)$; the last part of the theorem then follows from l'Hôpital's rule. ■

An example of the problem that can arise with the condition in part (a) of Theorem 4 is as follows. Figure 9.6 illustrates a case where $Q(F, 0.5)$ is non-robust. The population consists of two distinct equal-sized groups each of which has an underlying rectangular distribution so that $f(x) = c$ over $[a, b]$ and $[a', b']$ and 0 elsewhere.[12] We see that $Q(F, 0.5)$ is at b, and that a small amount of

[11] An example of a distribution which violates this condition and for which the influence function for $Q(\cdot, q)$ is unbounded as $q \to 1$ is the lognormal (Cowell and Victoria-Feser 1996b) .

[12] Note that these assumptions imply $b - a = b' - a'$.

contamination in the region $[a', \infty)$ would cause $Q(F, 0.5)$ in the mixture distribution to jump to a'. So Theorem 4 suggests that if first-order statistics are used to compare distributions then, as long as F is strictly increasing and the hazard rate has an appropriate property, then we can be reassured that the welfare comparison is robust in that a small amount of extreme values in the samples used to make the comparisons will not have any substantial effect on any of the $Q(\cdot, q)$ statistics.

However, we can say more than this if the first-order statistics are used jointly to make a welfare comparison – see Section 5.2.

5.1.2 Second-order statistics. Now consider the use of income-cumulations to give us information about parts of the income distribution. Again we consider the impact of data contamination as modelled in (9.4).

The C functional can be written as

$$C(F_\varepsilon^{(z)}; q) = \int_{\underline{x}}^{Q(F_\varepsilon^{(z)};q)} x dF_\varepsilon^{(z)}(x) \tag{9.16}$$

and the IF can be obtained by applying (9.5) to give:

$$\begin{aligned} \mathrm{IF}(z; C(\cdot, q), F) &= -\int_{\underline{x}}^{Q(F;q)} x dF(x) + Q(F;q) f(Q(F;q)) \mathrm{IF}(z; Q, F) \\ &\quad + \int_{\underline{x}}^{Q(F;q)} x dH^{(z)}(x) \\ &= qQ(F;q) - C(F;q) + \iota(q \geq F(z))[z - Q(F;q)]. \quad (9.17) \end{aligned}$$

The IF (9.17) is illustrated in Figure 9.7: we can see here that the influence of an infinitesimal amount of contamination on the GLC's q-group's income share can be large for high values of q. The IF can be unbounded at $q = 1$, if the income range extends to $+\infty$. On the other hand, if we suppose that the income range extends to $-\infty$, we can see from Figure 9.7 that the IF can be unbounded for any value of q. We may summarize thus:

Theorem 5. $\forall z \in \mathfrak{X}$ *and* $\forall F \in \mathfrak{F}$:

(a) $\forall q < 1$*:* $\mathrm{IF}(z; C(\cdot; q), F)$ *is bounded if* $\mathfrak{X}$ *is bounded below.*

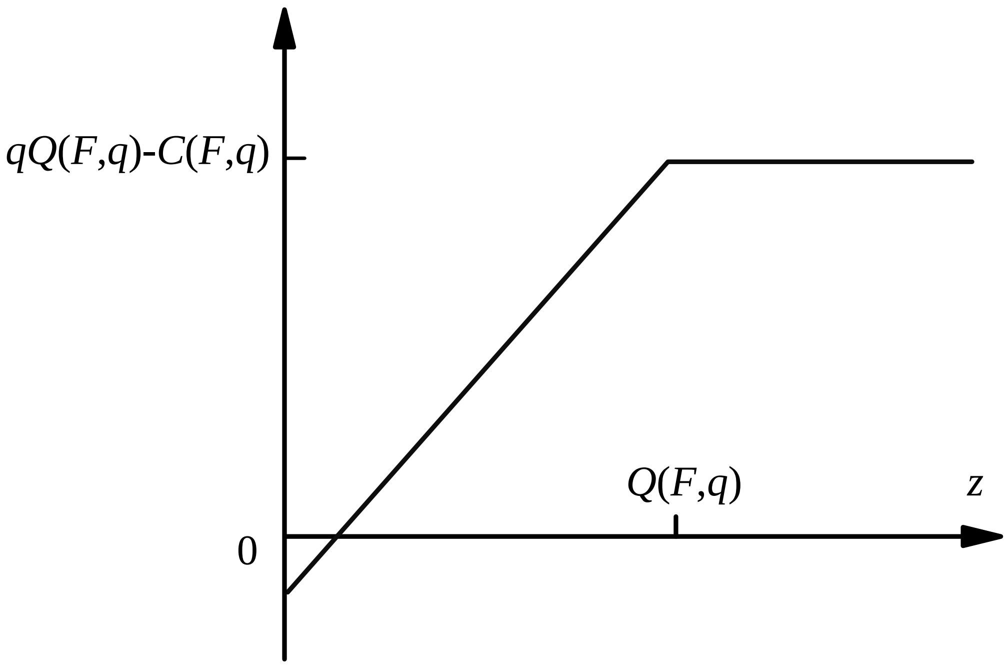

Figure 9.7: The influence function for the income cumulation $C(\cdot, q)$.

(b) $q = 1$, $\mathrm{IF}(z; C(\cdot; q), F)$ *is bounded if* $\mathfrak{X}$ *is bounded above and below.*

As we will see in section 5.2 this result has an important consequence for distributional ranking results.

5.2. Results on ranking

As we explained in the introduction, two of the principal reasons for focusing attention on analytical tools such as Pen's Parade and the Lorenz curve are the avoidance of ethical arbitrariness involved in a precommitment to specific inequality indices or SWFs, and the avoidance of the unsatisfactory properties associated with those statistics. However, this second reason may not be soundly based: as we have seen, there are conditions under which the key distributional statistics Q and C are non-robust, and this may have serious implications:

1. The ranking principle associated with a particular statistic may yield mistaken judgments if made under the influence of data contamination. As we

can see from section 3.1.2, given two income distributions and a ranking principle $\succeq$ there are obviously four possible outcomes: $F \succ G$, $G \succ F$, $F \sim G$ and $F \perp G$. This implies that there are, in principle, twelve types of errors that could be made in drawing welfare inferences from a pair of empirical distributions. However the chief problem arising from data contamination is that of mistaking $F \perp G$ for one of the other outcomes, or *vice versa.*

2. The welfare inferences based on these ranking principles could then be open to question.

To investigate these issues we will examine, in turn, the two major ranking principles associated with the statistics introduced in section 5.1.

5.2.1 . Could infinitesimal contamination change a first-order dominance result? As we have seen the IF of $Q(F;q)$ can be unbounded for cases where the underlying density $f(Q(F;q))$ vanishes or where the tails of the underlying distribution do not have appropriate limiting properties. It is also clear from (9.15) that, whether or not the IF is unbounded is independent of the point of contamination. However, these facts do *not* automatically imply that the Q-rankings of distributions are misleading, in the sense just explained. In fact, we may show:

Theorem 6. *The first-order dominance relation $\succeq_Q$ is robust.*

Proof: The result is established if it can be shown that, given $G \succ_Q F$, it is impossible that $F_\varepsilon^{(z)} \perp_Q G$ for infinitesimal ε. For this to be true there must be some q for which

$$Q(F_\varepsilon^{(z)};q) > Q(G;q) \geq Q(F;q). \tag{9.18}$$

Now everywhere $f(Q(F;q)) > 0$ (9.18) implies that, for some q,

$$Q\left(F;\frac{q-\iota(x_q \geq z)\varepsilon}{1-\varepsilon}\right) > Q(G;q) \tag{9.19}$$

But as $\varepsilon \to 0$, $\forall q$, $Q(F;q) > Q(G;q)$ which contradicts the hypothesis $Q(G;q) \geq Q(F;q)$. Next consider the exceptional cases (a)-(c) in Theorem 4:

(a) Let $I := [x^*, x^{**}] \subset \mathfrak{X}$ be a "dead" interval for F such that $f(x) = 0, \forall x \in I$, and let $q^* := F(x^*)$. For arbitrarily small ε we have $Q(F_\varepsilon^{(z)}; q) \to x^*$ as $q \uparrow q^*$ and $Q(F_\varepsilon^{(z)}; q) \to x^{**}$ as $q \downarrow q^*$. Given that Q is monotonic non-decreasing in q it is impossible for (9.18) to hold at q^*.

(b) If $G \succ_Q F$, then in particular

$$\lim_{x \to \sup \mathfrak{X}} G(x) \geq \lim_{x \to \sup \mathfrak{X}} F(x). \tag{9.20}$$

If $F_\varepsilon^{(z)} \perp_Q G$, it would have to be true that

$$\lim_{x \to \sup \mathfrak{X}} \left(F(x) + \frac{\varepsilon H^{(z)}(x)}{1-\varepsilon} \right) > \lim_{x \to \sup \mathfrak{X}} \frac{G(x)}{1-\varepsilon} \tag{9.21}$$

However, allowing ε to tend to zero in (9.21) immediately produces a contradiction with (9.20).

(c) A similar argument to part (b) applies for the case $x \to \inf \mathfrak{X}$. ■

Two clarifying remarks may be in order. First, the point in (a) is illustrated in Figure 9.8 using the example of Figure 9.6: here $x^* = b, x^{**} = a'$, and it is clear that even though there is an unbounded IF because of the dead interval, this does not affect the conclusion that $G \succ_Q F$. Second, note that the result is asymptotic: with a finite sample of size n, it is easy to see that contamination could, in principle, cause the first-order criterion to yield an inappropriate answer – if $G^{(n)} \succ_Q F^{(n)}$, then by extending the largest observation in $F^{(n)}$, one might have $G^{(n)} \perp_Q F^{(n)}$ because of the contaminated data.

5.2.2 The second-order dominance criterion. By analogy with the first-order case considered in 5.2.1, to understand the potential impact of contamination on second-order dominance we need to look at the consequences of the robustness properties of this fundamental statistic (9.8).

Theorem 7. *The second-order dominance relation $\succeq_C$ is non-robust.*

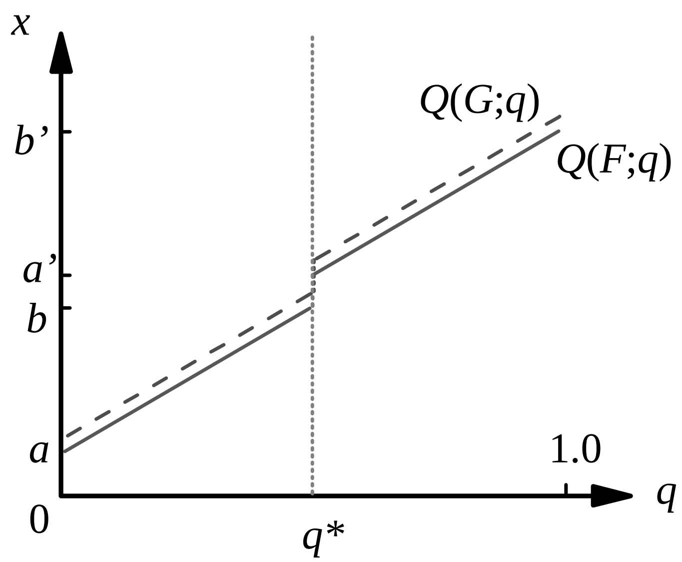

Figure 9.8: Q-rankings with a "dead" interval.

Proof: The result is established if it is the case that, for some $F, G \in \mathfrak{F}$ such that $G \succ_C F$, it is possible that $\lim_{\varepsilon \to 0} F_\varepsilon^{(z)} \perp_C G$. First, recall that $\mu(\cdot) = C(\cdot, 1)$ and note that

$$\mu\left(F_\varepsilon^{(z)}\right) = [1 - \varepsilon]\,\mu(F) + \varepsilon z \tag{9.22}$$

If $\mathfrak{X}$ is unbounded above then we may have $\lim_{\varepsilon \to 0} \lim_{z \to \infty} \varepsilon z =: k > 0$. So, if k is sufficiently large, we have

$$\lim_{\varepsilon \to 0} \lim_{z \to \infty} \mu\left(F_\varepsilon^{(z)}\right) > \mu(G) > \mu(F) \tag{9.23}$$

where the second inequality in (9.23) follows from $G \succ_C F$. But (9.23) implies $\lim_{\varepsilon \to 0} \lim_{z \to \infty} F_\varepsilon^{(z)} \perp_C G$. ■

The case used in the proof is shown in Figure 9.9 which depicts two distributions F, G such that $G \succ_C F$ and a mixture distribution $F_\varepsilon^{(z)}$. By definition of the ranking principle $\succ_C$ we have $\mu(G) > \mu(F)$, and $F_\varepsilon^{(z)}$ has been constructed as a mixture between F and a point mass distribution at z such that $\mu(F_\varepsilon^{(z)}) > \mu(G)$.

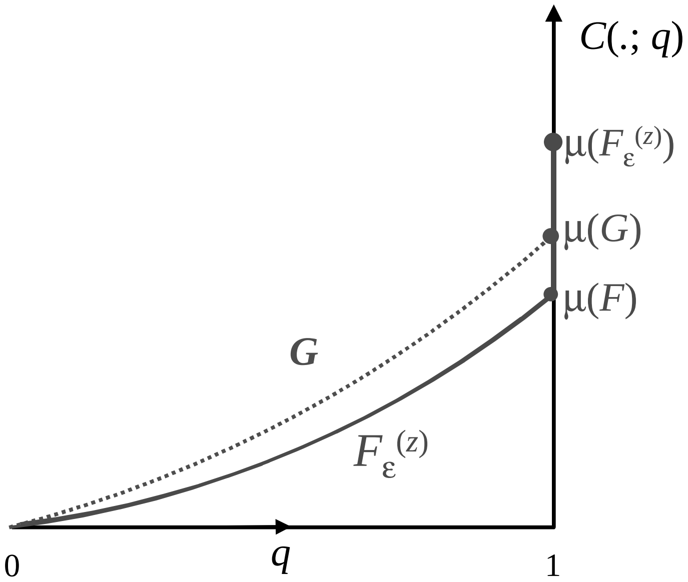

Figure 9.9: A small amount of contamination changes second-order dominance conclusions

5.2.3 Other second-order criteria. The behaviour discussed in 5.1.2 and 5.2.2 is inherited by other related distributional tools.

Theorem 8. *All RLC and ALC ordinates are non-robust.*

Proof: If we assume the contaminated distribution $F_\varepsilon^{(z)}$, we have

$$L(F_\varepsilon^{(z)}; q) = \frac{C(F_\varepsilon^{(z)}; q)}{\mu(F_\varepsilon^{(z)})} \tag{9.24}$$

$$A(F_\varepsilon^{(z)}; q) = C(F_\varepsilon^{(z)}; q) - \mu(F_\varepsilon^{(z)}) \cdot q. \tag{9.25}$$

and the IFs are given by

$$
\begin{aligned}
\text{IF}(z; L(\cdot; q), F) &= \frac{1}{\mu(F)^2}\left[\text{IF}(z; C(\cdot; q), F)\mu(F) - \text{IF}(z; \mu, F)C(F; q)\right] \\
&= \frac{Q(F;q)(q - \iota(q \geq F(z))}{\mu(F)} + \\
&\quad z\frac{\mu(F)\iota(q \geq F(z)) - C(F;q))}{\mu(F)^2}
\end{aligned} \tag{9.26}
$$

$$
\begin{aligned}
\text{IF}(z; A(\cdot; q), F) &= \text{IF}(z; C(\cdot; q), F) - \text{IF}(z; \mu, F)q \\
&= qQ(F; q) - C(F; q) + \iota(q \geq F(z))[z - Q(F; q)] + \\
&\quad (\mu(F) - z)q.
\end{aligned} \tag{9.27}
$$

where the second line in (9.26) follows from the fact that $\text{IF}(z; \mu, F) = z - \mu(F)$. Given that the last line in each of the expressions (9.26) and (9.27) is linear in z then, if z is unbounded, so too is IF. ■

So the IF for any RLC or ALC ordinate is unbounded. But here the result is stronger than in Theorem 5 – it applies for *all* values of q – and the reason is that we have to estimate the mean by the sample mean which is clearly not a robust estimator.

5.2.4 Higher-order dominance criteria. As we noted in section 4.4, the basic principles of monotonicity (first order) and transfers (second order) should be supplemented by others so as to generate third and higher order concepts of dominance. However, given that these criteria involve comparisons of an integral of the $C(\cdot, q)$ it is clear that the problems that occur with second-order dominance will necessarily occur with higher-order versions of distributional dominance.

5.3. Implications

Second-order ranking principles can have unbounded IFs, and the conditions under which they are unbounded correspond to phenomena that can reasonably be expected to arise in practical applications. A very small number of large outliers can give rise to serious problems for welfare analysis when using Lorenz-type tools.

We have already seen a typical example of this problem in connection with the GLC in Figure 9.9 where G dominates F for all incomes except for the highest one. Is it then reasonable to conclude that $G \perp F$ (which is what is actually observed) when it is clear that, had the highest income not been there, we would have concluded that $G \succ F$?

As a second example of misleading welfare inferences take the performance of the RLC in a simple "lottery winner" example. Suppose we have two populations of size n with discrete distributions F and G characterised by the income vectors $\mathbf{x}^F := (x_{[1]}, x_{[2]}, \ldots, x_{[n]})$ and $\mathbf{x}^G := (x_{[1]} - \gamma, x_{[2]}, \ldots, x_{[n]} + \gamma)$, $\gamma > 0$ (where $[i]$ is the ith order statistic). By construction we have $F \succ_C G$: in fact F dominates G in terms of RLC, GLC and ALC. Now suppose that a lottery is introduced in the first population, that everybody spends the same amount on the lottery, and that the winner is the richest, i.e. the person with income $x_{[n]}$. Then the corresponding income distribution becomes $\hat{F}$ where $\mathbf{x}^{\hat{F}} := (x_{[1]} - \delta, \ldots, x_{[n-1]} - \delta, x_{[n]} + (n-1)\delta)$. By imposing suitable mild conditions on γ, δ and n, it is easy to show that after the lottery has been introduced, $G \succ_C \hat{F}$.[13] Again we may ask whether the conclusion $F \succ_C G$ or $G \succ_C \hat{F}$ is appropriate.

Although the modifications to the income distribution in the above two cases are different, the issues raised for welfare ranking are similar – the Lorenz curves for the lottery example will be of the same form as the (relative) Lorenz curves corresponding to the example in Figure 9.9 – and so it makes sense to consider them together. There are three ways of looking at the GLC phenomenon in Figure 9.9:

1. The point mass z really belongs to the distribution and should be encoded

[13] As an example consider the income vectors
$\mathbf{x}^F = (50, 100, 100, 100, ..., 100, 150)$
$\mathbf{x}^G = (49, 100, 100, 100, ..., 100, 151)$
$\mathbf{x}^{\hat{F}} = (49, 99, 99, 99, ..., 99, 150 + [n-1])$. A similar problem arises in an example provided by Arnold (1987) of misreporting. Suppose x is true income and that individuals underreport by a fraction $1 - u$ so that reported income is $y := ux$; if x and u are independently distributed and $x \sim F(x)$, $y \sim G(y)$ then $F \succ G$ (Arnold 1987, page 51).

in the distributional ranking criterion along with all the other information.

2. The point mass z really belongs to the distribution but should be discarded as unimportant, so that the conclusion $G \succ F$ stands. The intersection of the GLCs at one end tells a different story from that which appears from the mass of the data. In this case the information in the upper tail can be interpreted as "hiding" the story from the rest of the data.

3. Point mass z is external contamination and should be discarded as irrelevant. The corresponding point in the lottery example would be the argument that in comparing $\hat{F}$ and G one would be using the "wrong" income concept – an *ex post* rather than an *ex ante* distribution.

The distinction between (1) and (2) is essentially a matter of economic judgment: what issue is it "appropriate" to address? Here appropriateness is to be judged by ethical criteria or by reference to pragmatic considerations of relevance: for example it is possible to imagine cases where it is appropriate to combine in the same distribution dramatically different subgroups – say a small rich group (Luxembourg?) and a large group with modest incomes (China?) – and cases where this composite distribution is inappropriate. There remains the issue of what can, or should, be done about (2) and (3). In case (2) it may be that the second-order welfare criterion that is being applied is inappropriately demanding, and should be replaced. In case (3) one wants to use information about the rest of the distribution to "work round" the problem caused by the contamination: this may involve a scientific rule for ignoring extreme values or a method for modelling the shape of the distribution. Practical methods of handling (2) and (3) will be tackled in section 6.

6. Robust approaches

Because second-order dominance criteria can be misleading in the presence of data contamination it is desirable to have a procedure that enables one to control systematically for suspect values that my distort the distributional comparisons. There are two main ways of approaching the problem. One is based on statistics that automatically remove from the sample data that are potentially troublesome.

The other relies on the specification of parametric models for the distribution of the data and uses robust estimators of the parameters. Each of these has a number of attractions. Consider first parametric approach: we do this in two steps – a fully-fledged application of parametric methods, and a kind of "half-way house" between full parametrisation and the non-parametric approach.

6.1. A full parametric approach

A parametric approach to robustness requires the specification of a functional form for modelling the data. One then estimates robustly the parameters of the model and uses the estimated distributions to compute the (estimated) Lorenz curves. To be more precise, suppose we choose F_θ as model for the data and estimate θ robustly by say $\hat{\theta}$, then robust estimates of the GLC, RLC and ALC are given by respectively

$$C(\hat{\theta}; q) = \int_{\underline{x}}^{Q(F_{\hat{\theta}}; q)} x dF_{\hat{\theta}}(x), \tag{9.28}$$

$$L(\hat{\theta}; q) = \frac{C(F_{\hat{\theta}}; q)}{\mu(F_{\hat{\theta}})}, \tag{9.29}$$

$$A(\hat{\theta}; q) = C(F_{\hat{\theta}}; q) - \mu(F_{\hat{\theta}}) \cdot q, \tag{9.30}$$

where $\mu(F_{\hat{\theta}}) = \int x dF_{\hat{\theta}}(x)$. The IF of the estimators of the Lorenz curves will then depend on the IF of the parameter's estimator. Indeed, the Lorenz curves depend on the data only through the estimator $\hat{\theta}$. If we write the latter as a functional of the contaminated distribution given in (9.4), i.e. $\hat{\theta}(F_\varepsilon^{(z)})$, then we have

$$\text{IF}(z; C, F_\theta) = \frac{\partial}{\partial \theta} C(F_\theta; q) \cdot \text{IF}(z; \hat{\theta}, F_\theta). \tag{9.31}$$

Then if the estimator is robust, or in other words if its IF is bounded, the Lorenz curve estimated through a parametric model is also robust.

Consider the standard approach to the estimation of a full parametric model of income distribution. This usually concentrates on an efficiency criterion: given a model F_θ with density function f_θ, the *maximum likelihood estimators* (MLE) are then obtained as the solution in θ of the m equations

$$\sum_{i=1}^{n} S(x_i;\theta) = 0 \tag{9.32}$$

where $m := \dim(\theta)$ and S is the scores function defined by

$$S(x;\theta) = \frac{\partial}{\partial\theta}\log f_\theta(x) \tag{9.33}$$

But, of course, the efficiency criterion alone takes no account of the of contamination problem: the MLE procedures would be optimal given the assumption that the data are generated by F_θ, but will be invalid for any variation around F_θ – as in equation (9.4) with $\varepsilon > 0$ (Hampel et al. 1986, Victoria-Feser 1993). To handle this requires an additional criterion that takes into account the robustness considerations outlined in section 5. In the robust approach to estimation, instead of applying (9.32 and 9.33) one requires an algorithm to filter outlying observations systematically. The MLE belong to a general class of so-called *M-estimators* which are defined as the solution in θ of

$$\sum_{i=1}^{n} \psi(x_i;\theta) = 0 \tag{9.34}$$

where ψ belongs to a very general class of functions. (Huber 1964) The robust approach consists of a search for the minimum (asymptotic) variance M-estimator with a bounded IF: efficiency is sacrificed to some extent in favour of robustness. There is a number of optimal estimators, depending on the exact method of bounding the IF. Take, for example the standardised *Optimal Bias-Robust Estimators (OBRE)* which also belong to (9.34); given a constant $c \in [\sqrt{m}, \infty)$ which plays the role of upper bound on the IF, the OBRE is defined as the solution in θ to

$$\textstyle\sum_{i=1}^{n} \psi_c^{\mathbf{A},\mathbf{a}}(x_i;\theta) = 0 \tag{9.35}$$

where

$$\psi_c^{\mathbf{A},\mathbf{a}}(x;\theta) = A(\theta)\left[\mathbf{S}(x;\boldsymbol{\theta}) - \mathbf{a}(\boldsymbol{\theta})\right] w_c(x;\theta)$$
$$w_c(x;\theta) = \min\left\{1; \frac{c}{\|A(\theta)\left[\mathbf{S}(x;\boldsymbol{\theta}) - \mathbf{a}(\boldsymbol{\theta})\right]\|}\right\}$$

$\mathbf{A}$ is an $m \times m$-matrix, and $\mathbf{a}$ is an m-vector; $\mathbf{A}$ and $\mathbf{a}$ are determined by:

$$E\left[\psi_c^{\mathbf{A},\mathbf{a}}(x;\boldsymbol{\theta})\psi_c^{\mathbf{A},\mathbf{a}}(x;\theta)^T\right] = \mathbf{I} \tag{9.36}$$
$$E\left[\psi_c^{\mathbf{A},\mathbf{a}}(x;\theta)\right] = 0 \tag{9.37}$$

$\mathbf{A}$ and $\mathbf{a}$ can be considered as Lagrange multipliers for restrictions (9.36) and (9.37); ψ is a modified and standardised scores function, weighted using w_c. The constant c may be selected as a "regulator" between the two statistical criteria, efficiency and robustness. Lower values of c yield more robust, but less efficient, estimators: the maximum-robustness estimator corresponds to the lower bound of the constant $c = \sqrt{m}$; on the other hand $c = \infty$ yields the MLE (Prieto-Alaiz and Victoria-Feser 1996, Victoria-Feser 1995).[14]

However, in the present context, a full parametric approach is inappropriate, even if it is undertaken using robust methods. This is because it forces the data into the "mould" of a functional form that may not be suitable for welfare comparisons. For example, if one supposes that the income data are Lognormal distributed, then a "parametric Lorenz" comparison of two distributions based on the Lognormal will always yield a strict dominance order! The parametric approach is therefore only appropriate provided that the postulated model is capable of yielding Lorenz curves that can cross: this may require specification of a complicated functional form that is difficult to estimate and to interpret.

6.2. A semi-parametric approach

In light of the above considerations, we suggest using a semi-parametric approach. As we have seen in Section 5, if the income range is bounded below (0 is a

[14]For more on optimal bounded-influence estimators see Hampel et al. (1986), Victoria-Feser and Ronchetti (1994, 1997). Software for robust estimation of parametric income-distribution models is provided in Cowell and Gomulka (1999).

typical value), the problems with contaminated data occur in the upper tail of the distribution. A case can therefore be made for using parametric modelling only in the upper tail, and estimating the parameter of the upper-tail model robustly.

A suitable model for the upper tail is the Pareto distribution given by

$$F_\theta = 1 - \left[\frac{x}{x_0}\right]^{-\theta}$$

A semi-parametric approach will combine a non-parametric RLC for say the $(1-\alpha)\%$ lower incomes and a parametric RLC based on the Pareto distribution for the $\alpha\%$ upper incomes. Let $F^{(n)}$ and $F_{\hat{\theta}}$ denote respectively the empirical distribution function and the estimated Pareto distribution. Usually, the estimates for θ are obtained through a simple linear regression of $\ln n(x)$, i.e. the natural logarithm of the number of x_is greater than x, and $\ln(x)$, for the αn upper values of the xs. In order to avoid "jumps" on the RLC at the $(1-\alpha)$ quantile, we propose here to re-center the Pareto distribution. That is, we put $x_0 = Q(F^{(n)}; 1-\alpha)$ and the regression model becomes

$$\ln\left(\frac{n(x)}{\alpha n}\right) = -\theta \ln\left(\frac{x}{x_0}\right) \tag{9.38}$$

from which we get $\hat{\theta}$ which is either a classical estimator such as the least squares estimator (LS), or a robust estimator. For the latter, the literature has a number of proposals (see e.g. Huber 1973, Hampel et al. 1986, Rousseeuw and Leroy 1987, Marazzi 1993) that can be classified into two broad groups, depending on how "robust" they are. Indeed, we can distinguish two types of robustness, namely global and infinitesimal robustness. The latter is based on the IF which measures the influence of an infinitesimal amount of contamination in the data upon the statistics of interest. A bounded IF means that the statistic is robust in the infinitesimal sense. The former is concerned with the proportion of contamination a statistic can withstand before it "breaks down", or in other words before its bias becomes arbitrarily large. High breakdown-point estimators such as the least median of squares of Rousseeuw (1984) which can withstand up to nearly 50% of contaminated data (but are less efficient) are robust in the global sense. For a robust estimator in the infinitesimal sense, we propose here to use a robust estimator of the so-called Mallows class of M-estimators. These generalise the maximum likelihood estimators which, for the normal regression model, are given by

$$\frac{1}{n}\sum \frac{r_i(\theta)}{\sigma} x_i = 0 \tag{9.39}$$

r_i being the residuals. In contrast to (9.39) the Mallows class is defined by

$$\frac{1}{n}\sum \psi\left(\frac{r_i(\theta)}{\sigma}\right) w(x_i) x_i = 0 \tag{9.40}$$

By choosing ψ and w in (9.40) appropriately, the influence of large residuals and extreme values in the xs is limited. There are different possible choice for ψ and w depending on several optimality criteria (see Hampel et al. 1986). Actually the differences lie mainly on the choice of $w(\cdot)$. A simple weighting scheme is based on the diagonal elements h_{ii} of the hat matrix $\mathbf{H} := \mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}'$ where $\mathbf{X}$ denotes the design matrix in the regression model. Without leverage points, $h_{ii} \approx p/n$ (p being the dimension of the independent variables) so that one downweights cases for which h_{ii} exceeds a value bp/n. Thus

$$w_b(x) = \min\left\{1; \frac{bp}{nh_{ii}}\right\}$$

A benchmark is given by $b = 1.5$. For the ψ function, an efficient choice is the Huber function

$$\psi_c\left(\frac{r}{\sigma}\right) = \frac{r}{\sigma}\min\left\{1; \frac{c}{|r/\sigma|}\right\}$$

which downweights standardized residuals lying far away, depending on the choice of the constant c. Choosing $c = \infty$ leads to no downweighting, and choosing $c = 1.345$ leads to an estimator which achieves 95% efficiency under the normal model. One also needs to estimate σ, and once again it should be robust. A relatively simple one is given by the median absolute deviation of residuals

$$\hat{\sigma} = k \text{ med}_i \left| r_i - \text{med}_i r_i \right|$$

where $k = 1.4826$ ensures consistency under the Gaussian model.[15]

The semi-parametric GLC is then given by

$$C(F_{\hat{\theta}}, F^{(n)}; q) = \begin{cases} \int_{\underline{x}}^{Q(F^{(n)};q)} x dF^{(n)}(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F^{(n)};1-\alpha)} x dF^{(n)}(x) & \\ +\alpha \int_{x_0}^{Q(F_{\hat{\theta}};q^*)} x dF_{\hat{\theta}}(x) & q > 1-\alpha \end{cases} \tag{9.41}$$

$$= \begin{cases} \int_{\underline{x}}^{Q(F^{(n)};q)} x dF^{(n)}(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F^{(n)};1-\alpha)} x dF^{(n)}(x) & \\ +\alpha \frac{\hat{\theta}}{1-\hat{\theta}} x_0 \left[[1-q^*]^{\frac{\hat{\theta}-1}{\hat{\theta}}} - 1 \right] & q > 1-\alpha \end{cases}$$

where $q^* = \frac{q-(1-\alpha)}{\alpha}$. The semi-parametric RLC is simply

$$L(F_{\hat{\theta}}, F^{(n)}; q) = \frac{C(F_{\hat{\theta}}, F^{(n)}; q)}{\mu(F_{\hat{\theta}}, F^{(n)})}$$

where

$$\mu(F_{\hat{\theta}}, F^{(n)}) = \int_{\underline{x}}^{Q(F^{(n)}:1-\alpha)} x dF^{(n)}(x) + \alpha \int_{x_0}^{\infty} x dF_{\hat{\theta}}(x) \tag{9.42}$$

$$= \int_{\underline{x}}^{Q(F^{(n)}:1-\alpha)} x dF^{(n)}(x) + \alpha \frac{\hat{\theta}}{\hat{\theta}-1} x_0$$

A question arises here about the choice of the proportion α of data to model: clearly it will not be the same value nor will it always be at the particular income level for all distributions. We may propose a simple rationale based on prior knowledge of the quality of the data. First it should be stressed that α should be as small as possible to avoid putting too much of the parametric approach

[15] As mentioned earlier, the choice for the robust estimator is not the most important feature of our approach and other robust estimators proposed in statistical programs like Splus are also possible choices.

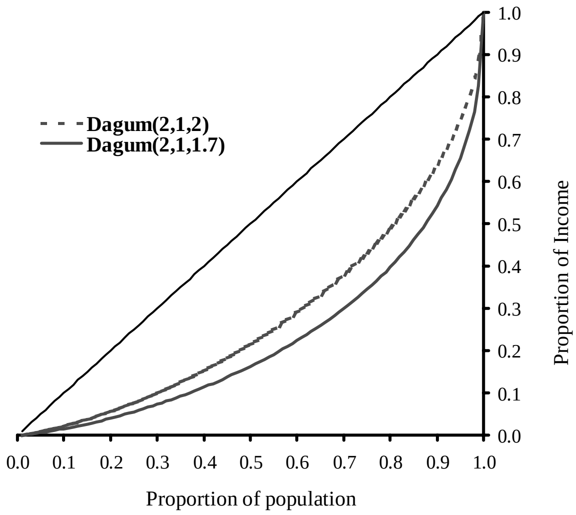

Figure 9.10: Uncontaminated Dagum-I distributions

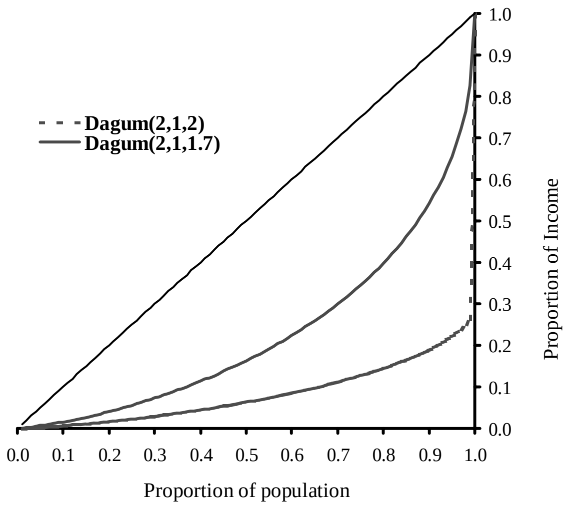

Figure 9.11: Contaminated Dagum-I distributions

into the ranking exercise for the reason we mentioned before. Second, α should be large enough so that a majority of data points in the upper tail subsample are uncontaminated data. Let ε be the (suspected) proportion of contaminated data in the whole sample, which should be relatively small. Suppose that the data analyst has a fairly good idea of that quantity which in general depends on the data source. We propose the adoption of a minimax approach in that we assume that the contamination will result in the worst senior, i.e. in extremely large incomes. To prevent the ranking exercise being completely determined by this proportion ε of contaminated data, one then should get the information on the upper tail through the estimation of the parameter of the Pareto distribution. If a high breakdown point estimator is used, then α could be taken just above $2 \cdot \varepsilon$, so that in the subsample of the upper tail, the proportion of contamination does not exceed 50%. If a more efficient estimator is chosen, then α is chosen so that the amount of contamination in the subsample of the upper tail should not exceed its breakdown point. For example, if the breakdown point of the Huber estimator is about 4%, then $\alpha = \frac{1}{0.04}\varepsilon$.

In order to test our semi-parametric RLC we performed the following simulation exercise. Two samples of 10 000 observations were simulated from a Dagum type I distribution given by

$$f(x; \beta, \lambda, \delta) = (\beta + 1)\lambda\delta x^{-(\delta+1)}(1 + \lambda x^{-\delta})^{-(\beta+1)} \tag{9.43}$$

(Dagum 1977).[16] The values of the parameters were chosen in order to get two distributions such that one exactly RLC-dominates the other. They are the Dagum(2,1,2) (i.e. $\beta = 2$, $\lambda = 1$, $\delta = 2$) and the Dagum(2,1,1.7). The RLCs for the two samples are given in Figure 9.10. We then contaminated the Dagum(2,1,2) by multiplying the three largest observations by 100. Note that although the multiplying factor is quite large, the proportion of contaminated data is very small (i.e. 0.03%). The RLC for the contaminated Dagum(2,1,2) and the Dagum(2,1,1.7) are given in Figure 9.11. We can see that the dominance or-

[16]The form (9.43) has the property that for large values of x, the distribution converges to the Pareto distribution. Note also that this model can be seen as a particular case of the generalized Beta distribution proposed by McDonald and Ransom (1979).

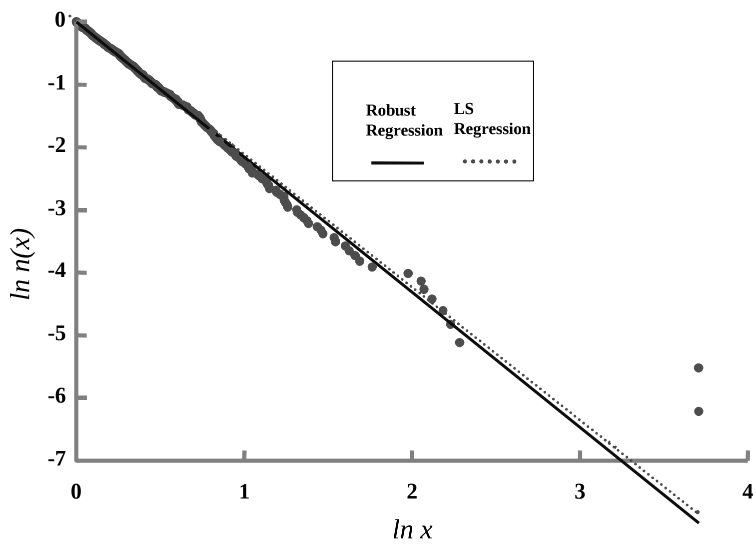

Figure 9.12: Alternative regression methods for a Pareto tail: Uncontaminated data

der is completely reversed because the contaminated Dagum(2,1,2) is completely determined by the three extreme observations introduced into the data.

In order to avoid this very misleading picture based on the non-parametric RLC, we modelled the upper tail of the distribution using the Pareto distribution as explained above. We used the classical LS estimator and the robust Mallows-type estimator with default parameters c=1.345, b=1.5, k=1.4826. We first estimated the parameter θ for the model given by (9.38) with $\alpha = 5\%$. The values of $\hat{\theta}$ for the non-contaminated sample are respectively $\hat{\theta} = 2.11$ for the LS estimator and $\hat{\theta} = 2.16$ for the robust estimator, whereas for the contaminated sample they are respectively $\hat{\theta} = 1.44$ for the LS estimator and $\hat{\theta} = 2.15$ for the robust estimator. The data and estimated regression lines are given in Figures 9.12 and 9.13. We can see that the classical LS estimator is influenced by the extreme observations, whereas the robust estimator remains very stable. If we then estimate the semi-parametric RLC using (9.41) and (9.42) and compare them to the non-parametric RLC using the non-contaminated sample, we get the picture given in Figure 9.14. We can see that the semi-parametric RLC on non-contaminated data and/or using a robust estimator give exactly the

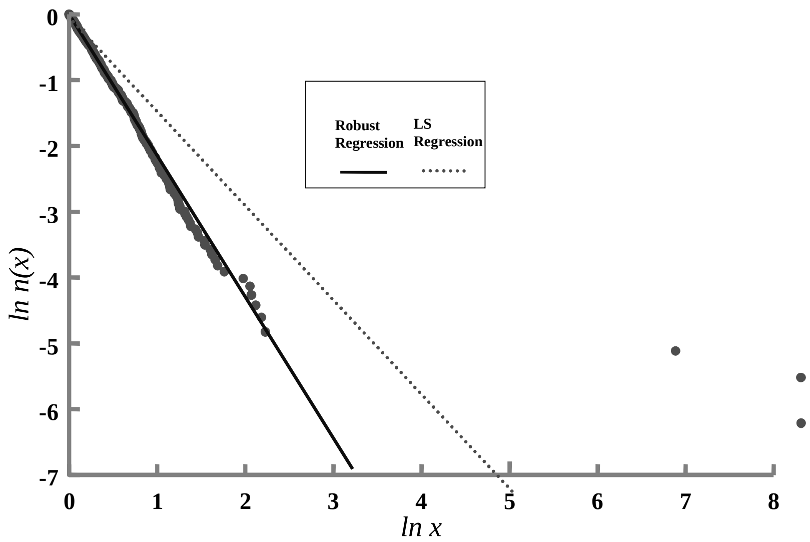

Figure 9.13: Alternative regression methods for a Pareto tail: Contaminated data

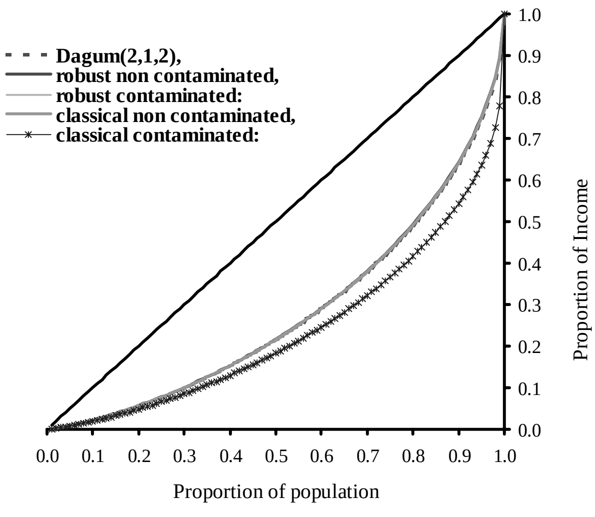

Figure 9.14: Semi-parametric approach RLCs

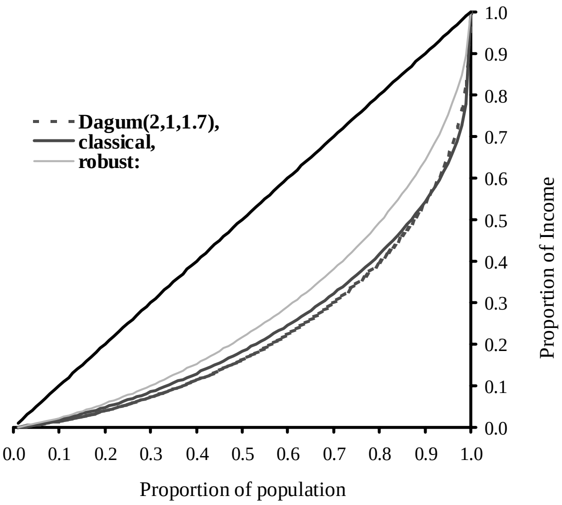

Figure 9.15: Semi-parametric Lorenz rankings: classical and robust

same RLC as the non-parametric RLC with non-contaminated data. However, when one uses a semi-parametric RLC with a classical estimator on contaminated data, the picture is distorted. It should be noted that it not as distorted as with the non-parametric RLC given in Figure 9.11. Finally, Figure 9.15 gives the RLC comparison of the semi-parametric RLC on the contaminated Dagum(2,1,2) compared with the non-parametric RLC on the Dagum(2,1,1.7). We can see that with the robust semi-parametric RLC, the dominance order is preserved, whereas with the classical semi-parametric RLC the curves cross, thus contradicting the original order.

6.3. Non-parametric methods: trimming

A natural approach would be to consider the use of *trimmed Lorenz Curves* as estimators of Lorenz curves. This concept builds upon an established tool – the so-called *trimmed mean.*

Consider a number $\alpha \in [0, \frac{1}{2})$ which we will call the *balanced trimming proportion.* The trimmed mean of distribution F with trimming parameter α, $\bar{X}_\alpha(F)$

is then given by

$$\begin{aligned}\bar{X}_\alpha(F) &= \tfrac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} ydF(y)\\ &= \tfrac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha} F^{-1}(t)dt.\end{aligned} \tag{9.44}$$

This estimator of location has intuitive appeal: one removes both the αn smallest and the αn largest observations in a sample of size n, and calculates the mean of the remaining observations: notice that $\lim_{\alpha\to 0.5}\bar{X}_\alpha(F) = Q(F, 0.5)$ – in the limiting case as α approaches 50% the trimmed estimate of the mean approaches the median.

To extend the idea to Lorenz curves, one has to interpret the quantile and income-cumulation functions (9.7) and (9.8). α-trimming the data means that $Q(F;q) \in (Q(F;\alpha), Q(F;1-\alpha))$ and thus $q \in (\alpha, 1-\alpha)$. Therefore, the α-trimmed generalized Lorenz, Lorenz and absolute Lorenz curves[17] are respectively given by

$$C_\alpha(F;q) = \tfrac{1}{1-2\alpha}\int_{Q(F;\alpha)}^{Q(F;q)} xdF(x), \tag{9.45}$$

$$L_\alpha(F;q) = \frac{C_\alpha(F;q)}{\bar{X}_\alpha(F)}, \tag{9.46}$$

$$A_\alpha(F;q) = (1-2\alpha)\cdot C_\alpha(F;q) - \bar{X}_\alpha(F)\cdot(q-\alpha). \tag{9.47}$$

– Cf equations (9.8), (9.9) and (9.10). From equations (9.45-9.47) we have $C_\alpha(F;\alpha) = 0$, $L_\alpha(F;\alpha) = 0$, $A_\alpha(F;\alpha) = 0$ and $C_\alpha(F;1-\alpha) = \bar{X}_\alpha(F)$, $L_\alpha(F;1-\alpha) = 1$, $A_\alpha(F;1-\alpha) = 0$.

The IFs of these trimmed Lorenz curves will be bounded for all q because extreme values in the data are automatically removed, for all $\alpha > 0$. Trimmed

[17] See the similar concept of restricted dominance discussed by Atkinson and Bourguignon (1989).

Lorenz curves can be thought of as Lorenz curves on a restricted sample and enable one to compare two distributions on the basis of $100(1 - 2\alpha)$ percent of the middle income-receiving units:[18] the $\succ_{C_\alpha}$-ranking criterion provides a parametrised extension of the standard $\succ_C$ method for a range of values of $\alpha \in [0, \frac{1}{2})$.

The above is predicated on the idea of a balanced trim of observations in both tails simultaneously, $\frac{1}{2}\alpha n$ observations from each. It makes sense to consider an extension of this method to single-tail trimming in the case where it is appropriate to form an *a priori* judgment about the nature of the contamination. If, for example, contamination is assumed to affect only the lower tail of the distribution then the bottom αn observations are removed in a 1-sided trim.

The appeal of the trimming procedure is not based solely upon intuition. It can, for example, be shown that if there are "rogue" values in a large sample then these can only significantly effect the outcome of a distirbutional comparison if they occur as extreme values of the sample – see Appendix 9.3. There remains the question of how many extreme values should be trimmed: how should one choose α? As with parametric methods there is a trade-off of robustness against efficiency and the example in section 7 shows how reasonable pragmatic choices may be made.

7. Empirical examples

The trimming approach offers a practical tool for the comparison of income distribution when one wants an explicit control for taking account of the influence of outliers. We use the analysis of section 6.3 to examine more carefully two aspects of conventional wisdom concerning comparisons of income distribution. In each case the data are taken from the LIS data-base and refer to real income per equivalent adult distributed amongst individuals (see Appendix 9.2).

[18] This is a practice that is sometimes adopted in pragmatic discussion of inequality trends. See also the discussion of related issues by Howes (1996).

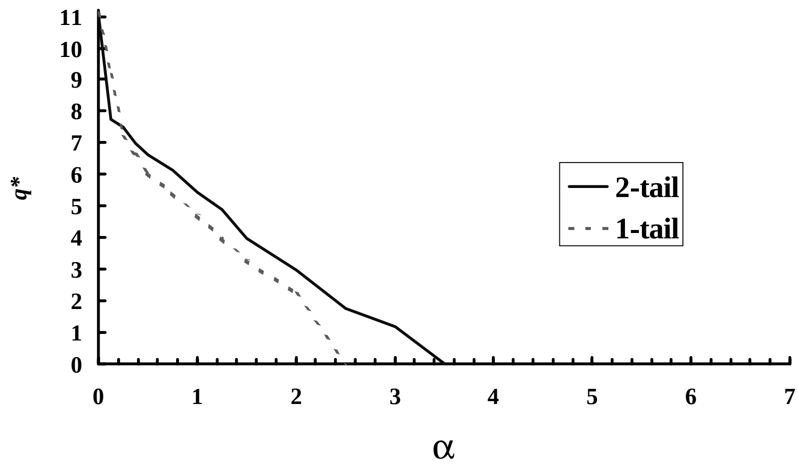

Figure 9.16: Is Sweden more equal than Germany?

7.1. Cross-country comparison: Sweden and Germany

The received wisdom suggests that 1980s Sweden is more equal than Germany. However, is this actually borne out by the data, and what are the implications for standard welfare comparisons? To investigate this we use data for Sweden 1981 and (West) Germany 1983. Given standard definitions it immediately appears that $F_{\text{GERMANY}} \succeq_C F_{\text{SWEDEN}}$ so that there is no question but that the German income-distribution second-order dominates that for Sweden: the generalised Lorenz curve for Germany is higher. However we also find $F_{\text{SWEDEN}} \perp_A F_{\text{GERMANY}}$ and $F_{\text{SWEDEN}} \succeq_{A_{0.005}} F_{\text{GERMANY}}$: given a very slight trim of both tails (a half of one percent) Sweden absolute-Lorenz dominates Germany.

What of inequality? As Figures 9.16 and 9.17 show there is an ambiguity for the raw data – $F_{\text{SWEDEN}} \perp_L F_{\text{GERMANY}}$ – which is due to a single intersection of the Lorenz curves. Figure 9.16 depicts the position of the switch-point (where the Lorenz curves intersect) for two types of trim for various values of α: $q^{**}(\alpha)$ for the balanced two-tail trim (solid curve), and $q^{*}(\alpha)$ for the one-sided lower-tail trim (dotted curve). Let the points where $q^{**}(.)$ and $q^{*}(.)$ become zero be denoted

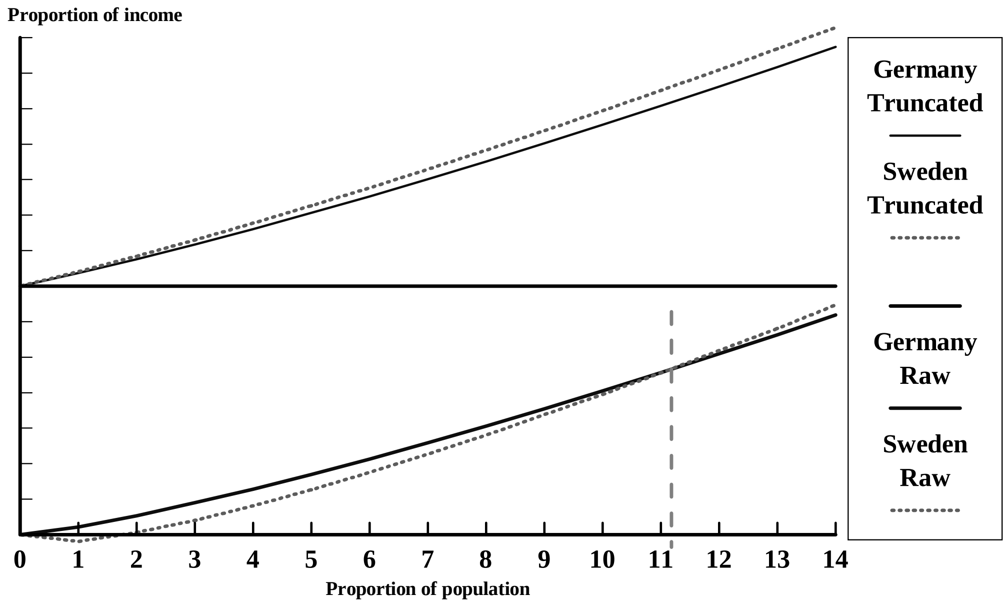

Figure 9.17: Germany vs Sweden: the effect of a 2.5% one-tail trim on the Lorenz comparison

α^{**} and α^* respectively. Then we have

$$\begin{aligned} q^{**}(0) &= q^*(0) = 0.11 \\ q^{**}(\alpha) &= 0,\ \alpha \geq \alpha^{**} = 0.031 \\ q^*(\alpha) &= 0,\ \alpha \geq \alpha^* = 0.025 \end{aligned}$$

We have $F_{\text{SWEDEN}} \succeq_{L_\alpha} F_{\text{GERMANY}}$ only if a trim of $2\frac{1}{2}\%$ of the observations is carried out on the lower tail, or a balanced trim of 3.1%.

Notice also that both q^{**} and q^* fall rapidly for α very close to zero and thereafter decrease more gently; the comparison is extremely sensitive to presence or absence the first few observations (in either the 1- or 2-tail case). However it appears to be unreasonable to suppose that the true picture is of strict Lorenz dominance in that at least 1000 observations would have to be discarded from the German data ($n \simeq 42,000$) in order for this conclusion to obtain.

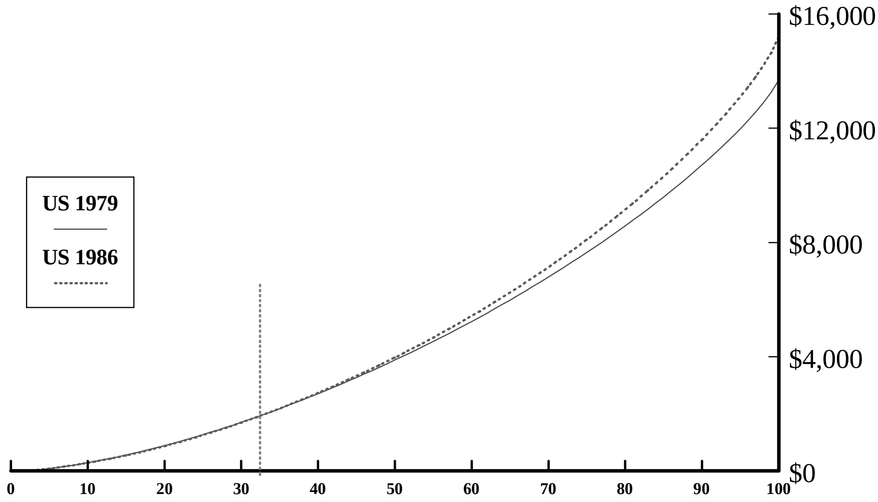

Figure 9.18: US 1986 does not second-order dominate US 1979

7.2. Inequality over time: the US in the 1980s

The same technique may of course be applied to comparisons within one country, but between two points in time. In the United States the conventional wisdom is perhaps even more sharp in its sketch of recent events – inequality rose over the 1980s. Again the fact is – perhaps surprisingly – that the raw data do *not* reveal an unambiguous increase in inequality, in the standard Pigou-Dalton sense. It might appear that this is principally due to the presence of negative incomes in the first centile group: as we will see this is not quite the whole story. Note first that $F_{\text{US86}} \bot_C F_{\text{US79}}$ – we do not have first- or second-order distributional dominance (see Figure 9.18 – the generalised Lorenz curves intersect three times at about $q = 0.02, 0.06, 0.32$), but $F_{\text{US79}} \succeq_A F_{\text{US86}}$.

The trimming procedure is more complex. The problem of negative incomes is disposed of by a very modest (less than 0.5%) trim; but there remains a problem of multiple intersections of the Lorenz curves at the bottom tail (there are intersections in the neighborhood of $q = 0.0258$ and $q = 0.0310$). Figure 9.20 plots $q^{**}(\alpha)$ and $q^{*}(\alpha)$ in this case: in view of the multiple intersections, these values are interpreted as the maximum switch point between the two Lorenz curves for each value of α. The outcome of the α-trimming procedure is interesting in that

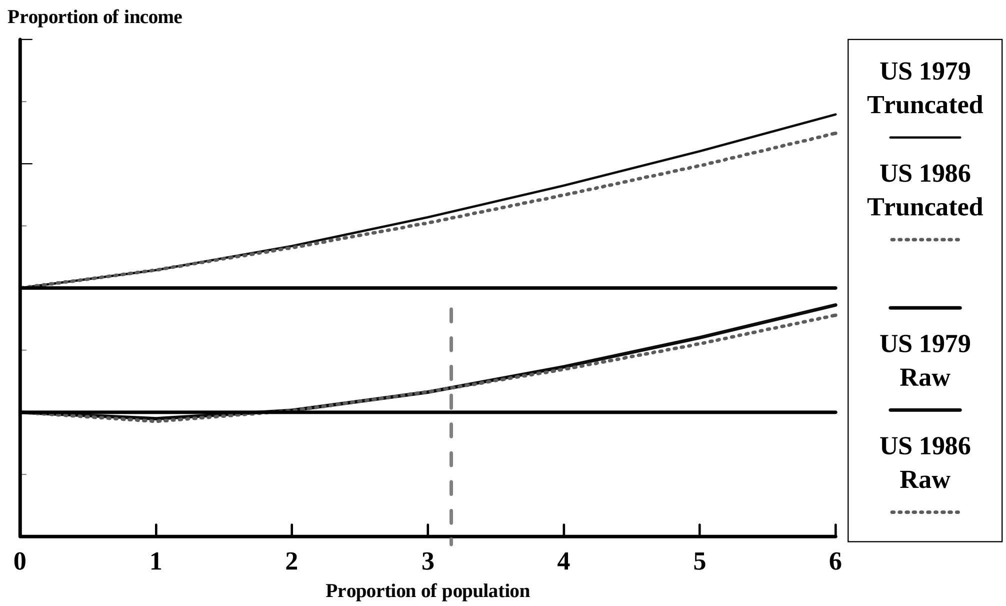

Figure 9.19: US inequality: the effect of a 2.5% one-tail trim on Lorenz comparisons

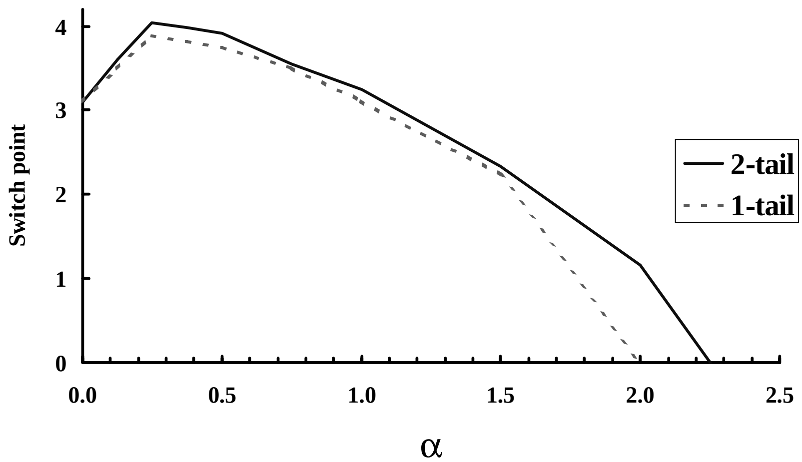

Figure 9.20: Did Inequality Rise in the US?

– by contrast to the Germany-versus Sweden example – neither $q^{**}(.)$ nor $q^{*}(.)$ is monotonic. By dropping some 200 to 300 observations (2 percent) in the single-tailed trim, or 600 to 700 observations ($4\frac{1}{2}$ percent of the whole sample) in the two tail trim one may then conclude that $F_{\text{US79}} \succeq_{L_\alpha} F_{\text{US86}}$.

However there are interesting points in common with the Germany-versus-Sweden example. First, for values of α in the range $[0, 0.01]$ one finds a relationship between the switch-point and α which is clearly different from the relationship that holds in the neighbourhood of the points α^{**} and α^{*}. Second, the shape of the two-trail trim graph follows closely that of the one-tail trim. Thirdly, all the action appears to come from the lower tail: in the distributional comparisons reported in subsections 7.1 and 7.2 an upper-tail trimming experiment has no effect on distributional rankings.

8. Conclusions

Using ranking criteria to make welfare inferences about income distributions is of immense theoretical advantage and practical convenience. In addition to avoiding the arbitrariness associated with the choice of specific welfare functions or inequality measures, it might be supposed that use of the distributional-ranking approach will also enable the empirically oriented researcher to avoid some of the pitfalls associated with sensitive inequality statistics. However, it has pitfalls, which may not be so readily apparent. These pitfalls can be set out in the form of a simplified story.

1. Except for some special cases "first-order" distributional statistics – the quantiles – are robust.

2. However, even these exceptions do not matter for first-order dominance results: the quantile *ranking* is robust.[19]

3. Second-order (and higher) statistics and stochastic dominance results are non-robust: contamination can seriously affect welfare conclusions when

[19] For further discussion of the statistical implementation of first order criteria see Ben Horim (1990) and Stein et al. (1987).

extreme values are present in the data: small amounts of data contamination in the wrong place can even reverse unambiguous welfare conclusions. This corresponds to other work on robust methods applied to income distribution: In Cowell and Victoria-Feser (1996a, 1996b) we showed that most inequality measures are non-robust, but that most poverty indices with exogenous poverty lines are robust – see also Monti (1991).

Finally, it is possible to implement practical "work-rounds" for cases where the stochastic dominance criteria are non-robust: in other words computational devices which can be used to draw restricted welfare inferences about the properties of distributional comparisons. Use of the semi-parametric approach to modelling income distribution enables one to control for the distortionary effect of upper-tail contamination in a systematic fashion. One-tail or two-tail trimming offers a way to extend the simple distributional-dominance criteria in a way that allows one to examine systematically the potential loss of information against robustness of the statistic. These procedures have already been used in an *ad hoc* fashion[20]: the above may serve as the basis for a more systematic treatment.

[20] Some empirical studies have concentrated upon a subset of the distribution delimited either by population subgroup (prime-age males) or by arbitrarily excluding some of the data in the tails (Gottschalk and Smeeding 2000).

9. Appendices

9.1. Quantile definition

In the text we followed a standard convention that determines the qth quantile from a (single-valued) functional Q of the distribution F. However, some would regard the definition of quantiles as being indeterminate in the case of "dead" intervals discussed above (Cf Kendall and Stuart (1977), pp. 39-41). An alternative approach is to suppose that the quantile is a correspondence. To do this define the following collection of subsets of $\mathfrak{X}$:

$$\Xi := \{\{x : a \leq x \leq b\} : a, b \in \mathfrak{X}\}$$

Then the quantile correspondence is

$$\tilde{Q} : \mathfrak{F} \times [0, 1] \rightarrow \Xi$$

such that

$$\tilde{Q}(F; q) = \{x : F(x) = q\}$$

Where F is strictly increasing $\tilde{Q}(F; q)$ yields a singleton set, and in other cases it may yield a singleton or a proper interval of values: the correspondence $\tilde{Q}$ is upper-semicontinuous.

Now consider the effect of contamination on the quantiles of the distribution. Wherever $f(x) > 0$ it is clear that $\tilde{Q}(F; q)$ and $Q(F; q)$ will be effectively identical. In other cases $Q(F; q) = \min\left(\tilde{Q}(F; q)\right)$. Take an arbitrary q_0 and let $\xi_0 := \tilde{Q}(F; q_0)$. In all cases we find that for arbitrarily small ε we have $\tilde{Q}(F_\varepsilon^{(z)}; q) \rightarrow \xi_0$.

9.2. Data Specification

The Luxembourg Income Study permits comparison of different countries' income distributions based on consistent international definitions of income and the income receiver. Accordingly the same basic specifications were used both the (Germany, Sweden) and the (US 1979, US 1986) comparisons in section 7. The sample sizes were:

Germany 1983	42,752
Sweden 1981	9,625
US 1979	15.928
US 1986	12,600

The income distributions are formed using the following concept of equivalised incomes (Buhmann, Rainwater, Schmaus, and Smeeding 1988) (Coulter, Cowell, and Jenkins 1992):

$$y = \frac{hhy}{hhsize^{\gamma}}.$$

where

- Household income, $hhy =$ net family (unit) income after tax.
- Household size, $hhsize =$ the number of persons in the family unit.
- $\gamma = 0.5$.

Each observation is then given a weight, $indwgt = hhsize * hweight$, to obtain distributions of income across individuals (Cowell 1984) (Danziger and Taussig 1979). The variable $hweight$, is the family unit sample weight.

For calculating distributions for different years and for conversion to dollars the following data from the IMF Year Book 1994 were used.

	1981	**1983**
Price level consumption		
Germany	106.3	115.6
Sweden	112.1	132.6
Dollar exchange rate		
Germany	2.26	2.553
Sweden	5.063	7.667

Consider a pair of empirical distributions

$$\mathbf{x} \; : \; = x_1, x_2, ..., x_n \tag{9.48}$$

$$\mathbf{y} \; : \; = y_1, y_2, ..., y_n \tag{9.49}$$

that have been sorted in ascending order of incomes; $\mathbf{x}$ and $\mathbf{y}$ are drawn, respectively from the mixtures

$$F_\varepsilon \; : \; = [1-\varepsilon]F + \varepsilon H \tag{9.50}$$

$$G_\varepsilon \; : \; = [1-\varepsilon]G + \varepsilon H \tag{9.51}$$

where the number ε is unknown but is the same for both distributions and where, for every pair (x_i, y_i), $i = 1, ..., n$, either both components are genuine observations (drawn from F and G respectively) or both are drawn from the same contamination distribution H. Further assume that $\mathbf{x}_{-i} \succeq_C \mathbf{y}_{-i}$ where

$$\mathbf{x}_{-i} \; : \; = x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n \tag{9.52}$$

$$\mathbf{y}_{-i} \; : \; = y_1, y_2, ..., y_{i-1}, y_{i+1}, ..., y_n \tag{9.53}$$

Defining

$$D_{i-1} \; : \; = \sum_{j=1}^{i-1} [x_j - y_j]$$

$$D_{i+1} \; : \; = D_{i-1} + x_{i+1} - y_{i+1}$$

then if $\mathbf{x} \perp_C \mathbf{y}$ we have[21]

$$\begin{aligned} D_{i-1} &\geq 0 \quad (9.54) \\ D_{i+1} &\geq 0 \quad (9.55) \\ y_i &> x_i + D_{i-1} \quad (9.56) \end{aligned}$$

The probability that (9.56) is true is given by

$$\Pr\{y_i > x_i + D_{i-1} |\, x_{i+1} \geq x_i \geq x_{i-1}; y_{i+1} \geq y_i \geq y_{i-1}\} \quad (9.57)$$

Expression (9.57) is 0 if [22]

$$x_{i+1} - x_{i-1} \leq D_{i+1} \quad (9.58)$$

and

$$\int_{x_{i-1}}^{x_{i+1}} \Pr\{y_i > x_i + D_{i-1} |\, x_i\} \frac{1}{x_{i+1} - x_{i-1}} dx_i \,. \quad (9.59)$$

otherwise. If x_i is rectangularly distributed in $[x_{i-1}, x_{i+1}]$ and y_i is rectangularly

[21] Conditions (9.54) to (9.56) are equivalent to

$$\begin{aligned} \sum_{j=1}^{i-1} x_j &\geq \sum_{j=1}^{i-1} y_j \\ \sum_{j=1}^{k} x_j &\geq \sum_{j=1}^{k} y_j, k = i+1, ..., n \\ \sum_{j=1}^{i} x_j &< \sum_{j=1}^{i} y_j \end{aligned}$$

[22] Note (9.61) is equivalent to

$$y_{i+1} \leq x_{i-1} + D_{i-1}$$

distributed in $[y_{i-1}, y_{i+1}]$, (9.59) simplifies to

$$\frac{\int_{x_{i-1}}^{x_{i+1}} \max\left\{\frac{y_{i+1}-D_{i-1}}{y_{i+1}-y_{i-1}}, 0\right\} dx_i}{x_{i+1}-x_{i-1}} = \int_{x_{i-1}}^{y_{i+1}-D_{i-1}} \frac{y_{i+1}-D_{i-1}-x_i}{y_{i+1}-y_{i-1}} \frac{1}{x_{i+1}-x_{i-1}} dx_i$$

$$= \frac{1}{2} \frac{[x_{i+1}-x_{i-1}-D_{i+1}]^2}{[y_{i+1}-y_{i-1}][x_{i+1}-x_{i-1}]} \tag{9.60}$$

Now consider the effect upon the distributional comparison of dropping observations from of the samples. Assume that the densities f and g exist and are strictly increasing. We may then show that in a large sample it is impossible to have an intermediate x which will change the result. Condition (9.58) can be rewritten

$$Q\left(\hat{F}; q+dq\right) - Q\left(\hat{F}; q\right) \leq C\left(\hat{F}; q+dq\right) - C\left(\hat{G}; q+dq\right) \tag{9.61}$$

where $i = 1 + nq$, $dq = 2/n$ and (9.60) becomes

$$\begin{aligned} \pi(q) \quad : &= \frac{1}{2} \frac{[x_{i+1}-x_{i-1}-D_{i+1}]^2}{[y_{i+1}-y_{i-1}][x_{i+1}-x_{i-1}]} \\ &= \frac{1}{2} \frac{\left[Q\left(\hat{F}; q+dq\right) - Q\left(\hat{F}; q\right) - \left[C\left(\hat{F}; q+dq\right) - C\left(\hat{G}; q+dq\right)\right]\right]^2}{\left[Q\left(\hat{G}; q+dq\right) - Q\left(\hat{G}; q\right)\right]\left[Q\left(\hat{F}; q+dq\right) - Q\left(\hat{F}; q\right)\right]} \\ &\simeq \frac{1}{2} \frac{g\left(Q(G;q)\right)}{f\left(Q(F;q)\right)} \left[1 - f\left(Q(F;q)\right) \frac{\Delta(q)}{dq}\right]^2 \end{aligned}$$

where f and g are the densities corresponding to F and G and $\Delta(q) := C(F;q) - C(G;q)$. Now consider what happens as $n \to \infty$. If F strictly dominates G somewhere in $[0, q)$ and $0 < q < 1$ then the right-hand side of (9.61) must be strictly positive, while the left-hand side goes to zero. So, under these conditions, only end-trimming can have alter the the distributional comparison.

Bibliography

Amiel, Y. and F. A. Cowell (1994). Monotonicity, dominance and the Pareto principle. *Economics Letters 45*, 447–450.

Arnold, B. C. (1987). *Majorization and the Lorenz Order: A Brief Introduction.* Heidelberg: Springer-Verlag.

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. and F. Bourguignon (1989). The design of direct taxation and family benefits. *Journal of Public Economics 41*, 3–29.

Ben Horim, M. (1990). Stochastic dominance and truncated sample data. *Journal of Financial Research 13*, 105–116.

Bishop, J. A., J. P. Formby, and P. D. Thistle (1991). Rank dominance and international comparisons of income distributions. *European Economic Review 35*, 1399–1409.

Buhmann, B., L. Rainwater, G. Schmaus, and T. Smeeding (1988). Equivalence scales, well-being, inequality and poverty: Sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database. *Review of Income and Wealth 34*, 115–142.

Coulter, F. A. E., F. A. Cowell, and S. P. Jenkins (1992). Equivalence scale relativities and the extent of inequality and poverty. *Economic Journal 102*, 1067–1082.

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

Cowell, F. A. (1988). 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. (2000). Measurement of inequality. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Chapter 2. Amsterdam: North Holland.

Cowell, F. A. and J. Gomulka (1999). *INeQ: Software for Distributional Analysis.* Software package. London School of Economics, London: STICERD.

Cowell, F. A. and M.-P. Victoria-Feser (1996a). Robustness properties of inequality measures. *Econometrica 64*, 77–101.

Cowell, F. A. and M.-P. Victoria-Feser (1996b). Welfare judgements in the presence of contaminated data. Distributional Analysis Discussion Paper 13, STICERD, London School of Economics, London WC2A 2AE.

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

Dalton, H. (1920). Measurement of the inequality of incomes. *Economic Journal 30*(9), 348–361.

Danziger, S. and M. K. Taussig (1979). The income unit and the anatomy of income distribution. *Review of Income and Wealth 25*, 365–375.

Dardanoni, V. and P. J. Lambert (1988). Welfare rankings of income distributions: a role for the variance and some insights for tax reform. *Social Choice and Welfare* (5).

Davies, J. B. and M. Hoy (1994). The normative significance of using third-degree stochastic dominance in comparing income distributions. *Journal of Economic Theory 64*, 520–530.

Davies, J. B. and M. Hoy (1995). Making inequality comparisons when Lorenz curves intersect. *American Economic Review 85*, 980–986.

Ebert, U. (1988). On the decomposition of inequality: Partitions into nonoverlapping sub-groups. In W. Eichhorn (Ed.), *Measurement in Economics.* Heidelberg: Physica Verlag.

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

Fishburn, P. C. and R. D. Willig (1984). Transfer principles in income redistribution. *Journal of Public Economics 25*, 323–328.

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

Gottschalk, P. and T. M. Smeeding (2000). Empirical evidence on income inequality in industrialized countries. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Chapter 3. Amsterdam: North Holland.

Hampel, F. R. (1968). *Contribution to the Theory of Robust Estimation*. Ph. D. thesis, University of California, Berkeley.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association 69*, 383–393.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley.

Heritier, S. and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association 89*(427), 897–904.

Howes, S. R. (1996). The influence of aggregation on the ordering of distributions. *Economica 63*, 253–272.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics 35*, 73–101.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics 1*, 799–821.

Huber, P. J. (1986). *Robust Statistics*. New York: John Wiley.

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.

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

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. (1974). Rectifiances et dominances intégrales de tous dégrés. Technical report, CEPREMAP, Paris.

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.

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

Marazzi, A. (1993). *Algorithms, Routines and S-Functions for Robust Statistics.* Belmont, California: Wadsworth and Brooks/Cole.

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

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

Monti, A. C. (1991). The study of the Gini concentration ratio by means of the influence function. *Statistica 51*, 561–577.

Moyes, P. (1987). A new concept of Lorenz domination. *Economics Letters 23*, 203–207.

Muliere, P. and M. Scarsini (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory 49*, 314–323.

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

Pigou, A. C. (1912). *Wealth and Welfare.* London: Macmillan.

Prieto-Alaiz, M. and M. P. Victoria-Feser (1996). Income distribution in Spain: A robust parametric approach. Distributional Analysis Discussion Paper 20, STICERD, London School of Economics, London WC2A 2AE.

Quirk, J. D. and R. Saposnik (1962). Admissibility and measurable utility functions. *Review of Economic Studies 29*, 140–146.

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

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association 79*, 871–880.

Rousseeuw, P. J. and A. M. Leroy (1987). *Robust Regression and Outlier Detection*. New York: John Wiley.

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.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica 50*, 3–17.

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

Stein, W. E., R. C. Pfaffenberger, and D. W. French (1987). Sampling error in first-order stochastic dominance. *Journal of Financial Research 10*, 259–269.

Van Praag, B. M. S., A. J. M. Hagenaars, and W. Van Eck (1983). The influence of classification and observation errors on the measurement of income inequality. *Econometrica 51*, 1093–1108.

Victoria-Feser, M.-P. (1993). *Robust Methods for Personal Income Distribution Models*. Ph. D. thesis, University of Geneva, Switzerland. Thesis no 384.

Victoria-Feser, M.-P. (1995). Robust methods for personal income distribution models with application to Dagum's model. In C. Dagum and A. Lemmi (Eds.), *Income Distribution, Social Welfare, Inequality and Poverty*. CT USA: JAI-Press of Greenwich.

Victoria-Feser, M.-P. (1997). A robust test for non-nested hypotheses. *Journal of the Royal Statistical Society, Series B 59*, 715–727.

Victoria-Feser, M.-P. and E. Ronchetti (1994). Robust methods for personal income distribution models. *Canadian Journal of Statistics 22*, 247–258.

Victoria-Feser, M.-P. and E. Ronchetti (1997). Robust estimation for grouped data. *Journal of the American Statistical Association 92*, 333–340.

Wiles, P. J. D. (1974). *Income Distribution, East and West.* Amsterdam: North Holland.

Wiles, P. J. D. and Markowski (1971). Income distribution under communism and capitalism. *Soviet Studies 22*, 344–369,485–511.