Welfare Rankings in the Presence of Contaminated Data

COWELL, Frank, VICTORIA-FESER, Maria-Pia

Abstract

Stochastic dominance criteria are commonly used to draw welfare-theoretic inferences about comparisons of income distribution as well as ranking probability distributions in the analysis of choice under uncertainty. However, just as some measures of location and dispersion can be catastrophically sensitive to extreme values in the data it is also possible that conclusions drawn from empirical implementations of dominance criteria are unduly influenced by data contamination. We show the conditions under which this may occur for a number of standard dominance tools used in welfare analysis.

Reference


DOI: 10.1111/1468-0262.00324
Welfare Rankings in the Presence of Contaminated Data

by

Frank A. Cowell
and

Maria-Pia Victoria-Feser

London School of Economics and Université de Genève

February 2001
Abstract

Stochastic dominance criteria are commonly used to draw welfare-theoretic inferences about comparisons of income distribution as well as ranking probability distributions in the analysis of choice under uncertainty. However, just as some measures of location and dispersion can be catastrophically sensitive to extreme values in the data it is also possible that conclusions drawn from empirical implementations of dominance criteria are unduly influenced by data contamination. We show the conditions under which this may occur for a number of standard dominance tools used in welfare analysis.

Keywords: Welfare dominance; Lorenz curve; robustness

JEL Classification: C13,D63

Correspondence to: Professor F. A. Cowell, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. F.Cowell@LSE.ac.UK.

The first author is partially supported by the ESRC grant #R 55 50 XY8. The second author is partially supported by the “Fond National Suisse pour la Recherche Scientifique”. This paper forms part of the research programme of the TMR network Living Standards, Inequality and Taxation [Contract No. ERBFMRXCT 980248] of the European Communities whose financial support is also gratefully acknowledged. We gratefully acknowledge helpful comments on an earlier version by Tony Atkinson, Robin Burgess, and referees of this journal.
1 Introduction

Ranking theorems based on concepts of stochastic dominance are fundamental to the analysis of income distributions as well as to finance theory and other aspects of the theory of choice under uncertainty. 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 sample data can be taken as a reasonable representation of the underlying income distributions which we want to compare. In practice the data may be contaminated by recording errors, measurement errors and the like and, if the data cannot be purged of these, 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. In any practical approach to the problem of distributional comparisons using contaminated data, two questions immediately arise.

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

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

For the first question there are three distinct approaches. The first, is to specify an explicit objective function such as a utility function or a social-welfare function (SWF): this has the attraction of simplicity but is rather restrictive. The second approach involves specifying a class of measures of dispersion, and using information about dispersion and mean income jointly to draw conclusions about welfare: but, this runs into a class of problems that arise in connection with the estimation of inequality measures and other measures of dispersion. 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. This is the approach adopted here. The discussion will cover both first-order and second-order dominance criteria, and also associated concepts as the Lorenz curve.

The second question is addressed in section 2, where we introduce a simple but powerful model of data contamination. The main formal results are contained in section 3: we show that considerable caution may be required in applying some commonly used ranking results based on stochastic dominance criteria.
2 The Approach

2.1 Notation and definitions

Let $\mathcal{F}$ be the set of continuous probability distributions. Let $X$ ("income") be a random variable with probability distribution $F \in \mathcal{F}$ and support $\mathcal{X} \subseteq \mathbb{R}$; let $\mathcal{X}$ denote the interior of $\mathcal{X}$. We shall write statistics of any distribution $F \in \mathcal{F}$ as a functional $T(F)$; in particular we write the mean as $\mu(F) := \int x dF(x)$.

A ranking principle is a partial order $\succeq$ on $\mathcal{F}$. Denote the strict dominance and equivalence parts of $\succeq$ by the symbols $\succ$ and $\sim$, and let the symbol $\perp$ denote cases where the two distributions cannot be ranked by $\succeq$ (for example where the Lorenz curves cross). The expression $F \succeq_T G$ is used to mean that distribution $F$ weakly dominates distribution $G$ according to the statistic $T$.

2.2 Data contamination

In order to represent the impact of contamination on an income distribution we need a specific model. Consider the elementary distribution $H(z)$ which has a unit point mass at $z$ and zero mass elsewhere:

$$H(z)(x) = \iota(x \geq z).$$

(1)

where $\iota$ is the indicator function $\iota(D) := \{1 \text{ if } D \text{ is true, } 0 \text{ otherwise}\}$. Now suppose that there is a small amount of undetectable contamination at point $z$ in the income distribution. Then the distribution that is actually observed will not be the true distribution $F$ but a mixture distribution:

$$F_\varepsilon(z)(x) := [1 - \varepsilon]F(x) + \varepsilon H(z)(x)$$

(2)

where the parameter $\varepsilon$ 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 $\varepsilon$ of being equal to $z$.

The central issue is then as follows. Suppose we wish to rank two distributions $F$ and $G \in \mathcal{F}$ 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.

For any $T$ this idea can be made more precise by introducing the influence function $IF$ (Hampel 1968, 1974), obtained by taking the derivative with respect to $\varepsilon$ of the statistic at $F_\varepsilon(z)$ when $\varepsilon \to 0$:

$$IF(z; T, F) := \frac{\partial}{\partial \varepsilon}T(F_\varepsilon(z)) \bigg|_{\varepsilon \to 0}.$$  

(3)
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. 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 $\varepsilon \to \sup \left| T(F^{(\varepsilon)}) - T(F) \right|$ when $\varepsilon \to 0$ (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 3.2.

3 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 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 weak form of it as

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

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

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 1 and (2) if two “true” distributions $F$ and $G$ are subjected simultaneously to exactly the same contamination: for any generally decomposable statistic $T$ if $G^{(\varepsilon)} \succeq_T F^{(\varepsilon)}$ 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 5 below. The second difficulty 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 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 influence function 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 3.1); then we consider the implications of this for the behaviour of ranking tools that use families of these statistics (section 3.2).

### 3.1 Robustness properties of distributional statistics

#### 3.1.1 First-order statistics

The first step is to introduce a basic concept that is required both for studying the robustness properties of first-order dominance criteria and as a building block for other ranking tools.

**Definition 2** For any \( q \in (0, 1] \) the 4\textsuperscript{th} quantile is the functional \( Q : \mathfrak{F} \times [0, 1] \mapsto \mathfrak{X} \) such that

\[
Q(F; q) = \inf \{ x | F(x) \geq q \} \tag{4}
\]

Where there is no ambiguity as to the distribution in question we will write \( Q(F; q) = x_q \).

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 (3) to find the influence function \( \text{IF}(z; Q(\cdot, q), F) \). We may write

\[
Q(F_\varepsilon(z); q) = Q \left( F; \frac{q - \iota(x_q \geq z)\varepsilon}{1 - \varepsilon} \right). \tag{5}
\]

Therefore,

\[
\text{IF}(z; Q(\cdot, q), F) = \frac{q - \iota(x_q \geq z)}{f(Q(F; q))} \tag{6}
\]

Note that for any \( F \in \mathfrak{F} \) and for all \( z : \iota(x_q \geq z) = \iota(q \geq F(z)) \). To interpret (6) it is convenient to introduce the concept of the hazard rate \( h(x) := \frac{f(x)}{1 - F(x)} \).

**Theorem 1** \( \forall z \in \mathfrak{X} \) and \( \forall F \in \mathfrak{F} \):

(a) \( \forall q \neq 0, 1 \): \( \text{IF}(z; Q(\cdot; q), F) \) is bounded if and only if \( f(x_q) > 0 \)
(b) If the hazard rate is non-decreasing for large $x$ then $IF(z; Q(\cdot; q), F)$ is bounded as $q \to 1$.

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

Proof. Part (a) is immediate from (6). (b) $\forall z < \sup X$, $\exists (\delta > 0)$ such that $(q = 1 - \delta)$ then $\nu(x_q \geq z) = 1$ and IF becomes $\lim_{x \to \sup X} \frac{1}{\nu(x)}$ and so the second part follows. (c) Likewise, $\forall z > \inf X$, $\exists (\delta > 0)$ such that $(q = \delta)$ then $\nu(x_q \geq z) = 0$ and IF becomes $\lim_{x \to \inf X} \frac{F(x)}{f(x)}$; 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 1 is as follows. Figure 1 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: $(-\infty, a)$, $[b, a’)$ and $[b’, \infty)$ are all “dead” intervals. We see that $Q(F, 0.5)$ is at $b$, and that a small amount of contamination in the region $[a’, \infty)$ would cause $Q(F, 0.5)$ in the mixture distribution to jump to $a’$. So Theorem 1 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,

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 1996).

1 Note that these assumptions imply $b - a = b’ - a’$. 

5
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.

That we have robustness under regularity conditions result is, perhaps, to be expected; but the unboundedness of the IF of individual first-order statistics in the absence of these conditions might appear to be worrying. However, we can say more than this if the first-order statistics are used jointly to make a welfare comparison – see Section 3.2.

### 3.1.2 Second-order statistics

Now consider the second key distributional concept to be derived from $F$.

**Definition 3** The $q^{th}$ cumulative income is the functional $C : \mathcal{F} \times [0, 1] \mapsto \mathcal{X}$: such that:

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

(7)

The importance of this concept in practical analysis of income distributions is considerable: note, for example, that the mean functional emerges as one particular case ($\mu(\cdot) = C(\cdot, 1)$) and the income share of the bottom $q$ of the population is given by $C(\cdot, q)/C(\cdot, 1)$. Again we consider the impact of data contamination as modelled in (2).

The $C$ functional can be written as

$$C(F^z; q) = \int_{Q(F^z; q)}^{Q(F^z ; q)} x dF^z(x)$$

(8)

and the IF can be obtained by applying (3) to give:

$$\text{IF}(z; C(\cdot, q), F) = -\int_{Q(F; q)}^{Q(F ; q)} x dF(x) + Q(F; q)f(Q(F; q))\text{IF}(z; Q, F)$$

$$+ \int_{Q(F; q)}^{Q(F ; q)} x dH(z)(x)$$

$$= qQ(F; q) - C(F; q) + \iota(q \geq F(z))[z - Q(F; q)].$$

(9)

We can see from (9) 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 $-\infty$ the IF can be unbounded for any value of $q$. We may summarize thus:

**Theorem 2** $\forall z \in \mathcal{X}$ and $\forall F \in \mathcal{F}$:

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

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

As we will see in section 3.2 this result has an important consequence for distributional ranking results.
3.2 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. 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 3.1.

3.2.1 The first-order dominance criterion

First we take the quantile function introduced in definition 2. The family of statistics $\{Q(\cdot; q) : q \in (0, 1]\}$ is in effect a formal description of Pen’s “parade of dwarfs and giants” (Pen 1971) and it induces the following principle:

**Definition 4** (a) (Q-ranking) For any $F, G \in \mathcal{F}$, $G \succeq Q F$ if and only if $\forall q \in (0, 1]: Q(G; q) \geq Q(F; q)$. (b) (Strict Q-ranking) For any $F, G \in \mathcal{F}$, $G \succ Q F$ if and only if $G \succeq Q F$ and $\exists q_0 \in (0, 1]: Q(G; q_0) > Q(F; q_0)$

This provides an appealing criterion for the welfare-ranking of income distributions: $G \succeq Q F$ if and only if welfare in distribution $G$ is at least as great as that in distribution $F$ for all decomposable SWFs that are monotonic increasing in income. (Quirk and Saposnik 1962, Saposnik 1981, Saposnik 1983).

Could infinitesimal contamination change a first-order dominance result? As we have seen the influence function 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 (6) 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, although we need to make allowance for a special, exceptional case. For any $F, G \in \mathfrak{F}$ the exceptional case arises where $F$ and $G$ have dead intervals with a nonempty intersection: i.e. there is some non-null interval $I \subset \mathcal{X}$ such that $\forall q \in I : f(Q(F; q)) = f(Q(G; q)) = 0$.

**Theorem 3**  If the exceptional case does not apply then the first-order dominance relations $\succ_Q$ and $\succeq_Q$ are robust.

**Proof.** We consider separately the cases without and with a dead interval for one distribution.

1. Assume $\forall q : f(Q(F; q)) > 0$ and $f(Q(G; q)) > 0$. For weak dominance suppose $G \succeq_Q F$ and let $F^{(z)}_\varepsilon$ be given by (2). Then, from definition 4:

$$\forall q \in [0, 1] : Q(G; q) \geq Q(F; q). \quad (10)$$

However because $\forall q f(Q(F; q)) > 0$ implies continuity of $Q(F; q)$ in $q$ we have

$$\forall q, \lim_{\varepsilon \to 0} Q(F^{(z)}_\varepsilon; q) = \lim_{\varepsilon \to 0} Q \left(F; \frac{q - \iota(x_q \geq z)\varepsilon}{1 - \varepsilon}\right) = Q(F; q). \quad (11)$$

Taking (10) and (11) together we have $\lim_{\varepsilon \to 0} Q(F^{(z)}_\varepsilon; q) \leq Q(G; q)$ and so, by definition 4, $\lim_{\varepsilon \to 0} F^{(z)}_\varepsilon \succeq_Q G$ which means that $\succeq_Q$ is robust. For strict dominance suppose $G \succ_Q F$. Then, from definition 4, there is some $q^* > 0$ such that:

$$Q(G; q^*) > Q(F; q^*). \quad (12)$$

Taking (11) and (12) together we have $\lim_{\varepsilon \to 0} Q(F^{(z)}_\varepsilon; q^*) < Q(G; q^*)$ and so, by definition 4, $\lim_{\varepsilon \to 0} F^{(z)}_\varepsilon \prec_Q G$ which means that $\prec_Q$ is robust. □

2. Let $I := [x^*, x^{**}) \subset \mathcal{X}$ be a dead interval for $F$ such that $f(x) = 0, \forall x \in I, I \neq \emptyset$ and $\forall q f(Q(G; q)) > 0$. First note that in this case weak dominance implies strict dominance: let $q^* := F(x^*)$, then

$$Q(F; q^*) = x^* < x^{**} \leq Q(G; q^*). \quad (13)$$

For arbitrarily small $\varepsilon$ we have $Q(F; q) \to x^*$ as $q \uparrow q^*$ and $Q(F; q) \to x^{**}$ as $q \downarrow q^*$ and by (13), the jump from $x^*$ to $x^{**}$ (or from $x^{**}$ to $x^*$) due to infinitesimal contamination does not affect the ranking.
Figure 2: $Q$-rankings with a “dead” interval for both distributions.

In the exceptional case it appears that we do not have robust $Q$-rankings. The situation is shown in Figure 2 where $[b, a')$ is the “dead” interval for $F$ such that $f(x) = 0, \forall x \in [b, a')$ and $q^* := F(b)$. Assume that there is contamination at $z \in (Q(G; q^*), \sup X)$. For $q = q^*$ we have $Q(G; q^*) < Q(F_{\varepsilon}^{(z)}; q^*) = \min(z; b)$ for arbitrarily small $\varepsilon$ although $Q(G; q^*) > Q(F_{\varepsilon}^{(z)}; q^*)$ for all other $q$. Furthermore it appears that this situation may actually occur in practice in the right tail of the distributions where dead intervals can happen after the largest observation. Indeed, if one has $G \succeq Q F$ then by extending the largest observation in $F$, one might have $G \perp Q F$ because of contaminated data.

However, it is arguable that the “exceptional case” is not as important as it might seem. First, it is of limited relevance in welfare terms since one is dealing with parts of the income distribution where there are, by assumption, no members of the population in either distribution. Second, the difficulty with this exceptional case would vanish if one were to use a slightly different definition of the quantile.

---

3Note that this issue did not arise in the standard welfare literature where the argument was pursued in terms of finite populations rather than the more appropriate general distributional formulation used here – see, for example, Hadar and Russell (1969) Saposnik (1981).

4For example some would regard the definition of quantiles as being indeterminate in the case of “dead” intervals discussed above (Cf Kendall and Stuart (1977), p. 39-41) in which case $Q$ is potentially multivalued and is determined by $Q(F; q) = \{x : F(x) = q\};$ the revised form of weak dominance is then $G \succeq Q F$ if $\forall q : \sup Q(G; q) \geq \inf Q(F; q)$, and the corresponding
3.2.2 The second-order dominance criterion

It can be argued that first-order welfare criteria are usually not satisfactory on their own. In practice one often finds the indecisive conclusion $F \perp_Q G$ and first-order criteria do not encode a principle that some would claim is basic to welfare comparisons of income distribution, the principle of transfers. For this reason it is commonly considered desirable to invoke second-order dominance criteria. Using (7), the family of statistics $\{C(\cdot; q) : q \in [0, 1]\}$ characterizes the generalized Lorenz curve (GLC) and induces the following principle:

**Definition 5**

(a) (C-ranking) For any $F, G \in \mathfrak{F}$, $G \succeq_C F$ if and only if $\forall q \in [0, 1] : C(G; q) \geq C(F; q)$. (b) (Strict C-ranking) For any $F, G \in \mathfrak{F}$, $G \succ_C F$ if and only if $G \succeq_C F$ and $\exists q_0 \in (0, 1] : C(G; q_0) > C(F; q_0)$.

A standard result states that $G \succeq_C F$ if and only if welfare in distribution $G$ is at least as great as that in distribution $F$ for all decomposable SWFs that are monotonic increasing and concave in income (Shorrocks 1983). By analogy with the first-order case considered in 3.2.1, 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 (7).

**Theorem 4** The second-order dominance relation $\succeq_C$ and $\succ_C$ are non-robust.

**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

$$\mu \left( F_{\varepsilon}^{(z)} \right) = [1 - \varepsilon] \mu(F) + \varepsilon z \quad (14)$$

If $X$ 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

$$\lim_{\varepsilon \to 0} \lim_{z \to \infty} \mu \left( F_{\varepsilon}^{(z)} \right) > \mu(G) > \mu(F) \quad (15)$$

where the second inequality in (15) follows from $G \succ_C F$. But (15) implies $\lim_{\varepsilon \to 0} \lim_{z \to \infty} F_{\varepsilon}^{(z)} \perp_C G$. ■

Note the fundamental difference between this and the first-order case. As (15) shows, here it is not true that $\forall q, \lim_{\varepsilon \to 0} C(F_{\varepsilon}^{(z)}; q) = C(F; q)$; contrast this with (11). The case used in the proof is shown in Figure 3 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 $\mu(F_{\varepsilon}^{(z)}) > \mu(G)$. concept of strict dominance is $G \succ_Q F$ iff $G \succeq_Q F$ and $\exists q : \inf Q(G; q) > \sup Q(F; q)$. 


3.2.3 Other second-order criteria

It might be remarked that there are other second-order dominance criteria that are frequently applied empirically, and that the properties of these in the face of data contamination should be also considered. In this regard two commonly-used tools – based on $C$ – are the relative Lorenz curve (RLC) and the absolute Lorenz curve (ALC). The typical ordinate of the RLC – the ordinary Lorenz curve – is just a standardized version of (7) given by

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

(16)

The typical ordinate of the ALC (Moyes 1987) is given by

$$A(F; q) = C(F; q) - \mu(F) \cdot q.$$  

(17)

However, the behaviour discussed in 3.1.2 and 3.2.2 is inherited by these distributional tools:

**Theorem 5** All RLC and ALC ordinates are non-robust.

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

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

(18)
\[ A(F^{(z)}; q) = C(F^{(z)}; q) - \mu(F^{(z)}) \cdot q. \] (19)

and the IFs are given by

\[
\text{IF}(z; L(\cdot; q), F) = \frac{1}{\mu(F)^2} \left[ \text{IF}(z; C(\cdot; q), F)\mu(F) - \text{IF}(z; \mu, F)C(F; q) \right]
\]

\[
= \frac{Q(F; q)(q - \iota(q \geq F(z)))}{\mu(F)} + z\frac{\mu(F)\iota(q \geq F(z)) - C(F; q)}{\mu(F)^2} \tag{20}
\]

\[
\text{IF}(z; A(\cdot; q), F) = \text{IF}(z; C(\cdot; q), F) - \text{IF}(z; \mu, F)q
\]

\[
= qQ(F; q) - C(F; q) + \iota(q \geq F(z))[z - Q(F; q)] + \mu(F) - z \right) q. \tag{21}
\]

where the second line in (20) follows from the fact that \(\text{IF}(z; \mu, F) = z - \mu(F)\).

Given that the last line in each of the expressions (20) and (21) 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 2 – 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.

3.2.4 Higher-order dominance criteria

It has been argued in the welfare literature that 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 (Fishburn and Willig 1984, Kolm 1976, Shorrocks and Foster 1987). 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.

3.3 Implications

Second-order ranking principles can have unbounded influence functions, 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 3 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

\[
x^F := (x_1^F, x_2^F, \ldots, x_n^F)
\]

and

\[
x^G := (x_1^G - \gamma, x_2^G, \ldots, x_n^G + \gamma), \quad \gamma > 0
\]

(\( x_i^F \) is the \( i \)th 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 \( x_{[n]}^F \). Then the corresponding income distribution becomes \( \hat{F} \) where

\[
x^\hat{F} := (x_1^F - \delta, \ldots, x_{[n-1]}^F - \delta, x_{[n]}^F + (n-1)\delta).
\]

By imposing suitable mild conditions on \( \gamma, \delta \) and \( n \), it is easy to show that after the lottery has been introduced, \( G \succ_C \hat{F} \).\(^5\) 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 3 – and so it makes sense to consider them together. There are three ways of looking at the GLC phenomenon in Figure 3:

1. The point mass \( z \) really belongs to the distribution and should be encoded 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. One should at least be aware that a solitary data point can mask the behaviour of the others. Note that standard procedures for handling sampling variability – confidence intervals for LC ordinates – will not help in eliminating the problem induced by the presence of \( z \).

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

\(^5\)As an example consider the income vectors

\[
x^F = (50, 100, 100, 100, \ldots, 100, 150)
\]

\[
x^G = (49, 100, 100, 100, \ldots, 100, 151)
\]

\[
x^\hat{F} = (49, 99, 99, 99, \ldots, 99, 150 + [n - 1]).
\]

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).
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) go beyond the scope of the present paper.

4 Conclusion

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 between the theoretical and the applied approaches to the distributional ranking there is a link that has been either missing or relatively neglected in many treatments of the subject. Neglect of this link can lead to pitfalls in distributional analysis.

The story on the statistics used for welfare rankings falls into three parts:

1. Apart from special cases in dead intervals and the tails “first-order” distributional statistics – the quantiles – are robust.

2. The quantile ranking is robust apart from the exceptional case when the distributions being compared have overlapping dead intervals.

3. Second-order statistics and stochastic dominance results are non-robust: contamination can seriously affect welfare conclusions when extreme values are present in the data: small amounts of data contamination in the wrong place can even reverse unambiguous welfare conclusions.

Two final remarks about these conclusions. First, the impact of contamination is only one of several potentially important statistical issues involved in comparing income distributions: one should also take sampling variability into account when making judgments on welfare comparisons. However, estimates of sampling variability are unfortunately not robust (see e.g. Cowell and Victoria-Feser 2001b) and therefore cannot be used to alleviate the problems induced by
data contamination. Second, 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. There are two potentially useful approaches: one based on statistics that automatically remove potentially troublesome outliers from the sample data and the other relying on the specification of a parametric model for the distribution and using robust estimators of the parameters. The first approach is based on the concept of trimmed Lorenz curves and its practical implementation can be found in Cowell and Victoria-Feser (2001a). For the parametric approach one can either consider parametric Lorenz curves computed though (robust) estimates of income distribution models or semi-parametric Lorenz curves for which the upper tail of the income distribution is (robustly) estimated using an appropriate model such as the Pareto distribution. Details can be found in Cowell and Victoria-Feser (2001b).

References


