Econometrics Journal (1999), volume **2**, pp. 49–75.

Inference for Lorenz curve orderings

VALENTINO DARDANONI[1], ANTONIO FORCINA[2]

[1]*Istituto di Scienze Finanziarie, Facoltà di Economia, Università di Palermo, 90132 Palermo, Italy*
E-mail: vdardano@unipa.it
[2]*Dipartimento di Scienze Statistiche, Facoltà di Economia, Università di Perugia, 06100 Perugia, Italy*
E-mail: forcina@fisher.stat.unipg.it

Received: December 1997

Summary In this paper we consider the issue of performing statistical inference for Lorenz curve orderings. This involves testing for an ordered relationship in a multivariate context and making comparisons among more than two population distributions. Our approach is to frame the hypotheses of interest as sets of linear inequality constraints on the vector of Lorenz curve ordinates, and apply order-restricted statistical inference to derive test statistics and their sampling distributions. We go on to relate our results to others which have appeared in recent literature, and use Monte Carlo analysis to highlight their respective properties and comparative performances. Finally, we discuss in general terms the issue and problems of framing hypotheses, and testing them, in the context of the study of income inequality, and suggest ways in which the distributional analyst could best proceed, illustrating with empirical examples.

Keywords: *Lorenz curve, Testing inequalities, Likelihood inference, Income inequality.*

1. INTRODUCTION

Suppose you were to draw 1000 observations from a population distribution, say a lognormal or Singh–Maddala, and then plot the decile Lorenz curve from this sample. Suppose you then draw another sample of 1000 observations *from the same population distribution*, and compare the two sample Lorenz curves. Suppose you then repeat the exercise many times. In what proportion of cases do you think you would find Lorenz dominance, of one sample curve over the other, even though both have come from the same population distribution? 5% of cases? 10%? The answer can be as high as 50% of cases.[1] This demonstrates the need for statistical inference procedures — and if we need them to test for equality of two empirical Lorenz curves, then, *a fortiori*, we also need them to test the hypothesis that one empirical curve comes from a Lorenz-dominant population, or that the underlying population Lorenz curves intersect.

This paper concerns: how to set up the null hypothesis, and the alternative(s), and how to say with confidence that, for example, inequality increased steadily in the 1980s in the USA (or that it did not). Namely, we shall consider three types of hypothesis: (1) that the Lorenz curves

[1]Details will be given later. This result, which may seem quite surprising at first sight, is simply due to the fact that empirical Lorenz curve ordinates are typically strongly positively correlated.

(of the underlying populations) are equal; (2) that there is a chain of dominance, of one curve over another, and the other over the next, and so on (allowing explicitly for multiple populations throughout); and (3) that the Lorenz curves are unrestricted. The relevant testing procedures will be described, and contrasted with related results already in the literature. Next, we summarize the results of some Monte Carlo experiments designed to compare size and power properties, and address the questions arising for the practitioner: which test or battery of tests should be engaged, and in what contexts? Finally, we give two illustrative applications, one of which compares US family incomes at two-year intervals throughout the 1980s, and another which compares the income inequality in two Italian regions. The final section contains concluding remarks and some practical advice for the applied researcher.

Since we wrote the first draft of this paper, two very valuable contributions of Davidson and Duclos (1997, 1998) have appeared. These authors significantly extend the seminal contribution of Beach and Davidson (1983) which derived the asymptotic distribution of the Lorenz curve sample ordinates. Davidson and Duclos derive the asymptotic sample distribution of quantile-based estimators needed for both the measurement of progressivity, redistribution and horizontal equity, and the ranking of distributions in terms of poverty, inequality and stochastic orders of any order. These results can obviously be applied in conjunction with the methods discussed in this paper: the researcher has simply to reshape the ordering of interest as a multivariate normal model subject to linear inequality constraints, and then apply the results presented in our paper to investigate these more general contexts.

2. NOTATION AND OVERVIEW

Let the income earned by individuals or families be a random variable Y, whose properties within a given population are described by the cumulative distribution function (cdf) $P(y)$; this function we assume to be strictly monotone and twice continuously differentiable. For any $p \in (0, 1)$, this implies the existence of the *quantile function* $Q(p) = P^{-1}(p)$; the corresponding Lorenz curve ordinate is defined as

$$\theta(p) = \mathrm{E}\{Y \mid Y \leq Q(p)\}/\mathrm{E}(Y).$$

Given a sample of size n, let $Y_{(j)}$ denote the jth order statistic; sample estimates of $Q(p)$ and $\theta(p)$ will be defined in the usual way

$$\hat{Q}(p) = Y_{(s)}, \qquad \text{where } s - 1 < np \leq s$$
$$\hat{\theta}(p) = \left\{\sum_{1}^{s} Y_{(j)}\right\} \Big/ (s\bar{Y})$$

where $\bar{Y}$ is the sample average.

Assume now that we want to compare m separate populations by taking independent samples, the one coming from the ith population having size n_i, $i = 1, \ldots, m$. Given a set of k population fractions $0 < p_1 < \cdots < p_k < 1$, write θ_{ij} for $\theta_i(p_j)$, that is, the Lorenz curve ordinate corresponding to the jth fraction in the ith population. We also write $\boldsymbol{\theta}_i$ for the $k \times 1$ vector of Lorenz curve ordinates in the ith population with $\hat{\boldsymbol{\theta}}_{\mathbf{i}}$ being the corresponding vector of sample estimates.

Beach and Davidson (1983) show that, under the above monotonicity and differentiability assumptions, the vector $\sqrt{n_i}(\hat{\boldsymbol{\theta}}_{\mathbf{i}} - \boldsymbol{\theta}_i)$ has an asymptotically joint normal distribution with covariance matrix, say, $\boldsymbol{\Sigma}_i$ whose entries are functions of quantiles, conditional expectations and variances.[2] They also show that a consistent estimate of $\boldsymbol{\Sigma}_i$ can be obtained by replacing the population quantiles, expectations and variances by the corresponding sample estimates (Beach and Davidson1983, p. 729). This is essentially a distribution-free result from which they derive a simple Wald statistic to test the hypothesis that two populations have the same Lorenz curve; the fact that such a hypothesis can be safely rejected may provide strong evidence about the conjecture that the structure of income (or wealth) inequality is essentially different in the two contexts.

However, in many cases the applied researcher in the field may be more interested in knowing whether one Lorenz curve lies above or below another, and may also want to compare simultaneously more than two populations, to establish that, for instance, due to certain economic policies, inequality has steadily increased during a sequence of years. In other circumstances economic theory may lead to expect a more complex pattern of change in the inequality of income or wealth so that the various Lorenz curves should have one or more points of intersection. One way of putting this more formally would be to say that we are faced with *three hypotheses*: H_0, which specifies that the m Lorenz curves of interest are equal; H_1, which specifies that the m Lorenz curves are completely ordered in a given direction, and H_2, which indicates that the m Lorenz curves are unrestricted.

We are interested in distribution-free inferential techniques for testing these hypotheses. These will be derived by using results from the theory of order restricted statistical inference, after reframing H_1 as a hypothesis concerning a set of inequality constraints on the parameters of interest, and put into the framework of distance tests. An alternative procedure will be derived from the idea of multiple comparison suitably adjusted.

3. HYPOTHESES TESTING

For the chosen set of proportions $0 < p_1 < \cdots < p_k < 1$, define first a discretized version of the partial order induced by comparison of two Lorenz curves. Given two vectors of Lorenz curve ordinates, say $\boldsymbol{\theta}_h$ and $\boldsymbol{\theta}_i$, we will say that $\boldsymbol{\theta}_h$ *Lorenz dominates* $\boldsymbol{\theta}_i$, written $\boldsymbol{\theta}_h \succeq \boldsymbol{\theta}_i$, whenever $\theta_{hj} \geq \theta_{ij}$ for $j = 1, \ldots, k$. Now, given an ordered set of populations, say $P_1, \ldots, P_m$, define:

- H_0: $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_m$, meaning that the m Lorenz curves are indistinguishable at the level of discretization adopted. It is important to notice that H_0 is a *composite* hypothesis, because it does not specify the true value of the common Lorenz curve ordinates. Thus, we assume the existence of a common true (up to an irrelevant scale parameter) income cdf, say $P_0 = P_1 = \cdots = P_m$;
- H_1: $\boldsymbol{\theta}_1 \succeq \cdots \succeq \boldsymbol{\theta}_m$, meaning that the m Lorenz curves are completely ordered;
- H_2 to indicate that the m Lorenz curves are unrestricted.[3]

[2] Convergence to normality holds for a fixed set of k population fractions when each n_i goes to infinity. This means that, whenever each n_i is not too small relative to k, one can expect the actual distribution to be reasonably close to normality. However, having both a finer discretization of the Lorenz curves (a larger k) and a greater accuracy in the normal approximation are conflicting requirements.

[3] It is worth noting that a convenient feature of this problem is that the hypothesis H_2 is always true. This implies that in our testing procedures we do not have to worry about the potential problem in the case that neither the null nor the alternative is true.

Sometimes it is convenient to work with a slightly different version of the same problem, so let $H_{1-0} = H_1 - H_0$ denote the hypothesis that at least one of the inequalities defining the partial order is strict and $H_{2-1} = H_2 - H_1$ the hypothesis that at least one of the same inequalities is not true, in which case we say that the curves *cross* (though they might actually be ordered in the direction opposite to the one specified by H_1). Our strategy will be to transform these hypotheses into systems of linear inequalities, and use results from the literature on statistical inference for testing multivariate inequalities.[4]

Let $\boldsymbol{\theta}$ be the $mk \times 1$ vector obtained by stacking the vectors $\boldsymbol{\theta}_i$ one below the other, and let the matrix $\mathbf{D}$ be a $(m-1) \times m$ *difference* matrix having 1 on the main diagonal, -1 on the diagonal above and 0 elsewhere. Let the $(m-1)k \times mk$ matrix $\mathbf{R} = (\mathbf{D} \otimes \mathbf{I}_k)$, with $\mathbf{I}_k$ being the k-dimensional identity matrix and $\otimes$ the Kronecker product. Define the parameter vector $\boldsymbol{\beta} \in \mathcal{R}^v$, with $v := (m-1)k$, as:

$$\boldsymbol{\beta} = \mathbf{R}\boldsymbol{\theta}.$$

The various hypotheses of interest can be written in terms of linear inequalities involving $\boldsymbol{\beta}$ and, from now on, they will also denote the corresponding parameter sets so that $H_0 = (\boldsymbol{\beta} : \boldsymbol{\beta} = \mathbf{0})$, $H_1 = (\boldsymbol{\beta} : \boldsymbol{\beta} \geq \mathbf{0}) = \mathcal{R}^v_+$, $H_2 = (\boldsymbol{\beta} \in \mathcal{R}^v)$, $H_{1-0} = (\boldsymbol{\beta} : \boldsymbol{\beta} \geq \mathbf{0}, \boldsymbol{\beta} \neq \mathbf{0})$, and $H_{2-1} = (\boldsymbol{\beta} : \min(\boldsymbol{\beta}) < 0)$.

Let $n = \sum n_i$ be the overall sample size and let $\mathbf{r} = (r_1, r_2, \ldots, r_m)$ be a $1 \times m$ vector such that $r_i = n_i/n$ is the relative size of the ith sample. The various hypotheses of interest specify restrictions on the mean of an asymptotically multivariate normal variable $\sqrt{n}\hat{\boldsymbol{\beta}} = \sqrt{n}\mathbf{R}\hat{\boldsymbol{\theta}} \sim N(\sqrt{n}\boldsymbol{\beta}, \boldsymbol{\Omega})$, where $\hat{\boldsymbol{\theta}}$ denotes the sample estimate of $\boldsymbol{\theta}$, and

$$\boldsymbol{\Omega} = \mathbf{R}\mathrm{diag}(\boldsymbol{\Sigma}_1/r_1, \ldots \boldsymbol{\Sigma}_m/r_m)\mathbf{R}'. \quad (1)$$

3.1. Tests based on distance statistics

Let $d(\mathbf{x}, S, \mathbf{V})$ denote the distance between a vector $\mathbf{x} \in \mathcal{R}^v$ and a set $S \subset \mathcal{R}^v$ in the metric of a $v \times v$ positive definite matrix $\mathbf{V}$,

$$d(\mathbf{x}, S, \mathbf{V}) = \inf_{\mathbf{y} \in S} (\mathbf{x} - \mathbf{y})' \mathbf{V}^{-1} (\mathbf{x} - \mathbf{y}).$$

In general terms, if $H_n,\ H_a \subset \mathcal{R}^v$ denote the parameter spaces under the null and alternative hypotheses respectively, the distance statistic D_{na} can be written as:

$$D_{na} = n\{d(\hat{\boldsymbol{\beta}}, H_n, \boldsymbol{\Omega}) - d(\hat{\boldsymbol{\beta}}, H_n \cup H_a, \boldsymbol{\Omega})\}$$

and is equal to zero whenever the sample value of $\boldsymbol{\beta}$ belongs to the null hypothesis. A sensible test procedure is then to reject the null hypothesis for large values of D_{na}.

The distance statistic for testing H_0 (equality) against H_2 (the unrestricted alternative) is equal to the usual chi-squared statistic

$$D_{02} = n\hat{\boldsymbol{\beta}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\beta}}.$$

[4]The subject of testing the hypothesis concerning a set of inequalities goes also under the titles of *order-restricted inference, one-sided testing, isotonic regression*; for a detailed introduction see Barlow *et al.* (1972,) and Robertson, Wright and Dykstra (1988). Most of the results needed in this paper are contained in Kodde and Palm (1986) and Wolak (1989).

The *one-sided* distance statistics for testing H_0 (equality) against H_1 (Lorenz ordering) and H_1 against H_2 (the unrestricted alternative) are easily derived from the definition above and are respectively equal to

$$D_{01} = n\hat{\boldsymbol{\beta}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\beta}} - \min_{\mathbf{y}\geq\mathbf{0}} n(\hat{\boldsymbol{\beta}} - \mathbf{y})'\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\beta}} - \mathbf{y}),$$
$$D_{12} = \min_{\mathbf{y}\geq\mathbf{0}} n(\hat{\boldsymbol{\beta}} - \mathbf{y})'\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\beta}} - \mathbf{y}).$$

Note that these statistics satisfy the equation $D_{02} = D_{01} + D_{12}$, and there are several efficient algorithms to solve these quadratic programming problems; see e.g. Dykstra (1983) or Goldman and Ruud (1993) and references therein.

In our context, however, $\boldsymbol{\Omega}$ is unknown and has to be replaced by a consistent estimate. Because the distributional results that we will be using (see Lemmas 1–3 below) hold under H_0, we replace $\boldsymbol{\Omega}$ with $\hat{\boldsymbol{\Omega}}_o$ where $\boldsymbol{\Omega}_o$ denotes the variance matrix of $\sqrt{n}\boldsymbol{\beta}$ under the true *cdf* P_0, and is obtained by letting $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_o$, say, in equation (1). A reasonable procedure to obtain a consistent estimate of $\boldsymbol{\Sigma}_o$ is as follows. Divide incomes within each sample by the corresponding arithmetic mean to remove possible scale effects, and then pool all samples together. An asymptotically equivalent estimate may be obtained by replacing each $\boldsymbol{\Sigma}_i$ in (1) with the corresponding estimate from the ith sample. This version of D_{02} for comparing two populations is exactly the statistic used by Beach and Davidson (1983, p. 731).[5]

The problem of testing H_0 against the alternative H_1 would make sense for example when economic theory predicts that the population Lorenz curves of interest should be ordered (due to, say, changes in taxation and benefit policies), so that the crossing of the Lorenz curves (i.e. H_{2-1}) is an unlikely event. Then we can effectively restrict the parameter space to H_1 and take a large value of D_{01} as reliable evidence to support the assumed population ordering.

In order to derive the asymptotic distribution of D_{01} under H_0, we recall that H_0 is a *composite* hypothesis. Because of the functional dependence of $\boldsymbol{\Sigma}_o$ on P_0, it follows that under H_0 there are infinitely many chi-bar-squared distributions which can describe the asymptotic distribution of D_{01}, one for each hypothesized value of P_0. In particular, as we show in the lemma below, the asymptotic distribution of D_{01} will depend on $\boldsymbol{\Omega}_o$. In the same lemma we also give the *least favourable distribution* of the test statistic, that is the distribution with the smallest rejection region. This is the distribution required in the formal approach to testing composite hypotheses (see, e.g. Lehmann, 1988).

Lemma 1. *Under the assumption that as n increases r_i, $i = 1, \ldots, m$, converges to a positive constant, the asymptotic distribution of D_{01} is such that, for any $x \geq 0$,*

$$\Pr(D_{01} > x \mid P_0) = \sum_{h=0}^{v} w_h(\boldsymbol{\Omega}_o)C_h(x) \leq \sum_{h=0}^{m-1} w_h\{\mathbf{D}\mathrm{diag}(\mathbf{r})^{-1}\mathbf{D}'\}C_{h+v-m+1}(x),$$

where $w_h(\boldsymbol{\Omega}_o)$ and $w_h\{\mathbf{D}\mathrm{diag}(\mathbf{r})^{-1}\mathbf{D}'\}$ are nonnegative probability weights that sum to 1 and $C_h(x)$ denotes the probability that a χ^2 with h degrees of freedom is greater than x.

Proof. The first equality follows trivially from the application of known results (see, for instance, Shapiro, 1988); in the interest of completeness, we provide a sketch of the proof in the Appendix.

[5] This is equivalent to estimate $\boldsymbol{\Omega}$ under H_2; the corresponding estimate under H_1 is a very difficult problem, beyond the scope of this paper.

In the Appendix we prove also the inequality defining the least favourable distribution, which essentially follows when the Lorenz vector ordinates under H_0 are perfectly correlated.

The result above states that the limiting distribution of the test statistic D_{01} is like that of a *chi-bar-squared* random variable, denoted $\bar{\chi}^2$. A chi-bar-squared random variable is essentially a mixture of chi-squared variables, where the mixing weight $w_h(\mathbf{V})$ is the probability that the projection of a $N(\mathbf{0}, \mathbf{V})$ onto the positive orthant has exactly h positive elements. The basic properties of this distribution are defined in the appendix; for a detailed treatment see Shapiro (1988) and Robertson, Wright and Dykstra (1988).

A useful result shown in the Appendix is that the weights $w_j\{\mathbf{D}\text{diag}(\mathbf{r})^{-1}\mathbf{D}'\}$, $j = 0, \ldots, m-1$, which determine the asymptotic least favourable distribution of D_{01}, are the same as the *level probabilities* used in the context of order-restricted inference (see Robertson, Wright and Dykstra (1988)). In the special case when the sample sizes n_i are equal in the m populations (*uniform margins*), the resulting weights may be easily calculated for an arbitrary number of populations by a recursive formula discussed, e.g. by Robertson, Wright and Dykstra (1988, p. 82); this allows an easy computation of the upper bound distribution.

In order to approach the least favourable distribution, P_0 should be such that the elements of $\hat{\boldsymbol{\theta}}_\mathbf{i}$ are almost perfectly correlated. Intuitively, this would happen when the variability of incomes within a given interval $[a, b]$ is negligible relative to the variability of incomes less then a, or equivalently if the density function increases exponentially at an extremely high rate. A distribution of this type is so unusual that in many practical cases there is sufficient evidence against using such a distribution. Under the least favourable distribution, however, computation of the probability weights is much easier so that one could start by computing the related upper bound to the critical value for this test and then, if the sample value of D_{01} is above this bound, the null can be safely rejected without further computations. In the case where the value of the test statistic falls below this critical value, an asymptotically correct procedure is then to compute the critical values using the weights $w_i(\hat{\boldsymbol{\Omega}}_o)$. Although exact formulae for the calculation of these weights exist only for $v \leq 4$, when $v > 4$, they can be estimated by projecting a reasonable number of $N(\mathbf{0}, \hat{\boldsymbol{\Omega}}_o)$ random vectors onto $\mathcal{R}_+^v$, and counting the proportion of times where exactly i coordinates of these projections are positive.

3.2. Joint distribution of D_{01} and D_{12}

It is actually possible to derive the joint asymptotic distribution of the two test statistics D_{01} and D_{12} under H_0 with little effort as in the following lemma.

Lemma 2. *Under the assumption that as n increases r_i, $i = 1, \ldots, m$, converges to a positive constant and that H_0 holds, it follows that the asymptotic joint distribution of D_{01} and D_{12} is such that, for any $x \geq 0$ and $y \geq 0$,*

$$\Pr(D_{01} > x, D_{12} > y \mid P_0) = \sum_0^v w_h(\boldsymbol{\Omega}_o) C_h(x) C_{v-h}(y).$$

Proof. This result also follows trivially from application of known results on the chi-bar-squared distribution. For the sake of completeness, however, we prove it in the Appendix.

Denote now by $F(x, y)$ the joint distribution function of of D_{01} and D_{12} under P_0 and by $G(x, y)$ the corresponding survival function, i.e.

$$G(x, y) = \Pr(D_{01} > x, D_{12} > y \mid P_0)$$

and recall the following identity which is useful below

$$F(x, y) = 1 - G(x, \infty) - G(\infty, y) + G(x, y),$$

where $G(x, \infty)$ and $G(\infty, y)$ denote respectively the two marginals of $G(x, y)$.

We now propose to use the joint distribution of D_{01} and D_{12} determined above in the construction of a testing procedure that partitions the sample space into three disjoint regions so that the error probabilities of rejecting H_0 in favour of H_{1-0} or of H_{2-1} may be controlled simultaneously. Our testing procedure is defined as follows:

- accept H_0 if $D_{01} \leq x$ and $D_{12} \leq y_1$ where $F(x, y_1) = 1 - \alpha_1 - \alpha_2$;
- reject H_0 in favour of H_{1-0} if $D_{01} > x$ and $D_{12} \leq y_0$, where $F(\infty, y_0) - F(x, y_0) = \alpha_1$;
- reject H_0 in favour of H_{2-1} when $D_{12} > y_1$ or when $D_{12} > y_0$ and $D_{01} > x$, with $1 - F(x, y_1) - F(\infty, y_0) + F(x, y_0) = \alpha_2$.

This procedure permits the setting of a different size for the error probabilities of rejecting H_0 in favour of H_{1-0} (α_1) and H_{2-1} (α_2). Moreover, to reduce rejection in favour of H_{1-0} in areas of the sample space outside $\mathcal{R}_+^v$ and far away from the origin, where acceptance of H_0 is ruled out, we allow the critical value y_0 to be much smaller than y_1. The procedure based on a value of y_0 as small as possible may be useful if one feels that, because the assumption of Lorenz dominance has a constructive content, it should be submitted to a stricter scrutiny. The price for this choice is the somehow *odd shape* of the rejection region towards H_{2-1}. This shape makes it possible that, if D_{01} increases while D_{12} remains constant or even decreases slightly, initial acceptance of H_0 changes into rejection towards H_{2-1} rather than H_{1-0}. Clearly this may be justified only if we consider acceptance of H_0 or rejection towards H_{2-1} to be much safer statements.

To take a conservative approach one should search for a P_0 which is *least favourable* in the sense that, under the corresponding joint distribution under H_0, the probability of rejecting in favour of H_{1-0} is minimized while that in favour of H_{2-1} is maximized. Note that this application of the idea of a least favourable distribution stems from the desire to be protected against rejecting in favour of H_{1-0} when it is not true. Alternatively, one could search for the greatest protection towards rejecting in favour of H_{2-1} when false, but this route is not pursued here.

Consider then the joint asymptotic survival function of D_{01} and D_{12} under H_0 obtained using the 'least favourable' weights' vector employed in Lemma 2:

$$G^+(x, y) = \sum_{h=0}^{m-1} w_h\{\mathbf{D}\mathrm{diag}(\mathbf{r})^{-1}\mathbf{D}'\}C_{h+v-m+1}(x)C_{m-1-h}(y).$$

The next lemma states that, under H_0, G^+ is also *least favourable* for the joint testing procedure.

Lemma 3. *For any* $x \geq 0$ *and* $y \geq 0$, *we have*

$$\Pr(D_{01} > x, D_{12} \leq y \mid H_0) \leq \Pr(D_{01} > x, D_{12} \leq y \mid G^+)$$

and

$$\Pr(D_{01} \leq x, D_{12} > y \mid H_0) \geq \Pr(D_{01} \leq x, D_{12} > y \mid G^+)$$

Proof. See the Appendix. Note that because the weights' vector involved in the computation of $G^+(x, y)$ is easily computable, this lemma allows simple calculation of conservative upper bounds to the joint test procedure.

3.3. Using crossing Lorenz curves as the null hypothesis

On the other hand, as mentioned before, the only way to effectively reduce the advantage given to H_{1-0} is to take the hypothesis H_{1^*} that $\min(\boldsymbol{\beta}) > 0$ as the *alternative*, and its complement H_{2-1^*}, that is $\min(\boldsymbol{\beta}) \leq 0$, as the *null*. This means that the researcher is not prepared to engage in a statement that the Lorenz curves are strictly ordered unless there is strong evidence in its favour. If we denote with $H_{0+} = \{\boldsymbol{\beta} : \min(\boldsymbol{\beta}) = 0\}$ the *boundary* of the null, D_{21}, the distance statistic for testing H_{2-1^*} against the alternative H_{1^*}, is equal to 0, unless $\hat{\boldsymbol{\beta}} > \mathbf{0}$, in which case

$$D_{21} = \inf_{\mathbf{y} \in H_{0+}} (\hat{\boldsymbol{\beta}} - \mathbf{y})' \boldsymbol{\Omega}^{-1} (\hat{\boldsymbol{\beta}} - \mathbf{y}).$$

Sasabuchi (1980, Lemma 3.1) shows that, for $\hat{\boldsymbol{\beta}} > \mathbf{0}$, the element of H_{0+} that minimizes D_{21} is actually among the projections of $\hat{\boldsymbol{\beta}}$ on each of the v subspaces $\{\boldsymbol{\beta} : \beta_i = 0\}$, $i = 1, \ldots, v$, though, clearly, some of these projections may not belong to H_{0+}. By exploiting the simple form of the projection onto a linear subspace, and under the assumption of a multivariate normal distribution, Sasabuchi shows (Theorem 3.1) that the critical region under the least favourable distribution has a surprisingly simple solution. Under asymptotic normality of $\hat{\boldsymbol{\beta}}$ his results apply immediately to our problem and we have:

Lemma 4. *Given the null hypothesis H_{2-1^*} and the alternative H_{1^*}, the critical region derived from the D_{21} statistic reduces to:*

$$\text{reject if } \hat{z}_i > z_\alpha, \forall\, i,$$

where $\hat{z}_i = \sqrt{n}\hat{\beta}_i / \sqrt{\hat{\Omega}_{oii}}$, and z_α denotes the α% critical value from the standard $N(0, 1)$ distribution.

Intuitively, because the *null* allows us to assign for the true $\boldsymbol{\beta}$ any value outside H_{1^*}, we should search for a critical region having rejection probability not greater than α for any choice of $\boldsymbol{\beta}$ along the boundary. Consider then any of the v extreme null hypotheses such that $\beta_i = 0$ for $i = j$ and $\beta_i > 0$ for $i \neq j$. It easily follows that $\Pr(\hat{z}_i > z_\alpha)$ equals α for $i = j$ and converges to 1 otherwise. Thus the test achieves the desired asymptotic size whenever the set of population Lorenz curves are strictly ordered, except for two curves that touch only at one point.

This test will have low power against alternatives of the form $z_i = \epsilon$, $i = 1, \ldots, v$, where ϵ is a suitably small quantity. This is a consequence of the fact that the actual rejection probability of the test is typically much smaller than its nominal level at H_0: for example, if the $\hat{\beta}_i$'s were independent the test would have a rejection probability equal to α^v, which for reasonably large v is a very small number for all conventional choices of α. Of course, with a composite null hypothesis, any test procedure will often have particular points with extremely low rejection probability. However, it is of concern that a case of special interest like H_0 is among such points.

Finally it is interesting to note that the rejection region defined by $\hat{z}_i > z_\alpha, \forall\, i$ coincides with that stemming from the application of the so-called *intersection-union* principle: according to this approach, which has been implemented by Howes (1994) for comparing two Lorenz curves, the null hypotheses can be written as the union of the v subhypothesis $\beta_i \leq 0$, while the alternative can be written as the intersection of the v subhypotheses $\beta_i > 0$. A result of Berger (1982) then states that the test that takes as the rejection region the intersection of rejection regions of the α-sized test for each of the v subhypotheses has under H_{2-1^*} overall rejection probability not greater than α.

3.4. The multiple comparison approach

The so-called *multiple comparison approach*, fruitfully employed for the case of testing for the Lorenz ordering of two populations in a series of papers by Bishop *et al.* (1991, 1992, 1994) and Beach *et al.*(1994) may be seen as a simplified version of our joint distance testing procedure. Our analysis can be adapted to extend the multiple comparison approach to the case of more than two populations, and to achieve a tighter control of some error probabilities of interest.

With the multiple comparison procedure, while the asymptotic normal distribution of $\hat{\boldsymbol{\beta}}$ under H_0 is still the starting point, one tries to make inferential statements that are correct irrespective of the correlation between different $\hat{\beta}_i$. Beach and Richmond (1985) utilize this approach to derive joint confidence intervals for the Lorenz curve ordinates and income shares; multiple comparison techniques have since been extensively employed in a series of papers investigating the Lorenz and related orderings in various settings (e.g. geographical inequality comparison (Bishop *et al.*, 1992), inequality trends in US over time (Bishop *et al.*, 1991), the effects of truncation bias (Bishop *et al.*, 1994), single-crossing Lorenz curves (Beach *et al.*, 1994), various stochastic dominance concepts (Anderson, 1996)).

The multiple comparison approach is based on deep and elegant results by Richmond (1982) and requires use of the *studentized maximum modulus* distribution. Given the asymptotic nature of the problem, in practice it can be reframed into the following procedure that is very simple to apply:

- accept H_0 if $\hat{\boldsymbol{\beta}} \in A_0 = \{\hat{\boldsymbol{\beta}} : \max(|\hat{z}_i|) \leq z_\delta\}$;
- reject H_0 in favour of H_{1-0} if $\hat{\boldsymbol{\beta}} \in A_{1-0} = \{\hat{\boldsymbol{\beta}} : \max(\hat{z}_j) > z_\delta$ and $\min(\hat{z}_i) \geq -z_\delta\}$;
- reject H_0 in favour of H_{2-1} otherwise;

where $\Pr(Z > z_\delta) = \delta$ and $(1 - 2\delta)^v = (1 - 2\alpha)$. This procedure relies on Sidak's inequality which states that the probability that a multivariate normal vector with zero mean and arbitrary correlations falls in a cube centred at the origin is always at least as large as the corresponding probability under independence (see Savin (1984) for a discussion of this and related inequalities) so that

$$\Pr(A_0 \mid H_0) = \Pr\{\max(|\hat{z}_i|\} \leq z_\delta \mid H_0) \geq \prod_1^v \Pr(|\hat{z}_i| \leq z_\delta \mid H_0) = (1 - 2\delta)^v;$$

hence the probability of erroneous rejections of H_0 is bounded above by 2α.

The consequence of the simplicity of this procedure is that, precisely because the information contained in the covariance matrix is not used, we cannot split the total error probability 2α into separate error probabilities of rejecting in favour of H_{1-0} or H_{2-1}. However we can establish that, say, $\alpha_1 = \Pr(\hat{\boldsymbol{\beta}} \in A_{1-0} \mid H_0)$ (which is the more crucial because A_{1-0} suggests a specific relationship among the m Lorenz curves) cannot be greater than α.

This follows from the symmetry of the normal distribution

$$\Pr(\hat{\boldsymbol{\beta}} \in A_{2-1} \mid H_0) = \Pr(\text{any } \hat{z}_i \leq -z_\delta \mid H_0) = \Pr(\text{any } \hat{z}_i \geq z_\delta \mid H_0) \geq \Pr\{(\text{any } \hat{z}_i \geq z_\delta) \text{ and } (\text{all } \hat{z}_i \geq -z_\delta) \mid H_0\} = \alpha_1.$$

4. ILLUSTRATION AND APPLICATIONS

4.1. Monte Carlo comparison

Properties of the power functions of the various procedures discussed in this paper could, at least in principle, be derived under the corresponding asymptotic distribution. Useful results in this direction have been derived by Goldberger (1992), who considers testing linear inequality constraints in the simple bivariate case. Goldberger's main conclusion is that $\bar{\chi}^2$-based procedures, by taking into account the covariance between the parameter estimates, can provide a clear improvement on the multiple comparison procedure mainly when these estimates are negatively correlated. It seems natural to expect that these results hold also for the general v-dimensional case. Because Lorenz curve ordinates are typically positively correlated, in many practical situations there will be quite a strong positive correlation among the elements of $\hat{\boldsymbol{\theta}}_{\mathbf{i}}$. However, given that $\boldsymbol{\Omega}_o$ is proportional to $\boldsymbol{\Sigma}_o$ when $m = 2$, while $\boldsymbol{\Omega}_o$ has blocks of negative elements for $m > 2$, we may expect that the relative performance of the various procedures should change dramatically according to whether we are comparing two or more than two populations.

Reasonably accurate estimates of power may be obtained from the relative frequency of rejection in favour of H_{1-0} and H_{2-1} in a set of suitably designed Monte Carlo experiments. Those briefly reported here are aimed to provide enough evidence about the conjecture outlined above, and also to assess how accurate the approximations are, as provided by the asymptotic distribution used to compute the critical values. Full details are contained in the Appendix.

Briefly, we drew repeated pairs of samples from Singh–Maddala (*SM*) income distributions, having selected benchmark parameter values to fit the US census data for 1980 (McDonald, 1984). To construct pairs of *SM* distributions that satisfy H_{1-0} or H_{2-1}, the results of Wilfling and Kramer (1993) may be used. Wilfling and Kramer state a simple necessary and sufficient condition for Lorenz curve dominance in terms of the two characterizing parameters of the *SM* distribution, which will be used in the following to generate samples of Lorenz curves under H_{1-0} and H_{2-1}.

The basic idea for H_{2-1} has been to construct situations with a single clear crossing of the Lorenz curves, and to use two different sets of parameters which correspond roughly to situations of low and high power (that is, corresponding to population Lorenz curves close to and far apart from each other respectively). Under H_{1-0} we also consider two different sets of parameters which guarantee that the corresponding population Lorenz curves are strictly ordered, in situations of low and high power, respectively.

In order to evaluate the relative performance of the different procedures, we recall that our inferential problem does not involve the usual dichotomy between a null and an alternative hypothesis, but three nested hypotheses. In the discussion below, when H_{1-0} is the true state of affairs, the power of each procedure will be evaluated as the probability of inferring H_{1-0}, while the error rate will be evaluated as the probability of inferring H_{2-1}. Alternatively, when H_{2-1} is the true state of affairs, the power of each procedure is the probability of inferring H_{2-1}, while the error rate is the probability of inferring H_{1-0}. Because of the asymmetric nature of the two hypotheses, and because H_{1-0} is actually the more 'assertive' one, particular attention should be paid to the relative performance of each procedure in terms of power under H_{1-0} and error rates under H_{2-1}.

For $m = 2$, the main conclusions were:

- A procedure that would infer dominance (H_{1-0}) of the first curve over the second whenever the sample estimate of the two curves were to be ordered in this direction results in extremely high error rates. For example, under H_0, almost 23% of the samples were actually ordered in the direction implied by H_1 and in about 45% of the cases the samples exhibited strict dominance in either direction.
- We tried two joint $\bar{\chi}^2$ procedures both with $\alpha_1 = \alpha_2 = 0.05$, one with $y_0 = y_1$ and the other using the smallest possible value for y_0. Of these procedures, the one with a smaller value of y_0 is superior in terms of power and of error rates when H_{2-1} is the true state of affairs. Under H_{1-0} the same procedure has again higher power, but a slightly larger error rate due to its effort to control for the advantages of H_1. On the whole the second procedure outperforms the first, and this result is confirmed by additional experiments based on different values of α_1 and α_2 which we omit reporting in detail for the sake of brevity. Hence we recommend that, after setting the relevant values for α_1 and α_2, one should always choose the smallest possible value of y_0.
- Under H_0 the rejection rates of the $\bar{\chi}^2$ procedures are very close to the nominal value 0.05, which indicates that the asymptotic approximation is reasonably accurate. On the other hand, the actual significance level for the multiple comparison procedure is well below its nominal value in both directions, and the $\bar{\chi}^2$ procedure under the least favourable distribution is quite biased in favour of H_2.
- The $\bar{\chi}^2$ procedure based on G^+ does very well when H_{2-1} is true. However, the larger power under H_{2-1} is achieved at the cost of a significantly reduced power under H_{1-0}. Nonetheless the computational simplicity of this procedure might make it quite appealing.
- The multiple comparison procedure, when considered against the best $\bar{\chi}^2$ procedure, appears to be inferior both in terms of power and error rates under H_{2-1}, while its higher power under H_{1-0} is not as substantial; hence the heavier computational burden of the $\bar{\chi}^2$ procedure may be worth the effort whenever greater precision is required.
- The best joint $\bar{\chi}^2$ procedure achieves higher power and lower error rates both under H_{1-0} and H_{2-1} than the procedure based on the marginal distribution of D_{01}, and thus is unambiguously better.
- As could have been easily predicted, the procedure based on D_{21}, because of its highly conservative nature towards H_{1-0}, brings down to negligible levels the probability of detecting erroneous evidence in favour of H_{1-0}. However, this is paid for with a large reduction in power when H_{1-0} is actually true.

As we had conjectured before, the results for the case $m = 4$ were quite different. The procedure which infers that the m true Lorenz curves are related exactly as the corresponding sample estimates now becomes much more conservative towards the ordering hypothesis, and is characterized by both low power and high error rates under H_{1-0} and high power and low error rates under H_{2-1}. The D_{21} test never rejects its null hypothesis. Thus this procedure in this case is useless as it pays no attention to sample evidence and is blindly conservative. As concerns the multiple comparison procedure, although its actual significance level remains at about half its nominal level, in this case it seems to have extremely low power in both directions. Hence we may tentatively conclude that there will be serious competitors to the joint $\bar{\chi}^2$ procedure only when the covariance matrix $\mathbf{\Omega}$ exhibits strong positive correlations, which is typically the case when $m = 2$.

4.2. A simple illustration of the rejection regions

Procedures for testing one-sided multivariate hypotheses are in general neither very well known nor treated in much detail in most textbooks in statistics and econometrics, even at the graduate level. Readers unfamiliar with the subject can achieve a direct appreciation of the properties and the functioning of these procedures by looking at plots of the rejection regions in the simple bivariate case. These are given in the Appendix. More detailed plots including the power functions are given by Goldberger (1992), who however considers only the marginal distributions of D_{01} and D_{12}. Of course, the bivariate nature of these examples means that these pictures are in general not directly applicable to our problem, but we believe that they are nevertheless very useful to appreciate the properties of these inferential procedures.

Figures B1–B6 compare rejection regions of the various procedures in the simplified situation where $\hat{\boldsymbol{\beta}}$ has dimension 2 and can be plotted on the plane. In particular, in Figures B1 and B2 we plot the rejection regions of the joint $\bar{\chi}^2$ procedure with $\alpha_1 = \alpha_2$ and y_0 set at its smallest possible value against that of the usual chi-squared D_{02} test with $\alpha = 0.10$. Figure B1 considers the case of positive correlation $(\rho = 0.77)$. This value for the correlation coefficient was chosen because it arises when comparing two benchmark (US 1980) *SM* Lorenz curves with $k = 2$ (hence $v = 2$). Any point in the positive orthant denotes a sample observation such that both empirical ordinates of the first Lorenz curve lie above those of the second, while the second and fourth quadrant denote a single crossing and the third quadrant denotes that the second curve lies above the first. Of course, a typical application would use a much higher value for k, but looking at the rejection region is instructive anyway.

In Figure B1, the boundary of the rejection region towards H_{1-0} is highlighted in darker grey. We see that the joint $\bar{\chi}^2$ procedure rejects H_0 in favour of H_{1-0} in virtually all the positive quadrant, minus the ellipsoid along the main diagonal. The boundary of the rejection region towards H_{2-1} is highlighted in a lighter shade of grey. We see that the bulk of the rejection region is contained in the second, third and fourth quadrant. For comparison, the rejection region for the usual chi-squared D_{02} test is given by the area outside the ellipsoid centred in the origin. In Figure B2 we consider the case of negative correlation, $\rho = -0.5$, which would arise for example when $m = 3, k = 1$ $(v = 2)$. Again, the darker and lighter shades of grey denote the rejection regions towards H_{1-0} and H_{2-1} respectively, and the ellipsoid centred along the main counterdiagonal defines the usual rejection region of the chi-squared test.

Recall from Section 3 that the joint $\bar{\chi}^2$ procedure may lead to reject H_0 in favour of H_{1-0} even if the sample estimates of the Lorenz curves are not actually ordered. However, the appropriate choice of y_0 prevents the corresponding rejection region from going much below the x-axis and to the left of the y-axis, under either positive or negative correlation. Finally, note that with both positive and negative correlation the $\bar{\chi}^2$ procedure accepts the null H_0 more often than the usual chi-squared in areas which are close to H_2.

In Figures B3 and B4 we compare the rejection regions of the joint chi-bar-squared and the multiple comparison procedure with $\alpha = 0.05$ for the cases of positive and negative correlation. The boundary of the rejection regions for the multiple comparison procedure are highlighted in the darkest grey (towards H_{2-1}) and the second darker shade of grey (towards H_{1-0}). Note that, precisely because the multiple comparison procedure does not use the covariance information, these rejection regions are identical in both figures. From looking at the figures, it emerges that the two procedures have substantial areas of disagreement, but in different directions according to the sign of the correlation. In particular, under positive correlation the multiple comparison procedure accepts H_0 in areas where the joint chi-bar-squared procedure rejects in favour of H_{2-1}, but also rejects in favour of H_{1-0} in areas where the joint chi-bar-squared rejects in favour of H_{2-1}. On the other hand, under negative correlation the multiple comparison procedure accepts H_0 in large areas near the origin where the joint chi-bar-squared procedure rejects in favour of H_{1-0} (in the positive orthant) or H_{2-1} (in the negative orthant), but also rejects in favour of H_{1-0} in large areas in the second and fourth quadrants were the joint chi-bar-squared rejects in favour of H_{2-1}.

Finally, in Figures B5 and B6 the darkest shade of grey highlights the boundary of the rejection region of the D_{21} procedure. It is interesting to note that while with positive correlation this rejection region has a slight intersection with the rejection region towards H_{1-0} of the joint chi-bar-squared procedure, under negative correlation it lies entirely inside it.

4.3. Illustrative examples

Income inequality in two Italian regions. In this section we compare the extent of income inequality as recorded by the Lorenz curve in two Italian regions, Veneto and Sicily. It is generally agreed that income inequality is greater in the less developed south as compared with the more developed north. We use a sample of net family incomes, contained in a survey carried out by the Bank of Italy in 1991. There are 359 families in the Venetian sample, and 476 in the Sicilian one. Sample estimates of the Lorenz ordinates and their covariance matrix are easily obtained from Beach and Davidson's (1983) results, using deciles ($k = v = 9$).

Plotting the two sample Lorenz curves, it emerges that the one for Veneto indeed lies above that for Sicily at all the decile points. This is why in this case the test statistic $D_{01} = 12.0472 = D_{02}$. Note that equality is definitely accepted against the unrestricted alternative as $\Pr(D_{02} \geq 12.0472) = 0.211$. The marginal chi-bar-squared procedure may be implemented computing the p-value as $\Pr(D_{01} > 12.0472)$. This is equal to 0.180 under the least favourable distribution and to 0.147 if the covariance matrix is estimated from the data. Finally, the p value corresponding to rejection towards H_{1-0} with the *best* joint chi-bar-squared procedure is equal to 0.073.

In this example the evidence towards the dominance hypothesis is not strong enough. However, it is worth noting that the *best* joint procedure implies a much smaller p value than the usual chi-squared procedure and this is an indirect indication of the gain in power that may be achieved by looking at the specific alternative of interest (H_{1-0}) rather than to all possible deviations from equality.

Finally, we note that in this sample $0.0656 \leq \hat{z}_i \leq 1.992$ ($i = 1, \ldots, 9$). Therefore, both the multiple comparison approach and the distance test D_{21} lead to the acceptance of their respective nulls at the 10% level.

Income inequality in the US in the 1980s. We consider the sample of US family incomes from 1979 to 1989 as contained in the CPS microdata (March tape). Because about 40% of households are in the sample in consecutive years, we use only data from alternate years. Consider then the hypothesis $H_1 : (L_{79} \succeq L_{81} \succeq L_{83} \succeq L_{85} \succeq L_{87} \succeq L_{89})$, that is, income inequality has progressively increased in the USA during the 1980s. Here $m = 6$ and we take again $k = 9$ so that $v = 45$. Each sample contains more than 40 000 observations.

The test statistics D_{01} and D_{12} are respectively equal to 588.08 and 4.61. Intuitively, the large value of D_{01} is in itself quite strong evidence against the assumption of unchanging income inequality while the relatively small value of the test statistic D_{12} seems to indicate that if the samples are not perfectly ordered, this may well be caused by random fluctuations. More formally, one can easily compute $\Pr(D_{01} \geq 588.08, D_{12} \leq 4.61 \mid G^+)$ which is less than 0.0001 and may be interpreted as the *p*-value under the least favourable distribution for the joint procedure discussed in Lemma 3 and thus provides strong evidence for increased inequality.

It is interesting to note that the lowest value of $\hat{z}_i (i = 1, \ldots, 45)$, is equal in this sample to -0.78, while its maximum is equal to 7.55. Thus, the multiple comparison approach also leads to the rejection of H_0 in favour of H_{1-0}. On the other hand, observing the sample of Lorenz curves, we note that there is some crossing among the later years (1985, 1987 and 1989). This implies that the distance test statistic D_{21} equals zero; thus, if the researcher is not prepared to accept the ordering hypothesis unless there is overwhelming evidence in its favour, the null hypothesis 'nothing can be said' will be maintained under this approach.

5. CONCLUDING REMARKS

Performing statistical inference for Lorenz curve orderings involves intricate statistical issues of testing for an ordered relationship in a multivariate context. Our approach has been to frame the hypotheses of interest as sets of linear inequality constraints on the vector of Lorenz curve ordinates, and apply order-restricted statistical inference to derive useful testing procedures.

One peculiarity of our inferential problem is that it does not involve the usual dichotomy between a null and an alternative hypothesis, but three nested hypotheses H_0, H_1 and H_2. Applying known results on the joint distribution of the statistics D_{01} and D_{12}, we proposed a new procedure which partitions the sample space into three regions, of acceptance of H_0, and rejection towards H_{1-0} and H_{2-1}. This procedure, which we called *joint chi-bar-squared*, gives the applied researcher some flexibility depending on which type of error he wants to be protected against, and on where he wishes to have greater power in detecting departures from H_0. If protection against erroneously believing that the sample Lorenz curves are ordered, when in fact they are not, is considered of primary importance, we also proposed the D_{21} testing procedure which simply takes the hypothesis of the absence of ordering as the null. Interestingly, this distance testing procedure coincides with that proposed by Howes (1994) in the context of ranking two Lorenz curves, which stems from the application of the intersection-union principle.

We also discussed the multiple comparison procedure which can be seen as a simplified version of the joint chi-bar-squared procedure. This procedure has been very popular in recent applications involving the comparison of two Lorenz curves in different contexts. We have shown how the procedure can be easily extended to consider more than two Lorenz curve at once, and how to achieve a tighter control of some probabilities of interest.

We have also performed a Monte Carlo experiment to compare the empirical performance of the three procedures, analysed their properties with the help of simple graphs of rejection regions, and given some illustrative examples of application. The general conclusions from the theoretical and empirical results appear to be the following.

- The greater difficulty of applying the joint chi-bar-squared procedure over the multiple comparison procedure seems to be worthwhile in terms of a substantially improved performance. This is really not surprising in view of the fact that the multiple comparison procedure ignores precisely the covariance information of the sample ordinates.
- The joint chi-bar-squared does not protect tightly and under all circumstances against the error of believing that the sample of Lorenz curves are ordered when actually they are not. If protection against this error is considered of overwhelming importance, then the D_{21} procedure is the obvious candidate. However, this choice could carry a price in terms of lower power in some regions of the parameter space. Thus, for example, if economic theory strongly predicts that a set of Lorenz curves should be ordered (say given some change in the taxation or benefit policy), so that the researcher could reasonably restrict the parameter space to the positive orthant, then the possible increase in power could make the joint chi-bar-squared a preferred alternative. The computational simplicity of D_{21} however could again tip the balance in its favour, at least in the case of comparing two Lorenz curves. However, when there are more than two Lorenz curves to compare, the joint chi-bar-squared procedure does seem to be the only viable alternative.

ACKNOWLEDGEMENTS

We would like to thank Peter Lambert, Neil Shephard and two anonymous referees for very helpful and detailed comments. Antonio Forcina thanks M.U.R.S.T. for financial support and Valentino Dardanoni thanks the Department of Economics at the University of California, San Diego for hospitality and M.U.R.S.T. for financial support.

REFERENCES

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica 64*, 1183–93.

Barlow, R. E., D. J. Bartholomew, J. M. Bremner and H. D. Brunk (1972). *Statistical Inference Under Order Restrictions*. New York: Wiley.

Beach, C. M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies 50*, 723–35.

Beach, C. M., R. Davidson and A. B. Slotsve (1994). Distribution-free statistical inference for inequality dominance with crossing Lorenz curves. Queen's University Discussion paper 912.

Beach, C. M. and J. Richmond (1985). Joint confidence intervals for income shares and Lorenz curves. *International Economic Review 26*, 439–50.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics 24*, 295–300.

Bishop, J. A., J.-R. Chiou, and J. P. Formby (1994). Truncation bias and the ordinal evaluation of income inequality. *Journal of Business and Economic Statistics 12*, 123–7.

Bishop, J. A., K. V. Chow and J. P. Formby (1994b). A large sample test for differences between Lorenz and concentration curves. *International Economic Review 35*, 479–88.

Bishop, J. A., J. P. Formby and W. J. Smith (1991). Lorenz dominance and welfare: Changes in the U.S. distribution of income, 1967–1986. *Review of Economic Statistics 73*, 134–9.

Bishop, J. A., J. P. Formby and P. D. Thistle (1992). Convergence of the south and non-south income distributions, 1969–79. *American Economic Review 82*, 262–72.

Dardanoni, V. and A. Forcina (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association 93*, 1112–23.

Davidson, R. and J.-Y. Duclos (1997). Statistical inference for the measurement of the incidence of taxes and transfers. *Econometrica 65*, 1453–65.

Davidson, R. and J.-Y. Duclos (1998). Statistical inference for stochastic dominance and the measurement of poverty and inequality. Université Laval, mimeo.

Dykstra, R. L. (1983). An algorithm for restricted least squares. *Journal of the American Statistical Association 78*, 837–42.

Goldberger, A. S. (1992). One-sided and inequality tests for a pair of means. In R. Bewley and T. Van Hoa (eds), *Contributions to Consumer Demand and Econometrics*, pp. 140–62, London: MacMillan.

Goldman, S. M. and P. A. Ruud (1993). Nonparametric multivariate regression subject to constraint. University of California, Berkley, mimeo.

Gourieroux, C., A. Holly and A. Monfort (1982). Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constraints in the regression parameters. *Econometrica 50*, 63–80.

Howes, S. (1994). Testing for dominance: Inferring population rankings from sample data. World Bank Discussion Paper.

Johnson, N. L. and S. Kotz (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Kodde, D. A. and F. C. Palm (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica 54*, 1243–8.

Lehmann, E. L. (1988). *Testing Statistical Hypothesis*. 2nd edn, New York: Wiley.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica 52*, 647–63.

Raubertas, R. F., C. C.-I. Lee and E. V. Nordheim (1986). Hypothesis tests for normal means constrained by linear inequalities. *Communications in Statistics–Theory and Methods 15*, 2809–33.

Richmond, J. (1982). A general method for constructing simultaneous confidence intervals. *Journal of the American Statistical Association 77*, 455–60.

Robertson, T., F. T. Wright and R. Dykstra (1988). *Order Restricted Statistical Inference*. New York: Wiley.

Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika 67*, 429–39.

Savin, N. E. (1984). Multiple hypothesis testing. In Z. Grilliches and M. D. Intrilligator (eds), *Handbook of Econometrics*, vol. 2, pp. 828–79, Amsterdam: North Holland.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistics Review 56*, 49–62.

Wilfling, B. and W. Kramer (1993). The Lorenz-ordering of Singh–Maddala income distributions. *Economic Letters 43*, 53–57.

Wolak, F. A. (1989). Testing inequality constraints in linear econometric models. *Journal of Econometrics 41*, 205–35.

A. APPENDIX. PROOF OF LEMMAS

A.1. Brief review of results on the $\bar{\chi}^2$ distribution

Let $\hat{\mathbf{y}}(\mathbf{V}, \mathcal{C})$ denote the projection of a vector $\hat{\mathbf{y}} \in \mathcal{R}^v$ onto a convex cone $\mathcal{C}$ in the $\mathbf{V}^{-1}$ metric; $\hat{\mathbf{y}}(\mathbf{V}, \mathcal{C})$ is the solution to the problem

$$\min_{\mathbf{y} \in \mathcal{C}} (\hat{\mathbf{y}} - \mathbf{y})' \mathbf{V}^{-1} (\hat{\mathbf{y}} - \mathbf{y}).$$

Under the assumption that $\hat{\mathbf{y}} \sim N(\mathbf{0}, \mathbf{V})$, the distribution of the chi-bar-squared random variable, defined as

$$\bar{\chi}^2(\mathbf{V}, \mathcal{C}) = \hat{\mathbf{y}}(\mathbf{V}, \mathcal{C})' \mathbf{V}^{-1} \hat{\mathbf{y}}(\mathbf{V}, \mathcal{C}) \tag{2}$$

is well known and depends on the cone $\mathcal{C}$ and the matrix $\mathbf{V}$; when $\mathcal{C}$ is the positive orthant, we simply write $\bar{\chi}^2(\mathbf{V})$. Also, let us define the *dual* of C in the $\mathbf{V}^{-1}$ metric as $\mathcal{C}^o = (\mathbf{y}^o\colon \mathbf{y}'\mathbf{V}^{-1}\mathbf{y}^o \leq 0, \forall\, \mathbf{y} \in \mathcal{C})$.

Convex cones share many duality properties with linear subspaces: the following is in a sense a version of Pythagoras' theorem (see e.g. Shapiro, 1988, p. 50)

$$\hat{\mathbf{y}}'\mathbf{V}^{-1}\hat{\mathbf{y}} - (\hat{\mathbf{y}} - \hat{\mathbf{y}}(\mathbf{V}, \mathcal{C}))'\mathbf{V}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{y}}(\mathbf{V}, \mathcal{C})) = \hat{\mathbf{y}}(\mathbf{V}, \mathcal{C})'\mathbf{V}^{-1}\hat{\mathbf{y}}(\mathbf{V}, \mathcal{C}). \tag{3}$$

For any $r \times v$ matrix $\mathbf{A}$ of full row rank let $\mathcal{C}(\mathbf{A})$ denote the polyhedral cone $\mathcal{C}(\mathbf{A}) = \{\mathbf{x} : \mathbf{Ax} \geq \mathbf{0}\}$ and $\mathbf{V}$ be a $v \times v$ positive definite matrix. By applying appropriate linear transformations (see e.g. Shapiro, 1988, pp. 54–55), the following identity holds

$$\bar{\chi}^2\{\mathbf{V}, \mathcal{C}(\mathbf{A})\} = \chi^2_{v-r} + \bar{\chi}^2(\mathbf{AVA}') \tag{4}$$

where χ^2_{v-r} is identically 0 when $\mathbf{A}$ is nonsingular and thus $v = r$.

Finally, given two convex cones $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{R}^v$ and any positive definite matrix $\mathbf{V}$, from the definition of the $\mathbf{V}$-projection it follows that for any real-valued x,

$$\Pr\{\bar{\chi}^2(\mathbf{V}, \mathcal{C}_1) \geq x\} \leq \Pr\{\bar{\chi}^2(\mathbf{V}, \mathcal{C}_2) \geq x\}. \tag{5}$$

A.2. Proof of Lemmas 1 and 2

Let $\hat{\mathbf{x}} = \sqrt{n}\hat{\boldsymbol{\beta}}$. Then under H_0, when P_0 is the true cdf, in the limit, when $\hat{\boldsymbol{\Omega}}_o \to \boldsymbol{\Omega}_o$ and $\hat{\mathbf{x}} \to\sim N(\mathbf{0}, \boldsymbol{\Omega}_o)$, D_{01} reduces to

$$\hat{\mathbf{x}}'\boldsymbol{\Omega}_o^{-1}\hat{\mathbf{x}} - \min_{\mathbf{x}\geq\mathbf{0}}(\hat{\mathbf{x}} - \mathbf{x})'\boldsymbol{\Omega}_o^{-1}(\hat{\mathbf{x}} - \mathbf{x}),$$

thus, from equations (2) and (3), it follows that the asymptotic distribution of D_{01} is $\bar{\chi}^2(\boldsymbol{\Omega}_o)$, which proves the first part of Lemma 1. By the same reasoning, D_{12} asymptotically reduces to

$$\min_{\mathbf{x}\geq\mathbf{0}}(\hat{\mathbf{x}} - \mathbf{x})'\boldsymbol{\Omega}_o^{-1}(\hat{\mathbf{x}} - \mathbf{x}),$$

and using equations (3.3) and (5.4) from Shapiro (1988) it follows that D_{12} will tend asymptotically to the random variable $\bar{\chi}^2(\boldsymbol{\Omega}_o, \mathcal{O}^o)$ where $\mathcal{O}^o$ denotes the dual of the positive orthant. The asymptotic joint distribution of D_{01} and D_{12} under P_0, considered in Lemma 2, follows by applying Theorem 3.4 in Raubertas *et al.*, (1986) or Theorem 1 in Wolak (1989).

To prove the second part of Lemma 1, which establishes the asymptotic least favourable distribution for D_{01}, notice that if we put $\boldsymbol{\Delta} = \mathbf{D}\text{diag}(\mathbf{r})^{-1}\mathbf{D}'$, the variance matrix $\boldsymbol{\Omega}_o$ can be written as

$$\boldsymbol{\Omega}_o = \boldsymbol{\Delta}\otimes\boldsymbol{\Sigma}_o = (\mathbf{I}_{m-1}\otimes\mathbf{L}_o)(\boldsymbol{\Delta}\otimes\mathbf{l}_k)(\mathbf{I}_{m-1}\otimes\mathbf{L}_o)'$$

where $\mathbf{L}_o$ denotes the Cholesky decomposition of $\boldsymbol{\Sigma}_o$, i.e. $\mathbf{L}_o\mathbf{L}_o{}' = \boldsymbol{\Sigma}_o$. Then, using equation (4), in the limit

$$D_{01} \sim \bar{\chi}^2(\boldsymbol{\Omega}_o) = \bar{\chi}^2\{\boldsymbol{\Delta}\otimes\mathbf{l}_k, \mathcal{C}(\mathbf{I}_{m-1}\otimes\mathbf{L}_o)\}. \tag{6}$$

When $\boldsymbol{\Sigma}_o$ tends to the case of perfect correlation, $\mathbf{L}_o$ tends to the limiting form $\boldsymbol{\sigma}\mathbf{e}_k{}'$ where $\boldsymbol{\sigma}$ is a vector of positive constants and $\mathbf{e}_k{}'$ is the $k \times 1$ vector $(1, 0, \ldots, 0)$. Note that the cone $\mathcal{C}(\boldsymbol{\sigma}\mathbf{e}_k{}')$ is the same as $\mathcal{C}(\mathbf{e}_k{}')$, and also that $\mathcal{C}(\mathbf{I}_{m-1}\otimes\mathbf{L}_o) \subset \mathcal{C}(\mathbf{I}_{m-1}\otimes\mathbf{e}_k{}')$ because the first (upper left) entry of $\mathbf{L}_o$ must be positive. Therefore, using equation (5), for any real-valued x

$$\Pr\{\bar{\chi}^2(\boldsymbol{\Omega}_o) \geq x\} = \Pr[\bar{\chi}^2\{\boldsymbol{\Delta}\otimes\mathbf{l}_k, \mathcal{C}(\mathbf{I}_{m-1}\otimes\mathbf{L}_o)\} \geq x] \leq \Pr[\bar{\chi}^2\{\boldsymbol{\Delta}\otimes\mathbf{l}_k, \mathcal{C}(\mathbf{I}_{m-1}\otimes\mathbf{e}_k{}')\} \geq x]$$

and the actual expression of the weights follows from (4) by putting $\mathbf{A} = \mathbf{I}_{m-1}\otimes\mathbf{e}_k{}'$.

A.3. Connection between probability weights and level probabilities

Use equation (4) again with $\mathbf{A} = \mathbf{D}$ to get:

$$w_{j-1}(\boldsymbol{\Delta}) = w_j\{\text{diag}(\mathbf{r})^{-1}, \mathcal{C}(\mathbf{D})\}, \qquad \text{for } j = 1, \ldots, m-1,$$

which is the probability that the projection of a normal vector $\mathbf{y} \sim N(\mathbf{0}, \text{diag}(\mathbf{r})^{-1})$ onto the cone of nonincreasing vectors $\mathcal{C}(\mathbf{D}) = \{\mathbf{x} : x_1 \geq \cdots \geq x_m\}$ falls on a subspace of dimension j, thus having j distinct elements. This is precisely the definition of level probabilities given in Robertson, Wright and Dykstra (1988, p. 69).

A.4. Proof of Lemma 3

Above we have essentially shown that, when $\hat{\boldsymbol{\Omega}}_o$ is replaced by its true value in the limit, D_{01} and D_{12} are equal to the square of the norm of the projection of a random vector $N(\mathbf{0}, \boldsymbol{\Delta} \otimes \mathbf{l}_k)$ respectively onto the convex cone $\mathcal{C}(\mathbf{I}_{m-1} \otimes \mathbf{L}_o)$ and its dual. Moreover, because of duality relations

$$\mathcal{C}(\mathbf{I}_{m-1} \otimes \mathbf{L}_o) \subset \mathcal{C}(\mathbf{I}_{m-1} \otimes \mathbf{e}_k{}') \quad \text{and} \quad \mathcal{C}^o(\mathbf{I}_{m-1} \otimes \mathbf{e}_k{}') \subset \mathcal{C}^o(\mathbf{I}_{m-1} \otimes \mathbf{L}_o),$$

under the least favourable distribution D_{01} is associated to the largest cone and D_{12} to the smallest, so that the result follows from direct geometrical considerations.

B. APPENDIX. REJECTION REGIONS

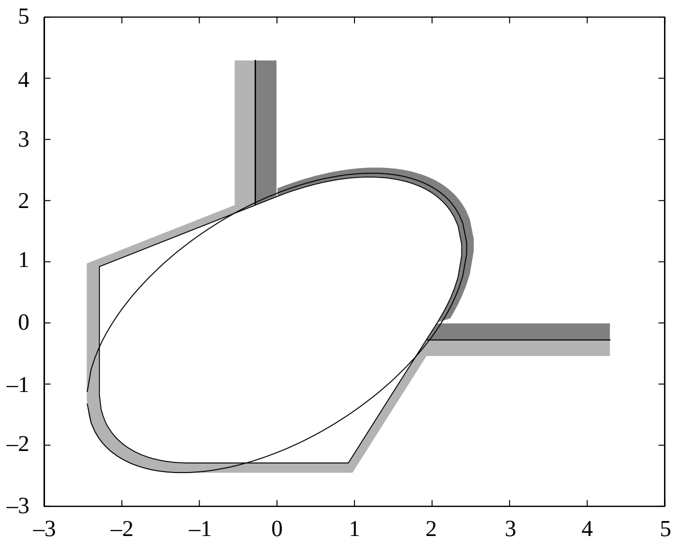

Figure B1. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and D_{02} test with $\alpha = 0.10$ and correlation $\rho = 0.77$.

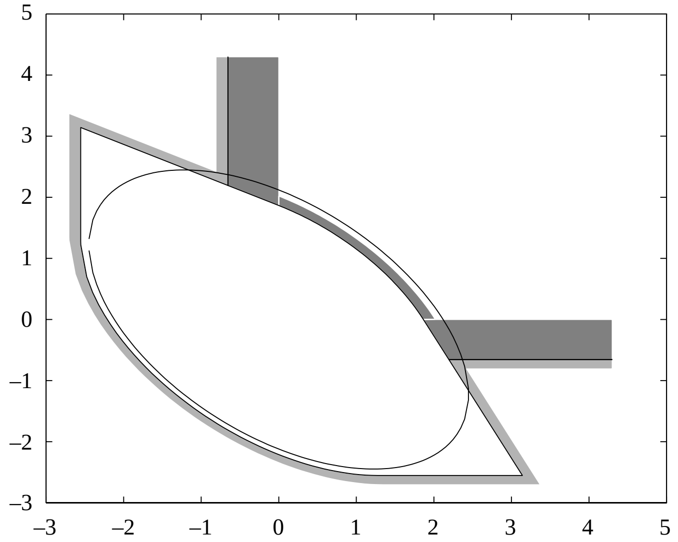

Figure B2. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and D_{02} test with $\alpha = 0.10$ and correlation $\rho = -0.5$.

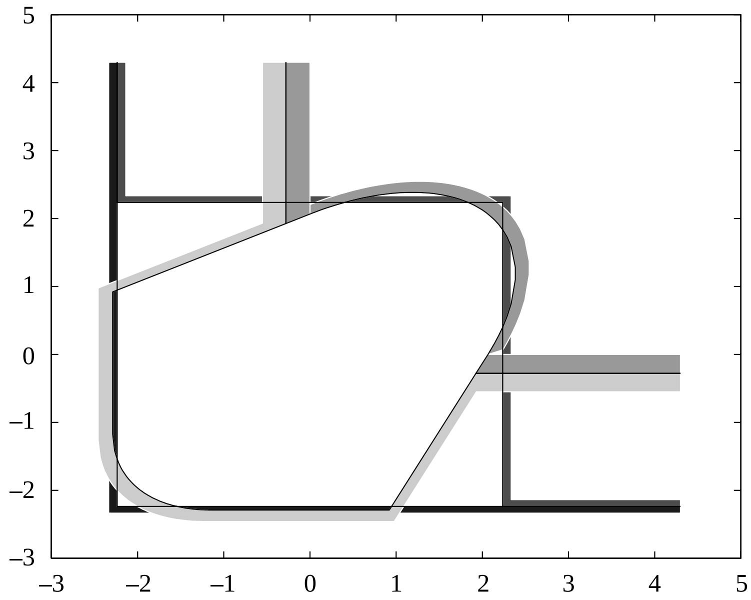

Figure B3. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and multiple comparison procedure with $\alpha = 0.05$ and correlation $\rho = 0.77$.

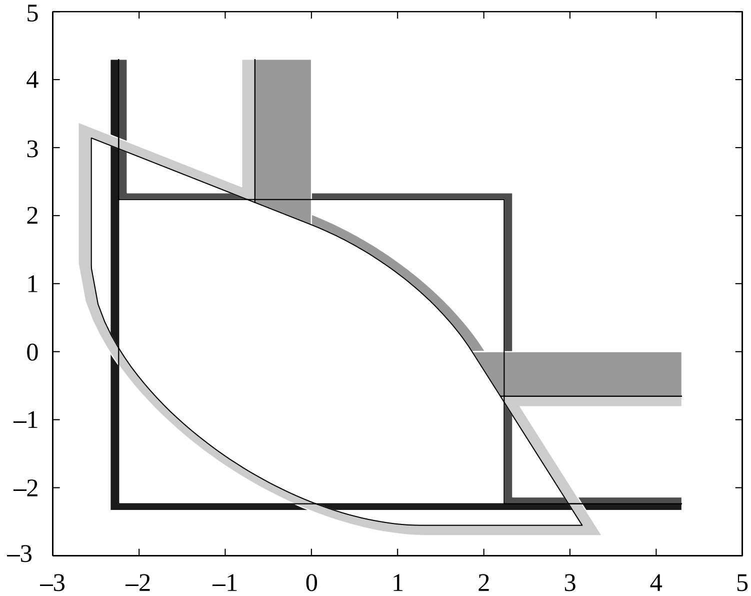

Figure B4. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and multiple comparison procedure with $\alpha = 0.05$ and correlation $\rho = -0.5$.

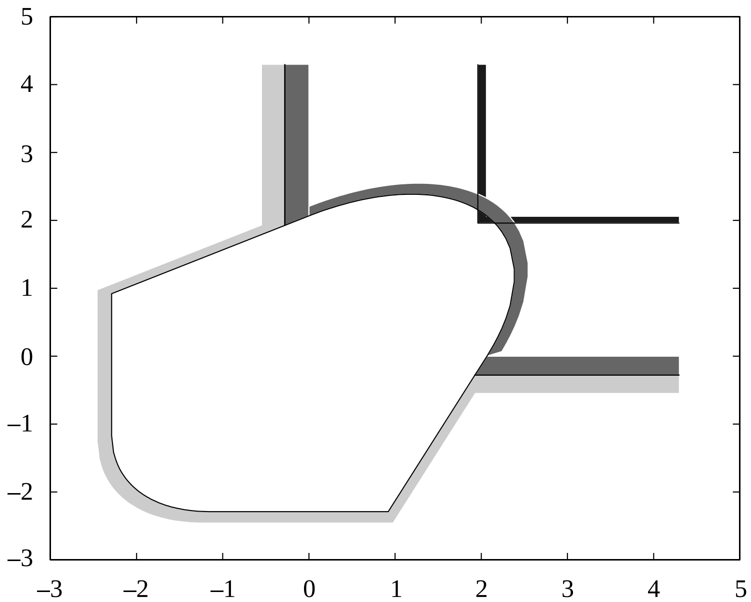

Figure B5. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and D_{21} with $\alpha = 0.05$ and correlation $\rho = 0.77$.

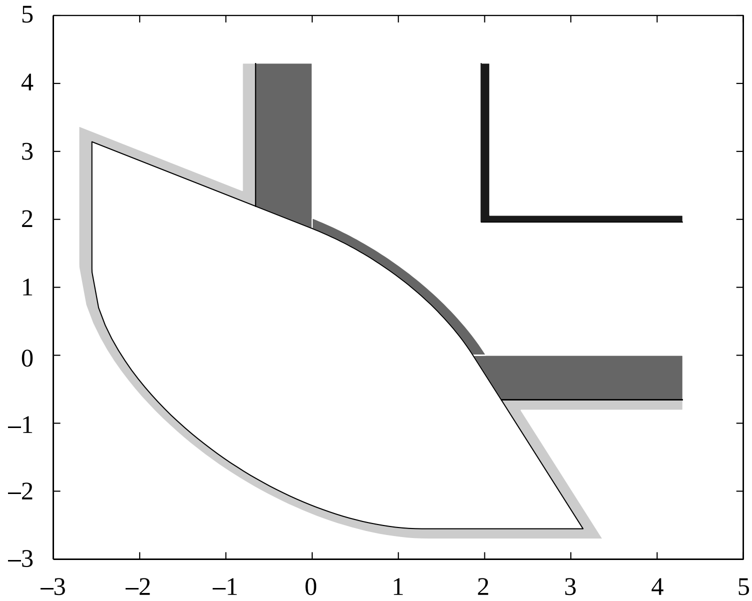

Figure B6. Rejection regions of $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$ and D_{21} with $\alpha = 0.05$ and correlation $\rho = -0.5$.

C. APPENDIX. THE MONTE CARLO DESIGN

A suitable class of income distribution functions from which to draw repeated samples should ideally provide a good fit to real income data, with a simple parametric form. The family of *SM* income distribution functions (also known as Burr XII in the statistics literature, see e.g. Johnson and Kotz (1972, p. 31)) has two essential parameters; we will say that $Y \sim SM(a, q)$ when its *cdf* can be written as:

$$P(y) = 1 - (1 + y^a)^{-q}, \qquad y \geq 0,$$

with $q > 1/a > 0$. In a detailed study on fitting several distribution functions to US income data, McDonald (1984) came to the conclusion that the *SM* performs very well in terms of goodness of fit. McDonald also fits the *SM* distribution to a sample of US census income data for 1980, and finds $a = 1.697$ and $q = 8.368$: we will take these values to define H_0 in our simulation.

A criterion to construct pairs of *SM* distributions that satisfy H_{1-0} or H_{2-1} is provided by Theorem 1 of Wilfling and Kramer (1993): let $SM(a_1, q_1)$ and $SM(a_2, q_2)$ be two *SM* distribution such that $a_1 \leq a_2$, then the first distribution Lorenz-dominates the second if and only if $a_1 q_1 \leq a_2 q_2$.

Because the *cdf* of a *SM* variable has a simple inverse

$$Y = P^{-1}(U) = \{-1 + (1 - U)^{-1/q}\}^{1/a}, \qquad \text{for } U \in (0, 1),$$

a sample from the *SM* distribution may be generated by transforming as above a random sample from the uniform distribution on $(0, 1)$. Moreover, the inverse function may also be used to compute accurate approximations of the true $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ in the populations of interest. For a reasonably large t, take t equally spaced points in $(0, 1)$, transform into the relevant *SM* distributions and compute any quantity of interest from these data. We have used this technique, with $t = 20\,000$, for the judicious choice of the parameters a_i and q_i as explained below, and also to obtain a very accurate approximation to the true weights under H_0.

Table C.1.

	$a_i, i = 1, 2, 3$			$q_i, i = 1, 2, 3$		
Low power	1.7970	1.8970	1.9970	4.6370	3.2190	2.4770
High power	1.8470	1.9970	2.1470	3.7240	2.4380	1.7950

Below we describe in detail our experiments with $m = 2$ and $m = 4$. In both cases we use a constant sample size $n_i = 2000$ and compare Lorenz curves at the level of deciles ($k = 9$). Each sample is replicated 10 000 times when $m = 2$ and 5000 times when $m = 4$.

C.1. Comparing two populations

Let the base population *cdf* P_0 be distributed as $SM(a_0, q_0)$. Under H_{1-0} we let P_1 be distributed as $SM(a_0 + c, q_0)$ with $c > 0$ and compare repeated pairs of samples of size 2000 drawn from P_0 and P_1 respectively. We used $c = 0.07$ and $c = 0.14$ which correspond approximately to $\max(z_i)$, $i = 1, \ldots, 9$ equal to 1.6 and 3.2, respectively. These can be thought of as situations of low and high power, and this impression is confirmed by the values of the statistic D_{02} in the populations which equal to 2.83 and 11.73, respectively.

The basic idea for H_{2-1} has been to construct situations with a single clear crossing of the Lorenz curves. Formally we let P_0 be as above and P_1 be $SM(a_1, q_1)$ with $a_1 > a_0$, and search for a q_1 such that $-\min z_i$ is approximately equal to $\max z_i$, $i = 1, \ldots, 9$. Here again we consider two versions of P_1, one having $a_1 = 1.817$ and $q_1 = 4.1996$ and the other with $a_1 = 2.057$ and $q_1 = 2.1397$. In both cases their respective Lorenz curves cross once with that of P_0, and $-\min z_i \approx \max z_i \approx 1$ and $-\min z_i \approx \max z_i \approx 3$ respectively. Thus, they correspond roughly to situations of low and high power. The true value of the statistic D_{02} is equal to 3.34 and 23.46, respectively.

C.2. Comparing four populations

The case of four populations is basically similar to the previous one. Given the $SM(a_0, q_0)$ base population, under H_{1-0} we let P_i, $i = 1, \ldots, 3$, be distributed as $SM(a_0 + ic, q_0)$ and compare repeated pairs of samples of size 2000 drawn from P_i, $i = 0, \ldots, 3$. We used $c = 0.03$ and 0.05 leading respectively to $\max(z_i) \approx 0.7$ and 1.5 with the true value of D_{02} being 5.25 and 22.14.

Under H_{2-1} we construct situations with a single clear crossing of each consecutive pair of Lorenz curves. More precisely, given P_0 as above and P_i distributed as $SM(a_i, q_i)$, $i = 1, \ldots, 3$, we search over possible values of a_i and q_i such that when comparing two consecutive *cdf*'s P_i and P_{i-1}, $i = 1, \ldots, 3$, $-\min z_i$ is about equal to $\max z_i$. Here again we consider two versions: one of low power, with $\max z_i \approx 0.8$, and $D_{02} = 19.14$, and a situation of high power with $\max z_i \approx 1.3$ and $D_{02} = 38.88$. The actual values of the parameters are summarized in Table C.1.

C.3. The computation of critical values

The critical values for the distance test D_{21} and those for the multiple comparison procedure may be derived easily from any table of the *cdf* of the standard normal distribution. The corresponding task for the joint chi-bar-squared procedures is not entirely trivial. First, one needs an estimate of the probability weights under H_0. For a given variance matrix $\boldsymbol{\Omega}_o$, the probability weights can be found by projecting a reasonable number

Table C.2.

Population	p_0	p_{20}	p_1	p_{21}	p_{22}
			Two populations		
H_0	0.0015	0.0002	0.0025	0.0012	0.0000
H_{1-0}	0.0020	0.0003	0.0036	0.0018	0.0000
H_{2-1}	0.0031	0.0004	0.0049	0.0018	0.0000
			Four populations		
H_0	0.0034	0.0009	0.0040	0.0007	0.0000
H_{1-0}	0.0030	0.0009	0.0036	0.0009	0.0000
H_{2-1}	0.0037	0.0009	0.0048	0.0008	0.0001

of $N(\mathbf{0}, \boldsymbol{\Omega}_o)$ random vectors on $\mathcal{R}^v_+$, and counting the proportion of times where exactly i coordinates of these projections are positive. Then one can use the following algorithm:

1. set the values of α_1, α_2;
2. set the value of α_{20} which should be much larger than α_2 but less than $1 - \Pr(D_{12} = 0)$;
3. compute y_0 as the solution to $F(\infty, y_0) = 1 - \alpha_{20}$;
4. compute x as the solution to $F(x, y_0) = 1 - \alpha_{20} - \alpha_1$;
5. compute y_1 as the solution to $F(x, y_1) = 1 - \alpha_1 - \alpha_2$;
6. if the last equation has no solution, decrease the value of α_{20} and go back to step 3.

We set $\alpha_1 = \alpha_2 = 0.05$ and $\alpha_{20} = 0.6$ for two populations and 0.55 for four populations. The equations above may be solved only by numerical inversion; the secant method which we implemented is reasonably fast and reliable but requires an initial guess and some supervision.

For the purposes of our simulation, the critical values for the least favourable distribution G^+ need to be computed only once because the corresponding probability weights are known (see Robertson, Wright and Dykstra 1988, p. 444). On the other hand, the critical values based on $\hat{\boldsymbol{\Omega}}_o$ have to be recomputed within each simulated sample. Due to the difficulty of making this automatic, we have used instead the critical values computed only once for each assumed population, using the weights estimated from the true $\boldsymbol{\Omega}_o$.

Dardanoni and Forcina (1998, p. 1118) show, in a similar context, that replacing unknown parameters with their ML estimates has negligible effects on the p-values of the $\bar{\chi}^2$ distribution. To examine whether a similar result holds in this context, we compare the true probabilities that determine the critical values of the joint $\bar{\chi}^2$ procedure, with the corresponding probabilities which result when using $w(\hat{\boldsymbol{\Omega}}_o)$, estimated with 500 replicates.

In Table C.2, we let $p_0 = \Pr(D_{01} \leq x, D_{12} \leq y_1)$, $p_1 = \Pr(D_{01} > x, D_{12} \leq y_0)$, $p_{20} = \Pr(D_{01} \leq x, D_{12} > y_1)$, $p_{21} = \Pr(D_{01} > x, y_0 < D_{12} \leq y_1)$ and $p_{22} = \Pr(D_{01} > x, D_{12} > y_1)$. For each of these cases, we calculated the distribution of the absolute error between the true and estimated probabilities, and reported the upper 95% quantiles. Note that H_{1-0} and H_{2-1} refer only to the situations of high power, as defined above.

These data seem to indicate that the critical values computed within each sample will differ, in most cases, only very slightly from the corresponding true values. Thus, we can conjecture that the probability that the sample falls exactly in between the two regions will be very small.

C.4. Results

We consider first the case of two populations. It is convenient to compare the performance of the various procedures in terms of power and error rates by concentrating first on those that, when they reject H_0, differentiate between the hypotheses H_{1-0} and H_{2-1}. Five such procedures are analysed in Figures C1 and C2, which display the rejection rates in favour of H_{1-0} on the left side and those in favour of H_{2-1} on the right. In Figures C1–C6, the *true state* is marked at the bottom of the histogram, and 'f' and 'c' represent situations far and close to the null respectively. The estimated probabilities plotted are to be interpreted as error rates to the left of H_0, and as power to the right.

In order to assess the performance of the D_{21} procedure, in Figure C3 below we compare its rejection rates with those achieved by the best joint $\bar{\chi}^2$ procedure and the simpler procedure based on the D_{01} statistics.

Finally, the results for the case of four populations are summarized in the same way as in Figures C4–C6.

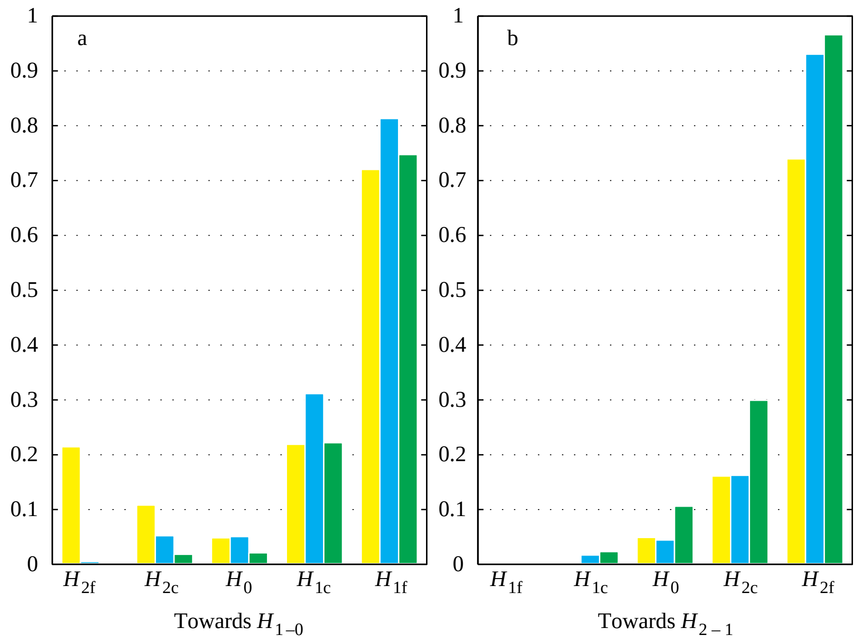

Figure C1. Rejection rates towards H_{1-0} and H_{2-1} of three joint chi-bar-squared procedures (from light to dark shade): a joint $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$; b as before but $\alpha_{20} = 0.60$; c joint $\bar{\chi}^2$ under G^+ and with $\alpha_{20} = 0.40$.

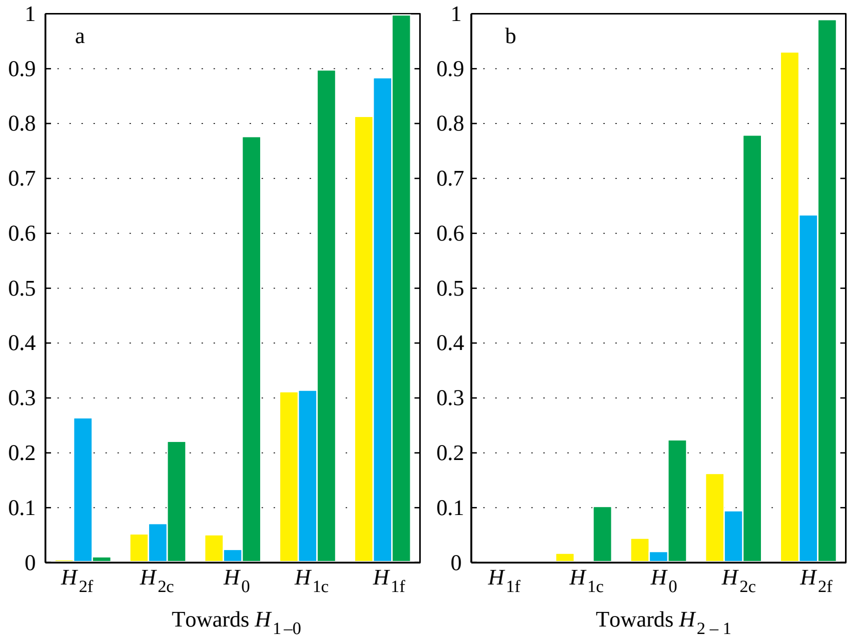

Figure C2. Rejection rates towards H_{1-0} and H_{2-1} of three procedures (from light to dark shade): a 'best' joint $\bar{\chi}^2$; b multiple comparisons with $\alpha = 0.05$; c sample comparisons.

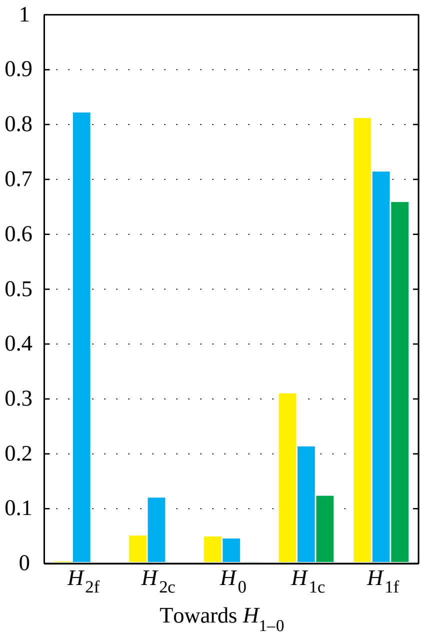

Figure C3. Rejection rates towards H_{1-0} of three procedures (from light to dark shade) a joint $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.60$; b D_{01} with $\alpha = 0.05$; c D_{21} with $\alpha = 0.05$.

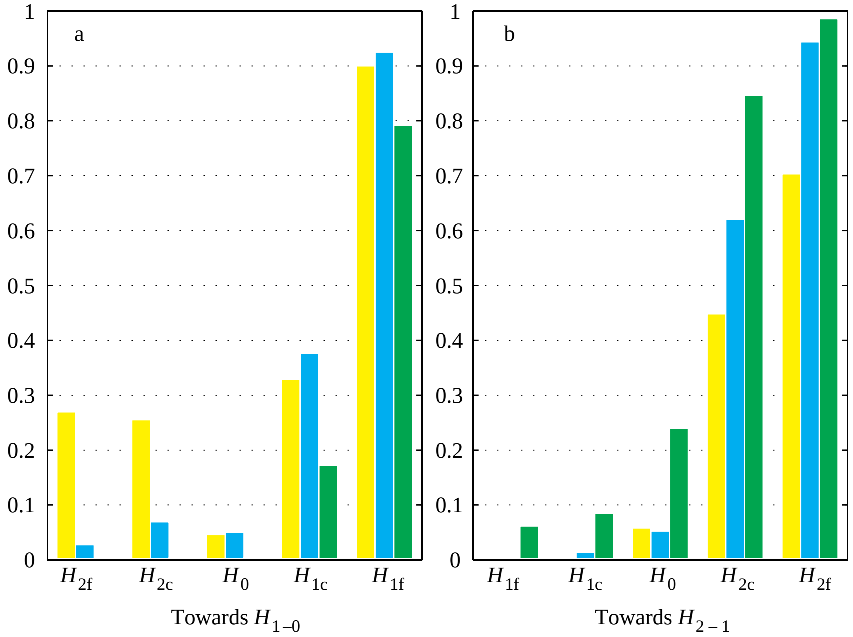

Figure C4. Rejection rates towards H_{1-0} and H_{2-1} of three procedures (from light to dark shade): a joint $\bar{\chi}^2$ with $\alpha_1 = \alpha_2 = \alpha_{20} = 0.05$; b as before but $\alpha_{20} = 0.60$; c joint $\bar{\chi}^2$ under G^+ and with $\alpha_{20} = 0.40$.

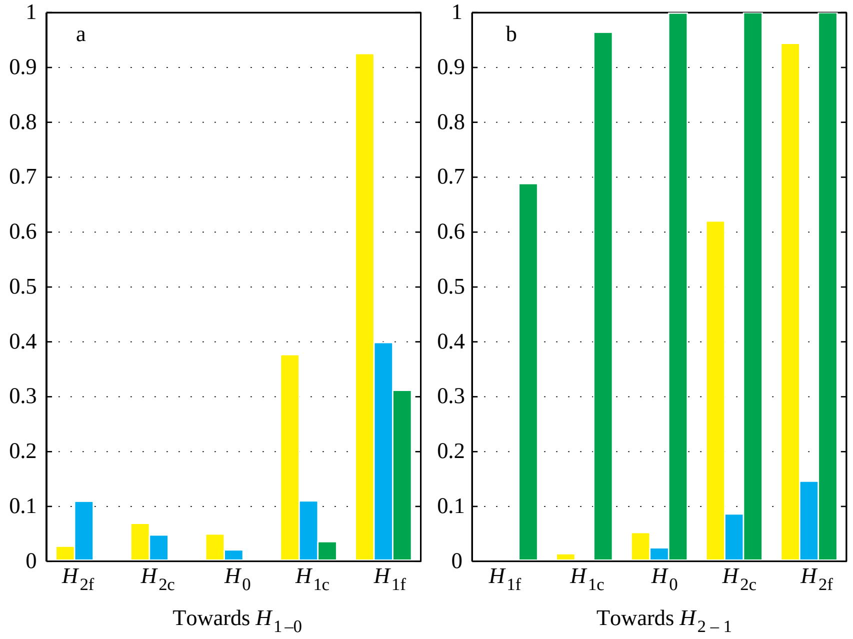

Figure C5. Rejection rates towards H_{1-0} and H_{2-1} of three procedures (from light to dark shade): a 'best' joint $\bar{\chi}^2$; b multiple comparisons with $\alpha = 0.05$; c sample comparisons.

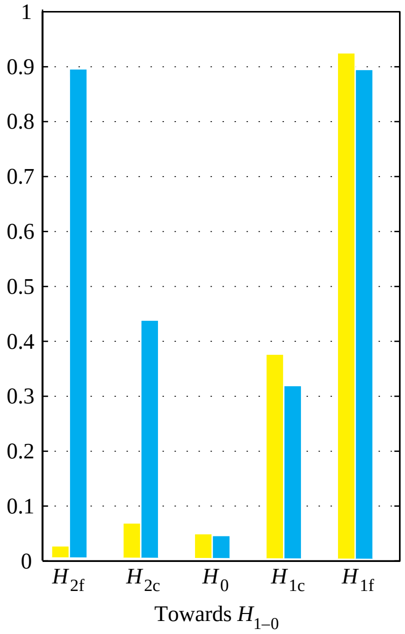

Figure C6. Rejection rates towards H_{1-0} of three procedures (from light to dark shade): a best joint $\bar{\chi}^2$; b D_{01} with $\alpha = 0.05$; c D_{21} with $\alpha = 0.05$.