Testing for Restricted Stochastic Dominance: Some Further Results

by

Russell Davidson

Department of Economics and CIREQ   GREQAM
McGill University   Centre de la Vieille Charité
Montréal, Québec, Canada   2 Rue de la Charité
H3A 2T7   13236 Marseille cedex 02, France

Abstract

Extensions are presented to the results of Davidson and Duclos (2007), whereby the null hypothesis of restricted stochastic non dominance can be tested by both asymptotic and bootstrap tests, the latter having considerably better properties as regards both size and power. In this paper, the methodology is extended to tests of higher-order stochastic dominance. It is seen that, unlike the first-order case, a numerical nonlinear optimisation problem has to be solved in order to construct the bootstrap DGP. Conditions are provided for a solution to exist for this problem, and efficient numerical algorithms are laid out. The empirically important case in which the samples to be compared are correlated is also treated, both for first-order and for higher-order dominance. For all of these extensions, the bootstrap algorithm is presented. Simulation experiments show that the bootstrap tests perform considerably better than asymptotic tests, and yield reliable inference in moderately sized samples.

August 2008

1. Introduction

In Davidson and Duclos (2006), henceforth DD, methods based on empirical likelihood are developed for testing stochastic dominance. The null hypotheses of the tests proposed postulate non dominance, the idea being that, if such a null hypothesis is rejected, it is reasonable to accept the only remaining alternative, which is dominance. It is shown, however, that a null of non dominance can never be rejected statistically if the distributions that are compared are continuous, and full account is taken of the tails of the distributions. In such circumstances, only a hypothesis of restricted dominance, that is, dominance restricted to some closed interval contained in the interior of the support of the distributions, can ever be rejected empirically.

The aim of this paper is to extend the results of DD beyond first order stochastic dominance to higher orders, and also to treat the case in which the samples drawn from the distributions that are to be compared are correlated. DD treated only the case of independent samples, but, in practice, correlated samples are common enough: One might wish the compare distributions of pre- and post-tax income, or the distributions of the separate incomes of married couples, for instance.

It is not difficult to set up asymptotic testing procedures, based on the intersection-union principle, for all the cases dealt with in this paper. It was shown by DD that not only are asymptotic t statistics and empirical likelihood-ratio statistics asymptotically equivalent under the null and local alternatives, but also that they are very close numerically in quite small samples under these same conditions. The disadvantage of tests based on these statistics is that they are typically seriously undersized, and lacking in power.

DD then show that bootstrap tests, which rely on the probabilities associated with maximising the empirical likelihood under the null in order to set up a bootstrap data-generating process (bootstrap DGP), can improve substantially on asymptotic tests regarding both size and power. This use of the probabilities generated by empirical likelihood maximisation is suggested by Brown and Newey (2002). See also Owen (2001) for a survey of empirical likelihood methods. There remains an ineradicable tendency to underreject, but it is greatly reduced relative to asymptotic inference. In this paper, these bootstrap methods are extended, and shown to share two main properties with the methods studied by DD, namely, the near numerical equivalence of the asymptotic t statistics and the empirical likelihood-ratio statistics, and the considerable improvement in the reliability of inference based on the bootstrap tests.

In section 2, we recall the definitions of higher-order stochastic dominance, and define an empirical likelihood-ratio statistic based on the difference between empirical loglikelihood functions computed unrestrictedly and computed under the constraint of the null hypothesis of non dominance. Then, in section 3, we study the problem of the existence of the statistic defined in the preceding section, and discuss numerical methods for its computation. The case in which the samples to be compared are correlated is treated in section 4, and the special case of first-order dominance, for which an analytic solution of the empirical likelihood problem exists, is treated in section 5.

In section 6, the relation between the likelihood-ratio statistic and the asymptotic t statistic of an intersection-union test is investigated, and it is seen how this relation makes it possible to have a more efficient implementation of the empirical likelihood problem. Bootstrap tests, with bootstrap DGPs defined in terms of the solution to that problem, are proposed in section 7, and the results of simulation experiments designed to study their performance are given in section 8. Some conclusions are presented in section 9.

2. Higher-Order Dominance

Recall that distribution A, characterised by a cumulative distribution function (CDF) F_A, is stochastically dominated at first order by distribution B, characterised by the CDF F_B, if, for all y in the joint support of the two distributions, $F_A(y) > F_B(y)$. If y is income, this means that the proportion of population A with income no greater than y is greater than in population B.

If we define the *dominance functions* D_K^S, $K = A, B$, recursively by the relations

$$D_K^{S+1}(y) = \int_0^y D_K^S(z)\,\mathrm{d}z, \qquad D_K^1(y) = F_K(y),$$

then B dominates A stochastically at order S if $D_A^S(y) > D_B^S(y)$ for all y in the joint support. Here we make the simplifying assumption that this support is entirely contained in the nonnegative real line. It is easy to show that the function D_K^S can be written explicitly as

$$D_K^S(y) = \frac{1}{(S-1)!} \int_0^y (y-z)^{S-1}\,\mathrm{d}F_K(z);$$

see Davidson and Duclos (2000), where the link between higher-order stochastic dominance and inequality and poverty indices is explained.

We are interested in the hypothesis that B does not dominate A at order S. Suppose that we have two samples, here supposed IID and mutually independent, one of size N_A, with typical observation y_i^A, drawn from the distribution A, another of size N_B, with typical observation y_i^B, from B. We can define empirical distribution functions (EDFs) for each sample as follows:

$$\hat{F}_K(y) = \frac{1}{N_K} \sum_{i=1}^{N_K} \mathrm{I}(y_i^K \le y), \quad K = A, B, \tag{1}$$

where I denotes an indicator function, equal to 1 if its Boolean argument is true, and to 0 otherwise. The EDF evaluated at y is thus the proportion of the observations less than or equal to y.

Similarly, we define empirical versions of the dominance functions:

$$\hat{D}_K^S(y) = \frac{1}{(S-1)!} \int_0^y (y-z)^{S-1} \, \mathrm{d}\hat{F}_K(z)$$

$$= \frac{1}{N_K(S-1)!} \sum_{i=1}^{N_K} (y - y_i^K)^{S-1} \, \mathrm{I}(y_i^K \le y) = \frac{1}{N_K(S-1)!} \sum_{i=1}^{N_K} (y - y_i^K)_+^{S-1},$$

$$\tag{2}$$

where for convenience we write z_+ for $\max(0, z)$. We say that B dominates A at order S in the sample if $\hat{D}_A^S(y) > \hat{D}_B^S(y)$ for all y in the joint support. It is clear that the hypothesis of non dominance in the underlying distributions should not be rejected unless there is dominance in the sample. Procedures for testing for restricted dominance can therefore proceed under the assumption that, for all y in the interval of interest, $\hat{D}_A^S(y) > \hat{D}_B^S(y)$.

An empirical likelihood-ratio test statistic is computed, just like an ordinary likelihood-ratio statistic, as twice the difference between the values of the empirical loglikelihood maximised under the alternative hypothesis and under the null. The empirical log-likelihood function (ELF) depends on, and is maximised with respect to, a set of probabilities, one each assigned to the observations of the sample. Since here there are two independent samples, the joint ELF is the sum of the ELFs of the individual samples. As in DD, it is convenient to use a notation that can apply equally well to samples drawn from continuous or discrete distributions. We denote by Y^K the set of distinct values in the sample drawn from distribution K, and then, for each $y_t^K \in Y^K$, we let n_t^K denote the number of sample points equal to y_t^K. If the underlying distribution is continuous, then, with probability 1, $n_t^K = 1$ for each t, but with a discrete distribution, higher integer values are possible.

If a probability p_t^K is assigned to each $y_t^K \in Y^K$, the empirical loglikelihood function for the sample drawn from distribution K is

$$\sum_{y_t^K \in Y^K} n_t^K \log p_t^K.$$

If this is maximised with respect to the p_t^K under the constraint that their sum is equal to 1, the maximising probabilities are given by $p_t^K = n_t^K / N_K$, and the maximised value of the ELF is

$$\sum_{y_t^K \in Y^K} n_t^K (\log n_t^K - \log N_K) = \sum_{y_t^K \in Y^K} n_t^K \log n_t^K - N_K \log N_K. \tag{3}$$

The first term on the right-hand side above vanishes if each $n_t^K = 1$.

Any assignment of probabilities to the points of the observed sample implicitly defines a set of weighted dominance functions, by the relation

$$\tilde{D}_K^S(y) = \frac{1}{(S-1)!} \sum_{y_t^K \in Y^K} p_t^K (y - y_t^K)_+^{S-1}, \tag{4}$$

In order to maximise the ELF under the hypothesis of non dominance, given that there is dominance in the sample, we impose the requirement that $\tilde{D}_A^S(y) = \tilde{D}_B^S(y)$ for some $y \in Y$ where Y is the union of Y^A and Y^B. One reason for having to limit attention to restricted dominance in statistical work is immediately visible: If y is the greatest observation in the pooled sample, then, for $S = 1$, $\hat{D}_A^1(y) = \hat{D}_B^1(y) = 1$, whatever the relation between the two distributions may be. If the definition (1) were formulated with strict inequality in the indicator functions, the same problem would arise at the smallest observation.

The device proposed by DD to overcome this problem is to limit attention to some interval that is strictly inside the interval determined by the smallest and greatest observations in the pooled sample, without concern for what happens outside this interval. The null hypothesis is thereby changed from one of global non dominance to the stronger hypothesis of restricted non dominance, which requires that the inequality $D_A^S(y) > D_B^S(y)$ should be violated, not just somewhere in the joint support, but somewhere in the chosen interval. Of course, rejection of this stronger hypothesis allows only a weaker conclusion than the one possible if global non dominance is rejected: all we can conclude is restricted dominance.

The problem of maximising the ELF subject to the null of restricted non dominance can be formulated as follows:

$$\max_{p_t^A, p_t^B, y \in Y} \sum_{y_t^A \in Y^A} n_t^A \log p_t^A + \sum_{y_s^B \in Y^B} n_s^B \log p_s^B$$

$$\text{subject to} \quad \sum_t p_t^A = 1, \quad \sum_s p_s^B = 1, \tag{5}$$

$$\text{and} \quad \sum_t p_t^A (y - y_t^A)_+^{S-1} = \sum_s p_s^B (y - y_s^B)_+^{S-1},$$

with an obvious notation for sums over t and s. The last condition here implies that $\tilde{D}_A^S(y) = \tilde{D}_B^S(y)$. For the moment we fix the point y of contact between the two dominance functions. We discuss later the problem of determining the value of y that solves the maximisation problem. A suitable Lagrangean for the maximisation with respect to the probabilities p_t^A and p_s^B is

$$\sum_t n_t^A \log p_t^A + \sum_s n_s^B \log p_s^B + \lambda_A (1 - \sum_t p_t^A) + \lambda_B (1 - \sum_s p_s^B)$$

$$- \mu \left(\sum_t p_t^A (y - y_t^A)_+^{S-1} - \sum_s p_s^B (y - y_s^B)_+^{S-1} \right).$$

The signs of the Lagrange multipliers λ_A, λ_B, and μ are chosen so that all three are nonnegative when A is dominated by B in the sample. The first-order necessary conditions for a maximum are given by the three constraints, the conditions

$$\frac{n_t^A}{p_t^A} - \lambda_A - \mu (y - y_t^A)_+^{S-1} = 0 \quad \text{for all } t, \tag{6}$$

$$- 4 -$$

and a similar set of conditions for B. Solving (6) gives

$$p_t^A = \frac{n_t^A}{\lambda_A + \mu(y - y_y^A)_+^{S-1}}. \tag{7}$$

The corresponding solution for the probabilities in distribution B is

$$p_s^B = \frac{n_s^B}{\lambda_B - \mu(y - y_s^B)_+^{S-1}}. \tag{8}$$

The requirement that $\sum_t p_t^A = 1$ implies that

$$\begin{aligned}
\lambda_A &= \sum_t \frac{\lambda_A n_t^A}{\lambda_A + \mu(y - y_t^A)_+^{S-1}} \\
&= \sum_t n_t^A \frac{\lambda_A + \mu(y - y_t^A)_+^{S-1}}{\lambda_A + \mu(y - y_t^A)_+^{S-1}} - \mu \sum_t \frac{n_t^A (y - y_t^A)_+^{S-1}}{\lambda_A + \mu(y - y_t^A)_+^{S-1}} \\
&= N_A - \mu \sum_t p_t^A (y - y_t^A)_+^{S-1}. \tag{9}
\end{aligned}$$

Similarly,

$$\lambda_B = N_B + \mu \sum_s p_s^B (y - y_s^B)_+^{S-1}. \tag{10}$$

Now let

$$\sum_t p_t^A (y - y_t^A)_+^{S-1} = \sum_s p_s^B (y - y_s^B)_+^{S-1} = \nu. \tag{11}$$

Then (9) and (10) can be written as

$$\lambda_A = N_A - \mu\nu \quad \text{and} \quad \lambda_B = N_B + \mu\nu,$$

while (7) and (8) become

$$p_t^A = \frac{n_t^A}{N_A - \mu(\nu - (y - y_t^A)_+^{S-1})} \quad \text{and} \quad p_s^B = \frac{n_s^B}{N_B + \mu(\nu - (y - y_S^B)_+^{S-1})}. \tag{12}$$

The two unknowns in (12) are μ and ν. Two equations that can be solved to find their values are

$$\sum_t \frac{n_t^A (y - y_t^A)_+^{S-1}}{N_A - \mu(\nu - (y - y_t^A)_+^{S-1})} - \sum_s \frac{n_s^B (y - y_s^B)_+^{S-1}}{N_B + \mu(\nu - (y - y_s^B)_+^{S-1})} = 0,$$

$$\text{and} \quad \sum_s \frac{n_s^B (y - y_s^B)_+^{S-1}}{N_B + \mu(\nu - (y - y_s^B)_+^{S-1})} = \nu, \tag{13}$$

which follow from substituting (12) into (11). It is obvious that, if the probabilities are given by (12) with μ and ν solutions of (13), then the third constraint is satisfied. A little algebra shows that the constraints that the two sets of probabilities each sum to 1 are also satisfied; see Lemma 1 in the Appendix.

It is unfortunately not possible to find an analytic solution to the equations (13). It is however not at all difficult to solve them numerically, as we will see in the next section.

3. Existence of the Constrained Solution

The equations (13) all have the same algebraic structure, whatever the values of y and S, and so the same considerations apply to their solutions. In this section, it is convenient to simplify notation a little by writing $k_t^A = (y - y_t^A)_+^{S-1}$ and $k_s^B = (y - y_s^B)_+^{S-1}$. Observe that $k_t^A \geq 0$, $k_s^B \geq 0$ for all t and s. We may then define two functions of the unknowns μ and ν as follows:

$$f_A(\mu, \nu) = \sum_A \frac{n_t^A k_t^A}{N_A - \mu(\nu - k_t^A)} \qquad \text{and} \qquad f_B(\mu, \nu) = \sum_B \frac{n_s^B k_s^B}{N_B + \mu(\nu - k_s^B)}. \qquad (14)$$

Here the sums indicated schematically by the notation mean that, for each sample, we sum over only those observations for which k_t^A or k_s^B is nonzero. In this particular case, the values of these sums are of course the same as they would be if we summed over all observations in the sample. Equations (13) become

$$f_A(\mu, \nu) - f_B(\mu, \nu) = 0 \qquad \text{and} \qquad f_B(\mu, \nu) = \nu \qquad (15)$$

and the probabilities (12) become

$$p_t^A = \frac{n_t^A}{N_A - \mu(\nu - k_t^A)} \qquad \text{and} \qquad p_s^B = \frac{n_s^B}{N_B + \mu(\nu - k_s^B)}. \qquad (16)$$

In order for all of these probabilities to be nonnegative, we require that

$$\mu\nu \leq N_A + \mu k_t^A \qquad \text{and} \qquad \mu\nu \geq -N_B + \mu k_s^B \qquad (17)$$

for *all* t and s, not just those included in the sums in (14). If y is restricted to a set in the interior of the joint support, we can be sure that there will be observations t in the sample from A for which $k_t^A = 0$, which means that the first inequality of (17) is satisfied for all t if and only if $\nu \leq N_A/\mu$. The second inequality is satisfied for all s if and only if $\nu \geq -N_B/\mu + k_+^B$, where $k_+^B = \max_s k_s^B$. The admissible region in (μ, ν)–space is thus bounded above by the hyperbola with equation $\nu = N_A/\mu$ and below by the hyperbola with equation $\nu = -N_B/\mu + k_+^B$; see Figure 1. The figure shows only the region in which $\mu \geq 0$, since only solutions satisfying that requirement can solve the constrained maximisation problem, if sample B dominates sample A. The dotted red line in the figure is another hyperbola, displaced upwards from the upper bound; its equation is $\nu = N_A/\mu + k_-^A$, where k_-^A is the smallest nonzero value in the set of the k_t^A.

It is easy to check that the functions f_A and f_B are positive in the region bounded by the lower hyperbola and the upwards-displaced upper one. If $k_-^A < k_+^B$, the two hyperbolas intersect; otherwise this region is unbounded to the right. On the other hand, the hyperbolas with equations $\nu = N_A/\mu$ and $\nu = -N_B/\mu + k_+^B$ always intersect for some finite μ.
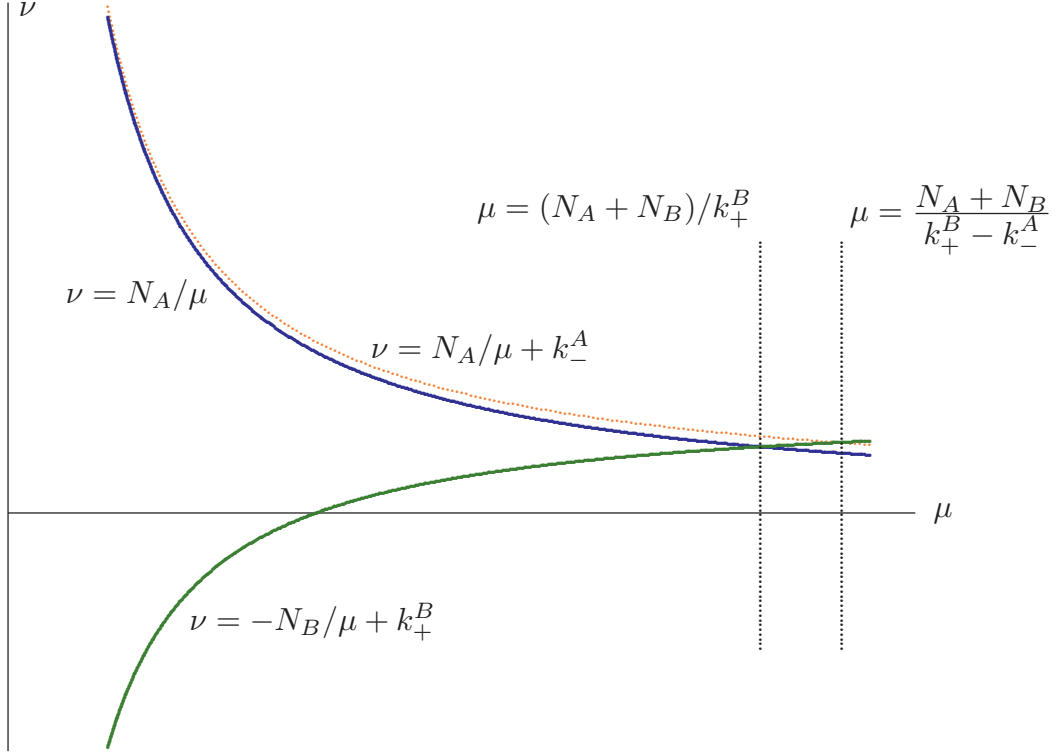
Figure 1. Region of Nonnegative Probabilities

The function f_A is monotonically increasing in ν, and tends to infinity as ν tends upwards to $N_A/\mu + k_-^A$ on the displaced upper hyperbola. Similarly, f_B is monotonically decreasing in ν and tends to infinity as ν tends downwards to $-N_B/\mu + k_+^B$ on the lower hyperbola. Consequently, for any given positive μ less than the value at which the displaced upper hyperbola and the lower hyperbola intersect, $f_A - f_B$ has a unique zero for the given μ, for some ν in the interval $[-N_B/\mu + k_+^B, N_A/\mu + k_-^A]$. We denote by $\nu(\mu)$ the value of ν such that $f_A(\mu, \nu) - f_B(\mu, \nu) = 0$.

A pair (μ, ν) that solves the equations (15) is therefore such that $\nu = \nu(\mu)$ and $f_B(\mu, \nu(\mu)) = \nu(\mu)$. Define the function $f(\mu)$ as the common value of $f_A(\mu, \nu(\mu))$ and $f_B(\mu, \nu(\mu))$. If we can find μ such that $f(\mu) = \nu(\mu)$, we have a solution to (15).

If sample B dominates sample A at order S, then, for any value of y, $\hat{D}_A^S(y) > \hat{D}_B^S(y)$. In our present notation, this means that, for any (irrelevant) value of ν, we have $f_A(0, \nu) > f_B(0, \nu)$. Thus the first equation of (15) cannot be satisfied with $\mu = 0$. Consequently, the function f is not defined at $\mu = 0$, and $\nu(0)$ does not exist. However, the product $\mu\nu(\mu)$ tends to a nonzero, finite, limit as $\mu \to 0$, obtained by solving the first equation of (15) with $\mu = 0$, $\mu\nu \neq 0$, that is,

$$\frac{N_A f_A(0,0)}{N_A - (\mu\nu)} = \frac{N_B f_B(0,0)}{N_B + (\mu\nu)}.$$

From this we see that the limit of $\mu\nu(\mu)$ as $\mu \to 0$ is

$$-\frac{N_A N_B (f_A(0,0) - f_B(0,0))}{N_A f_A(0,0) + N_B f_B(0,0)},$$

and from this it follows that the limit of $f(\mu)$ as $\mu \to 0$ is

$$\lim_{\mu \to 0} f(\mu) = \frac{N_A f_A(0,0) + N_B f_B(0,0)}{N_A + N_B},$$

that is, a weighted average of $f_A(0,0)$ and $f_B(0,0)$, and thus a positive quantity. It follows as well that $\nu(\mu) \to -\infty$ as $\mu \to 0$. Thus for positive values of μ close enough to 0, $f(\mu) > \nu(\mu)$.

Denote by μ_+ the value of μ at which the upper and lower hyperbolas between which all probabilities are nonnegative intersect. We see that $\mu_+ = (N_A + N_B)/k_+^B$, and that $N_A/\mu_+ = -N_B/\mu_+ + k_+^B$. Since $\nu(\mu_+) > -N_B/\mu_+ + k_+^B$, the graph of the function $\nu(\mu)$ must cut the hyperbola $\nu = N_A/\mu$ at least once from below for a value of μ less than μ_+. Let the greatest value of μ for which $\nu(\mu) = N_A/\mu$ be denoted as μ_{lim}. For any $\mu > \mu_{\text{lim}}$, the value of $\nu(\mu)$ is greater than N_A/μ, and so must generate some negative probabilities. We are therefore interested only in solutions with $\mu \le \mu_{\text{lim}}$. See Figure 2 for an illustration of the case in which the lower and displaced upper hyperbolas intersect when $\mu = \mu_{\text{max}}$.
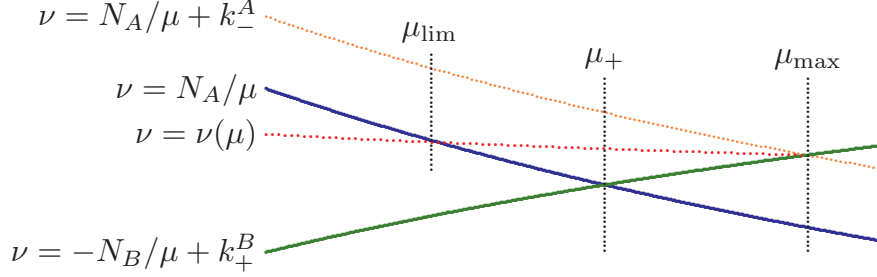

Figure 2. Approach to Upper Limit of μ

Let us calculate the value of $f(\mu_{\text{lim}})$. Since $\nu(\mu_{\text{lim}}) = N_A/\mu_{\text{lim}}$, this value is $f_A(\mu_{\text{lim}}, N_A/\mu_{\text{lim}})$. From (14) we find that

$$f(\mu_{\text{lim}}) = \sum_A \frac{n_t^A k_t^A}{\mu_{\text{lim}} k_t^A} = \frac{1}{\mu_{\text{lim}}} \sum_A n^A < \frac{N_A}{\mu_{\text{lim}}} = \nu(\mu_{\text{lim}}).$$

The inequality above is strict because the sum is over only those observations for which k_t^A is nonzero. Since for μ in the neighbourhood of 0, $\nu(\mu) < 0 < f(\mu)$, and since the functions f and ν are continuous in μ, there must exist a $\mu \in]0, \mu_{\text{lim}}[$ where $f(\mu) = \nu(\mu)$. This proves the existence of a solution to equations (15) in the region in which all the probabilities (16) are nonnegative.

There remains the question of the uniqueness of the solution. I have not as yet discovered either a proof of uniqueness or a counterexample, but it is clear that one may choose among multiple solutions on the basis of the maximised ELF.

A suitable way of finding the solution(s) to the equations (15) is to use Newton's method. The starting point of the algorithm should of course be chosen to lie inside

the region bounded by the undisplaced upper hyperbola and the lower one. Experience shows that it is also desirable to check at each iteration of Newton's method that the new (μ, ν) pair is still inside the admissible region, and, if not, put it back in. The derivatives of the functions f_A and f_B with respect to μ are often very much smaller than those with respect to ν, and so it can be helpful to rescale them, in order that the Jacobian of the left-hand sides of (15) with respect to μ and ν should not be nearly singular. If these precautions are observed, Newton's method seems to find a solution after a small number of iterations – always fewer than 10 in the cases so far examined.

4. Correlated Samples

If the samples drawn from the distributions A and B are correlated, an "observation" must be thought of as a couple (y_t^A, y_t^B) of correlated drawings. There is but one sample size, N say, in this case. The ELF now ascribes probabilities p_t to each couple, so that, if n_t is the number of drawings equal to (y_t^A, y_t^B), the ELF is $\sum_t n_t \log p_t$. If it is maximised with respect to the p_t subject only to the constraint that $\sum_t p_t = 1$, the maximising probabilities are $p_t = n_t/N$ and the maximised ELF is $\sum_t n_t \log n_t - N \log N$. As before, if each $n_t = 1$, the first term vanishes.

In order to test for restricted non dominance, we now wish to impose the condition that, for some y in the restricted interval of interest,

$$\sum_t p_t\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big) = 0, \tag{18}$$

which means that $\tilde{D}_A^S(y) = \tilde{D}_B^S(y)$, where as usual the weighted empirical dominance functions are given by (4). For any given y, the problem of maximising the ELF with respect to the p_t is characterised by the Lagrangean

$$\sum_t n_t \log p_t + \lambda(1 - \sum_t p_t) - \mu\Big(\sum_t p_t\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)\Big). \tag{19}$$

Equating the partial derivative of this Lagrangean with respect to p_t to zero gives the first-order condition

$$p_t = \frac{n_t}{\lambda + \mu\big((y - y_t^A)_+^{S-1} - (y - y_y^B)_+^{S-1}\big)} \tag{20}$$

The constraint (18) that the dominance functions touch at y becomes

$$0 = \sum_t \frac{n_t\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)}{\lambda + \mu\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)}. \tag{21}$$

Similarly, the constraint that $\sum_t p_t = 1$ becomes

$$1 = \sum_t \frac{n_t}{\lambda + \mu\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)}. \tag{22}$$

Adding together λ times (22) and μ times (21) gives

$$\lambda = \sum_t \frac{n_t\big[\lambda + \mu\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)\big]}{\lambda + \mu\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)} = N.$$

There is only one unknown left, namely μ, and it is determined by the constraint

$$\sum_t \frac{n_t\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)}{N + \mu\big((y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}\big)} = 0. \tag{23}$$

In order to analyse the set of solutions for μ to the equation (23), make as before the definitions $k_t^A = (y - y_t^A)_+^{S-1}$ and $k_t^B = (y - y_t^B)_+^{S-1}$. Further, let $k_t = k_t^A - k_t^B$ and $\gamma = \mu/N$. The left-hand side of (23) becomes

$$\frac{1}{N} \sum_t \frac{n_t k_t}{1 + \gamma k_t} = \frac{1}{N} \sum_{\{t : k_t \neq 0\}} \frac{n_t}{\gamma + 1/k_t}. \tag{24}$$

This expression has poles at the points $\gamma = -1/k_t$ for $k_t \neq 0$. Since the derivative with respect to γ of the expression is

$$-\frac{1}{N} \sum_{\{t : k_t \neq 0\}} \frac{n_t}{(\gamma + 1/k_t)^2} < 0,$$

it follows that, between poles, the function decreases continuously from $+\infty$ to $-\infty$ as γ increases. Consequently, (23) has a unique solution (for μ) between each successive pair of points of the form $-N/k_t$. However, in order that the probabilities (20) should all be nonnegative, we require that $1 + \gamma k_t > 0$ for all t. The condition is trivially satisfied if $k_t = 0$. If $k_t > 0$, the condition is that $\gamma > -1/k_t$; if $k_t < 0$, that $\gamma < -1/k_t$. If there are some negative and some positive k_t, it follows that γ must be greater than $-1/k_t$ for the largest positive k_t and smaller than $-1/k_t$ for the negative k_t that is greatest in absolute value. Since these are the locations of adjacent poles, it follows that there is exactly one admissible solution for μ. This solution can be found easily by one-dimensional search methods, with the admissible region defined by the two bounds, which therefore bracket the desired solution. See Press, Flannery, Teukolsky, and Vetterling (1986) for discussion of search methods that rely on bracketing.

We assumed earlier that A is dominated by B in the sample. This requirement is incompatible with a situation in which all the k_t are nonpositive and some negative. Consider then the case in which they are all nonnegative. This implies that every term of the left-hand side of (23) is nonnegative, so that the only way in which the constraint can be satisfied is by setting $p_t = 0$ for every observation for which $k_t \neq 0$. Formally, nonnegative probabilities now require only that γ should be greater than the least negative $-1/k_t$. In the interval from this lower bound to $+\infty$, the expression (24) decreases monotonically from $+\infty$ at the pole to 0 at $+\infty$. The only solution to (23)

in that interval thus corresponds to $\gamma = \infty$, or, equivalently, $\mu = \infty$. By (20), we see that all observations for which $k_t > 0$ have zero probability, which means in turn, of course, that the ELF diverges to minus infinity. It is still possible to maximise the ELF restricted to the contributions from observations for which $k_t = 0$, in which case all the nonzero probabilities p_t are proportional to the corresponding n_t, with values n_t/N^+, where $N^+ = \sum_{\{t:k_t=0\}} n_t$. The set of t with $k_t = 0$ includes all observations for which both y_t^A and y_t^B are greater than y. By considering only values of y within a restricted interval, we can ensure that the set is always non-empty, so that the constraint can be satisfied. In cases like this, no special numerical methods are needed to obtain the solution.

5. Special Case: First-Order Dominance

With uncorrelated samples, the problem of first-order dominance has been exhaustively treated in DD. With correlated samples, however, it is useful to specialise the results of the preceding section to the case in which $S = 1$, because then an analytic solution exists, which of course greatly simplifies numerical implementations.

When $S = 1$, the expression $(y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}$ is zero whenever both $y_t^A \leq y$ and $y_t^B \leq y$ or both $y_t^A \geq y$ and $y_t^B \geq y$. If $y_t^A < y$ and $y_t^B > y$, then the expression is equal to $+1$; if $y_t^A > y$ and $y_t^B < y$, it is equal to -1. Let us denote by N^{++} the number of observed couples with $y_t^A \leq y$ and $y_t^B \leq y$, by N^{--} the number with $y_t^A > y$ and $y_t^B > y$, by N^{+-} the number with $y_t^A \leq y$ and $y_t^B > y$, and by N^{-+} the number with $y_t^A > y$ and $y_t^B \leq y$. Then the condition (23) becomes

$$\frac{N^{+-}}{N + \mu} - \frac{N^{-+}}{N - \mu} = 0. \tag{25}$$

If both N^{+-} and N^{-+} are zero, condition (25) is vacuously satisfied, naturally enough, since in this case $F_A(y) = F_B(y) = N^{++}/N$. If $N^{+-} = 0$ and $N^{-+} > 0$, then A is not dominated by B in the sample, contrary to our assumption. If $N^{-+} = 0$ and $N^{+-} > 0$, the situation is like the one considered in the previous section in which all the k_t are nonnegative. Indeed, (25) can be satisfied only in the limit when $\mu \to \infty$. This implies that only the observations in the sets counted by N^{++} and N^{--} have (equal) positive probability.

In the general case in which N^{+-} and N^{-+} are both nonzero, we may solve (25) to find that

$$\mu = \frac{N(N^{+-} - N^{-+})}{N^{+-} + N^{-+}}.$$

We can then see from (20) that, for those observations counted by N^{++} or N^{--}, $p_t = n_t/N$, while, for those counted by N^{+-},

$$p_t = \frac{n_t}{N + \mu} = \frac{n_t(N^{+-} + N^{-+})}{2NN^{+-}}, \tag{26}$$

and, for those counted by N^{-+},

$$p_t = \frac{n_t}{N - \mu} = \frac{n_t(N^{+-} + N^{-+})}{2NN^{-+}}.$$

This constitutes an exact solution to the problem of maximising the ELF in this special case, analogous to that presented for uncorrelated samples in DD. In particular, the solution always exists and is unique.

6. Relation to Intersection-Union Test

In Kaur, Prakasa Rao, and Singh (1994), henceforth KPS, an intersection-union test is proposed for testing restricted stochastic dominance at first order. See also Howes (1993). The test statistic is the minimum over all points of the pooled sample of an asymptotic t statistic for the hypothesis that the values of the CDFs of the two distributions are equal at that point. DD show that, with uncorrelated samples, under the tested null and local alternatives to it, this minimum t statistic is asymptotically equivalent to the signed square root of the empirical likelihood-ratio (ELR) statistic given by twice the difference between the unconstrained maximum of the ELF and the ELF resulting from solving the problem (5). Indeed, they find that, in the cases they consider, the t statistic and the (square-root) ELR statistic are very close numerically.

It is natural to extend the KPS procedure to higher orders of stochastic dominance by replacing the CDFs by dominance functions for the chosen order. In order to compute the asymptotic t statistics needed for the KPS test, it is necessary to estimate the variance of the values of the two empirical dominance functions at all points of the pooled sample. Distribution-free estimates of the variances and covariances of empirical distribution functions, and also empirical dominance functions, are given in Davidson and Duclos (2000); see also the proof of Lemma 2 in the Appendix. These can be used to provide convenient implementations of an intersection-union test for which the statistic is the minimum t statistic of the form

$$t(y) = \frac{\hat{D}_A^S(y) - \hat{D}_B^S(y)}{\left(\widehat{\text{Var}}(\hat{D}_A^S(y)) + \widehat{\text{Var}}(\hat{D}_B^S(y))\right)^{1/2}} \tag{27}$$

as y varies over all points of the pooled sample Y.

The maximisation in problem (5) over y, the point at which the two CDFs are equal, is solved by DD by a brute-force search over all the points of the pooled sample. With higher-order dominance, such a search would be reasonably costly, since a nonlinear optimisation problem must be solved for each point. But it is not obvious that there is a much better way of minimising the t statistic (27) than searching over all points of the pooled sample. However, these t statistics are easy to compute and involve no nonlinear optimisation, and so a brute-force search is no more costly than the similar searches used by DD.

As shown in Lemma 2 of the Appendix, at or near a point y at which $D_A^S(y) = D_B^S(y)$ for the two underlying distributions, the signed square root of the ELR statistic and the statistic (27), computed using DD's distribution-free variance estimates, are asymptotically equivalent. This fact suggests a simple approach to minimising the ELR statistic over y, a problem that is equivalent to the maximisation over y in problem (5). First, one finds, by brute-force search (or otherwise), the point $y^* \in Y$ at which the right-hand side of (27) attains its minimum. We ignore the zero-probability possibility that the minimum is attained at more than one point. Then one solves the part of problem (5) involving the choice of the probabilities p_t^A and p_s^B. Next, the estimated dominance functions (4) for the two distributions are compared over their full range. If $\hat{D}_A^S(y) > \hat{D}_B^S(y)$ for all $y \in Y$ except for y^*, at which the inequality becomes an equality, then stop. If there is a point $y_* \neq y^*$ at which $\hat{D}_A^S(y) < \hat{D}_B^S(y)$, then set $y^* = y_*$, re-solve the minimisation (5) with respect to the probabilities at the new y^*, and so on until the desired condition is satisfied. One may expect that at most two or three iterations of this procedure will be necessary. Since the number of points in Y is finite, the number of iterations cannot exceed the size of the pooled sample.

A by-product of the proof of Lemma 2 is approximate expressions for the quantities μ and ν that solve the equations (15). It is easy to estimate these approximate expressions using the observed data, and using the result as the starting point for Newton's method can speed the convergence of the method.

A similar procedure for finding the point y^* can be used with correlated samples. In such cases, it is necessary in the denominator of the t statistic to take account of the covariance of the estimated dominance functions. Estimates of such a covariance are also provided in the Davidson and Duclos (2000) paper. Lemma 3 in the Appendix proves the asymptotic equivalence of the t statistic with estimated covariance and the ELR statistic found by solving the problem with Lagrangean (19).

7. Bootstrap Tests

The main use of the empirical likelihood approach, here as in DD, is not the computation of the ELR statistic itself, but rather the computation of the probabilities that make the weighted empirical dominance functions satisfy the constraint of non dominance. The process of minimising the ELR statistic over points of the pooled sample leads to empirical dominance functions for which dominance fails at just one point. They therefore represent a state of affairs on the frontier of the null hypothesis. It is shown in DD that the minimum t statistic and the ELR statistic are asymptotically pivotal on this frontier, and the same is true here, for the same reason, namely that the asymptotic distribution of the minimum t statistic on the frontier is standard normal. DD also show that the rejection probability of the test based on either statistic, at any reasonable nominal level, is lower for a DGP that lies in the interior of the null hypothesis than for another one that lies on the frontier. That result, too, carries over to the testing situations examined in this paper.

These remarks justify extending the bootstrap procedure developed in DD to these new testing situations. We now outline the way in which a bootstrap test for the null hypothesis that distribution B does not dominate distribution A can be carried out.

- If the distributions that are to be compared for stochastic dominance are continuous, select an interval in the interior of the joint support such that there is at least one point in each sample above its upper limit and at least one other below its lower limit. The null of the test is then one of non dominance restricted to that interval.

- Compute the usual, unweighted, empirical dominance functions, $\hat{D}_K^S$, $K = A, B$, and evaluate them at all points $y \in Y^\circ$, where Y° is the set of points in the pooled sample inside the interval for which restricted non dominance is to be tested. Unless $\hat{D}_A^S(y) > \hat{D}_B^S(y)$ for all $y \in Y^\circ$, the null is not rejected: Stop here, setting the bootstrap P value equal to 1.

- Compute the t statistics $t(y)$ for all $y \in Y^\circ$, either as in (27) if the samples are uncorrelated, or else with the same numerator as in (27), but with denominator the square root of

$$\widehat{\mathrm{Var}}(\hat{D}_A^S(y)) + \widehat{\mathrm{Var}}(\hat{D}_B^S(y)) - 2\widehat{\mathrm{cov}}(\hat{D}_A^S(y), \hat{D}_B^S(y)). \tag{28}$$

Locate the $y^* \in Y^\circ$ at which $t(y)$ attains its minimum.

- Solve problem (5) for uncorrelated samples, or else the problem with Lagrangean (19) for correlated samples. Construct the weighted empirical dominance functions $\tilde{D}_K^S(y)$, $K = A, B$, for $y \in Y^\circ$, as in (4). Check whether $\tilde{D}_A^S(y) > \tilde{D}_B^S(y)$ for all $y \in Y^\circ$ except y^*. If so, skip the next step.

- Replace y^* by the point y_* at which $\tilde{D}_A^S(y) - \tilde{D}_B^S(y)$ is minimised, and repeat the previous step. Move on to the next step only when the check of the previous step is satisfied at y^*.

- Construct a bootstrap DGP as follows. For uncorrelated samples, bootstrap samples are drawn from the distributions for which the CDFs are the weighted empirical distribution functions $\tilde{D}_K^1$, $K = A, B$. That is, points in a bootstrap sample are drawn by resampling from the observed sample, with unequal probabilities as specified by the solution to the empirical likelihood problem. For correlated samples, a bootstrap sample is drawn by resampling pairs from the observed correlated samples, again with the unequal probabilities given by the empirical likelihood problem.

- For each of B bootstrap replications, draw bootstrap samples of sizes N_A and N_B, and compute the minimum t statistic t_j^*, $j = 1, \ldots, B$, using the bootstrap data, in exactly the same way as $t(y^*)$ was computed using the observed data. Set t_j^* to zero unless there is dominance in the bootstrap data.

- The bootstrap P value is the proportion of the t_j^* that are greater than the original statistic $t(y^*)$. Reject the null of (restricted) non dominance if the bootstrap P value is smaller than the desired significance level.

Unlike most conventional bootstrap P values, the one computed using the above algorithm is not asymptotically distributed uniformly on $[0, 1]$ under the null that the DGP is on the frontier of non dominance. This is because there is a positive probability, asymptotically one-half, that the P value is equal to 1. However, the asymptotic distribution conditional on a P value of less than 1 is uniform on $[0, 0.5]$. Thus, for any conventional significance level, the usual rejection rule, as specified in the last step above, is asymptotically justified.

Most of the statements in the above paragraph follow from standard bootstrap theory applied to a case in which the test statistic is asymptotically pivotal under the null, see Beran (1988). As we saw above, that is the case here for the null that the DGP is on the frontier of non dominance. If the DGP is inside that frontier, then all tests, asymptotic or bootstrap, are asymptotically conservative. The statements regarding the asymptotic distribution of P values greater than a half follow from the fact that, if y_0 is the single point of contact of two dominance functions one of which dominates the other everywhere except at y_0, then $\hat{D}_A^S(y_0) - \hat{D}_B^S(y_0)$ is asymptotically normal with zero expectation. The probability that it takes one sign rather than the other thus tends to a half as the sample size grows. At all other points, the probability of a "wrong" sign tends to zero. Thus the probability of failing to reject on account of finding non dominance in the data tends to a half. The statements above follow from this fact, applied both to the original test and the bootstrap tests.

8. A Few Simulation Experiments

In this section, results of some simulation experiments are reported. The set of experiments is representative of interesting cases, but is far from complete. One reason for this is that the results are very similar indeed to those found by DD for the case of first-order dominance.

Just two configurations are considered. The first is one situated on the frontier of non dominance at second order, where the function D_A^2 and D_B^2 touch at exactly one point in the interior of the joint support. This means, of course that there is no first-order dominance, and the configuration is thus in the interior of the set of configurations in which B does not dominate A at first order. The second configuration has B dominating A at second order, but not at first order. In both cases, the null hypothesis is non dominance of A by B at second order.

Unlike first-order dominance, which is invariant to monotonically increasing transformations applied to both distributions, higher-order dominance is invariant only to increasing affine transformations. Thus a relation of second-order dominance is globally scale invariant, but not locally so – the intensity of poverty matters for second-order relations. The choice made here to consider distributions defined on the $[0, 1]$ interval is thus harmless, but the precise location of crossings or tangencies of the dominance functions matter.

In both configurations studied, each of the distributions A and B has a piecewise linear CDF. Distribution A is in fact just a uniform distribution on $[1/9, 1]$. For

distribution B, there are 8 segments, $[i/9, (i+1)/9]$, $i = 1, \ldots, 8$, to which a total probability of p_i is assigned, the distribution within each segment being uniform, which is what makes the CDF piecewise linear. For the first setup, the probabilities are $p_1 = 0.075$, $p_2 = 0.125$, $p_3 = 0.175$, $p_4 = 0.225$, $p_5 = 0.025$, $p_6 = 0.025$, $p_7 = 0.045$, $p_8 = 0.305$. The CDFs of A and B are shown in the left panel of Figure 3, and the second-order dominance functions in the right panel. Note that the CDFs are equal at the point of tangency of the second-order functions.
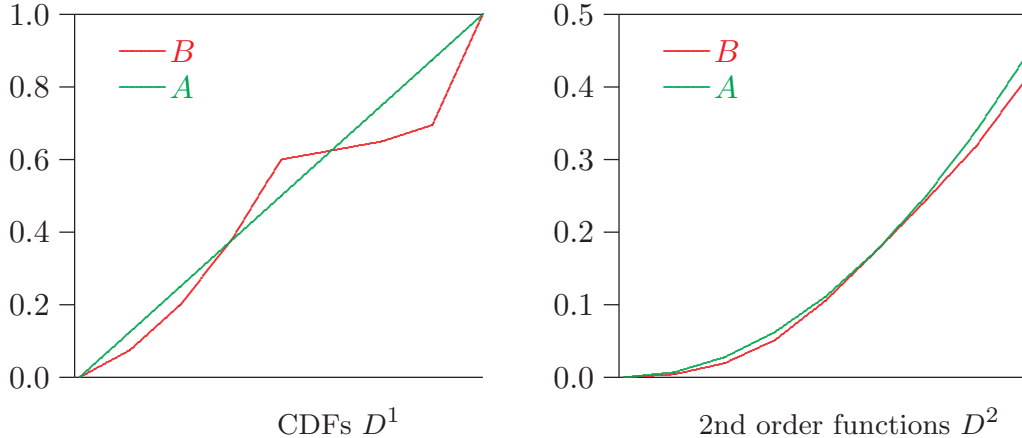

CDFs D^1 2nd order functions D^2

Figure 3. A configuration on the boundary of second-order non dominance

Rejection frequencies were computed for various sizes of independent samples. In all cases, the null hypothesis is that B does not dominate A at second order over the restricted interval $[0.2, 0.9]$. The experiments all comprised 10,000 repetitions, and for the bootstrap tests, 399 bootstrap repetitions were used. Figure 4 shows P value plots of the sort proposed by Davidson and MacKinnon (1998) for $N_A = 32$ and $N_B = 27$ on the left, and for $N_A = 64$ and $N_B = 57$ on the right. These plots show the observed rejection frequency, a simulation-based estimate of the true rejection probability, as a function of the nominal level. A P value plot below the $45°$ line corresponds to under-rejection, above to over-rejection. The sample sizes are chosen to be of comparable orders of magnitude, but unequal. If the sample sizes are the same, largely meaningless discrete jumps appear in the plots. These simulation artefacts are eliminated if the sample sizes are relatively prime.

Although rejection frequencies for nominal levels much above 0.1 are not of great interest in themselves, the P value plot is a useful way to characterise the complete distribution of a statistic. Here, we go out only as far as a level of 0.5, because, beyond that, we run into the problem mentioned in the previous section.

Plots are shown, first, for the asymptotic test based on the minimum t statistic, with critical values given by the right-hand tail of the standard normal distribution. The other plot is for the bootstrap test described above.

The phenomenon of under-rejection remarked in DD is plainly visible with these small sample sizes. However, although for $N_A = 54$ the asymptotic test is still quite noticeably undersized, the bootstrap test is already providing reasonably reliable inference.

In Figure 5, the pattern set for smaller sample sizes continues for larger ones. For $N_A = 512$, there is even some evidence of over-rejection by the bootstrap test at nominal levels greater than those usually of interest.
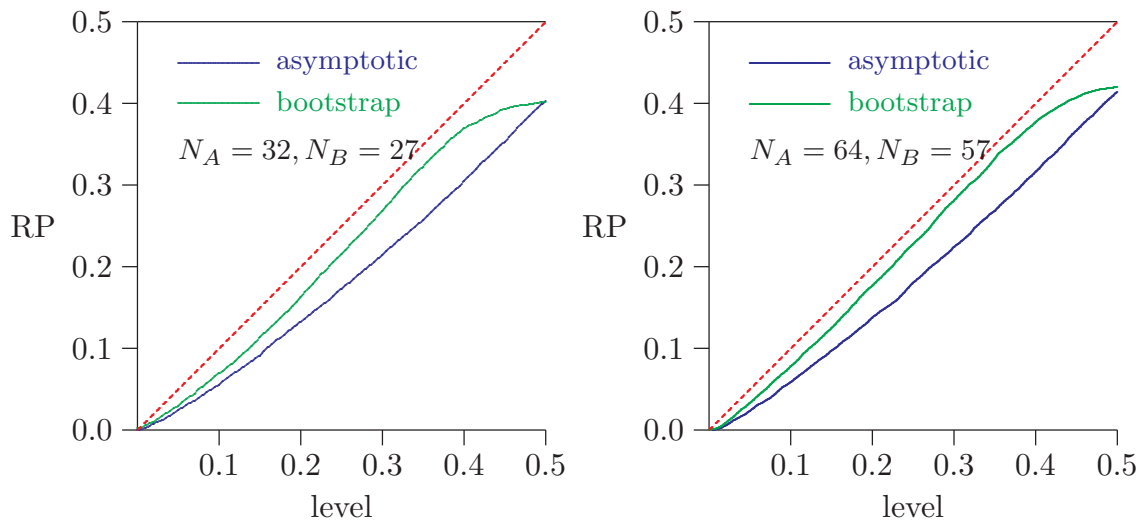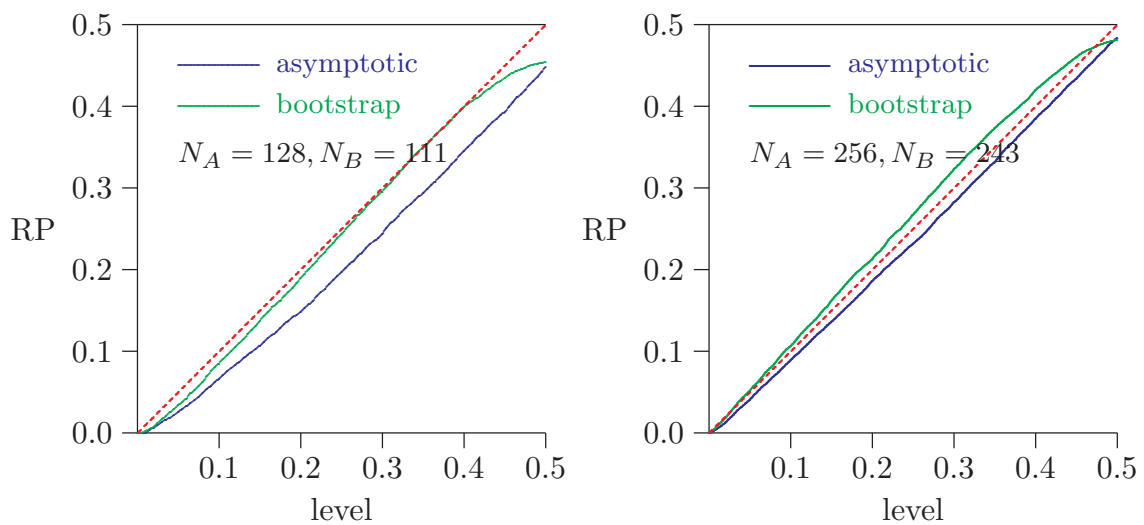

Figure 4. *P* value plots

Figure 5. *P* value plots for larger sample sizes

The next set of experiments looks at power. The configuration of the two distributions is shown in Figure 6, which, like Figure 3, shows the CDFs and the second-order dominance functions for the two distributions. There is second-order dominance of A by B, but not first-order dominance. A desirable test rejects the null of (restricted) second-order non dominance with high probability.
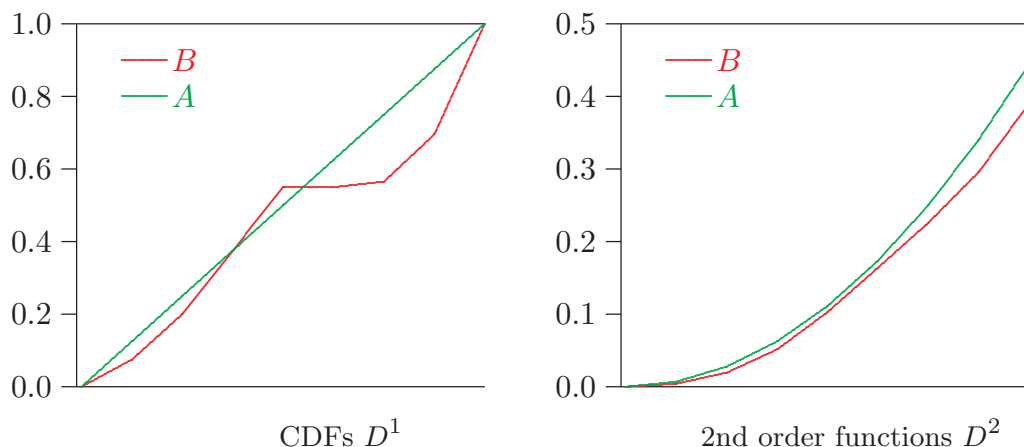
– 17 –

Figure 6. A configuration with second-order dominance

Figure 7 shows P value plots for the sample sizes shown in Figure 4, and Figure 8 for those in Figure 5. For the small sample sizes of Figure 7, there is very little useful power, and the asymptotic test still rejects with probability smaller than the nominal level, even though the null hypothesis is not true. With the larger sample sizes of Figure 8, both tests acquire useful power, but the bootstrap test is considerably more powerful.
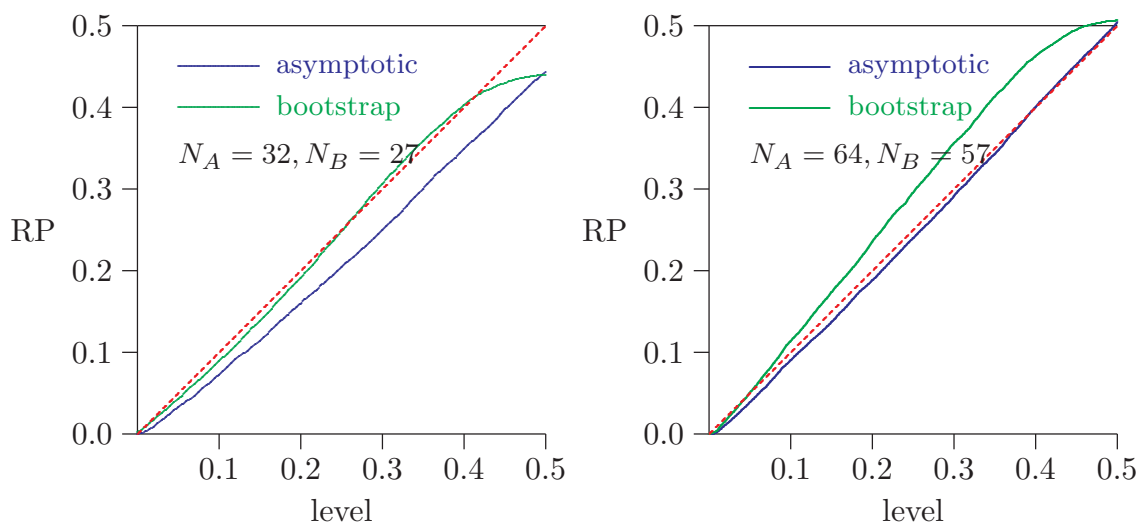

Figure 7. P value plots with second-order dominance

It is of interest to see how well the bootstrap algorithm worked in the experiments just described. The table below shows a number of statistics for the configurations of the experiments. First, the number of times (out of the 10,000 repetitions) that the value y^* at which the asymptotic t statistic attained its minimum did not satisfy the requirement that $\tilde{D}_A^S(y) > D_B^S(y)$ for all $y \neq y^*$, so that other values of y had to be checked – this number is called "moves" in the table. Next the maximum number of iterations needed for the nonlinear optimisation routine, and finally the number of occasions
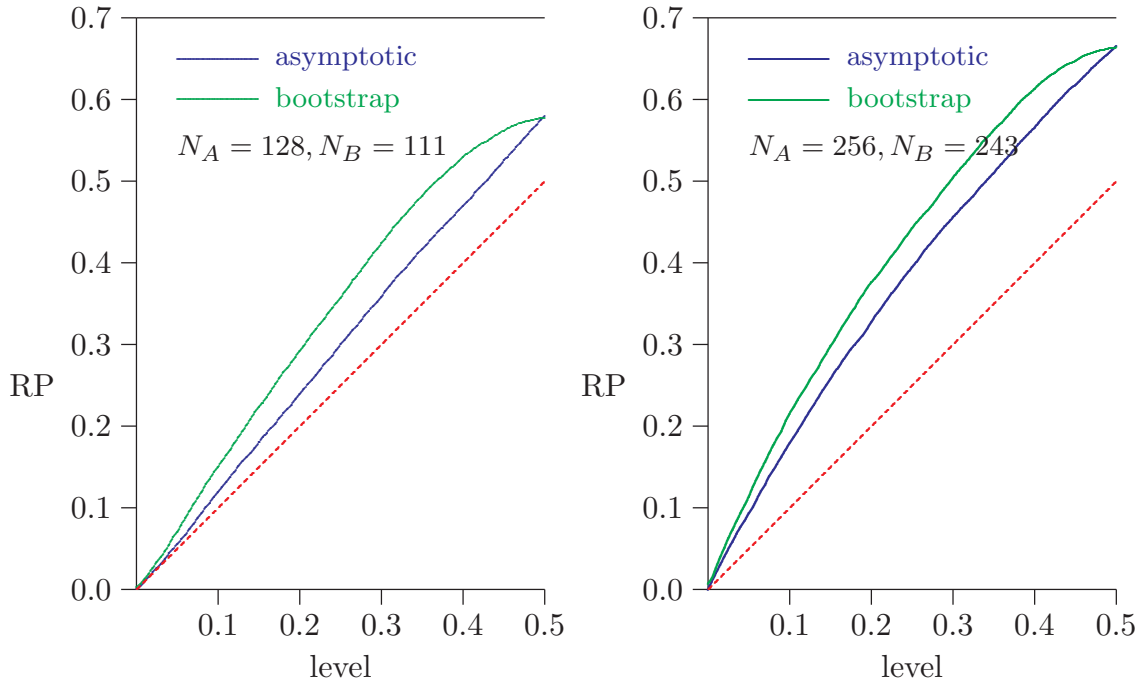
Figure 8. *P* **value plots for larger samples with second-order dominance**

on which a pair (μ, ν) generated by this routine was outside the admissible zone for nonnegative probabilities. The figures given in the table were found when Newton's method was initialised as described above using the approximate expressions in the proof of Lemma 2.

Sample sizes	Null true	Moves	Max iterations	outside zone
32/27	yes	105	7	0
64/57	yes	43	7	1
128/111	yes	25	6	0
256/243	yes	17	2	0
32/27	no	155	8	0
64/57	no	101	8	3
128/111	no	47	8	3
256/243	no	52	3	0

It emerges clearly that the algorithm functions quite efficiently. In the overwhelming majority of cases, minimising the t statistic gives the right value of y at which to construct a bootstrap DGP on the frontier of the null hypothesis, and Newton's method strays outside the admissible zone only very rarely indeed. In none of the experiments was it necessary to try more than six values of y before satisfying the requirement for the bootstrap DGP, and again in the overwhelming majority of cases only one try was enough. The rapid convergence of Newton's method is evident, especially for larger samples.

9. Conclusion

This paper extends the approach of Davidson and Duclos (2006) for testing for stochastic dominance. The null hypothesis of tests based on this approach is that one distribution does not dominate another over some restricted interval. The restriction is necessary whenever the distributions are continous – the tails must not be taken into account for a dominance or non-dominance relation.

It is shown how to test a null hypothesis of the type considered for an arbitrary order of stochastic dominance, and for correlated as well as uncorrelated samples. For orders greater than the first, a nonlinear optimisation problem must be solved, and an algorithm based on Newton's method is given in order to do so. The bootstrap tests proposed can be implemented efficiently, using the fact that a nonlinear problem need be solved only a few times, often just once, in order to set up a suitable bootstrap DGP. Simulations show that the method is useful, in that the bootstrap tests are superior to asymptotic tests as regards both size and power.

Appendix

Lemma 1

We show here that, if the probabilities p_t^A and p_s^B are given by (12) with μ and ν solutions of the equations (13), then $\sum_t p_t^A = 1$ and $\sum_s p_s^B = 1$. From (12) we see that

$$
\begin{aligned}
\sum_t p_t^A &= \sum_t \frac{n_t^A}{N_A - \mu(\nu - (y - y_t^A)_+^{S-1})} \\
&= \frac{1}{N_A - \mu\nu} \sum_t \frac{n_t^A \left[N_A - \mu(\nu - (y - y_t^A)_+^{S-1}) - \mu(y - y_t^A))_+^{S-1} \right]}{N_A - \mu(\nu - (y - y_t^A)_+^{S-1})} \\
&= \frac{1}{N_A - \mu\nu} \left(N_A - \mu \sum_t \frac{n_t^A (y - y_t^A)_+^{S-1}}{N_A - \mu(\nu - (y - y_t^A)_+^{S-1})} \right) \\
&= \frac{1}{N_A - \mu\nu} (N_A - \mu\nu) = 1.
\end{aligned}
$$

The proof for distribution B is exactly similar. ∎

Lemma 2

As the size $N = N_A + N_B$ of the pooled sample Y tends to infinity in such a way that $N_A/N = r$, $0 < r < 1$, then, at any point y in the interior of Y at which $D_A^S(y) = D_B^S(y)$, the signed ELR statistic and the asymptotic t statistic (27) are asymptotically equivalent, in the sense that the difference between them tends to zero in probability as $N \to \infty$. The same result holds if $D_A^S(y) - D_B^S(y)$ is nonzero, but is of the order of $N^{-1/2}$ as $N \to \infty$.

Proof:

In this proof, we will forget the factorial in the definition (2) of $\hat{D}_K^S$, $K = A, B$, in order to lighten notation. It will be easy to see that the statistics we consider are unchanged when this factor, the same for both A and B, is omitted. We write then that $\hat{D}_A^S(y) = (1/N_A) \sum_t n_t^A k_t^A$. The observations are IID, and so this expression is the average of the random variables k_i^A, $i = 1, \ldots, N_A$, where we consider observations singly instead of grouping them by their values. The variance of $\hat{D}_A^S(y)$ is thus the variance of k_i divided by N_A. This variance can be estimated using the formula

$$
\begin{aligned}
N_A \widehat{\mathrm{Var}}(\hat{D}_A^S(y)) &= \frac{1}{N_A} \sum_{i=1}^{N_A} (k_i^A)^2 - \left(\frac{1}{N_A} \sum_{i=1}^{N_A} k_i^A \right)^2 \\
&= \frac{1}{N_A} \sum_t n_t^A (k_t^A)^2 - \left(\frac{1}{N_A} \sum_t n_t^A k_t^A \right)^2.
\end{aligned}
$$

The formula for B is exactly similar. Make the definitions

$$\hat{K}_A = \frac{1}{N_A} \sum_t n_t^A (k_t^A)^2 \qquad \text{and} \qquad \hat{K}_B = \frac{1}{N_B} \sum_s n_s^B (k_s^B)^2.$$

Then the square of the t statistic (27) can be written as

$$t^2 = \frac{(\hat{D}_A^S - \hat{D}_B^S)^2}{\dfrac{1}{N_A}(\hat{K}_A - (\hat{D}_A^S)^2) + \dfrac{1}{N_B}(\hat{K}_B - (\hat{D}_B^S)^2)}$$
$$= \frac{r(1-r)N(\hat{D}_A^S - \hat{D}_B^S)^2}{(1-r)(\hat{K}_A - (\hat{D}_A^S)^2) + r(\hat{K}_B - (\hat{D}_B^S)^2)}, \tag{29}$$

where we omit explicit dependence on y to avoid notational clutter.

Denote by D the limit in probability as $N \to \infty$ of $D_A^S(y)$, and so also of $D_B^S(y)$. Then, since $\hat{D}_K^S(y)$, $K = A, B$, is a root-n consistent estimate of D, we have $\hat{D}_K^S(y) = D + O_p(N^{-1/2})$. It follows as well that $d \equiv N^{1/2}(\hat{D}_A^S(y) - \hat{D}_B^S(y)) = O_p(1)$. Let K_A and K_B be the probability limits as $N \to \infty$ of $\hat{K}_A$ and $\hat{K}_B$ respectively. Then (29) can be written as

$$t^2 = \frac{r(1-r)d^2}{(1-r)K_A + rK_B - D^2} + O_p(N^{-1/2}). \tag{30}$$

From (3) we see that the unconstrained maximum of the ELF is equal to

$$\sum_t n_t^A \log n_t^A + \sum_s n_s^B \log n_s^B - N_A \log N_A - N_B \log N_B.$$

When the constraint that $\sum_A n_t^A k_t^A = \sum_B n_s^B k_s^B$ is imposed, the maximised ELF becomes

$$\sum_t n_t^A \log p_t^A + \sum_s n_s^B \log p_s^B,$$

where the probabilities p_t^A and p_s^B are given by (16). This constrained maximum is therefore equal to

$$\sum_t n_t^A \log n_t^A + \sum_s n_s^B \log n_s^B - \sum_t n_t^A \log(N_A - \mu(\nu - k_t^A)) - \sum_s n_s^B \log(N_B + \mu(\nu - k_s^B))$$
$$= \sum_t n_t^A \log n_t^A + \sum_s n_s^B \log n_s^B - N_A \log N_A - N_B \log N_B$$
$$- \sum_t n_t^A \log\left(1 - \frac{\mu(\nu - k_t^A)}{N_A}\right) - \sum_s n_s^B \log\left(1 + \frac{\mu(\nu - k_s^B)}{N_B}\right).$$

The difference between the unconstrained and constrained maxima is twice the ELR statistic:

$$\tfrac{1}{2}\text{ELR} = \sum_t n_t^A \log\left(1 - \frac{\mu(\nu - k_t^A)}{N_A}\right) + \sum_s n_s^B \log\left(1 + \frac{\mu(\nu - k_s^B)}{N_B}\right). \qquad (31)$$

Now $\tilde{D}_A^S(y)$, as defined in (4), is also a root-n consistent estimator of D, so that $\hat{D}_A^S(y) - \tilde{D}_A^S(y) = O_p(N^{-1/2})$. Aside from the factorial factor, this difference is

$$\sum_t n_t^A k_t^A \left(\frac{1}{N_A} - \frac{1}{N_A - \mu(\nu - k_t^A)}\right) = -\frac{\mu}{N_A} \sum_t \frac{n_t^A k_t^A (\nu - k_t^A)}{N_A - \mu(\nu - k_t^A)}.$$

By (11), the quantity ν is a weighted average of either N_A or N_B quantities, and so is of order 1 in probability. In fact, (11) implies that $\nu = D + O_p(N^{-1/2})$. Therefore the sum on the right-hand side above is also of order 1 in probability, from which it follows that $\mu/N = O_p(N^{-1/2})$, or, equivalently, $\mu = O_p(N^{1/2})$. If we write $m = N^{-1/2}\mu$, then it follows that $m = O_p(1)$.

If we Taylor expand the first logarithm in (31), we find that

$$\sum_t n_t^A \log\left(1 - \frac{\mu(\nu - k_t^A)}{N_A}\right) = \sum_t n_t^A \left(-\frac{\mu(\nu - k_t^A)}{N_A} - \frac{\mu^2(\nu - k_t^A)^2}{2N_A^2}\right) + O_p(N^{-1/2})$$

$$= -\mu\nu + \mu\hat{D}_A^S - \frac{\mu^2\nu^2}{2N_A} + \frac{\mu^2\nu}{N_A}\hat{D}_A^S - \frac{\mu^2}{2N_A}\hat{K}_A + O_p(N^{-1/2})$$

$$= -\mu\nu + \mu\hat{D}_A^S - \frac{m^2}{2r}D^2 + \frac{m^2}{r}D^2 - \frac{m^2}{2r}K_A + O_p(N^{-1/2})$$

$$= -\mu\nu + \mu\hat{D}_A^S - \frac{m^2}{2r}(K_A - D^2) + O_p(N^{-1/2}).$$

An exactly similar calculation shows that

$$\sum_s n_s^B \log\left(1 + \frac{\mu(\nu - k_s^B)}{N_B}\right) = \mu\nu - \mu\hat{D}_B^S - \frac{m^2}{2(1-r)}(K_B - D^2) + O_p(N^{-1/2}).$$

Thus (31) becomes

$$\tfrac{1}{2}\text{ELR} = md - \frac{m^2}{2r(1-r)}\left((1-r)K_A + rK_B - D^2\right). \qquad (32)$$

We now wish to express the random variable m in terms of d. For this, note that, from (11),

$$\nu = \frac{1}{N_A}\sum_t n_t^A k_t^A \left(1 - \frac{\mu(\nu - k_t^A)}{N_A} + O_p(N^{-1})\right)$$

$$= \hat{D}_A^S\left(1 - N^{-1/2}\frac{m\nu}{r}\right) + N^{-1/2}\frac{m}{r}\hat{K}_A + O_p(N^{-1}) \qquad (33)$$

But we also have

$$\nu = \hat{D}_B^S \left(1 + N^{-1/2}\frac{m\nu}{1-r}\right) - N^{-1/2}\frac{m}{1-r}\hat{K}_B + O_p(N^{-1}), \tag{34}$$

so that, on subtracting (34) from (33) and multiplying by $N^{1/2}$, we obtain

$$0 = d - \frac{1}{r(1-r)}m(D^2 - (1-r)K_A - rK_B) + O_p(N^{-1/2}),$$

whence

$$m = \frac{r(1-r)}{(1-r)K_A + rK_B - D^2}d + O_p(N^{-1/2}). \tag{35}$$

Substituting (35) into (32) gives

$$\tfrac{1}{2}\text{ELR} = \frac{r(1-r)d^2}{(1-r)K_A + rK_B - D^2} - \frac{1}{2}\frac{r(1-r)d^2}{(1-r)K_A + rK_B - D^2} + O_p(N^{-1/2})$$

so that

$$\text{ELR} = \frac{r(1-r)d^2}{(1-r)K_A + rK_B - D^2} + O_p(N^{-1/2}). \tag{36}$$

From (30) and (36) we see that the difference between t^2 and ELR is $O_p(N^{-1/2})$, which is the asymptotic equivalence we wished to demonstrate. ∎

Lemma 3

As the sample size N tends to infinity, at any point y in the interior of the joint support of the distributions A and B at which $D_A^S(y) - D_B^S(y) = O(N^{-1/2})$, the signed ELR statistic and the asymptotic t statistic with denominator the square root of (28), based on a correlated sample jointly drawn from the two distributions, are asymptotically equivalent in the same sense as in Lemma 2.

Proof:

The proof follows lines similar to that of Lemma 2, but it is much simpler. We use the notation of section 4, with $k_t = (y - y_t^A)_+^{S-1} - (y - y_t^B)_+^{S-1}$. The squared t statistic can then be written as follows, where as before we omit the explicit dependence on y.

$$t^2 = \frac{(\hat{D}_A^S - \hat{D}_B^S)^2}{\widehat{\text{Var}}(\hat{D}_A^S) + \widehat{\text{Var}}(\hat{D}_B^S) - 2\widehat{\text{cov}}(\hat{D}_A^S, \hat{D}_B^S)}. \tag{37}$$

Ignoring the factorial factor as before, we have that

$$\hat{D}_A^S - \hat{D}_B^S = \frac{1}{N}\sum_t n_t k_t,$$

and so the denominator of (37) can be obtained by use of the formula

$$N\widehat{\text{Var}}(\hat{D}_A^S - \hat{D}_B^S) = \frac{1}{N}\sum_t n_t k_t^2 - \left(\frac{1}{N}\sum_t n_t k_t\right)^2.$$

Let $N^{1/2}(\hat{D}_A^S - \hat{D}_B^S) \equiv d$. Then $d = O_p(1)$ as $N \to \infty$, because D_A^S and D_B^S are root-n consistent. Further, let K be the limit in probability of $1/N\sum_t n_t k_t^2$. Then (37) becomes

$$t^2 = d^2/K + O_p(N^{-1/2}). \tag{38}$$

The unconstrained maximum of the ELF is $\sum_t n_t \log n_t - N \log N$. The constrained maximum is $\sum_t n_t \log p_t$, with the p_t given by (20). We have $p_t = n_t/(N + \mu k_t)$, so that the constrained maximum is

$$\sum_t n_t \log n_t - \sum_t n_t \log(N + \mu k_t) = \sum_t n_t \log n_t - N \log N - \sum_t n_t \log\left(1 + \frac{\mu k_t}{N}\right).$$

It is easy to show, as in the proof of Lemma 2, that $\mu = O_p(N^{1/2})$, and so we see that

$$\text{ELR} = 2\sum_t n_t \log\left(1 + \frac{\mu k_t}{N}\right) = 2\frac{\mu}{N}\sum_t n_t k_t - \frac{\mu^2}{N^2}\sum_t n_t k_t^2 + O_p(N^{-1/2})$$

$$= 2md - Km^2 + O_p(N^{-1/2}), \tag{39}$$

where $m \equiv N^{-1/2}\mu = O_p(1)$.

The constraint (23) can be written as $\sum_t n_t k_t/(N + \mu k_t) = 0$. Taylor expansion of this gives

$$0 = N^{-1/2}d - N^{-1/2}Km + O_p(N^{-1/2}), \qquad \text{or} \qquad d = Km + O_p(N^{-1/2}).$$

Consequently (39) becomes

$$\text{ELR} = d^2/K + O_p(N^{-1/2}).$$

With (38), this completes the proof ∎

References

Beran, R. (1988). "Prepivoting test statistics: A bootstrap view of asymptotic refinements", *Journal of the American Statistical Association*, **83**, 687–697.

Brown, B. W. and W. Newey (2002). "Generalized method of moments, efficient bootstrapping, and improved inference", *Journal of Business and Economic Statistics*, **20**, 507–517.

Davidson, R. and J.-Y. Duclos (2000). "Statistical Inference for Stochastic Dominance and the for the Measurement of Poverty and Inequality", *Econometrica*, **68**, 1435–1465.

Davidson, R., and J.-Y. Duclos (2006). "Testing for Restricted Stochastic Dominance", working paper, Department of Economics, McGill University.

Davidson, R. and J. G. MacKinnon (1998). "Graphical Methods for Investigating the Size and Power of Hypothesis Tests" *The Manchester School*, **66**, 1–26.

Howes, S. (1993). "Asymptotic Properties of Four Fundamental Curves of Distributional Analysis", Unpublished paper, STICERD, London School of Economics.

Kaur, A., B. L. Prakasa Rao, and H. Singh (1994). "Testing for Second-Order Stochastic Dominance of Two Distributions", *Econometric Theory*, **10**, 849–866.

Owen, A. B. (2001). *Empirical Likelihood*, Chapman and Hall.

Press, W. H., B. P. Flannery, S. A. Teukolsky, et W. T. Vetterling (1986). *Numerical Recipes*, Cambridge University Press, Cambridge.