A Monte Carlo Investigation of Some Tests for Stochastic Dominance

Y.K. Tse
School of Economics and Social Sciences
Singapore Management University
Singapore 259756

and

X.B. Zhang
Department of Econometrics and Business Statistics
Monash University
Clayton VIC 3800
Australia

July 2002

Abstract: This paper compares the performance of several tests for stochastic dominance up to order three using Monte Carlo methods. The tests considered are the Davidson and Duclos (2000) test, the Anderson test (1996) and the Kaur, Rao and Singh (1994) test. We find that the Davidson-Duclos test appears to be the best. The Kaur-Rao-Singh test is overly conservative and does not compare favorably against the Davidson-Duclos and Anderson tests in terms of power.

Keywords: Burr distribution, Income distribution, Monte Carlo method, Portfolio investment, Stochastic dominance, Union-intersection test

JEL Classification: C12, D31, G11

Corresponding author: Y.K. Tse, School of Economics and Social Sciences, Singapore Management University, 469 Bukit Timah Road, Singapore 259756, Tel: +65 6822 0257, Fax: +65 6822 0833, Email: yktse@smu.edu.sg.

1 Introduction

The comparison of distribution functions has long been a research topic of major interests with applications to income distribution and portfolio investment. Studies on income distribution focuses on the issue of income inequality as measured by rank dominance and Lorenz dominance. In investment decision comparison between asset classes can be made under the criteria of stochastic dominance. Comparisons of the distribution functions and their cumulative functions over the support of the distributions are required to draw conclusions concerning the existence of dominance.

The early work of Beach and Davidson (1983) examined dominance at the first order. Later work by Bishop, Formby and Thistle (hereafter BFT) (1992) extended the comparison to the second order. Recently, several methods have been proposed for testing for stochastic dominance. These tests may be broadly divided into two groups. The first group relies on the comparison of the distributions at a finite number of grid points. The papers by Anderson (1996) and Davidson and Duclos (2000) are along this line. The second group propose the use of the inf or sup statistics over the support of the distributions. The tests due to McFadden (1989) and Kaur, Rao and Singh (hereafter KRS) (1994) belong to this group.

Although there have been a number of different test procedures for stochastic dominance in the literature, little is known about the relative performance of these tests. For the implementation of these tests there are often some questions that remain to be answered. For example, many tests are based on the values of certain statistics calculated at some grid points. The selection of such grid points is, however, often by rule of thumb. Also, the critical values are usually based on the assumption of independence of the grid statistics. As the grid statistics are often computed from the same sample they are not independent. Thus, the empirical size of the test may not be reliable. Finally, little is known about the relative power of these tests, which is an important issue for empirical applications. In this paper we consider some of the tests for stochastic dominance available in the literature. Specifically, we consider the Davidson-Duclos test, the Anderson test and the KRS test. We compare the performance of these tests using Monte Carlo methods, which will hopefully throw some light on the above issues.

The balance of this paper is as follows. In Section 2 we describe the application of the tests examined in this paper. The Monte Carlo results are reported in Section 3. The Davidson-Duclos test appears to be the best. The KRS test is overly conservative and does not compare favorably against the Davidson-Duclos test in power. Some conclusions are provided in Section 4.

2 Tests for Stochastic Dominance

Consider a random sample of N_Y observations y_i, $i = 1, ..., N_Y$, from a population with distribution function $F_Y(.)$. Define $D_Y^1(x) = F_Y(x)$ and let[1]

$$D_Y^s(x) = \int_0^x D_Y^{s-1}(u)\, du,$$

for any integer $s \geq 2$. Suppose z_i, $i = 1, ..., N_Z$, are a random sample from another population with distribution function $F_Z(.)$. We define $D_Z^s(x)$ analogously. It can be shown that

$$D_i^s(x) = \frac{1}{(s-1)!} \int_0^x (x-u)^{s-1} dF_i(u), \quad i = Y, Z.$$

Y is said to dominate Z stochastically at order s if $D_Z^s(x) \geq D_Y^s(x)$ for all $x \geq 0$, with strict inequality for some x.[2] If this is true, we write $Y \succ_s Z$.

If the correlation between Y and Z is to be taken into account, we assume $N_Y = N_Z = N$ and let (y_i, z_i) be a paired observation (such as returns of two funds in the

[1] Without loss of generality we assume the observations are nonnegative.

[2] We use Y and Z to denote the random variables representing the distributions of the populations. If Y and Z are returns in investments, "Y stochastically dominates Z" means that Y is preferred to Z under some conditions. If Y and Z are incomes, the first-order dominance is equivalent to rank dominance and the second-order dominance is equivalent to generalized-Lorenz dominance.

same period). Davidson and Duclos (2000) considered the following sample statistics:[3]

$$\hat{D}_Y^s(x) = \frac{1}{N(s-1)!}\sum_{i=1}^{N}(x-y_i)_+^{s-1},$$

$$\hat{D}_Z^s(x) = \frac{1}{N(s-1)!}\sum_{i=1}^{N}(x-z_i)_+^{s-1},$$

$$\hat{V}_Y^s(x) = \frac{1}{N}\left[\frac{1}{((s-1)!)^2}\frac{1}{N}\sum_{i=1}^{N}(x-y_i)_+^{2(s-1)} - \hat{D}_Y^s(x)^2\right],$$

$$\hat{V}_Z^s(x) = \frac{1}{N}\left[\frac{1}{((s-1)!)^2}\frac{1}{N}\sum_{i=1}^{N}(x-z_i)_+^{2(s-1)} - \hat{D}_Z^s(x)^2\right],$$

$$\hat{V}_{YZ}^s(x) = \frac{1}{N}\left[\frac{1}{((s-1)!)^2}\frac{1}{N}\sum_{i=1}^{N}(x-y_i)_+^{s-1}(x-z_i)_+^{s-1} - \hat{D}_Y^s(x)\hat{D}_Z^s(x)\right],$$

and proposed the following normalized statistic:

$$T^s(x) = \frac{\hat{D}_Y^s(x) - \hat{D}_Z^s(x)}{\sqrt{\hat{V}^s(x)}},$$

where

$$\hat{V}^s(x) = \hat{V}_Y^s(x) + \hat{V}_Z^s(x) - 2\hat{V}_{YZ}^s(x),$$

for testing the equality of $D_Z^s(x)$ and $D_Y^s(x)$. They showed that, under $H_0 : D_Z^s(x) = D_Y^s(x)$, $T^s(x)$ is asymptotically distributed as a standard normal variate.

When unpaired observations are independently drawn from two populations, we allow $N_Y \neq N_Z$. In this case, $\hat{V}^s(x)$ is computed as $\hat{V}_Y^s(x) + \hat{V}_Z^s(x)$ and the asymptotic normality result still holds. In this paper we only consider the case where the observations from the two populations are independent.[4]

To test for stochastic dominance, H_0 has to be examined for the full support. This, of course, is empirically impossible. A compromise is to test H_0 for a selected finite number of values of x. As multiple hypotheses are involved, tests based on multiple

[3] u_+ is defined as $\max(u, 0)$. To simplify the notation we adopt the convention $u_+^s = (u_+)^s$.

[4] The purpose of this assumption is to scale down the Monte Carlo experiment. Many empirical studies in the literature are based on independent samples. See, for example, Beach and Davidson (1983).

comparison have to be adopted. A multiple comparison procedure was proposed by Bishop, Formby and Thistle (1992) by employing the union-intersection test. Following BFT we consider fixed values $x_1, x_2, ..., x_K$, and their corresponding statistics $T^s(x_i)$, for $i = 1, ..., K$. The following hypotheses are defined:[5]

1. $H_0 : D^s_Y(x_i) = D^s_Z(x_i)$ for all x_i,
2. $H_A : D^s_Y(x_i) \neq D^s_Z(x_i)$ for some x_i,
3. $H_{A1} : Y \succ_s Z$,
4. $H_{A2} : Z \succ_s Y$.

The overall null hypothesis H_0 is the logical intersection of several hypotheses (one for each x_i) and the overall alternative hypothesis H_A is the logical union of the corresponding alternative hypotheses. To control for the probability of rejecting the overall null hypothesis, BFT suggested using the studentized maximum modulus statistic with K and infinite degrees of freedom, denoted by M^K_∞. We denote the $(1-\alpha)$ percentile of M^K_∞ by $M^K_{\infty,\alpha}$, which was tabulated by Stoline and Ury (1979). The following decision rules are adopted:

1. If $|T^s(x_i)| < M^K_{\infty,\alpha}$ for $i = 1, ..., K,$ do not reject H_0.
2. If $-T^s(x_i) > M^K_{\infty,\alpha}$ for some i and $T^s(x_i) < M^K_{\infty,\alpha}$ for all i, accept H_{A1}.
3. If $T^s(x_i) > M^K_{\infty,\alpha}$ for some i and $-T^s(x_i) < M^K_{\infty,\alpha}$ for all i, accept H_{A2}.
4. If $T^s(x_i) > M^K_{\infty,\alpha}$ for some i and $-T^s(x_i) > M^K_{\infty,\alpha}$ for some i, accept H_A.

As the test above is based on the asymptotic distribution result of $T^s(x)$ derived by Davidson and Duclos (2000), we call it the DD test. This procedure was earlier

[5] Note that the equality and inequality conditions for $D^s_Y(x)$ and $D^s_Z(x)$ under the different hypotheses are now defined with respect to the finite set of values x_i.

proposed by BFT, in which a different normalized statistic for the point-by-point comparison was used. In this paper, we adopt the statistic due to Davidson and Duclos (2000) due to its generality.

Anderson (1996) suggested an alternative method of estimating the $D^s(x)$ functions by applying the trapezoidal rule to approximate the required integrals. Suppose we divide the data into $K+1$ mutually exclusive intervals with d_i as the length of the ith interval.[6] We define the following matrices:

$$I_f = \begin{pmatrix} 1 & 0 & . & . & . & 0 \\ 1 & 1 & 0 & . & . & 0 \\ 1 & . & 1 & 0 & . & 0 \\ . & . & . & . & . & . \\ 1 & 1 & 1 & . & 1 & 0 \end{pmatrix}$$

and

$$I_F = 0.5 \times \begin{pmatrix} d_1 & 0 & . & . & . \\ d_1+d_2 & d_2 & 0 & . & . \\ d_1+d_2 & d_2+d_3 & d_3 & 0 & . \\ . & . & . & . & . \\ d_1+d_2 & d_2+d_3 & d_3+d_4 & . & d_K \end{pmatrix},$$

where I_f is $K \times (K+1)$ and I_F is $K \times K$. Let p_{ji} be the probability of finding an observation in the ith interval for the jth population, for $j = Y, Z$. We write $p_j = (p_{j1}, ..., p_{j,K+1})'$. To test for the first-order stochastic dominance, we set up the following hypotheses:

$$H_0^1 : I_f(p_Y - p_Z) = 0,$$

$$H_1^1 : I_f(p_Y - p_Z) \leq 0.$$

If H_0^1 is rejected against H_1^1, we conclude that $Y \succ_1 Z$.[7] For the second-order dominance, we consider

$$H_0^2 : I_F I_f(p_Y - p_Z) = 0,$$

$$H_1^2 : I_F I_f(p_Y - p_Z) \leq 0.$$

[6] Thus, the first interval is $(0,\, d_1)$ and the ith interval is $(d_1 + .. + d_{i-1}, d_1 + .. + d_i)$ for $i = 2, ..., K+1$, and the full interval of the data is $(0, d_1 + d_2 + ... + d_{K+1})$.

[7] Note that H_0^1 and H_1^1 state the equality and inequality, respectively, of two K-vectors. The same is true for H_0^2, H_1^2, H_0^3 and H_1^3 below.

Thus, if H_0^2 is rejected against H_1^2, we conclude that $Y \succ_2 Z$. Likewise, for the third-order dominance, we consider

$$H_0^3 : I_F I_F I_f (p_Y - p_Z) = 0,$$

$$H_1^3 : I_F I_F I_f (p_Y - p_Z) \leq 0,$$

from which $Y \succ_3 Z$ if H_0^3 is rejected against H_1^3.[8]

We define $n_j = (n_{j1}, n_{j2}, ..., n_{j,K+1})'$, for $j = Y, Z$, as the vector of frequencies of the sample from population j in the $K+1$ categories. Thus, $\sum_{i=1}^{K+1} n_{ji} = N_j$. Under the null hypothesis that $F_Y = F_Z$, we let $p_j = p = (p_1, ..., p_{K+1})'$, for $j = Y, Z$. We denote

$$v = \frac{n_Y}{N_Y} - \frac{n_Z}{N_Z}$$

and

$$\Omega = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & . & . & -p_1 p_{K+1} \\ -p_2 p_1 & p_2(1-p_2) & . & . & -p_2 p_{K+1} \\ . & . & . & . & . \\ . & . & . & . & . \\ -p_{K+1} p_1 & -p_{K+1} p_2 & . & . & p_{K+1}(1-p_{K+1}) \end{pmatrix}.$$

Thus, n_j/N_j is asymptotically distributed as $N(p, \Omega/N_j)$, for $j = Y, Z$.[9] From this we have[10]

$$v \underset{D}{\rightarrow} N(0, m\Omega),$$

where

$$m = \frac{N_Y + N_Z}{N_Y N_Z}.$$

Writing H_s for I_f, $I_F I_f$ or $I_F I_F I_f$, respectively, for the tests for the first-, second- and third-order dominance, we obtain

$$H_s v \underset{D}{\rightarrow} N(0, m H_s \Omega H_s').$$

[8]Higher-order stochastic dominance can be tested similarly. But we shall stop at the third order.

[9]This result holds when $N_j \to \infty$, with $N_j p_i > 5$ for $i = 1, ..., K+1$. See Anderson (1996) for the details.

[10]We write $\underset{D}{\rightarrow}$ to denote convergence in distribution.

Ω can be estimated by replacing p by

$$\hat{p} = \frac{n_Y + n_Z}{N_Y + N_Z}.$$

We denote this estimate by $\hat{\Omega}$. Denoting $H_s \upsilon(i)$ as the ith element of $H_s \upsilon$ and $mH_s\hat{\Omega}H_s'(i,i)$ as the ith diagonal element of $mH_s\hat{\Omega}H_s'$, we define

$$A_i^s = \frac{H_s \upsilon(i)}{\sqrt{mH_s\hat{\Omega}H_s'(i,i)}}, \qquad i = 1, ..., K,$$

which is asymptotically distributed as $N(0,1)$. Allowing A_i^s to take the role of $T^s(x_i)$, we obtain an alternative set of tests for stochastic dominance. The decision rules stated above for the $T^s(x_i)$ statistics can be applied to A_i^s. We shall call this test the A test. As before, for each order of stochastic dominance there are four possible outcomes. We shall adopt the same terminology for the various hypotheses (namely, H_0, H_A, H_{A1} and H_{A2}) as in the case of the DD test.[11]

The DD and A tests depend on a predetermined set of grid points x_i. The arbitrariness of the grid points and the choice of the number of points K are undesirable features of the tests. Are the powers of the tests affected by the choice of K? As the A tests are based on trapezoidal approximations of the integrals, are the size and power of the tests compromised? It should be noted that the use of the M_∞^K statistic in providing the critical values for the DD (or the A) tests is valid only when the $T^s(x_i)$ (or A_i^s) statistics are independent. As pointed out by Sidak (1967) and Hochberg (1974), the studentized maximum modulus test is conservative (i.e., the nominal size overstates the actual size) if these statistics are not independent. As both the $T^s(x_i)$ and A_i^s statistics are correlated across different grid points, it is important to investigate how the performance of the tests is affected. In the next section we shall provide some answers to these questions through a Monte Carlo experiment.

[11]It should be noted that the A test applies only to independent samples from two different populations. It does not have the flexibility of the DD test to incorporate paired observations.

Kaur, Rao and Singh (1994) proposed a test that does not require the set-up of grid points. Their test can be used to compare the following hypotheses: $H_0 : Y \not\succ_s Z$ against $H_1 : Y \succ_s Z$. By reversing Y and Z, the following hypotheses can also be compared: $H_0' : Z \not\succ_s Y$ against $H_1' : Z \succ_s Y$.

Assuming $[0, b]$ is the support of Y and Z, we define

$$T_M^s = \inf_{0 \le x \le b} \{T^s(x)\},$$

$$T_{M^*}^s = \inf_{0 \le x \le b} \{-T^s(x)\}.$$

Thus, these statistics are evaluated over the full support of the population rather than at certain grid points. Denoting z_α as the $(1 - \alpha)$ percentile of the standard normal distribution, the following decision rules can be applied at the level of significance α:

1. If $T_{M^*}^s < z_\alpha$ do not reject H_0.
2. If $T_{M^*}^s > z_\alpha$ accept H_1.
3. If $T_M^s < z_\alpha$ do not reject H_0'.
4. If $T_M^s > z_\alpha$ accept H_1'.

We call this test procedure the KRS test.[12] KRS showed that their test is consistent and has an upper bound α on the asymptotic size. It appears to be rather stringent to require the inf statistic to exceed z_α for the rejection of H_0 and H_0'. Thus, it would be interesting to see how close the empirical size of the test is to its nominal value. Whether the power of the test is compromised due to the stringent rejection criterion is also an important issue. We shall examine these questions in the next section.

[12] Kaur, Rao and Singh (1994) considered the special case of $s = 2$, i.e., the case of the second-order stochastic dominance. The proposed inf is taken over a statistic that is asymptotically equivalent to $T^2(x)$. An algorithm was proposed to find the minimum of the statistics over the support of the distributions. An extension of their algorithm to $s = 3$ is, however, intractable. In this paper, we use grid search to find the minimum. Following KRS's convention we take $T^s(x)$ as zero when x is less than the minimum observation of the combined sample.

We conclude this section by making references to some other tests for stochastic dominance not studied in our Monte Carlo experiment. First, a test based on $\sup_x\{\hat{D}_Y^s(x) - \hat{D}_Z^s(x)\}$ was proposed by McFadden (1989). However, the asymptotic distribution of the test statistic for $s \geq 2$ is analytically intractable, although Monte Carlo methods may be used to provide some estimates for the critical values. Maasoumi and Heshmati (2000) applied McFadden's approach to study income distribution in Sweden. They estimated the critical values of their test statistics using bootstrap method. Second, a recent paper by Dardanoni and Forcina (1999) examined the use of the distance statistic for making comparisons among more than two populations. The proposed statistic involves testing for joint inequality. As shown by Wolak (1989), the statistic is asymptotically distributed as a chi-bar-squared random variable.[13] This approach has been applied earlier in studies on stochastic dominance by Xu (1997), Xu and Osberg (1998) and Fisher, Wilson and Xu (1998). A recent paper by Barret and Donald (2001) provides some limited Monte Carlo results where the p-values of the joint-inequality tests were estimated using simulation. Due to the computational complexity of these two methods, they will not be included in our Monte Carlo study.

3 Monte Carlo Results

To sample observations from a known population we need to select a distribution function. Following Dardanoni and Forcina (1999) we adopt the Burr distribution, which has been found to fit empirical income data and has a simple parametric form. The distribution function of a Burr distribution with parameters α and β, denoted by $B(\alpha, \beta)$, is given by

$$F(x) = 1 - (1 + x^\alpha)^{-\beta}, \qquad x \geq 0,$$

[13] See Shapiro (1985, 1988) for the properties of the chi-bar-squared distributions.

with $\beta > 1/\alpha > 0$.[14] The inverse of $F(.)$ is given by

$$F^{-1}(u) = [-1 + (1-u)^{-1/\beta}]^{1/\alpha}, \qquad 0 \leq u < 1.$$

Thus, random numbers of the Burr distribution can be generated easily from random numbers of the uniform distribution.

As the support of the Burr distribution does not have an upper limit, it presents an undesirable feature in the simulation. To obtain distributions with finite support, we set the upper limit of the distribution at a point x^*. The modified distribution function of the truncated distribution, denoted by $F^*(.)$, is

$$F^*(x) = \begin{cases} F(x)/F(x^*) & x \leq x^* \\ 1 & x > x^*. \end{cases}$$

Figure 1 shows the distribution functions of two truncated Burr distributions, namely, $B(4.7,\ 0.55)$ and $B(2,\ 0.65)$.[15] Thus, there is no dominance for these two populations at the first order. The $D^s(.)$ functions for these two distributions are presented in Figures 2 and 3, respectively, for $s = 2$ and 3. Thus, $B(4.7,\ 0.55)$ dominates $B(2,\ 0.65)$ at the third order, but not at the second order. These two distributions form the basis of our Monte Carlo experiments.

The number of grid points used in empirical studies is often decided by rule of thumb. Many studies used the deciles for lack of clear guidance. In this paper we consider $K = 6, 10$ and 15. For these numbers the critical values of the studentized maximum modulus statistics are given by Stoline and Ury (1979). Two methods of determining the grid points are examined, namely, grid points that divide the samples into equal number of observations and grid points that are evenly spaced in the range of the sample.[16] The sample size N is taken to be 500, 1000, 2000 and 3000.

[14] This distribution is also known as the Burr Type XII distribution. The simpler terminology, however, is adopted here. See Johnson, Kotz and Balakrishnan (1994) for more details of the properties of this distribution.

[15] We truncate the distribution at approximately the 98th percentile.

[16] In the simulation it is the combined sample from the two populations that is used for determining the grid points. For the equal-probability scheme, we fix x_K as the 99th percentile of the combined sample.

As the DD and A tests depend on the asymptotic distributions of the $T^s(x_i)$ and A_i^s statistics being standard normal, we first examine this approximation. We generate observations from the truncated $B(4.7,\ 0.55)$ distribution. A set of grid points are fixed. These points divide the support of the truncated $B(4.7, 0.55)$ distribution into eleven intervals such that x_{10} is the 99th percentile of the truncated distribution and the first ten intervals are of equal length. Independent samples of the T^s and A^s statistics are generated at each grid point. For each of the samples of T^s and A^s statistics, eight groups are used to construct the χ^2 goodness-of-fit test. We use 5,000 Monte Carlo samples to compute the χ^2.[17] The results are summarized in Table 1. It can be seen that the convergence to normality is rather slow for $s = 1$, especially for the ending grid points. The normality approximation, however, appears to work well for $s \geq 2$ with N larger than 2000.

Table 2 summarizes the empirical relative frequencies of accepting H_0 when the equal-probability scheme is used to fix the grid points.[18] The observations from populations Y and Z are generated from the same distribution. $B(4.7,\ 0.55)$ and $B(2.0,\ 0.65)$ are considered. In the experiments, the numbers of cases of accepting each of the four hypotheses are recorded. We use 10,000 Monte Carlo runs for each experiment. From the results we find that the number of cases where H_A is accepted is negligible, while the acceptance of H_{A1} and H_{A2} are almost equally divided. Thus, only the acceptance of H_0 is given in the table. Note that one minus the figures in Table 2 gives the empirical sizes of the tests. We can see that the DD and A tests are too conservative.[19] In comparison, the understatement in the size of the A test is more serious. The variation

Then $x_1, ..., x_{K-1}$ are chosen to divide the combined sample into K groups of equal probability. For the equal-interval scheme, x_K is fixed similarly. However, $x_1, ..., x_{K-1}$ are chosen to divide the interval $(0, x_K)$ into K equal subintervals.

[17] Note that independent samples are used at each grid point so that the reported χ^2 statistics are independent.

[18] Tables 2 through 5 report the cases when the same sample observations are used to calculate the grid statistics at all grid points. In Tables 6 through 8 below we report results for the cases when different samples are used to compute the grid statistics at different grid points.

[19] Recall that this result agrees with the findings of Sidak (1967) and Hochberg (1974).

in the empirical size with respect to K seems to be small, although it appears that the empirical size is slightly closer to the nominal value when K is smaller. This is true for both tests. Also, the nominal size is better approximated when s is smaller. Again, this is true for both tests.

To examine the power of the tests, we generate Z from $B(2.0, 0.65)$. We consider the following distributions for Y: $B(4.5, 0.55)$, $B(4.6, 0.55)$ and $B(4.7, 0.55)$. For each case of Y there is no dominance between Y and Z at the first and second order (i.e., H_A is true for $s = 1, 2$). However, $Y \succ_3 Z$, so that H_{A1} is true for $s = 3$. Table 3 summarizes the empirical relative frequencies of accepting H_{A1} when the equal-probability scheme is used to fix the grid points. We note that the numbers of cases of H_0 and H_{A2} being accepted are almost zero, so that one minus the figures in the table gives the relative frequencies of accepting H_A. It can be seen that for $s = 1$, the tests are able to make the right decision almost all the time, even for a small sample of 500. For $s = 2$, however, there are high frequencies of making the wrong decision that Y dominates Z. The A test makes the wrong decision more frequently than the DD test. As expected, the wrong decision is made less frequently when the sample size N increases. Rather interestingly, wrong decisions are made less frequently when K is small.[20] This regularity applies to both tests and for all pairs of populations. Finally, for $s = 3$, both tests have no difficulty identifying the dominance correctly.

We repeat the experiments using the equal-interval scheme to fix the grid points. The results are presented in Tables 4 and 5. Comparing Tables 2 and 4, we can see that the under-rejection for the equal-interval scheme is more serious.[21] On the other hand, from Tables 3 and 5 we can see that the probability of making the wrong conclusion $Y \succ_2 Z$ has reduced when the equal-interval scheme is used. The improvement is

[20]Note that this finding also applies to the A test, notwithstanding the fact that the trapezoidal approximation is expected to improve with large K.

[21]It is noted that the under-rejection is due to the correlations between the grid-point statistics. The equal-probability scheme has the effect of spacing out the grid-points to reduce the correlations. On the other hand, when the equal-interval scheme is used the correlation between two adjacent grid-point statistics may be very high in a region where the functions D^s are flat.

especially significant for the A test.[22] Indeed, when the equal-interval scheme is used, the A test is only marginally inferior to the DD test in wrongly concluding that $Y \succ_2 Z$.

The fact that the T^s and A^s statistics use the same sample observations for different x_i renders these statistics correlated across different x_i. This causes the tests to under-reject when the null is correct. One method to overcome this problem is to use independent samples for different x_i. We run Monte Carlo experiments with the same set-up as above, except that an independent sample is used for each x_i.[23] We find that the empirical size becomes very close to the nominal size. To save space, the results are not summarized here.[24]

To investigate how the powers of the tests are affected by the independent sampling method, we conduct Monte Carlo experiments following the same set-up as Tables 3 and 5. The results are summarized in Tables 6 and 7, for the equal-probability and equal-interval schemes, respectively. From the results we can see that, as before, the decisions concerning $s = 1$ and $s = 3$ are made with high accuracy. For $s = 2$, there is improvement in the reduction of the probability of making the wrong decision. While the improvement is marginal for the equal-probability scheme, it is very significant for the equal-interval scheme.[25] As independent sampling requires more data, it is important to see if such a sampling scheme is worthwhile. As far as the size of the test is concerned, the improvement in the independent sampling is impressive. However, the improvement in avoiding wrongly accepting H_{A1} is not significant and does not warrant the additional observations.

We now turn to the KRS test. The same Monte Carlo set-up is used. The results are summarized in Table 8. Two sets of hypotheses are compared. These are

[22] The improvement in the A test may be due to better results in the trapezoidal approximation when the equal-interval scheme is used. For the DD test, the differences are small and the improvement in the equal-interval scheme is not uniform.

[23] Hence the total number of observations in the sample is KN.

[24] The accuracy of the empirical size is indeed very impressive. For the complete set of experiments conducted, the empirical probabilities of accepting H_0 range from 94.30% to 96.06%.

[25] As the tests are no longer conservative under independent sampling, the improvement in power is expected.

$H_0 : Y \not\succ_s Z$ against $H_1 : Y \succ_s Z$, and $H_0' : Z \not\succ_s Y$ against $H_1' : Z \succ_s Y$. The figures in Table 8 are the empirical relative frequencies of accepting H_0 (against H_1) and H_0' (against H_1'). Five pairs of populations are considered. These are labelled PD1 through PD5. The figures in Table 8 should be interpreted as follows. For PD1 and PD2, both H_0 and H_0' are correct. Thus, one minus the figures in the table represents the empirical size of the tests. This applies to all s. For PD3, PD4 and PD5, there is no dominance between Y and Z for $s = 1, 2$, while Y stochastically dominates Z for $s = 3$. Thus, for $s = 1, 2$, H_0 and H_0' are correct. For $s = 3$, H_0 is incorrect, while H_0' is correct.

The following conclusions emerge from Table 8. Firstly, the true size of the KRS test is very much understated by the nominal size. Indeed, we can see that the empirical size of the test is zero for many of the cases simulated. Secondly, considering H_0 with $s = 3$ for PD3, PD4 and PD5 (which is a wrong hypothesis), we can see that the empirical power of the test is rather low. The lack in power is clear when we compare the results against those of the DD and A tests in Tables 3, 5, 6 and 7. Both DD and A tests correctly select the hypothesis with very high probability.

There is always a danger in generalizing results of Monte Carlo experiments. Notwithstanding this caveat, however, our results tend to suggest the following findings. The DD test appears to be the best procedure to adopt in examining stochastic dominance. If the studentized maximum modulus statistic is used to provide critical values, there is under-rejection of the null hypothesis when it is true. The under-rejection is corrected when independent sampling scheme is adopted to calculate the grid statistics. The improvement in power, however, is not impressive. It appears one may as well adopt the single sampling scheme. While the DD and A tests are similar in execution, the DD test appears to be superior in both size and power. As far as the choice of the number of grid points is concerned, the tests appear to perform well with small number of grid points such as six, although the literature often adopts ten points. The KRS test is excessively conservative. When it comes to the power, the KRS test does not compare favorably against the DD and A tests. Furthermore, the KRS test is computationally more intensive than the DD and A tests.

4 Conclusions

We have reported some Monte Carlo results for the performance of several tests for stochastic dominance, namely, the Davidson-Duclos (2000) test, the Anderson (1996) test and the Kaur-Rao-Singh (1994) test. The KRS test is found to be too conservative and has weaker power compared to the DD and A tests. The DD test appears to dominate the A test in terms of better size and higher power. While the problem of under-rejection may be corrected by using independent samples, the improvement in power is not significant. Overall, of the tests examined, the DD test is the one to be recommended.

References

[1] Anderson, G., 1996, "Nonparametric tests of stochastic dominance in income distributions", *Econometrica*, 64, 1183 – 1193.

[2] Barrett, G. and S. Donald, 2001, "Consistent tests for stochastic dominance", mimeo.

[3] Beach, C.M. and R. Davidson, 1983, "Distribution-free statistical inference with Lorenz curves and income shares", *Review of Economic Studies*, 50, 723 – 735.

[4] Bishop, J.A., J.P. Formby and P.D. Thistle, 1992, "Convergence of the south and non-south income distribution, 1969 – 1979", *American Economic Review*, 82, 262 – 272.

[5] Dardanoni, V. and A. Forcina, 1999, "Inference for Lorenz curve orderings", *Econometrics Journal*, 2, 49 – 75.

[6] Davidson, R. and J.-Y. Duclos, 2000, "Statistical inference for stochastic dominance and for the measurement of poverty and inequality", *Econometrica*, 68, 1435 – 1464.

[7] Fisher, G., D. Wilson and K. Xu, 1998, "An empirical analysis of term premium using significance tetss for stochastic dominance", *Economics Letters*, 60, 195 – 203.

[8] Hochberg, Y., 1974, "Some generalizations of the T-method in simultaneous inference", *Journal of Multivariate Analysis*, 4, 224 – 234.

[9] Johnson, N.L., S. Kotz and N. Balakrishnan, 1994, *Continuous Univariate Distributions*, Vol 1, 2nd edition, John Wiley: New York.

[10] Kaur, A., B.L.S.P. Rao and H. Singh, 1994, "Testing for second-order stochastic dominance of two distributions", *Econometric Theory*, 10, 849 – 866.

[11] Maasoumi, E. and A. Heshmati, 2000, "Stochastic dominance amongst Swedish income distribution", *Econometric Reviews*, 19, 287 – 320.

[12] McFadden, D., 1989, "Testing for stochastic dominance", in *Studies in the Economics of Uncertainty*, T.B. Fomby and T.K. Seo, eds., Springer Verlag: New York.

[13] Shapiro, A., 1985, "Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints", *Biometrika*, 72, 133 – 144.

[14] Shapiro, A., 1988, "Towards a unified theory of inequality constrained testing in multivariate analysis", *International Statistical Review*, 56, 49 – 62.

[15] Sidak, Z., 1967, “Rectangular confidence regions for the means of multivariate normal distributions”, *Journal of the American Statistical Association*, 62, 626 – 633.

[16] Stoline, M.R. and H.K. Ury, 1979, “Tables of the studentized maximum modulus distribution and an application to multiple comparison among means”, *Technometrics*, 21, 87 – 93.

[17] Wolak, F.A., 1989, “Testing inequality constraints in linear econometric models”, *Journal of Econometrics*, 41, 205 – 235.

[18] Xu, K., 1997, “Asymptotically distribution-free statistical test for generalized Lorenz curves: An alternative approach”, *Journal of Income Distribution*, 7, 45 – 62.

[19] Xu, K. and L. Osberg, 1998, “A distribution-free test for deprivation dominance”, *Econometric Reviews*, 17, 415 – 429.

Table 1 χ^2 Goodness-of-Fit Test for the normality of the T^s and A^s Statistics

N	s	Statistic	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
500	1	T^s	658.51	30.65	14.99	23.88	25.75	68.03	56.96	114.72	242.00	354.26
		A^s	70.72	8.30	11.47	6.85	5.92	26.71	25.80	29.04	30.87	1962.92
	2	T^s	110.73	3.07	7.19	4.77	5.66	9.25	7.55	7.52	6.11	2.52
		A^s	66.08	3.60	4.93	8.80	3.78	10.52	8.67	18.44	7.05	2.79
	3	T^s	471.94	5.57	8.03	8.68	4.40	10.52	6.69	7.55	6.96	4.95
		A^s	67.39	3.53	2.30	6.50	2.41	8.39	6.47	6.80	8.52	4.29
1000	1	T^s	290.89	24.79	11.76	13.66	7.09	24.45	36.99	60.23	99.43	209.50
		A^s	48.95	1.94	24.27	8.54	16.68	7.72	35.82	2.75	37.43	213.55
	2	T^s	27.42	9.87	9.87	5.04	1.91	9.40	9.80	13.61	7.84	8.15
		A^s	59.80	7.67	2.39	10.60	10.26	8.84	2.60	14.30	3.64	4.12
	3	T^s	119.88	7.44	6.84	3.80	2.52	3.70	10.92	4.62	9.46	10.90
		A^s	75.71	1.61	7.72	2.50	2.52	5.50	6.68	4.01	12.80	5.33
2000	1	T^s	174.63	11.86	23.72	5.96	22.05	15.46	17.08	31.62	45.73	62.86
		A^s	52.48	11.42	6.04	6.32	11.00	7.95	11.47	4.00	16.10	447.27
	2	T^s	4.44	11.21	9.74	4.17	1.48	5.44	7.28	6.36	1.12	5.73
		A^s	37.51	10.40	14.77	3.78	4.16	10.09	3.24	8.22	4.45	3.81
	3	T^s	21.62	5.78	5.35	9.63	6.11	4.17	4.45	4.68	5.83	2.60
		A^s	30.58	6.43	6.31	13.05	7.02	4.64	2.56	4.83	6.91	4.81
3000	1	T^s	102.00	4.93	10.15	16.93	11.50	9.63	29.26	17.16	42.54	51.13
		A^s	37.97	1.57	8.22	11.25	10.26	12.11	4.61	8.57	18.83	174.64
	2	T^s	3.56	1.86	5.69	4.63	4.25	8.04	5.36	2.27	13.24	3.44
		A^s	20.55	7.74	7.93	4.66	9.52	4.69	6.22	4.08	5.81	2.40
	3	T^s	14.24	4.39	7.81	7.82	3.23	8.69	12.04	5.10	9.12	15.32
		A^s	17.21	7.62	7.30	10.60	6.80	10.42	12.65	3.32	9.64	17.30

Notes: The figures are the χ^2 statistics for the normality of T^s and A^s computed at the grid points $x_i, i = 1, ...10$. The grid points are fixed and divide the support of the truncated $B(4.7, 0.55)$ distribution into eleven intervals such that x_{10} is the 99th percentile and the first ten intervals are of equal length. For each of the T^s and A^s statistic, eight groups are used to construct the goodness-of-fit test. The critical value of χ^2_8 is 14.1 at the 5% level.

Table 2 Empirical Relative Frequency of Accepting H_0 for the DD and A Tests: The Case of Dependent Samples with Equal-Probability Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$	500	6	96.34	96.39	97.24	97.38	97.47	97.85
$Z \sim B(4.7, 0.55)$		10	96.32	96.36	97.91	98.13	98.63	98.73
		15	96.34	96.53	98.50	98.82	99.07	99.24
	1000	6	95.56	95.53	97.34	97.74	97.55	98.19
		10	96.51	96.49	98.00	98.28	98.28	98.62
		15	96.80	96.92	98.68	98.80	98.71	98.99
	2000	6	95.83	95.84	97.65	98.00	97.74	98.28
		10	96.17	96.17	98.07	98.17	98.44	98.71
		15	96.87	96.82	98.64	98.82	98.92	99.10
	3000	6	95.76	95.70	97.50	97.79	97.57	98.19
		10	96.50	96.56	98.29	98.59	98.54	98.89
		15	96.53	96.71	98.77	98.94	98.86	98.99
$Y \sim B(2.0, 0.65)$	500	6	96.21	96.32	97.04	97.24	97.02	97.63
$Z \sim B(2.0, 0.65)$		10	96.47	96.45	97.72	98.05	98.05	98.23
		15	96.64	96.87	98.51	98.68	98.63	98.74
	1000	6	94.95	94.95	96.76	96.95	97.19	97.73
		10	96.29	96.27	97.68	97.92	98.09	98.30
		15	96.76	96.97	98.51	98.51	98.29	98.64
	2000	6	95.83	95.82	97.25	97.38	97.42	97.70
		10	96.15	96.18	97.76	98.00	98.03	98.36
		15	96.96	96.92	98.26	98.32	98.61	98.86
	3000	6	95.94	95.88	97.37	97.74	97.47	98.11
		10	96.04	96.16	97.88	98.04	98.35	98.51
		15	96.55	96.64	98.27	98.51	98.57	98.76

Notes: The figures (in percentage) are the empirical relative frequencies of accepting $H_0 : D_Y^s(x_i) = D_Z^s(x_i)$. N is the sample size and K is the number of grid points.

Table 3 Empirical Relative Frequency of Accepting H_{A1} for the DD and A Tests: The Case of Dependent Samples with Equal-Probability Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$	500	6	4.77	5.17	90.20	93.07	99.88	99.96
$Z \sim B(2.0, 0.65)$		10	3.48	3.40	92.55	94.17	99.97	99.98
		15	3.08	3.33	94.37	95.47	99.95	99.99
	1000	6	0.04	0.04	76.55	84.12	99.97	100.00
		10	0.01	0.01	81.24	86.05	99.99	100.00
		15	0.01	0.01	84.61	87.37	100.00	100.00
	2000	6	0.00	0.00	46.89	63.78	99.98	100.00
		10	0.00	0.00	53.65	64.36	100.00	100.00
		15	0.00	0.00	59.62	66.99	100.00	100.00
	3000	6	0.00	0.00	24.36	43.56	100.00	100.00
		10	0.00	0.00	30.12	42.30	100.00	100.00
		15	0.00	0.00	34.15	43.77	100.00	100.00
$Y \sim B(4.6, 0.55)$	500	6	5.94	6.17	88.43	91.76	99.87	99.95
$Z \sim B(2.0, 0.65)$		10	3.68	3.53	91.82	93.52	99.91	99.97
		15	3.82	4.13	93.48	94.56	99.92	99.95
	1000	6	0.02	0.04	73.50	81.41	99.96	99.99
		10	0.01	0.01	78.31	83.23	99.96	100.00
		15	0.02	0.02	81.83	84.68	99.98	100.00
	2000	6	0.00	0.00	41.05	57.49	99.96	100.00
		10	0.00	0.00	48.30	58.34	99.99	100.00
		15	0.00	0.00	52.35	59.75	99.99	99.99
	3000	6	0.00	0.00	18.32	35.55	99.99	100.00
		10	0.00	0.00	23.74	34.48	99.99	100.00
		15	0.00	0.00	27.98	36.03	100.00	100.00
$Y \sim B(4.5, 0.55)$	500	6	6.77	7.15	86.84	90.06	99.75	99.85
$Z \sim B(2.0, 0.65)$		10	4.42	4.33	89.85	91.71	99.87	99.96
		15	4.32	4.55	92.44	93.53	99.96	99.99
	1000	6	0.11	0.11	69.22	77.09	99.83	100.00
		10	0.01	0.02	75.38	80.03	99.91	99.98
		15	0.04	0.03	79.45	82.91	99.91	99.98
	2000	6	0.00	0.00	35.74	51.49	99.92	100.00
		10	0.00	0.00	40.95	50.82	99.90	100.00
		15	0.00	0.00	47.41	53.86	99.97	100.00
	3000	6	0.00	0.00	14.46	28.49	99.93	100.00
		10	0.00	0.00	18.13	27.25	99.98	100.00
		15	0.00	0.00	22.18	29.10	99.99	100.00

Notes: The figures (in percentage) are the empirical relative frequencies of accepting H_{A1} : $Y \succ_s Z$. For all pairs of populations, there is no dominance between Y and Z for $s = 1, 2$, while Y dominates Z for $s = 3$. Thus, H_{A1} is true for $s = 3$, but not for $s = 1, 2$. N is the sample size and K is the number of grid points.

Table 4 Empirical Relative Frequency of Accepting H_0 for the DD and A Tests: The Case of Dependent Samples with Equal-Interval Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$ $Z \sim B(4.7, 0.55)$	500	6	96.07	96.30	97.54	98.25	97.86	98.41
		10	96.30	96.56	97.99	98.36	98.68	98.63
		15	96.98	97.14	98.73	98.78	99.09	99.02
	1000	6	95.66	95.75	97.36	97.88	97.72	98.12
		10	95.89	96.15	97.98	98.25	98.46	98.35
		15	96.28	96.46	98.58	98.53	99.02	98.86
	2000	6	95.26	95.31	97.42	97.79	97.58	98.14
		10	95.86	95.87	98.18	98.30	98.29	98.42
		15	96.68	96.64	98.64	98.70	98.78	98.82
	3000	6	96.02	96.13	97.54	97.67	97.90	98.24
		10	96.24	96.24	98.31	98.24	98.46	98.48
		15	96.43	96.40	98.41	98.58	98.70	98.80
$Y \sim B(2.0, 0.65)$ $Z \sim B(2.0, 0.65)$	500	6	96.75	96.84	98.29	98.47	98.26	98.63
		10	96.76	96.92	98.56	98.70	98.61	99.03
		15	97.30	97.39	98.87	98.97	98.72	99.08
	1000	6	95.61	95.83	98.22	98.40	98.31	98.54
		10	96.89	96.97	98.45	98.84	98.52	98.88
		15	96.91	96.95	98.92	99.09	98.96	99.04
	2000	6	95.95	95.90	98.00	98.45	98.30	98.64
		10	96.64	96.67	98.83	98.99	98.69	98.84
		15	97.28	97.30	99.02	99.06	98.86	99.21
	3000	6	96.13	96.16	98.18	98.47	98.34	98.60
		10	96.58	96.59	98.30	98.35	98.82	98.90
		15	97.36	97.35	98.96	99.10	98.86	99.09

Notes: The figures (in percentage) are the empirical relative frequencies of accepting $H_0 : D_Y^s(x_i) = D_Z^s(x_i)$. N is the sample size and K is the number of grid points.

Table 5 Empirical Relative Frequency of Accepting H_{A1} for the DD and A Tests: The Case of Dependent Samples with Equal-Interval Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	1.71	1.71	89.53	90.78	99.91	99.85
		10	1.50	1.59	92.97	92.99	99.96	99.97
		15	1.89	1.94	94.57	94.69	99.97	99.97
	1000	6	0.00	0.00	75.69	79.38	99.97	99.98
		10	0.00	0.01	81.12	81.13	99.98	99.98
		15	0.00	0.00	84.25	84.26	99.98	99.98
	2000	6	0.00	0.00	46.26	54.80	100.00	99.99
		10	0.00	0.00	53.39	54.54	99.98	99.98
		15	0.00	0.00	58.77	59.31	99.99	99.99
	3000	6	0.00	0.00	24.63	33.12	99.99	100.00
		10	0.00	0.00	29.69	30.43	100.00	100.00
		15	0.00	0.00	34.79	35.16	100.00	100.00
$Y \sim B(4.6, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	2.17	2.12	89.12	90.03	99.87	99.84
		10	1.95	2.11	91.16	91.44	99.96	99.92
		15	2.19	2.29	93.66	93.81	99.95	99.94
	1000	6	0.00	0.00	73.45	77.27	99.93	99.92
		10	0.00	0.00	78.46	78.67	99.97	99.96
		15	0.00	0.00	81.54	81.96	99.96	99.96
	2000	6	0.00	0.00	41.13	49.54	99.99	99.97
		10	0.00	0.00	47.26	48.06	100.00	100.00
		15	0.00	0.00	53.18	53.59	99.98	99.98
	3000	6	0.00	0.00	18.37	26.29	99.99	99.98
		10	0.00	0.00	23.31	23.83	99.99	99.99
		15	0.00	0.00	28.15	28.55	99.99	99.99
$Y \sim B(4.5, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	2.21	2.25	86.63	87.82	99.81	99.70
		10	2.17	2.22	89.99	90.03	99.86	99.86
		15	2.57	2.80	92.27	92.44	99.91	99.91
	1000	6	0.00	0.00	69.57	73.40	99.77	99.76
		10	0.01	0.01	74.49	74.65	99.93	99.92
		15	0.01	0.01	78.98	79.23	99.89	99.88
	2000	6	0.00	0.00	35.15	42.60	99.89	99.92
		10	0.00	0.00	41.92	42.54	99.91	99.92
		15	0.00	0.00	47.29	47.74	99.93	99.93
	3000	6	0.00	0.00	14.14	20.86	99.97	99.96
		10	0.00	0.00	18.01	18.79	99.93	99.89
		15	0.00	0.00	21.82	22.30	99.96	99.96

Notes: The figures (in percentage) are the empirical relative frequencies of accepting H_{A1} : $Y \succ_s Z$. For all pairs of populations, there is no dominance between Y and Z for $s = 1, 2$, while Y dominates Z for $s = 3$. Thus, H_{A1} is true for $s = 3$, but not for $s = 1, 2$. N is the sample size and K is the number of grid points.

Table 6 Empirical Relative Frequency of Accepting H_{A1} for the DD and A Tests: The Case of Independent Samples with Equal-Probability Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	3.05	3.21	90.43	93.03	99.92	99.97
		10	0.27	0.25	92.10	94.09	99.98	100.00
		15	0.01	0.01	93.28	94.47	99.99	100.00
	1000	6	0.00	0.00	76.54	83.85	99.95	99.99
		10	0.00	0.00	81.29	85.85	99.98	99.99
		15	0.00	0.00	82.94	86.39	99.97	99.99
	2000	6	0.00	0.00	46.55	63.00	100.00	100.00
		10	0.00	0.00	53.33	63.91	99.99	100.00
		15	0.00	0.00	55.59	63.78	100.00	100.00
	3000	6	0.00	0.00	24.23	43.95	99.99	100.00
		10	0.00	0.00	30.60	43.56	100.00	100.00
		15	0.00	0.00	32.09	41.83	100.00	100.00
$Y \sim B(4.6, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	3.43	3.65	88.98	91.64	99.83	99.93
		10	0.39	0.42	91.12	92.75	99.92	99.96
		15	0.05	0.05	91.15	92.63	99.95	99.99
	1000	6	0.01	0.01	72.78	80.71	99.95	100.00
		10	0.00	0.00	77.92	82.66	99.98	100.00
		15	0.00	0.00	78.87	83.01	99.95	99.97
	2000	6	0.00	0.00	39.38	56.37	99.96	100.00
		10	0.00	0.00	47.83	57.74	99.98	100.00
		15	0.00	0.00	48.64	56.30	99.99	100.00
	3000	6	0.00	0.00	18.19	35.65	99.98	100.00
		10	0.00	0.00	23.46	34.92	99.99	100.00
		15	0.00	0.00	24.73	33.23	100.00	100.00
$Y \sim B(4.5, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	4.06	4.30	87.79	91.02	99.81	99.93
		10	0.40	0.39	89.79	91.61	99.85	99.94
		15	0.10	0.10	91.01	92.36	99.91	99.94
	1000	6	0.02	0.02	68.66	77.35	99.85	99.96
		10	0.00	0.00	73.26	78.42	99.91	99.99
		15	0.00	0.00	76.18	79.85	99.94	99.98
	2000	6	0.00	0.00	35.66	50.96	99.90	99.99
		10	0.00	0.00	40.39	49.56	99.93	100.00
		15	0.00	0.00	42.84	50.38	99.93	99.97
	3000	6	0.00	0.00	13.71	28.55	99.90	100.00
		10	0.00	0.00	18.41	28.00	99.95	99.98
		15	0.00	0.00	19.13	26.81	99.99	100.00

Notes: The figures (in percentage) are the empirical relative frequencies of accepting H_{A1} : $Y \succ_s Z$. For all pairs of populations, there is no dominance between Y and Z for $s = 1, 2$, while Y dominates Z for $s = 3$. Thus, H_{A1} is true for $s = 3$, but not for $s = 1, 2$. N is the sample size and K is the number of grid points.

Table 7 Empirical Relative Frequency of Accepting H_{A1} for the DD and A Tests: The Case of Independent Samples with Equal-Interval Groups

Population Distribution	N	K	$s=1$ DD	$s=1$ A	$s=2$ DD	$s=2$ A	$s=3$ DD	$s=3$ A
$Y \sim B(4.7, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	0.02	0.01	83.66	85.43	99.92	99.90
		10	0.00	0.00	82.48	82.53	99.95	99.90
		15	0.00	0.00	81.77	82.39	99.92	99.92
	1000	6	0.00	0.00	64.17	69.91	99.98	99.97
		10	0.00	0.00	61.67	62.15	99.99	100.00
		15	0.00	0.00	57.70	58.63	99.97	99.97
	2000	6	0.00	0.00	29.60	39.10	100.00	100.00
		10	0.00	0.00	25.51	26.31	100.00	99.99
		15	0.00	0.00	19.75	20.32	100.00	100.00
	3000	6	0.00	0.00	10.79	19.34	100.00	99.99
		10	0.00	0.00	7.57	8.15	100.00	100.00
		15	0.00	0.00	3.97	4.05	100.00	100.00
$Y \sim B(4.6, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	0.11	0.10	80.29	92.19	99.87	99.83
		10	0.00	0.00	79.55	79.83	99.84	99.85
		15	0.05	0.05	78.64	78.96	99.92	99.91
	1000	6	0.00	0.00	57.98	64.58	99.88	99.91
		10	0.00	0.00	55.79	55.87	99.96	99.95
		15	0.00	0.00	51.21	51.56	99.98	99.98
	2000	6	0.00	0.00	22.65	31.89	99.93	99.95
		10	0.00	0.00	17.41	17.99	99.98	99.98
		15	0.00	0.00	12.76	13.47	100.00	100.00
	3000	6	0.00	0.00	6.74	12.82	99.97	99.98
		10	0.00	0.00	3.53	3.77	100.00	100.00
		15	0.00	0.00	1.88	2.02	100.00	100.00
$Y \sim B(4.5, 0.55)$ $Z \sim B(2.0, 0.65)$	500	6	0.11	0.11	77.67	79.40	99.75	99.67
		10	0.00	0.00	76.18	76.29	99.79	99.74
		15	0.00	0.00	74.71	74.95	99.82	99.80
	1000	6	0.00	0.00	53.08	58.71	99.85	99.80
		10	0.00	0.00	48.08	48.77	99.91	99.89
		15	0.00	0.00	44.23	44.65	99.94	99.92
	2000	6	0.00	0.00	16.57	24.37	99.89	99.90
		10	0.00	0.00	12.11	12.41	99.96	99.95
		15	0.00	0.00	8.14	8.26	99.95	99.94
	3000	6	0.00	0.00	3.98	8.37	99.91	99.93
		10	0.00	0.00	1.58	1.81	99.93	99.90
		15	0.00	0.00	0.67	0.72	99.94	88.93

Notes: The figures (in percentage) are the empirical relative frequencies of accepting H_{A1} : $Y \succ_s Z$. For all pairs of populations, there is no dominance between Y and Z for $s = 1, 2$, while Y dominates Z for $s = 3$. Thus, H_{A1} is true for $s = 3$, but not for $s = 1, 2$. N is the sample size and K is the number of grid points.

Table 8 Empirical Relative Frequencies of Accepting H_0 and $H_0^{'}$ for the KRS Test: The Case of Independent Samples

Population Distribution	N	s = 1 H_0	s = 1 $H_0^{'}$	s = 2 H_0	s = 2 $H_0^{'}$	s = 3 H_0	s = 3 $H_0^{'}$
PD1:	500	100.00	100.00	99.75	99.70	99.75	99.25
$Y \sim B(4.7, 0.55)$	1000	99.95	100.00	99.80	99.45	99.60	99.60
$Z \sim B(4.7, 0.55)$	2000	100.00	100.00	99.50	99.80	99.45	99.65
PD2:	500	100.00	100.00	99.90	99.75	99.80	99.90
$Y \sim B(2.0, 0.65)$	1000	100.00	99.95	99.85	99.90	99.85	99.65
$Z \sim B(2.0, 0.65)$	2000	100.00	100.00	99.85	99.85	99.85	99.75
PD3:	500	100.00	100.00	99.55	100.00	54.90	100.00
$Y \sim B(4.7, 0.55)$	1000	100.00	100.00	99.90	100.00	28.55	100.00
$Z \sim B(2.0, 0.65)$	2000	100.00	100.00	100.00	100.00	7.40	100.00
PD4:	500	100.00	100.00	99.75	100.00	60.35	100.00
$Y \sim B(4.6, 0.55)$	1000	100.00	100.00	100.00	100.00	37.30	100.00
$Z \sim B(2.0, 0.65)$	2000	100.00	100.00	100.00	100.00	13.90	100.00
PD5:	500	100.00	100.00	99.60	100.00	64.50	100.00
$Y \sim B(4.5, 0.55)$	1000	100.00	100.00	100.00	100.00	45.10	100.00
$Z \sim B(2.0, 0.65)$	2000	100.00	100.00	100.00	100.00	21.65	100.00

Notes: H_0 is $Y \succ_s Z$ and $H_0^{'}$ is $Z \succ_s Y$. For PD1 and PD2, both H_0 and H_0' are correct. Thus, one minus the figures in the table represents the empirical size of the tests. This applies to all s. For PD3, PD4 and PD5, there is no dominance between Y and Z for $s = 1, 2$, while Y stochastically dominates Z for $s = 3$. Thus, for $s = 1, 2$, H_0 and H_0' are correct. For $s = 3$, H_0 is incorrect, while H_0' is correct.

Figure 1: Distribution Functions of Truncated B(alpha, beta)

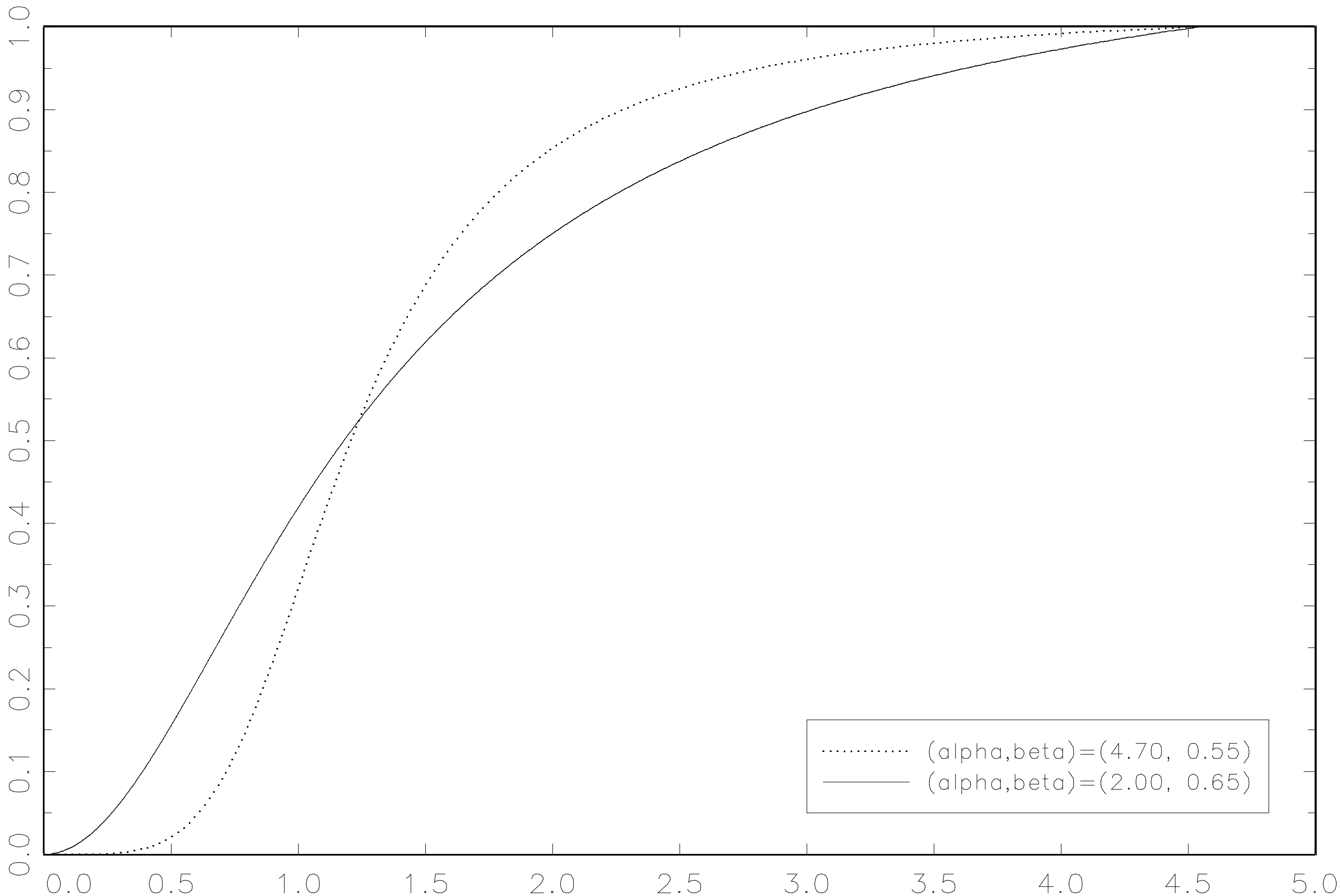

Figure 2: D^2(.) Functions of Truncated B(alpha, beta)

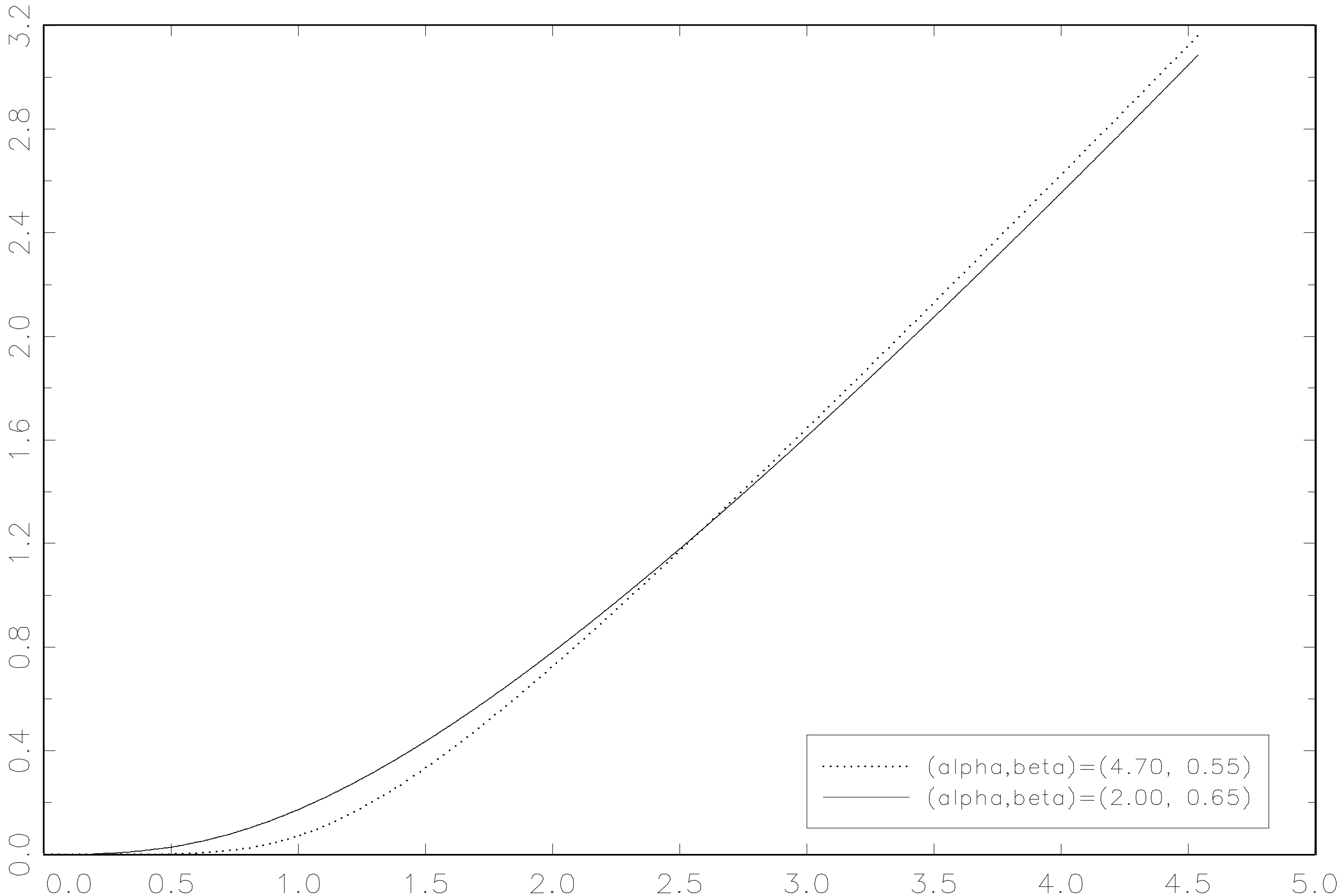

Figure 3: D^3(.) Functions of Truncated B(alpha, beta)