Testing for Bivariate Stochastic Dominance Using Inequality Restrictions*

Thanasis Stengos† and Brennan S. Thompson‡

May 31, 2011

Abstract

In this paper, we propose of a test of bivariate stochastic dominance using a generalized framework for testing inequality constraints. Unlike existing tests, this test has the advantage of utilizing the covariance structure of the estimates of the joint distribution functions. The performance of our proposed test is examined by way of a Monte Carlo experiment. We also consider an empirical example which utilizes household survey data on income and health status.

Keywords: Stochastic dominance, inequality restrictions, multidimensional welfare

*The authors wish to thank Russell Davidson, Michael Hoy, Alex Maynard, and Yiguo Sun for their helpful comments and suggestions.

†Department of Economics, University of Guelph, Guelph, ON N1G 2W1. E-mail: tstengos@uoguelph.ca

‡Department of Economics, Ryerson University, Toronto, ON M5B 2K3. E-mail: brennan@ryerson.ca

1 Introduction

In the past two decades, a number of statistical tests of stochastic dominance have been put forth in the literature. These tests can broadly be divided into two broad categories. Tests in category one, which include those proposed by Anderson (1996), Fisher et al. (1998), Davidson and Duclos (2000), and Davidson and Duclos (2007), all of which are applicable only to univariate distributions, involve evaluating each CDF at a finite number of points.[1] Tests in category two, on the other hand, are based on evaluations over the entire support of each CDF. This category includes the univariate tests of McFadden (1989), Kaur et al. (1994), Maasoumi and Heshmati (2000), Barrett and Donald (2003), Linton et al. (2005), and Horvath et al. (2006), as well as the multivariate tests of McCaig and Yatchew (2007), hereafter MY, and Anderson (2008).

Tests in category one have the disadvantage of requiring the researcher to specify a set of arbitrary evaluation points. As suggested by Davidson and Duclos (2000) and Barrett and Donald (2003), these tests might, as a result, be inconsistent. However, these tests have the advantage of making use of the covariances between the estimates made at each of the evaluation points (see Davidson and Duclos, 2000). Tests in category two ignore this covariance structure.

[1] Related tests in this category are the tests of Lorenz dominance by Beach and Davidson (1983), Beach and Richmond (1985), Bishop et al. (1993), and Dardanoni and Forcina (1999), as well as the test of distribution dominance by Xu and Osberg (1998).

In this paper, we propose a test for bivariate stochastic dominance which involves evaluating each CDF at a finite number of points (i.e., over a grid of points). This test, belonging to category one, can be seen as a simple extension of the methods of Fisher et al. (1998) and Davidson and Duclos (2000) to the bivariate case. While a partial extension of these methods was considered by Duclos et al. (2006), these authors do not utilize the covariance structure between the estimates at each grid point in their hypothesis tests. We are able to do by using the general methods of Kodde and Palm (1986) and Wolak (1989) for testing vectors of inequality constraints.

The remainder of this paper is organized as follows. Section 2 provides formal definitions and discuss how stochastic dominance relations can be estimated. In Section 3 we propose a hypothesis test based on the asymptotic distribution of the estimates introduced in the previous section, and contrast this test with that of MY. These two tests are then compared in a Monte Carlo simulation in Section 4. In Section 5, we present an empirical example using Canadian household survey data on income and health status. Section 6 concludes.

2 Estimation and inference

Let F_A and F_B denote two right-continuous d-dimensional distribution functions. We say that distribution F_A (weakly) dominates distribution F_B stochastically at order s (an integer) if $D^s_A(\mathbf{z}) \leq D^s_B(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^d_+$, where, for $K = A, B$, $D_K^1(\mathbf{z}) = F_K(\mathbf{z})$ and $D_K^s(\mathbf{z})$ is defined recursively as

$$D_K^s(\mathbf{z}) = \int_0^{\mathbf{z}} D_K^{s-1}(\mathbf{u})d\mathbf{u}, \qquad s \geq 2.$$

In what follows, we will denote this relation by $F_A \succeq_s F_B$.

Following Davidson and Duclos (2000), it will be convenient, in the bivariate case, to rewrite $D_K^s(\mathbf{z}) = D_K^s(z_x, z_y)$ as

$$D_K^s(z_x, z_y) = \frac{1}{(s-1)!}E[(z_x - X_K)_+^{(s-1)}(z_y - Y_K)_+^{(s-1)}], \qquad s \geq 1,$$

where $\phi_+ = \max(0, \phi)$, and the random vector (X_K, Y_K) has distribution function F_K.

Letting $\{(x_{K,i}, y_{K,i})\}_{i=1}^{n_K}$ denote a sample of n_K independent and identically distributed (IID) observations drawn from F_K, a natural estimator of $D_K^s(z_x, z_y)$ is

$$\hat{D}_K^s(z_x, z_y) = \frac{1}{n_K(s-1)!}\sum_{i=1}^{n_K}(z_x - x_{K,i})_+^{s-1}(z_y - y_{K,i})_+^{s-1}. \tag{1}$$

In what follows, we wish to estimate both $D_A^s(x, y)$ and $D_B^s(x, y)$ on the same $J \times J$ grid of arbitrary evaluation points. Specifically, let $\lambda_{X,1}, \ldots, \lambda_{X,J}$ denote a set of points on the combined support of X_A and X_B, and $\lambda_{Y,1}, \ldots, \lambda_{Y,J}$ denote a set of points on the combined support of Y_A and Y_B. Next, let

$$\lambda \;=\; ((\lambda_{X,1}, \lambda_{Y,1}), (\lambda_{X,1}, \lambda_{Y,2}), \ldots, (\lambda_{X,1}, \lambda_{Y,J}),$$

$$\ldots, (\lambda_{X,J}, \lambda_{Y,1}), (\lambda_{X,J}, \lambda_{Y,2}) \ldots, (\lambda_{X,J}, \lambda_{Y,J}))$$

denote the J^2-vector of unique evaluation points.

Since each of our estimates is just a sum of IID random variables, we can apply a multivariate central limit theorem to find its asymptotic distribution. Specifically, letting the population moment of order $2s - 2$ of the random vector (X_K, Y_K) exist, $\sqrt{n_K}[\hat{D}_K^s(\lambda) - D_K^s(\lambda)] \xrightarrow{D} N(0, \Sigma_K)$, where Σ_K has typical element

$$\begin{aligned}
&\lim_{n_K \to \infty} n_K \mathrm{Cov}[\hat{D}_K^s(\lambda_{X,j}, \lambda_{Y,k}), \hat{D}_K^s(\lambda_{X,l}, \lambda_{Y,m})] \\
&= \frac{1}{[(s-1)!]^2} E[(\lambda_{X,j} - X_K)_+^{s-1}(\lambda_{Y,k} - Y_K)_+^{s-1}(\lambda_{X,l} - X_K)_+^{s-1}(\lambda_{Y,m} - Y_K)_+^{s-1}] \\
&\quad - D_K^s(\lambda_{X,j}, \lambda_{Y,k}) D_K^s(\lambda_{X,l}, \lambda_{Y,m}),
\end{aligned}$$

with $j, k, l, m = 1, \ldots, J$. These results follow directly from Davidson and Duclos (2000) and Duclos et al. (2006).

A consistent estimate of $\mathrm{Cov}[\hat{D}_K^s(\lambda_{X,j}, \lambda_{Y,k}), \hat{D}_K^s(\lambda_{X,l}, \lambda_{Y,m})]$ can be obtained using

$$\begin{aligned}
&\hat{\mathrm{Cov}}[\hat{D}_K^s(\lambda_{X,j}, \lambda_{Y,k}), \hat{D}_K^s(\lambda_{X,l}, \lambda_{Y,m})] \\
&= \frac{1}{n_K[(s-1)!]^2} \sum_{i=1}^{n} [(\lambda_{X,j} - x_i)_+^{s-1}(\lambda_{Y,k} - y_i)_+^{s-1}(\lambda_{X,l} - x_i)_+^{s-1}(\lambda_{Y,m} - y_i)_+^{s-1}] \\
&\quad - \hat{D}^s(\lambda_{X,j}, \lambda_{Y,k}) \hat{D}^s(\lambda_{X,l}, \lambda_{Y,m}).
\end{aligned}$$

In the following section, we show how these results can be used to test for bivariate stochastic dominance between two populations.

3 Hypothesis testing

To test for bivariate stochastic dominance, we use the general approach to testing multivariate inequality restrictions of Kodde and Palm (1986) and Wolak (1989). This approach has also been used for tests for of univariate stochastic dominance by Fisher et al. (1998) and Davidson and Duclos (2000).

Specifically, we are interested in testing hypotheses of the form

$$H_0 : F_A \succeq_s F_B$$

against an unrestricted alternative. Letting $\Delta = D_B^s(z_X, z_Y) - D_A^s(z_X, z_Y)$, we can rewrite the null hypothesis above as

$$H_0 : \Delta \geq 0.$$

The unrestricted estimate of Δ is $\hat{\Delta} = \hat{D}_B^s(\lambda) - \hat{D}_A^s(\lambda)$, where $\hat{D}_K(\lambda)$ is the estimator given in the previous section for population $K = A, B$. The restricted estimate of Δ can be found as the solution to the following minimization problem:

$$\min_{\Delta \geq 0} (\hat{\Delta} - \Delta)' \hat{\Omega}^{-1} (\hat{\Delta} - \Delta), \tag{2}$$

where $\hat{\Sigma}$ is an estimate the asymptotic covariance matrix of $\hat{\Delta}$. Under the assumption that A and B represent two independent samples, we have

$$\hat{\Omega} = \hat{\Sigma}_A/n_A + \hat{\Sigma}_B/n_B,$$

where $\hat{\Sigma}_K$ is an estimate of the asymptotic covariance matrix of $\sqrt{n_K}(\hat{D}^s_K - D^s_K)$, for $K = A, B$ (see Section 2 for details).[2]

Solving for Δ in (2) is a straightforward quadratic programming (QP) problem. Denoting the solution by $\tilde{\Delta}$, we have the Wald-type test statistic

$$W = (\hat{\Delta} - \tilde{\Delta})'\hat{\Omega}^{-1}(\hat{\Delta} - \tilde{\Delta}).$$

As shown by Kodde and Palm (1986), under the null, W will converge in distribution to a mixture of χ^2 distributions.

To avoid the complexities associated with computing the critical values for W (see Wolak, 1989 for a more complete discussion), we suggest using the bootstrap. Specifically, we combine samples of observations on each population into pooled sample (which is of length $n_A + n_B$). Resampling (in pairs) n_K observations from this pooled sample produces the bootstrap sample $\{(x^*_{K,i}, y^*_{K,i})\}_{i=1}^{n_K}$, for $K = A, B$. Using the two bootstrap samples, we calculate the bootstrap test statistic, W^*, in a matter analogous to that for the original test statistic, W. Repeating this process some large number

[2]See Duclos et al. (2006) for a discussion on how to estimate the covariance matrix in the case of dependent samples.

of times, the bootstrap p-value for W is the proportion of times that W^* exceeds W.

We now briefly contrast this approach with that of MY, who consider tests of multivariate stochastic dominance of the category two type. Their test statistic is

$$T = \left\{ \int [\psi^s(\mathbf{u})]^2 d\mathbf{u} \right\}^{1/2}$$

where $\psi^s(\mathbf{u}) = \max\{D_A^s(\mathbf{u}) - D_B^s(\mathbf{u}), 0\}$. Of course, when the null is true, T is equal to zero.

In practice, this test involves estimating T and testing whether it is statistically different from zero. Specifically, in the bivariate case, MY estimate T by

$$\hat{T} = \left\{ \sum_{j=1}^{J} \sum_{k=1}^{J} [\hat{\psi}^s(\lambda_{X,j}, \lambda_{Y,k})]^2 \right\}^{1/2},$$

where

$$\hat{\psi}^s(\lambda_{X,j}, \lambda_{Y,k}) = \max\{\hat{D}_A^s(\lambda_{X,j}, \lambda_{Y,k}) - \hat{D}_B^s(\lambda_{X,j}, \lambda_{Y,k}), 0\},$$

and $\hat{D}_A^s(\lambda_{X,j}, \lambda_{Y,k})$ and $\hat{D}_B^s(\lambda_{X,j}, \lambda_{Y,k})$ are obtained using the estimator in (1). As in our approach, $\lambda_{X,1}, \ldots, \lambda_{X,J}$ denote a set of points on the combined support of X_A and X_B, $\lambda_{Y,1}, \ldots, \lambda_{Y,J}$ denote a set of points on the combined support of Y_A and Y_B. Thus, in practice, this test would seem to fall in the same category as our proposed one. However, there is nothing inhibiting the use of an extremely large number of grid points (perhaps every unique point supported by the combined sample). That being said, MY use $J = 25$ in their simulations and empirical applications. While this number of grid points would be quite computationally demanding for our approach (requiring, e.g., the inverse of a $25^2 \times 25^2$ covariance matrix to be computed), it would not be out of the question given current processing power.

Finally, as MY note, $\hat{T}$ does not have a known asymptotic distribution. Accordingly, they suggest the use of a bootstrap procedure which is analogous to the one we have described above for our proposed test statistic, W.

4 Simulation evidence

We now present the results of some simple Monte Carlo experiments. Each of these experiments involves generating 100,000 sets of two independent samples, one from distribution A and one from distribution B, and testing the null hypothesis $H_0 : F_A \succeq_1 F_B$. The distributions used are various parameterizations of the bivariate lognormal distribution (see Table 1), some of which were also used by MY. The size of the samples are $n_A = n_B = n = 50$ and 500.

Table 1: Parameters for simulated data

	$E(\log X)$	$E(\log Y)$	$\text{Var}(\log X)$	$\text{Var}(\log Y)$	$\text{Cov}(\log X, \log Y)$
D1	0.85	0.85	0.36	0.36	0.20
D2	0.60	0.60	0.64	0.64	0.20
D3	0.85	0.60	0.36	0.64	0.20

Note: In each case, $\log X$ and $\log Y$ are bivariate normal.

We consider three different cases. In the first case, distribution D1 is used to generate samples for both A and B. Since the null is (weakly) true in this case, we would expect to reject it at the nominal level of the test. In the second case, distribution D2 is used to generate samples for A and distribution D1 is used to generate samples for B. In this case, the null is clearly false ($F_B \succeq_1 F_A$), so the rejection frequencies can give us an idea of the relative power of the tests. In the third case, distribution D3 is used to generate samples for A and distribution D1 is used to generate samples for B. Letting $F_{K,X}$ and $F_{K,Y}$ denote the marginal distribution functions of X and Y, respectively, for population $K = A, B$, we have, in this case, $F_{A,X} \succeq_1 F_{B,X}$ but $F_{B,Y} \succeq_1 F_{A,Y}$. Thus, the null hypothesis $H_0 : F_A \succeq_1 F_B$ is false, but not so clearly as in the second case. Hence, the rate or rejection should be lower in this case.

The simulated rejection frequencies for tests based on the W and $\hat{T}$ statistics at the 10%, 5%, and 1% nominal levels are reported in Table 2. For both test statistics we use $J = 9$, so that the total number of grid points is 81.[3] These points are chosen along each dimension so as to divide the combined sample into 10 intervals which contain an equal number of observations. We use 99 bootstrap replications.

[3] We also computed $\hat{T}$ using $J = 25$ (the number used by MY in their simulations), and found no material difference in the rejection frequencies.

Table 2: Rejection frequencies for simulated data

				$n = 50$		$n = 500$	
Case	F_A	F_B	Level	W	$\hat{T}$	W	$\hat{T}$
1	D1	D1	10%	0.1065	0.1019	0.1009	0.1011
			5%	0.0537	0.0515	0.0505	0.0502
			1%	0.0115	0.0104	0.0099	0.0010
2	D2	D1	10%	0.7655	0.6620	1.0000	1.0000
			5%	0.6307	0.5005	1.0000	1.0000
			1%	0.3179	0.2080	1.0000	0.9976
3	D3	D1	10%	0.5181	0.3891	1.0000	0.9994
			5%	0.3671	0.2430	1.0000	0.9945
			1%	0.1348	0.0696	0.9991	0.8910

Notes: The null hypothesis in each case is $H_0 : F_A \succeq_1 F_B$.

Based on the results for the first case, it is clear that the sizes of the tests based on both statistics are extremely close to their nominal levels, particularly for $n = 500$. However, as evidenced by the rejection rates in the second and third cases, tests based the W statistic seem to have substantially higher power for both sample sizes than those based on $\hat{T}$.

5 Empirical example

For illustrative purposes, we now consider an empirical application which uses income and health status data for two subgroups of the Canadian population: those born in Canada, and those born outside. The data for this example is obtained from the *Joint Canada/United States Survey of Health* conducted in 2002-2003.[4]

In order to reduce the level of heterogeneity within the sample, only unattached individuals living in Canada are included. There are 568 Canadian-born individuals and 99 foreign-born individuals for which we have data on income, as measured by income from all sources, and health status, as measured by the Health Utilities Index Mark 3 (HUI3). The HUI3 is part of the Comprehensive Health Status Measurement System developed at McMaster University's Centre for Health Economics and Policy Analysis, and is designed to measure an individual's overall functional health. It is based on eight attributes: vision, hearing, speech, mobility, dexterity, cognition, emotion, and pain and discomfort; see Furlong et al. (1998) for more details. Summary statistics for the data are provided in Table 3.

Table 3: Summary statistics for empirical example

	Canadian-born	Foreign-born
Mean of income	35,033	34,688
Std. dev. of income	24,063	27,900
Mean of health status	0.8362	0.8586
Std. dev of health status	0.2249	0.2012
Correlation	0.2714	0.2369
No. of observations	568	99

Using this data, we conduct tests for first-order stochastic dominance based on both the W and the $\hat{T}$ test statistics discussed above. For both test statistics we use $J = 19$, so that the total number of grid points is 361.

[4]For the purposes of this example, we ignore the complex sampling scheme of the survey data; see Davidson and Duclos (2000).

Analogous to what was done in the simulation described above, these points are chosen along each dimension so as to divide the combined sample into 20 intervals which contain an equal number of observations. Here, we use 999 bootstrap replications.

Testing for first-order bivariate stochastic dominance of Canadian-born individuals over foreign-born individuals, the W and $\hat{T}$ test statistics are 35.1351 and 0.4593, respectively, and the bootstrap p-values are 0.0390 and 0.4244, respectively. Thus, we can reject the null of first-order stochastic dominance (at, say, the 5% level) using the W test statistic, but can not do so using the $\hat{T}$ test statistic. On the other hand, testing for first-order stochastic dominance of foreign-born individuals over Canadian-born individuals, the W and $\hat{T}$ test statistics are 4.0143 and 0.5274, respectively, and the bootstrap p-values are 0.5295 and 0.3854, respectively.

6 Conclusion

In this paper, we have proposed a test for bivariate stochastic dominance which involves evaluating each CDF at a finite number of points (i.e., over a set of grid points). Simulation evidence presented here suggests that the proposed test has substantially higher power than the test of MY. This conclusion is borne out by the results of our empirical example; using the test of MY we are unable to obtain any clear inference, while our proposed test leads us to suggest that the joint distribution of income and health status for foreign-born individuals dominates that of Canadian-born individuals stochastically at first-order.

References

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica*, 64(5):1183–1193.

Anderson, G. (2008). The empirical assessment of multidimensional welfare, inequality and poverty: Sample weighted multivariate generalizations of the Kolmogorov-Smirnov two sample tests for stochastic dominance. *Journal of Economic Inequality*, 6(1):73–87.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

Beach, C. M. and Davidson, R. (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies*, 50(4):723–735.

Beach, C. M. and Richmond, J. (1985). Joint confidence intervals for income shares and Lorenz curves. *International Economic Review*, 26(2):439–450.

Bishop, J. A., Formby, J. P., and Smith, W. J. (1993). International comparisons of welfare and poverty: Dominance orderings for ten countries. *Canadian Journal of Economics*, 26(3):707–726.

Dardanoni, V. and Forcina, A. (1999). Inference for Lorenz curve orderings. *Econometrics Journal*, 2(1):49–75.

Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and the measurement of povery and inequality. *Econometrica*, 68(6):1435–1464.

Davidson, R. and Duclos, J.-Y. (2007). Testing for restricted stochastic dominance. Unpublished manuscript.

Duclos, J.-Y., Sahn, D., and Younger, S. D. (2006). Robust multidimensional poverty comparisons. *Economic Journal*, 116(514):943–968.

Fisher, G., Wilson, D., and Xu, K. (1998). An empirical anlysis of term premiums using significance tests for stochastic dominance. *Economics Letters*, 60(2):195–203.

Furlong, W., Feeny, D., Torrance, G. W., Goldsmith, C., DePauw, S., Boyle, M., Denton, M., and Zhu, Z. (1998). Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: A technical report. McMaster University Centre for Health Economics and Policy Analysis Working Paper No. 98-11.

Horvath, L., Kokoszka, P., and Zitikis, R. (2006). Testing for stochastic dominance using the weighted Mcfadden-type statistic. *Journal of Econometrics*, 133(1):191–205.

Kaur, A., Prakasa Rao, B. L. S., and Singh, H. (1994). Testing for second-order stochastic dominance of two distributions. *Econometric Theory*, 10(5):849–866.

Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54(5):1243–1248.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, 72(3):735–765.

Maasoumi, E. and Heshmati, A. (2000). Stochastic dominance amongst Swedish income distributions. *Econometric Reviews*, 19(3):287–320.

McCaig, B. and Yatchew, A. (2007). International welfare comparisons and nonparametric testing of multivariate stochastic dominance. *Journal of Applied Econometrics*, 22(5):951–969.

McFadden, D. (1989). Testing for stochastic dominance. In Fomby, T. B. and Seo, T. K., editors, *Studies in the Economics of Uncertainty*. Springer-Verlag, New York.

Wolak, F. A. (1989). Testing inequality constraints in linear econometric models. *Journal of Econometrics*, 41(2):205–235.

Xu, K. and Osberg, L. (1998). A distribution-free test for deprivation dominance. *Econometric Reviews*, 17(4):415–429.