Consistent Testing for Stochastic Dominance under General Sampling Schemes[*]

Oliver Linton[†]

London School of Economics

Esfandiar Maasoumi[‡]

Southern Methodist University

Yoon-Jae Whang[§]

Ewha University

February 2003 (First Version April 2001)

Abstract

We propose a procedure for estimating the critical values of the extended Kolmogorov-Smirnov tests of First and Second Order Stochastic Dominance in the general K-prospect case. We allow for the observations to be serially dependent and, for the first time, we can accommodate *general* dependence amongst the *prospects* which are to be ranked. Also, the prospects may be the residuals from certain conditional models, opening the way for *conditional* ranking. We also propose a test of Prospect Stochastic Dominance. Our method is subsampling; we show that the resulting tests are consistent and powerful against some $N^{-1/2}$ local alternatives even when computed with a data-based subsample size. We also propose some heuristic methods for selecting subsample size and demonstrate in simulations that they perform reasonably. We show that our test is asymptotically similar on the entire boundary of the null hypothesis, and

is unbiased. In comparison, any method based on resampling or simulating from the least favor-able distribution does not have these properties and consequently will have less power against some alternatives.

1   Introduction

There is considerable interest in uniform weak ordering of investment strategies, welfare outcomes (income distributions, poverty levels), and in program evaluation exercises. Partial *strong* orders are commonly used on the basis of specific utility (loss) functions. This is the predominant form of evaluation and is done when one employs indices of inequality or poverty in welfare, mean-variance analysis in finance, or performance indices in program evaluation. By their very nature, strong orders do not command consensus. The most popular *uniform* order relations are the Stochastic Dominance (SD) relations of various orders, based on the expected utility paradigm and its mathematical reg-ularity conditions. These relations are defined over relatively large classes of utility functions and represent "majority" preferences.

In this paper we propose an alternative procedure for estimating the critical values of a suitably extended Kolmogorov-Smirnov test for first and second order stochastic dominance in the general K-prospect case. Alternative implementations of this test have been examined by several authors including McFadden (1989), Klecan, McFadden, and McFadden (1991), and Barrett and Donald (2002). Our method is based on subsampling. We prove that the resulting test is consistent against all (nonparametric) alternatives. Our sampling scheme is quite general: for the first time in this literature, we allow for general dependence amongst the prospects, and for the observations to be autocorrelated over time. Accommodating generic dependence between the variables which are to be ranked is especially necessary in substantive empirical settings where income distributions, say, are compared before and after taxes (or some other policy decision), or returns on different funds are compared in the same or interconnected markets. We are not aware of any evidence suggesting either that such prospects are independent or exchangeable. Indeed such assumptions appear to be patently false in most empirical settings of consequence to policy analysis.

We also allow the prospects themselves to be residuals from some estimated model. This latter generality is very important for policy makers where one wishes to control for certain characteristics before comparing outcomes. For instance, one may wish to "purge" incomes from the influence of age and/or education, thereby isolating both their influence and the separate contribution of other factors (collectively) on the income distribution. For example, in a recent study, Maasoumi and Millimet (2002) control for the influence of 'growth' on the distribution of several pollutants in the US. This is done by comparing the results of unconditional dominance tests with tests of dominance amongst

the residual distributions.[1] Based on their SD tests, they are able to infer that incomes contribute positively while other factors collectively have a negative influence on environmental quality. See also Abadie (2001) for comments about the desirability of controlling for observables before applying such tests. Style analysis [Sharpe (1992)] is currently a popular method amongst practitioners for ranking the performance of investment funds after taking account of their 'style', e.g., value or growth funds. This involves a comparison of features of the residuals from a linear regression. Finally, given the recent credible challenges to the standard risk aversion and expected utility paradigm, we propose a 'new' test of Prospect Stochastic Dominance and propose consistent critical values using subsampling.

The finite sample performance of our method is investigated on simulated data and found to be quite good provided the sample sizes are appropriately large for distributional rankings. Our simulation designs include the Burr distributions examined by Tse and Zhang (2000), the lognormal distribution recently employed by Barrett and Donald (2000), and the multivariate normal with exchangeable and correlated prospects as in Klecan et al. (1991). Suggestive results on subsample size are provided, and some power comparisons with other methods are given. In addition, we describe an empirical application to Dow Jones and S&P daily returns which demonstrates the potential of these tests and concludes the paper. The following brief definitions will be useful:

Let X_1 and X_2 be two variables (incomes, returns/prospects) at either two different points in time, or for different regions or countries, or with or without a program (treatment). Let X_{ki}, $i = 1, \ldots, N; k = 1, 2$ denote the not necessarily i.i.d. observations. Let $\mathcal{U}_1$ denote the class of all von Neumann-Morgenstern type utility functions, u, such that $u' \geq 0$, (increasing). Also, let $\mathcal{U}_2$ denote the class of all utility functions in $\mathcal{U}_1$ for which $u'' \leq 0$ (strict concavity), and $\mathcal{U}_3$ denote a subset of $\mathcal{U}_2$ for which $u''' \geq 0$. Let $X_{(1p)}$ and $X_{(2p)}$ denote the p-th quantiles, and $F_1(x)$ and $F_2(x)$ denote the cumulative distribution functions, respectively.

Definition 1 X_1 *First Order Stochastic Dominates* X_2, *denoted* $X_1 \succeq_{FSD} X_2$, *if and only if:*
(1) $E[u(X_1)] \geq E[u(X_2)]$ *for all* $u \in \mathcal{U}_1$, *with strict inequality for some* u; *Or*
(2) $F_1(x) \leq F_2(x)$ *for all* x *with strict inequality for some* x; *Or*
(3) $X_{(1p)} \geq X_{(2p)}$ *for all* $0 \leq p \leq 1$, *with strict inequality for some* p.

Definition 2 X_1 *Second Order Stochastic Dominates* X_2, *denoted* $X_1 \succeq_{SSD} X_2$, *if and only if one of the following equivalent conditions holds:*
     (1) $E[u(X_1)] \geq E[u(X_2)]$ *for all* $u \in \mathcal{U}_2$, *with strict inequality for some* u; *Or:*

---

[1] The regression method for purging of the dependent variable from certain conditioning variables is well understood. If these conditioning variables are the only ones relevant to the 'true' data generating process, the residuals will have zero means. The residuals will normally be orthogonal to the conditioning variables by construction. Neither this fact, nor the possibly zero means for the residuals precludes the existence of dominance relations between their distributions.

(2) $\int_{-\infty}^{x} F_1(t)dt \leq \int_{-\infty}^{x} F_2(t)dt$ *for all x with strict inequality for some x; Or:*

(3) $\Phi_1(p) = \int_0^p X_{(1t)}dt \geq \Phi_2(p) = \int_0^p X_{(2t)}dt$ *for all $0 \leq p \leq 1$, with strict inequality for some value(s) p.*

Weak orders of SD obtain by eliminating the requirement of strict inequality at some point. When these conditions are not met, as when either Lorenz or Generalized Lorenz Curves of two distributions cross, unambiguous First and Second order SD is not possible. Any partial ordering by specific *indices* that correspond to the utility functions in $\mathcal{U}_1$ and $\mathcal{U}_2$ classes, will not enjoy general consensus. Whitmore introduced the concept of third order stochastic dominance (TSD) in finance, see (e.g.) Whitmore and Findley (1978). Shorrocks and Foster (1987) showed that the addition of a "transfer sensitivity" requirement leads to TSD ranking of income distributions. This requirement is stronger than the Pigou-Dalton principle of transfers since it makes regressive transfers less desirable at lower income levels. Higher order SD relations correspond to increasingly smaller subsets of $\mathcal{U}_2$. Davidson and Duclos (2000) offer a very useful characterization of these relations and their tests. See Post (2002) for a recent application of stochastic dominance in finance.

In this paper we shall also consider the concept of prospect stochastic dominance. Kahneman and Tversky (1979) mounted a critique of expected utility theory and introduced an alternative theory, called prospect theory. They argued that their model provided a better rationalization of the many observations of actual individual behavior taken in laboratory experiments. Specifically, they proposed an alternative model of decision making under uncertainty in which: (a) gains and losses are treated differently; (b) individuals act as if they had applied monotonic transformations to the underlying probabilities before making payoff comparisons.[2] Taking only part (a), individuals would rank prospects according to the expected value of S-shaped utility functions $u \in \mathcal{U}_P \subseteq \mathcal{U}_1$ for which $u''(x) \leq 0$ for all $x > 0$ but $u''(x) \geq 0$ for all $x < 0$. These properties represent risk seeking for losses but risk aversion for gains. This leads naturally to the concept of Prospect Stochastic Dominance.

Definition 3 *X_1 Prospect Stochastic Dominates X_2, denoted $X_1 \succeq_{PSD} X_2$, if and only if one of the following equivalent conditions holds:*

(1) $E[u(X_1)] \geq E[u(X_2)]$ *for all $u \in \mathcal{U}_P$, with strict inequality for some u; Or:*

(2) $\int_y^x F_1(t)dt \leq \int_y^x F_2(t)dt$ *for all pairs (x,y) with $x > 0$ and $y < 0$ with strict inequality for some (x,y); Or:*

(3) $\int_{p_1}^{p_2} X_{(1t)}dt \geq \int_{p_1}^{p_2} X_{(2t)}dt$ *for all $0 \leq p_1 \leq F_1(0) \leq F_2(0) \leq p_2 \leq 1$, with strict inequality for some value(s) p.*

---

[2]In Tversky and Kahneman (1992) this idea is refined to make the cumulative distribution function of payoffs the subject of the transformation. Thus, individuals would compare the distributions $F_k^* = T(F_k)$, where T is a monotonic decreasing transformation that can be interpreted as a subjective revision of probabilities that varies across investors.

Now consider the second component of prospect theory, (b), the transformation of probabilities. One question is whether stochastic dominance [of first, second, or higher order] is preserved under transformation, or rather what is the set of transformations under which an ordering is preserved. Levy and Wiener (1998) show that the PSD property is preserved under the class of monotonic transformations that are concave for gains and convex for losses. Therefore, if one can verify that a prospect is dominated according to (2), this implies that it will be dominated even after transforming the probabilities for a range of such transformations.

Econometric tests for the existence of SD orders involve composite hypotheses on inequality restrictions. These restrictions may be equivalently formulated in terms of distribution functions, their quantiles, or other conditional moments. Different test procedures may also differ in terms of their accommodation of the inequality nature (information) of the SD hypotheses. The literature also divides according to whether the tests are designed to be consistent against all alternatives or whether the class of alternatives against which the test has power is essentially finite dimensional. Most of the large literature works with tests that have the more limited objective. Even in that case the statistical problems discussed in the opening of this paragraph are quite formidable. See for example Anderson (1996), Davidson and Duclos (2000), Kaur et al. (1994), Dardanoni and Forcina (2000), Bishop et al. (1998), and Xu, Fisher, and Wilson (1995), and Crawford (1999). Maasoumi (2001) contains an extensive discussion of these alternative approaches. Tse and Zhang (2000) provide some Monte Carlo evidence on the power of some of these *alternative* tests. There are just a handful of papers that have pursued the more general objective of consistency against all alternatives, which we next discuss.

McFadden (1989) proposed a generalization of the Kolmogorov-Smirnov test of First and Second order SD among K (≥ 1) prospects (distributions) based on i.i.d. observations *and independent* prospects. Klecan, McFadden, and McFadden (1991) extended these tests allowing for dependence in observations, and replacing independence with a general exchangeability amongst the competing prospects. Since the asymptotic null distribution of these tests depends on the unknown distributions, they proposed a Monte Carlo permutation procedure for the computation of critical values that relies on an exchangeability property. In fact, although they derived the asymptotic distribution of the test statistics allowing for time series dependence the proof that their critical values were consistent is only valid in the i.i.d. over time case. Barrett and Donald (1999) propose an alternative simulation method based on an idea of Hansen (1996b) for deriving critical values in the case where the prospects are mutually independent, and the data are i.i.d. We note that the methods relying on standard bootstrap or simulation typically try to mimic the asymptotic null distributions in the least favorable case of the equal distribution functions for the prospects. However, the null hypothesis of stochastic dominance is a composite hypothesis and hence the tests based on the approximation of the least

5

favorable case are not asymptotically similar on the boundary of the null hypothesis. On the other hand, our test based on a subsampling procedure which approximates the true sampling distribution under the composite null hypothesis is asymptotically similar on the boundary. Consequently, our test might be asymptotically more powerful than the bootstrap (or simulation)-based tests for some local alternatives, see Section 4 for details.

2   The Test Statistics

We shall suppose that there are K prospects $X_1, \ldots, X_k$ and let $\mathcal{A} = \{X_k : k = 1, \ldots, K\}$. Let $\{X_{ki} : i = 1, \ldots, N\}$ be realizations of X_k for $k = 1, \ldots, K$. To subsume the empirically important case of "conditional" dominance, we suppose that $\{X_{ki} : i = 1, \ldots, N\}$ might depend on an unknown finite dimensional parameter $\theta_{k0} \in \Theta_k \subset \mathbb{R}^{L_k}$:

$$X_{ki} = Y_{ki} - Z'_{ki}\theta_{k0}, \tag{1}$$

where the random variables $Y_{ki} \in \mathbb{R}$ and $Z_{ki} \in \mathbb{R}^{L_k}$ satisfy the linear regression relationship

$$Y_{ki} = \mu_{k0} + Z'_{ki}\theta_{k0} + \varepsilon_{ki}, \ E(\varepsilon_{ki}|Z_{ki}) = 0 \text{ a.s.} \tag{2}$$

for $\mu_{k0} \in \mathbb{R}$, $i = 1, \ldots, N$ and $k = 1, \ldots, K$. Therefore, X_{ki} can be viewed as an "intercept-adjusted" regression error with mean μ_{k0}. We allow for serial dependence of the realizations and for mutual correlation across prospects. Let $X_{ki}(\theta) = Y_{ki} - Z'_{ki}\theta$, $X_{ki} = X_{ki}(\theta_{k0})$, and $\widehat{X}_{ki} = X_{ki}(\widehat{\theta}_k)$, where $\widehat{\theta}_k$ is some sensible estimator of θ_{k0} whose properties we detail below, i.e., the prospects can be estimated from the data. (When the prospects do not depend on estimated parameters, i.e., $X_{ki}(\theta) = X_{ki}$, results analogous to those given below can be established using a substantially simpler arguments than ours.) Since we have a linear regression model, there are many possible ways of obtaining consistent estimates of the unknown parameters. The motivation for considering estimated prospects is that when data is limited one may want to use a model to adjust for systematic differences. Common practice is to group the data into subsets, say of families with different sizes, or by educational attainment, or subgroups of funds by investment goals, and then make comparisons across homogenous populations. When data are limited this can be difficult. In addition, the preliminary regressions may identify "causes" of different outcomes which may be of substantive interest and useful to control for.

For $k = 1, \ldots, K$, define

$$F_k(x, \theta) = P\left(X_{ki}(\theta) \leq x\right) \text{ and}$$

$$\overline{F}_{kN}(x, \theta) = \frac{1}{N}\sum_{i=1}^{N} 1\left(X_{ki}(\theta) \leq x\right).$$

6

We denote $F_k(x) = F_k(x, \theta_{k0})$ and $\overline{F}_{kN}(x) = \overline{F}_{kN}(x, \theta_{k0})$, and let $F(x_1, \ldots, x_k)$ be the joint c.d.f. of $(X_1, \ldots, X_k)'$. Now define the following functionals of the joint distribution

$$d^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} [F_k(x) - F_l(x)] \tag{3}$$

$$s^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} \int_{-\infty}^{x} [F_k(t) - F_l(t)]\, dt \tag{4}$$

$$p^* = \min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} \int_{y}^{x} [F_k(t) - F_l(t)]\, dt, \tag{5}$$

where $\mathcal{X}$ denotes a given set contained in the union of the supports of X_{ki} for $k = 1, \ldots, K$ and $\mathcal{X}_+ = \{x \in \mathcal{X}, \ x > 0\}$. Without loss of generality we assume that the supports are bounded, as do Klecan et al. (1991). The hypotheses of interest can now be stated as:

$$H_0^d : d^* \leq 0 \ vs. \ H_1^d : d^* > 0 \tag{6}$$

$$H_0^s : s^* \leq 0 \ vs. \ H_1^s : s^* > 0 \tag{7}$$

$$H_0^p : p^* \leq 0 \ vs. \ H_1^p : p^* > 0. \tag{8}$$

The null hypothesis H_0^d implies that the prospects in $\mathcal{A}$ are not first-degree stochastically maximal, i.e., there exists at least one prospect in $\mathcal{A}$ which first-degree dominates the others. Likewise for the second order and prospect stochastic dominance test.

The test statistics we consider are based on the empirical analogues of (3)-(5). They are defined to be:

$$D_N = \min_{k \neq l} \sup_{x \in \mathcal{X}} \sqrt{N} \left[\overline{F}_{kN}(x, \widehat{\theta}_k) - \overline{F}_{lN}(x, \widehat{\theta}_l) \right]$$

$$S_N = \min_{k \neq l} \sup_{x \in \mathcal{X}} \sqrt{N} \int_{-\infty}^{x} \left[\overline{F}_{kN}(t, \widehat{\theta}_k) - \overline{F}_{lN}(t, \widehat{\theta}_l) \right] dt$$

$$P_N = \min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} \sqrt{N} \int_{y}^{x} \left[\overline{F}_{kN}(t, \widehat{\theta}_k) - \overline{F}_{lN}(t, \widehat{\theta}_l) \right] dt.$$

The first two are the same as the Klecan et al. (1991) test statistics except that we have allowed the prospects to have been estimated from the data.

We next discuss the issue of how to compute the supremum in D_N, S_N and P_N, and the integrals in S_N and P_N. There have been a number of suggestions in the literature that exploit the step-function nature of $\overline{F}_{kN}(t, \theta)$. The supremum in D_N can be (exactly) replaced by a maximum taken over all the distinct points in the combined sample. Regarding the computation of S_N, Klecan et al. (1991) propose a recursive algorithm for exact computation of S_N, see also Barratt and Donald (1999) for an extension to third order dominance statistics. Integrating by parts we have

$$\int_{-\infty}^{x} F_k(t)dt = E[\max\{0, x - X_k\}],$$

7

which holds for all x provided $E \|X_k\| < \infty$. Motivated by this, Davidson and Duclos (1999) have proposed computing the empirical analogue

$$\frac{1}{N} \sum_{i=1}^{N} (x - X_{ki}(\theta)) 1\left(X_{ki}(\theta) \leq x\right).$$

The computation of P_N can be based on the fact that

$$\int_y^x \overline{F}_{kN}(t, \theta) dt = \int_{-\infty}^x \overline{F}_{kN}(t, \theta) dt - \int_{-\infty}^y \overline{F}_{kN}(t, \theta) dt$$

for all x, y.

To reduce the computation time, it may be preferable to compute approximations to the suprema in D_N, S_N, P_N based on taking maxima over some smaller grid of points $\mathcal{X}_J = \{x_1, \ldots, x_J\}$, where $J < n$. This is especially true of P_N, which requires a grid on $\mathbb{R}_+ \times \mathbb{R}_-$. Thus, we might compute

$$P_N^J = \min_{k \neq l} \max_{0 < x, 0 > y \in \mathcal{X}_J} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{(x - X_{ki}(\widehat{\theta}_k)) 1(X_{ki}(\widehat{\theta}_k) \leq x) - (y - X_{li}(\widehat{\theta}_l)) 1(X_{li}(\widehat{\theta}_l) \leq y)\}.$$

Theoretically, provided the set of evaluation points becomes dense in the joint support, the distribution theory is unaffected by using this approximation.

3   Asymptotic Null Distributions

3.1   Regularity Conditions

We need the following assumptions to analyze the asymptotic behavior of our test statistics:

Assumption 1: (i) $\{(Y_{ki}, Z_{ki}) : i = 1, \ldots, N\}$ is a strictly stationary and α- mixing sequence with $\alpha(m) = O(m^{-A})$ for some $A > \max\{(q-1)(q+1), 1+2/\delta\}$ for $k = 1, \ldots, K$, where q is an even integer that satisfies $q > 3(L_{\max} + 1)/2$, $L_{\max} = \max\{L_1, \ldots, L_K\}$ and δ is a positive constant that also appears in Assumption 2(ii) below. (ii) $E \|Z_{ki}\|^2 < \infty$ for all $k = 1, \ldots, K$, for all $i \geq 1$. (iii) The conditional distribution $H_k(\cdot|Z_{ki})$ of X_{ki} given Z_{ki} has bounded density with respect to Lebesgue measure a.s. for $k = 1, \ldots, K$, for all $i \geq 1$.

Assumption 2: (i) The parameter estimator satisfies $\sqrt{N}(\widehat{\theta}_k - \theta_{k0}) = (1/\sqrt{N}) \sum_{i=1}^{N} \Gamma_{k0} \psi_k(Y_{ki}, Z_{ki}, \theta_{k0}) + o_p(1)$, where Γ_{k0} is a non-stochastic matrix for $k = 1, \ldots, K$; (ii) The function $\psi_k(y, z, \theta) : \mathbb{R} \times \mathbb{R}^{L_k} \times \Theta_k \to \mathbb{R}^{L_k}$ is measurable and satisfies (a) $E\psi_k(Y_{ki}, Z_{ki}, \theta_{k0}) = 0$ and (b) $E \|\psi_k(Y_{ki}, Z_{ki}, \theta_{k0})\|^{2+\delta} < \infty$ for some $\delta > 0$ and for $k = 1, \ldots, K$, for all $i \geq 1$.

Assumption 3: (i) The function $F_k(x, \theta)$ is differentiable in θ on a neighborhood Θ_{k0} of θ_{k0} for $k = 1, \ldots, K$; (ii) For all sequence of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$,

$\sup_{x \in \mathcal{X}} \sup_{\theta:\|\theta-\theta_{k0}\| \leq \xi_N} \|\partial F_k(x,\theta)/\partial\theta - \Delta_{k0}(x)\| \to 0$ for $k = 1, \ldots, K$, where $\Delta_{k0}(x) = \partial F_k(x,\theta_{k0})/\partial\theta$; (iii) $\sup_{x \in \mathcal{X}} \|\Delta_{k0}(x)\| < \infty$ for $k = 1, \ldots, K$.

For the tests S_N and P_N we need the following modification of Assumptions 1 and 3:

Assumption 1*: (i) $\{(Y_{ki}, Z_{ki}) : i = 1, \ldots, n\}$ is a strictly stationary and α- mixing sequence with $\alpha(m) = O(m^{-A})$ for some $A > \max\{rq/(r-q), 1 + 2/\delta\}$ for $k = 1, \ldots, K$ and some $r > q \geq 2$, where q satisfies $q > L_{\max}$ and δ is a positive constant that also appears in Assumption 2(ii). (ii) $E\|Z_{ki}\|^r < \infty$ for $k = 1, \ldots, K$, for all $i \geq 1$.

Assumption 3* : (i) Assumption 3(i) holds; (ii) For $k = 1, \ldots, K$ and for all sequence of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$, $\sup_{x \in \mathcal{X}} \sup_{\theta:\|\theta-\theta_{k0}\| \leq \xi_N} \left\|(\partial/\partial\theta) \int_{-\infty}^{x} F_k(t,\theta)dt - \Lambda_{k0}(x)\right\| \to 0$, where $\Lambda_{k0}(x) = (\partial/\partial\theta) \int F_k(y,\theta_{k0})dy$; (iii) $\sup_{x \in \mathcal{X}} \|\Lambda_{k0}(x)\| < \infty$ for $k = 1, \ldots, K$.

Assumption 3** : (i) Assumption 3(i) holds; (ii) For $k = 1, \ldots, K$ and for all sequence of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$, $\sup_{x, -y \in \mathcal{X}_+} \sup_{\theta:\|\theta-\theta_{k0}\| \leq \xi_N} \left\|(\partial/\partial\theta) \int_{y}^{x} F_k(t,\theta)dt - \Xi_{k0}(x,y)\right\| \to 0$, where $\Xi_{k0}(x,y) = (\partial/\partial\theta) \int_{y}^{x} F_k(t,\theta_{k0})dt$; (iii) $\sup_{x, -y \in \mathcal{X}_+} \|\Xi_{k0}(x,y)\| < \infty$ for $k = 1, \ldots, K$.

REMARKS.

1. The mixing condition in Assumption 1 is stronger than the condition used in Klecan et. al. (1991, Theorem 6). This assumption, however, is needed to verify the stochastic equicontinuity of the empirical process (for a class of bounded functions) indexed by estimated parameters, see proof of Lemma 1(a). Assumption 1* introduces a trade-off between mixing and moment conditions. This assumption is used to verify the stochastic equicontinuity result for the (possibly) unbounded functions that appear in the test S_N (or P_N), see proof of Lemma 1(b)(or (c)). Without the estimated parameters, weaker conditions on the dependence can be assumed.

2. Assumptions 3 and 3* (or 3**) differ in the amount of smoothness required. For second order (or prospect) stochastic dominance, less smoothness is required.

3. When there are no estimated parameters: we do not need any moment conditions for D_N and only a first moment for S_N, P_N, and the smoothness conditions on F are redundant.

3.2 The Null Distributions

In this section, we derive the asymptotic distributions of our test statistics under the null hypothesis.

To help understanding of the reader, we first introduce a heuristic argument for the test D_N in the simple setting where there are no estimated parameters and $K = 2$. Suppose that $F_1(x) \leq F_2(x)$ for all $x \in \mathcal{X}$ but $F_1(x) = F_2(x)$ for $x \in \mathcal{B}$ ($\subset \mathcal{X}$). Assume that $\mathcal{B}$ is nonempty, which implies $d^* = 0$. Let $\overline{A}_N(x) = \sqrt{N}[\overline{F}_{1N}(x) - \overline{F}_{2N}(x)]$, $A_N(x) = \sqrt{N}[F_1(x) - F_2(x)]$, and $\widetilde{A}_N(x) = \overline{A}_N(x) - A_N(x)$. By an empirical process CLT, the "centered" process $\widetilde{A}_N(\cdot)$ will converge weakly to a mean zero Gaussian process, say $v(\cdot)$, under suitable regularity conditions. Since $A_N(x) = 0$ for $x \in \mathcal{B}$ but

$A_N(x) \to -\infty$ for $x \notin \mathcal{B}$, it is easy to see that the supremum of the uncentered process $\overline{A}_N(x)$ ($=$ $\widetilde{A}_N(x) + A_N(x)$) over $x \in \mathcal{X}$ is approximately equal to the supremum of the centered process $\widetilde{A}_N(x)$ over $x \in \mathcal{B}$ for N sufficiently large. On the other hand, $\sup_{x \in \mathcal{X}}[-\overline{A}_N(x)]$ will diverge to infinity. Therefore, it follows that the asymptotic distribution of $D_N = \min\{\sup_{x \in \mathcal{X}}[\overline{A}_N(x)], \sup_{x \in \mathcal{X}}[-\overline{A}_N(x)]\}$ will be determined by $\sup_{x \in \mathcal{X}}[\overline{A}_N(x)]$, and the latter will converge weakly to $\sup_{x \in \mathcal{B}}[v(x)]$ as discussed above. Clearly, if $F_1(x) < F_2(x)$ for all $x \in \mathcal{X}$ and hence $\mathcal{B}$ is empty, then D_N will diverge to minus infinity.

We now turn to our general setting and make the above heuristic statement more rigorous. Define the empirical processes in x, θ

$$\nu_{kN}^d(x, \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[1\left(X_{ki}(\theta) \leq x\right) - F_k(x, \theta) \right]$$

$$\nu_{kN}^s(x, \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\int_{-\infty}^{x} 1\left(X_{ki}(\theta) \leq t\right) dt - \int_{-\infty}^{x} F_k(t, \theta) dt \right]$$

$$\nu_{kN}^p(x, y, \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\int_{y}^{x} 1\left(X_{ki}(\theta) \leq t\right) dt - \int_{y}^{x} F_k(t, \theta) dt \right]. \tag{9}$$

Let $\left(\begin{array}{ccc} \widetilde{d}_{kl}(\cdot) & \nu'_{k0} & \nu'_{l0} \end{array}\right)'$ be a mean zero Gaussian process with covariance functions given by

$$C^d(x_1, x_2) = \lim_{N \to \infty} E \left(\begin{array}{c} \nu_{kN}^d(x_1, \theta_{k0}) - \nu_{lN}^d(x_1, \theta_{l0}) \\ \sqrt{N} \overline{\psi}_{kN}(\theta_{k0}) \\ \sqrt{N} \overline{\psi}_{lN}(\theta_{l0}) \end{array} \right) \left(\begin{array}{c} \nu_{kN}^d(x_2, \theta_{k0}) - \nu_{lN}^d(x_2, \theta_{l0}) \\ \sqrt{N} \overline{\psi}_{kN}(\theta_{k0}) \\ \sqrt{N} \overline{\psi}_{lN}(\theta_{l0}) \end{array} \right)', \tag{10}$$

where $\overline{\psi}_{kN}(\theta_{k0}) = (1/N) \sum_{i=1}^{N} \psi_k(Y_{ki}, Z_{ki}, \theta_{k0})$ for all k. We analogously define $\left(\begin{array}{ccc} \widetilde{s}_{kl}(\cdot) & \nu'_{k0} & \nu'_{l0} \end{array}\right)'$ and $\left(\begin{array}{ccc} \widetilde{p}_{kl}(\cdot, \cdot) & \nu'_{k0} & \nu'_{l0} \end{array}\right)'$ to be mean zero Gaussian processes with covariance functions given by $C^s(x_1, x_2)$ and $C^p(x_1, y_1, x_2, y_2)$ respectively. The limiting null distributions of our test statistics are given in the following theorem.

Theorem 1. (a) *Suppose Assumptions 1-3 hold. Then, under the null H_0^d, we have*

$$D_N \Rightarrow \begin{cases} \min_{(k,l) \in \mathcal{I}^d} \sup_{x \in \mathcal{B}_{kl}^d} \left[\widetilde{d}_{kl}(x) + \Delta_{k0}(x)' \Gamma_{k0} \nu_{k0} - \Delta_{l0}(x)' \Gamma_{l0} \nu_{l0} \right] & \text{if } d^* = 0 \\ -\infty & \text{if } d^* < 0, \end{cases}$$

where $\mathcal{I}^d = \{(i,j) | i \neq j, \sup_{x \in \mathcal{X}} [F_i(x) - F_j(x)] = 0\}$ and $\mathcal{B}_{kl}^d = \{x \in \mathcal{X} : F_k(x) = F_l(x)\}$.

(b) *Suppose Assumptions 1^*, 2 and 3^* hold. Then, under the null H_0^s, we have*

$$S_N \Rightarrow \begin{cases} \min_{(k,l) \in \mathcal{I}^s} \sup_{x \in \mathcal{B}_{kl}^s} \left[\widetilde{s}_{kl}(x) + \Lambda_{k0}(x)' \Gamma_{k0} \nu_{k0} - \Lambda_{l0}(x)' \Gamma_{l0} \nu_{l0} \right] & \text{if } s^* = 0 \\ -\infty & \text{if } s^* < 0, \end{cases}$$

10

where $\mathcal{I}^s = \{(i,j)|i \neq j, \sup_{x \in \mathcal{X}} \int_{-\infty}^x [F_i(t) - F_j(t)] \, dt = 0\}$ and $\mathcal{B}_{kl}^s = \{x \in \mathcal{X} : \int_{-\infty}^x F_k(t)dt = \int_{-\infty}^x F_l(t)dt\}$.

(c) Suppose Assumptions 1*, 2 and 3** hold. Then, under the null H_0^p, we have

$$P_N \Rightarrow \begin{cases} \min_{(k,l) \in \mathcal{I}^p} \sup_{x \in \mathcal{B}_{kl}^p} [\widetilde{p}_{kl}(x,y) + \Xi_{k0}(x)'\Gamma_{k0}\nu_{k0} - \Xi_{l0}(x)'\Gamma_{l0}\nu_{l0}] & \text{if } p^* = 0 \\ -\infty & \text{if } p^* < 0, \end{cases}$$

where $\mathcal{I}^p = \{(i,j)|i \neq j, \sup_{x,-y \in \mathcal{X}_+} \int_y^x [F_i(t) - F_j(t)] \, dt = 0\}$ and $\mathcal{B}_{kl}^p = \{(x,y) : x \in \mathcal{X}_+, -y \in \mathcal{X}_+ \text{ and } \int_y^x F_k(t)dt = \int_y^x F_l(t)dt\}$.

The asymptotic null distributions of D_N, S_N and P_N depend on the "true" parameters $\{\theta_{k0} : k = 1, \ldots, K\}$ and distribution functions $\{F_k(\cdot) : k = 1, \ldots, K\}$. This implies that the asymptotic critical values for D_N, S_N, P_N can not be tabulated once and for all. However, we define below various procedures to estimate them from the data.

4 Critical Values

We next describe our main method for obtaining critical values, the subsampling approach. We derive its asymptotic properties and propose various practical methods for selecting subsample size. We then discuss an alternative approach based on the bootstrap and a recentered test statistic.

4.1 Subsampling

In this section, we consider the use of subsampling to approximate the *asymptotic* null distributions of our test statistics. As was pointed out by Klecan et. al. (1991), even when the data are i.i.d. the standard bootstrap does not work because one needs to impose the null hypothesis in that case, which is difficult because it is defined by a complicated system of inequalities, see below for more discussion. The mutual dependence of the prospects and the time series dependence in the data also complicate the issue considerably. The subsampling method is very simple to define and yet provides consistent critical values in a very general setting. In contrast to the simulation approach of Klecan et. al. (1991), our procedure does not require the assumption of generalized exchangeability of the underlying random variables. Indeed, we require no additional assumptions beyond those that have already been made.

We now discuss the asymptotic validity of the subsampling procedure for the test D_N (The argument for the tests S_N and P_N is similar and hence is omitted). Let $W_i = \{(Y_{ki}, Z_{ki}) : k = 1, \ldots, K\}$ for $i = 1, \ldots, N$. With some abuse of notation, the test statistic D_N can be re-written as a function of the data $\{W_i : i = 1, \ldots, N\}$:

$$D_N = \sqrt{N} d_N(W_1, \ldots, W_N),$$

where

$$d_N(W_1, \ldots, W_N) = \min_{k \neq l} \sup_{x \in \mathcal{X}} \left[\overline{F}_{kN}(x, \widehat{\theta}_k) - \overline{F}_{lN}(x, \widehat{\theta}_l) \right]. \tag{11}$$

Let

$$G_N(w) = P\left(\sqrt{N} d_N(W_1, \ldots, W_N) \leq w \right) \tag{12}$$

denote the distribution function of D_N. Let $d_{N,b,i}$ be equal to the statistic d_b evaluated at the sub-sample $\{W_i, \ldots, W_{i+b-1}\}$ of size b, i.e.,

$$d_{N,b,i} = d_b(W_i, W_{i+1}, \ldots, W_{i+b-1}) \text{ for } i = 1, \ldots, N - b + 1.$$

This means that we have to recompute $\widehat{\theta}_l(W_i, W_{i+1}, \ldots, W_{i+b-1})$ using just the subsample as well. We note that each subsample of size b (taken without replacement from the original data) is indeed a sample of size b from the true sampling distribution of the original data. Hence, it is clear that one can approximate the sampling distribution of D_N using the distribution of the values of $d_{N,b,i}$ computed over $N - b + 1$ different subsamples of size b. That is, we approximate the sampling distribution G_N of D_N by

$$\widehat{G}_{N,b}(w) = \frac{1}{N - b + 1} \sum_{i=1}^{N-b+1} 1\left(\sqrt{b} d_{N,b,i} \leq w \right).$$

Let $g_{N,b}(1 - \alpha)$ denote the $(1 - \alpha)$-th sample quantile of $\widehat{G}_{N,b}(\cdot)$, i.e.,

$$g_{N,b}(1 - \alpha) = \inf\{w : \widehat{G}_{N,b}(w) \geq 1 - \alpha\}.$$

We call it the *subsample critical value* of significance level α. Thus, we reject the null hypothesis at the significance level α if $D_N > g_{N,b}(1 - \alpha)$. The computation of this critical value is not particularly onerous, although it depends on how big b is. The subsampling method has been proposed in Politis and Romano (1994) and is thoroughly reviewed in Politis, Romano, and Wolf (1999). It works in many cases where the standard bootstrap fails: in heavy tailed distributions, in unit root cases, in cases where the parameter is on the boundary of its space, etc.

We now show that our subsampling procedure works under a very weak condition on b. In many practical situations, the choice of b will be data-dependent, see the next section for some methodology for choosing b. To accommodate such possibilities, we assume that $b = \widehat{b}_N$ is a data-dependent sequence satisfying

Assumption 4: $P[l_N \leq \widehat{b}_N \leq u_N] \to 1$ where l_N and u_N are integers satisfying $1 \leq l_N \leq u_N \leq N$, $l_N \to \infty$ and $u_N/N \to 0$ as $N \to \infty$.

The following theorem shows that our test based on the subsample critical value has asymptotically correct size.

Theorem 2. *Suppose Assumptions 1-4 hold. Then, under the null hypothesis H_0^d, we have when $d^* = 0$ that*

$$(a)\ g_{N,\hat{b}_N}(1-\alpha) \xrightarrow{p} g(1-\alpha)$$

$$(b)\ P[D_N > g_{N,\hat{b}_N}(1-\alpha)] \to \alpha$$

as $N \to \infty$, where $g(1-\alpha)$ denotes the $(1-\alpha)$-th quantile of the asymptotic null distribution of D_N which is given in Theorem 1(a).

REMARKS.

1. Results analogous to Theorems 2 hold for the test S_N (P_N) under Assumptions 1^*, 2, $3^*(3^{**})$ and 4. The proof is similar to that of Theorem 2.

4.1.1 Asymptotic Power Properties

In this section, we investigate power properties of our tests. We first establish that the test D_N is consistent against the fixed alternative hypothesis H_1^d. Analogous results can be established for the tests S_N and P_N using similar arguments as below. For brevity, we do not provide the details.

Theorem 3. *Suppose Assumptions 1-4 hold. Then, under the alternative hypothesis H_1^d, we have*

$$P\left[D_N > g_{N,\hat{b}_N}(1-\alpha)\right] \to 1 \text{ as } N \to \infty.$$

Next, we determine the power of the D_N test against a sequence of contiguous alternatives converging to the null at the rate $N^{-1/2}$. Consider the following sequence of local alternative distribution functions:

$$F_{kN}(x) = F_k(x) + \frac{\delta_k(x)}{\sqrt{N}} \text{ for } k = 1, \ldots, N; N = 1, 2, \ldots, \tag{13}$$

where $\delta_k(\cdot)$ are real functions and the distribution functions $F_k(\cdot)$ satisfy

$$d^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} [F_k(x) - F_l(x)] \leq 0.$$

To analyze the asymptotic behavior of the test under local alternatives, we need to modify Assumptions 1-3. That is, we assume:

Assumption 1-lc: (i) $\{(Y_{Nki}, Z_{Nki}) =: (Y_{ki}, Z_{ki}) : i \geq 1,\ N \geq 1\}$ is an α- mixing array with $\alpha(m) = O(m^{-A})$ for some $A > \max\{(q-1)(q+1), 1+2/\delta\}$ for $k = 1, \ldots, K$, where q is an even integer that satisfies $q > 3(L_{\max} + 1)/2$, $L_{\max} = \max\{L_1, \ldots, L_K\}$ and δ is a positive constant that also appears in Assumption 2-lc (ii) below. (ii) $\sup_{N \geq 1} E\|Z_{ki}\|^2 < \infty$ for all $k = 1, \ldots, K$, for all $i \geq 1$. (iii) The conditional distribution $H_k(\cdot|Z_{ki})$ of X_{ki} given Z_{ki} has a density with respect to Lebesgue measure a.s. for all $k = 1, \ldots, K$, for all $i \geq 1$ which is bounded uniformly over $N \geq 1$.

Assumption 2-lc : (i) The parameter estimator satisfies $\sqrt{N}(\widehat{\theta}_k - \theta_{k0}) = (1/\sqrt{N}) \sum_{i=1}^{N} \Gamma_{k0} \psi_k(Y_{ki}, Z_{ki}, \theta_{k0}) + o_p(1)$, where Γ_{k0} is a non-stochastic matrix for $k = 1, \ldots, K$; (ii) The function $\psi_k(y, z, \theta) : \mathbb{R} \times \mathbb{R}^{L_k} \times \Theta_k \to \mathbb{R}^{L_k}$ is measurable and satisfies (a) $\sqrt{N} E \psi_k(Y_{ki}, Z_{ki}, \theta_{k0}) \to m_{k0}$ and (b) $\sup_{N \geq 1} E \|\psi_k(Y_{ki}, Z_{ki}, \theta_{k0})\|^{2+\delta} < \infty$ for some $\delta > 0$ and for $k = 1, \ldots, K$, for all $i \geq 1$.

Assumption 3-lc: (i) The function $F_{kN}(x, \theta)$ is differentiable in θ on a neighborhood Θ_{k0} of θ_{k0} for $k = 1, \ldots, K$; (ii) For all sequence of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$, $\sup_{x \in \mathcal{X}} \sup_{\theta : \|\theta - \theta_{k0}\| \leq \xi_N} \|\partial F_{kN}(x, \theta)/\partial \theta - \Delta_{k0}(x)\| \to 0$ for $k = 1, \ldots, K$, where $\Delta_{k0}(x) = \partial F_{kN}(x, \theta_{k0})/\partial \theta$; (iii) $\sup_{N \geq 1} \sup_{x \in \mathcal{X}} \|\Delta_{k0}(x)\| < \infty$ for $k = 1, \ldots, K$.

Note that Assumption 2-lc implies that the asymptotic distribution of $\sqrt{N}(\widehat{\theta}_k - \theta_{k0})$ has mean m_{k0} which might be non-zero under local alternatives Then, the asymptotic distribution of D_N under the local alternatives is given in the following theorem:

Theorem 4. *Suppose Assumptions 1-lc, 2-lc and 3-lc hold. Then, under the local alternatives, we have*

$$D_N \Rightarrow \begin{cases} L_D & \text{if } d^* = 0 \\ -\infty & \text{if } d^* < 0, \end{cases}$$

where

$$
\begin{aligned}
L_D &= \min_{(k,l) \in \mathcal{I}^d} \sup_{x \in \mathcal{B}_{kl}^d} \left[\widetilde{d}_{kl}(x) + \Delta_{k0}(x)'\Gamma_{k0}\nu_{k0} - \Delta_{l0}(x)'\Gamma_{l0}\nu_{l0} + \mu_{kl}(x) \right], \\
\mu_{kl}(x) &= \Delta_{k0}(x)'\Gamma_{k0}m_{k0} - \Delta_{l0}(x)'\Gamma_{l0}m_{l0} + \delta_k(x) - \delta_l(x),
\end{aligned}
$$

$\mathcal{I}^d$ *and* $\mathcal{B}_{kl}^d$ *are defined as in Theorem 1 and* $(\ \widetilde{d}_{kl}(\cdot)\ \ \nu_{k0}'\ \ \nu_{l0}'\ )'$ *is the Gaussian process defined in Section 3.2.*

This result implies that asymptotic local power of the D_N test based on the subsample critical value is given by the following Corollary:

Corollary 5. *Suppose Assumptions 1-lc, 2-lc,3-lc and 4 hold. Then, under the local alternatives, we have when* $d^* = 0$ *that*

$$P\left[D_N > g_{N,\widehat{b}_N}(1 - \alpha)\right] \to P\left[L_D > g(1 - \alpha)\right]$$

as $N \to \infty$, *where* $g_{N,\widehat{b}_N}(1 - \alpha)$ *and* $g(1 - \alpha)$ *are defined in Section 4.*

REMARKS.

1. Theorem 4 implies that our D_N test is asymptotically locally unbiased, i.e.

$$\lim_{N \to \infty} P[D_N > g_{N,\widehat{b}_N}(1 - \alpha)] \geq \alpha \tag{14}$$

when $d^* = 0$ under the local alternatives. This follows because, by Anderson's lemma (e.g., see Bickel

et al. (1993, p.466), we can show that

$$
P\left[\min_{k \neq l} \sup_{x \in \mathcal{B}_{kl}^d} \left[\widetilde{d}_{kl}(x) + \Delta_{k0}(x)'\Gamma_{k0}\nu_{k0} - \Delta_{l0}(x)'\Gamma_{l0}\nu_{l0} + \mu_{kl}(x)\right] > g(1-\alpha)\right] \tag{15}
$$

$$
\geq \quad P\left[\min_{k \neq l} \sup_{x \in \mathcal{B}_{kl}^d} \left[\widetilde{d}_{kl}(x) + \Delta_{k0}(x)'\Gamma_{k0}\nu_{k0} - \Delta_{l0}(x)'\Gamma_{l0}\nu_{l0}\right] > g(1-\alpha)\right] = \alpha
$$

and the left-hand-sides of (14) and (15) are equal by Theorem 4.

2. Corollary 5 shows that the asymptotic local power of the D_N test against the local alternatives (13) is $P\left[L_D > g(1-\alpha)\right].$

4.1.2 Choice of Subsample Size

In practice, the choice of b is important and rather difficult. It is rather akin to choosing bandwidth in tests of parametric against nonparametric hypotheses, an issue that has been notoriously difficult to resolve. Politis, Romano, and Wolf (1999) discuss various methods for selecting subsample size. Delgado, Rodriguez-Poo, and Wolf (2001) propose a method for selecting b to minimize size distortion in the context of hypothesis testing within the maximum score estimator, although no optimality properties of this method were proven. The main problem with this general approach is that usually the b that is good for size distortion is not good for power and vice a versa. In any case, their approach can only be implemented in situations where there is enough structure that a standard bootstrap could be applied.

We propose a number of criteria for choosing b and investigate below how well they do in practice. We suppose that there is a set $B_N = \{b_{N1} < b_{N2} < \cdots < b_{Nr_N}\}$ of candidate subsample sizes, where $b_{N1} \to^p \infty$ and $b_{Nr_N}/N \to^p 0$, while r_N is allowed to increase with N and that our methods will select a sequence of subsample values from B_N [hence the conditions of our theorems 2-4 are satisfied by such a sequence]. For each significance level α we obtain the sample of estimated critical values $\{g_{N,b_{Nj}}(1-\alpha), j = 1, \ldots, r_N\}$.

Politis, Romano, and Wolf (1999) suggest the 'minimum volatility' method. This involves computing the local (in b) standard deviation of $g_{N,b}$ and then taking the subsample $\widehat{b}_{MV}$ that minimizes this volatility measure. The idea is that when b is in the right range the critical values should be relatively stable.

A second approach is to use the mean or median critical value:

$$
\overline{g}_N(1-\alpha) = \frac{1}{r_N} \sum_{j=1}^{r_N} g_{N,b_{Nj}}(1-\alpha) \tag{16}
$$

$$
g_N^{Med}(1-\alpha) = \text{med}\{g_{N,b_{Nj}}(1-\alpha) : j = 1, \ldots, r_N\} \tag{17}
$$

15

and reject when $D_N > \overline{g}_N(1 - \alpha)$ in the first case and reject when $D_N > g_N^{Med}(1 - \alpha)$ in the second case.[3] The idea in the median case is that each critical value reflects a standard of evidence from a different 'court of opinion'. Taking the median critical value is like taking the majority outcome of a vote by all critical values on accept or reject.

In applications, we favor computing a plot of p-values against subsamples for a range of subsamples. If the p-value is insensitive to subsample sizes within a 'reasonable' range, then inferences are likely to be robust, and whatever automatic method is chosen will yield similar results. We illustrate this method below.

4.2   Alternative Approaches

We next discuss alternative approaches to obtaining critical values such as bootstrap or multiplier simulation procedures. In contrast to subsampling, one now has to impose the null hypothesis either explicitly or implicitly in the resampling schemes. In practice, one has to impose the least favorable case

$$LF = \{F : F_1(x) = \cdots = F_K(x) \text{ for all } x \in \mathcal{X}\}, \tag{18}$$

which is a strict subset of the (boundary of the) null hypothesis. When the prospects are mutually independent and independent over time and there are no estimated parameters it is easy to resample imposing that the data come from LF - you just pool the data into a common distribution and draw from that empirical distribution with standard replacement bootstrap. In an innovative paper, Klecan et al. showed that with suitable modification this idea can be applied to the case where the prospects are mutually dependent as long as the dependence is of a specific variety called generalized exchangeable.[4] More generally though it has not been possible to simultaneously impose (18) and correctly account for the mutual dependence in the outcomes. There is an alternative: instead of imposing (18) in the resample one should resample in a way appropriate to capture the mutual and temporal dependence and then 'recentre' the test statistic, as has recently been suggested in Chernozhukov (2002). It should be noted that this approach also imposes (18) implicitly. We next describe one such method for the case where the data are mutually dependent but are independent over time.

Let $\widehat{\varepsilon}_{ki} = Y_{ki} - \widehat{\mu}_{k0} - Z_{ki}'\widehat{\theta}_k$ denote the residual computed using the original sample $\{W_i : i = 1, \ldots, N\}$, where $W_i = \{(Y_{ki}, Z_{ki}) : k = 1, \ldots, K\}$. Let $\{\varepsilon_i^* = (\varepsilon_{1i}^*, \ldots, \varepsilon_{Ki}^*) : i = 1, \ldots, N\}$ be the

---

[3] This corresponds to some implicit subsample size. Instead of doing a formal test we can equivalently report the mean or median p-value across the sample of tests with different $b \in B_n$.

[4] This structure is necessary to their method. It is also clear that they require time series independence in the proofs of their Theorem 7.

bootstrap residual drawn with replacement from the centered residual $\{\widehat{\varepsilon}_i^c = (\widehat{\varepsilon}_{1i} - \overline{\widehat{\varepsilon}}_1, \ldots, \widehat{\varepsilon}_{Ki} - \overline{\widehat{\varepsilon}}_K) : i = 1, \ldots, N\}$, where $\overline{\widehat{\varepsilon}}_k = \sum_{i=1}^{N} \widehat{\varepsilon}_{ki}/N$.[5] Then compute $Y_{ki}^* = \widehat{\mu}_{k0} + Z_{ki}'\widehat{\theta}_k + \varepsilon_{ki}^*$. Using the bootstrap sample $\{W_i^* : i = 1, \ldots, N\}$ where $W_i^* = \{(Y_{ki}^*, Z_{ki}) : k = 1, \ldots, K\}$, compute $\widehat{\theta}_k^*$. These steps will take care of the effect of the parameter estimation error in the bootstrap distribution described below.[6] Define $X_{ki}^*(\theta) = Y_{ki}^* - Z_{ki}'\theta$ and the empirical distributions

$$\overline{F}_{kN}^*(x, \theta) = \frac{1}{N} \sum_{i=1}^{N} 1\left(X_{ki}^*(\theta) \leq x\right)$$

for $k = 1, \ldots, K$. Define the recentred empirical $\overline{F}_{kN}^{*c}(x) = \overline{F}_{kN}^*(x, \widehat{\theta}_k^*) - \overline{F}_{kN}(x, \widehat{\theta}_k)$ and the centred bootstrap test statistic

$$D_N^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} \sqrt{N} \left[\overline{F}_{kN}^{*c}(x) - \overline{F}_{lN}^{*c}(x)\right].$$

We then compute the distribution of D_N^* conditional on the original sample and take the critical value from this distribution. That is, we approximate the sampling distribution H_N of D_N by

$$\widehat{H}_N(w) = \frac{1}{M} \sum_{i=1}^{M} 1\left(D_{N,i}^* \leq w\right)$$

where M is the number of bootstrap samples. Let $h_N(1 - \alpha)$ denote the $(1 - \alpha)$-th sample quantile of $\widehat{H}_{N,b}(\cdot)$, i.e.,

$$h_N(1 - \alpha) = \inf\{w : \widehat{H}_N(w) \geq 1 - \alpha\}.$$

We call it the *bootstrap critical value* of significance level α. Thus, we reject the null hypothesis at the significance level α if $D_N > h_N(1 - \alpha)$. It can be shown that this test is consistent; we investigate the finite sample behaviour below. The recentering in D_N^* is crucial and is used to impose the restriction (18). The idea of recentering has also been suggested in other contexts by Hall and Horowitz (1999) and Whang (2001) and in this context in a recent paper by Chernozhukov (2002).[7]

In the time series case, the resampling should be modified to account for the dependence, see Horowitz (2000) or Härdle, Horowitz and Kreiss (2001) for a survey of bootstrap methods for time series. We briefly describe the non-overlapping (viz., Carlstein (1986)) and overlapping (viz., Künsch (1989)) block bootstrap procedures that can be used in our context. The observations to be

---

[5] The centering is redundant of course when the model includes a constant term and the parameters are estimated by OLS.

[6] When there are no estimated parameters, i.e., when $X_{ki}(\theta) = X_{ki}$, the bootstrap sample $\{(X_{1i}^*, \ldots, X_{Ki}^*) : i = 1, \ldots, N\}$ are defined to be a random draw (with replacement) from the empirical (joint) distribution of $\{(X_{1i}, \ldots, X_{Ki}) : i = 1, \ldots, N\}$.

[7] Chernozhukov (2002) actually combines recentering with subsampling in his application.

bootstrapped are the centered residuals $\{\widehat{\varepsilon}_i^c : i = 1, \ldots, N\}$. Let L denote the length of the blocks satisfying $L \propto N^\gamma$ for some $0 \leq \gamma \leq 1$.[8] With non-overlapping blocks, block 1 is observations $\{\widehat{\varepsilon}_j^c : j = 1, \ldots, L\}$, block 2 is observations $\{\widehat{\varepsilon}_{L+j}^c : j = 1, \ldots, L\}$, and so forth. There are B different blocks, where $BL = N$. With overlapping blocks, block 1 is observations $\{\widehat{\varepsilon}_j^c : j = 1, \ldots, L\}$, block 2 is observations $\{\widehat{\varepsilon}_{1+j}^c : j = 1, \ldots, L\}$, and so forth. There are $N - L + 1$ different blocks. The bootstrap residuals $\{\varepsilon_i^* : i = 1, \ldots, N\}$ are obtained by sampling B blocks randomly with replacement from either the B non-overlapping blocks or the $N - L + 1$ overlapping blocks and laying them end-to-end in the order sampled. The remaining steps are the same as the independent case described above.[9]

We now compare the alternative approaches with our subsampling approach. Under the least favorable case (18), it is not difficult to show that the asymptotic distribution of D_N, say, is given by $L_B =: \min_{k \neq l} \sup_{x \in \mathcal{X}} [\widetilde{d}_{kl}(x) + \Delta_{k0}(x)' \Gamma_{k0} \nu_{k0} - \Delta_{l0}(x)' \Gamma_{l0} \nu_{l0}]$ using an argument similar to the proof of Theorem 1. Any valid bootstrap (or simulation)-based test hence would try to approximate L_B. However, our test statistic D_N has a non-degenerate limit distribution on the boundary "$d^* = 0$" of our null hypothesis H_0^d. Note that "$d^* = 0$" is in fact a composite hypothesis and includes the least favorable case (18) as a special case. Therefore, when (18) fails to hold but $d^* = 0$ is true[10], then the test based on the bootstrap (or simulation) critical value would not have asymptotic size α. This implies that the latter test is *not asymptotically similar on the boundary,* which in turn implies that the test is biased, i.e., there exist alternatives under which acceptance of the hypothesis is more likely than in some cases in which the hypothesis is true, see Lehmann (1986, Chapter 4) for the concept of similarity and unbiasedness. On the other hand, our test based on the subsample critical value is unbiased and *asymptotically similar on the boundary* since the subsampling distribution mimics the true sampling distribution everywhere on the boundary. Note that, in general, an asymptotically similar test is more powerful than an asymptotically non-similar test for some local alternatives near the boundary, see, e.g., Hansen (2001).

---

[8]It is also possible to sample L randomly from the geometric distribution and use the overlapping blocks. This procedure amounts to the stationary bootstrap of Politis and Romans (1993) and it guarantees that the resulting bootstrap data series is stationary.

[9]When there are no estimated parameters, the observations to be bootstrapped are $\{(X_{1i}, \ldots, X_{Ki}) : i = 1, \ldots, N\}$ and we can apply the block bootstrap methods directly to them to get the bootstrap sample $\{(X_{1i}^*, \ldots, X_{Ki}^*) : i = 1, \ldots, N\}$.

[10]For example, if $K = 3$, this happens if $F_1(x) = F_2(x)$ for all $x \in \mathcal{X}$ but $F_3(x)$ crosses with F_1 (and F_2). More generally, this happens if $F_k(x) \leq F_l(x)$ with equality holding for $x \in B_{kl}(\subset \mathcal{X})$ for some pair (k, l) but there are crossings of the distributions (i.e., no FSD relationship) for the other pairs.

5   Numerical Results

5.1   Simulations

We examined three sets of designs: the Burr distributions most recently examined by Tse and Zhang (2000), the lognormal distributions most recently studied by Barrett and Donald (1999), and the exchangeable normal processes of Klecan et al. (1991). The first two sets have mutually independent and temporally independent prospects, while the third designs are both mutually and temporally dependent. By choosing already published designs we are able to compare our procedures with those of the authors' in regard to size and power. We do not recompute their tests, but refer the reader to their tables to make comparison. We have also carried out simulations in the case where there are upto 10 prospects; full details of this are available from the authors.

We first give some general details common to the simulations. In computing the suprema in D_N, S_N, we took a maximum over an equally spaced grid of size n on the range of the pooled empirical distribution. We experimented with a variety of such grids [this is quite important in practice], but found our approach worked adequately. We chose a total of twenty different subsamples for each sample size $n \in \{50, 500, 1000\}$. In earlier work we tried fixed rules of the form $b(n) = c_j n^{a_j}$, but found it did not work as well. Instead, we took an equally spaced grid of subsample sizes: for $n = 50$, the subsample sizes are $\{20, 21, \ldots, 40\}$; for $n = 500$ the subsample sizes are $\{50, 65, \ldots, 350\}$; for $n = 1000$ the subsample sizes are $\{100, 120, \ldots, 500\}$. This grid of subsamples are then used to implement the automatic methods of sections 7.[11] We report the results of the automatic methods here and comment also on the results for fixed subsamples [which are available from the authors]. In computing the suprema in each $d_{N,b,i}$ we took the same grid of points as was used in the original test statistic. In addition to the subsampling method we also computed the 'recentered bootstrap' method; we used a total of 200 bootstrap repetitions in each case. In each experiment we did $1,000$ replications.

The overall impression is that the (automatic) subsample methods and the recentered full sample bootstrap method work reasonably well in samples above 500. The full sample method works slightly better under the null hypothesis, while the subsample method works better under the alternative. In the cases where the full sample method works better, this advantage effectively disappears in the larger sample sizes, but in cases [1c,1d,1e, and 2d below] where the subsample method is superior, that superiority can be quite substantial even in the larger sample. This is consistent with our theory.[12] Also, we note that in the smallest sample size, the recentred bootstrap does better for all

---

[11] The subsample grid is quite crude in its coverage, and perhaps better results would be obtained by considering more subsamples.

[12] We have also obtained results for the uncentered bootstrap and the recentred subsample method but do not report

designs. It is worth reminding the reader that these designs, especially the first two settings, favor the alternative methods which are designed for iid observations on independent or exchangeable prospects.

5.1.1 Tse and Zhang (2000): Burr Type Distributions

In the context of independent prospects and i.i.d. observations, Tse and Zhang (2000) have provided some Monte Carlo evidence on the power of the alternative tests proposed by Davidson and Duclos (2000), the "DD test", and Anderson (1996). They also shed light on the convergence to the Gaussian limiting distribution of these tests. The evidence on the latter issue is not very encouraging except for very large sample sizes, and they conclude that the DD test has better power than the Anderson test for the cases they considered.

Tse and Zhang (2000) investigated the Burr Type XII distribution, $B(\alpha, \beta)$, which is often an empirically plausible candidate in the income distribution field. This is a two parameter family defined by:

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

where $E(X) < \infty$ if $\beta > 1/\alpha > 0$. This distribution has a convenient inverse: $F^{-1}(v) = [(1 - v)^{-\frac{1}{\beta}} - 1]^{\frac{1}{\alpha}}$, $0 \leq v < 1$. We investigated the five different Burr designs of Tse and Zhang (2000), which are given below along with the population values of d^*, s^* :

Design	X_1	X_2	d^*	s^*
1a	$B(4.7, 0.55)$	$B(4.7, 0.55)$	$0.000(FSD)$	$0.0000(SSD)$
1b	$B(2.0, 0.65)$	$B(2.0, 0.65)$	$0.0000(FSD)$	$0.0000(SSD)$
1c	$B(4.7, 0.55)$	$B(2.0, 0.65)$	0.1395	0.0784
1d	$B(4.6, 0.55)$	$B(2.0, 0.65)$	0.1368	0.0773
1e	$B(4.5, 0.55)$	$B(2.0, 0.65)$	0.1340	0.0761

The first two designs are in the null hypothesis, while the remaining three are in our alternative. Note that Tse and Zhang (2000) actually report results for the complementary hypotheses, so that only their first two tables are directly comparable with ours. We report our results in Tables 1F and 1S, for cases 1a-e below.

The first two designs are useful for an evaluation of the size characteristics of our tests, but only in the the "least favorable" case of equality of the two distributions. The estimated CDFs "kiss" at

them here. The uncentered bootstrap is inconsistent and the simulations strongly support this - the rejection frequencies are zero or close to zero for all designs. The recentred subsample method although consistent was considerably inferior to both the approaches we do report.

many more points than do the integrated CDFs. As a result, large sample sizes will be needed for accurate size of FSD, as well as relatively large subsamples. For SSD, however, the accuracy is quite good for moderate sample sizes. Given the nature of the testing problem, sample sizes less than 100 are very small indeed. In such cases the tests will over-reject at conventional levels. Even in this demanding case, however, one is led to the correct decision that the two (equal) prospects here do not dominate each other. The accuracy of size estimation for SSD is rather impressive. Regarding the automatic subsample methods, the Mean and Median methods seem to work similarly and better than the MinVol method, especially for $n = 50$. MinVol overestimates size with very small sample sizes.

In the last three designs (Tables 1F and 1S, cases 1c-1e), the power of our tests are forcefully demonstrated. This is so even at relatively small samples sizes. Even with a sample of size 50 there is appreciable power, especially for the recentred bootstrap method. There is not much to choose between the performance of the three automatic methods. Regarding the fixed subsample size methods (available from the authors): the power declines as the number of subsamples declines (the subsample size increases). This seems to indicate that larger *number of subsamples* are needed for more accurate estimation especially when moderate size samples are available. The performance of the fixed subsample tests in these cases is quite satisfactory.

5.1.2 Barrett and Donald (1999): Lognormal Distributions

The lognormal distribution is a long celebrated case in both finance and income and wealth distribution fields. It was most recently investigated in Barrett and Donald (1999) in a Monte Carlo study of the Klecan et al. tests along with some of its competitors. Let,

$$X_j = \exp(\mu_j + \sigma_j Z_j),$$

where Z_j are standard normal and mutually independent.

Design	X_1	X_2	d^*	s^*
2a	$LN(0.85, 0.6^2)$	$LN(0.85, 0.6^2)$	$0.0000(FSD)$	$0.0000(SSD)$
2b	$LN(0.85, 0.6^2)$	$LN(0.7, 0.5^2)$	$0.0000(FSD)$	$0.0000(SSD)$
2c	$LN(0.85, 0.6^2)$	$LN(1.2, 0.2^2)$	0.0834	$0.0000(SSD)$
2d	$LN(0.85, 0.6^2)$	$LN(0.2, 0.1^2)$	0.0609	0.0122

These designs are clearly favorable to the independent samples assumption in Barret and Donald (1999). The results shown in Tables 2F and 2S, cases a-d correspond exactly to cases 1,2,3, and 4 of Barrett and Donald (1999). We note that the comparison of the automatic selection methods is similar to the previous example.

21

The first two designs are in the null and the next two (2c-2d) are in the alternative for FSD, borderline null for SSD in design 2c, and in the alternative for SSD in design 2d. The first design is a "least favorable" case and, at least for the FSD test, it demonstrates the demand for higher sample sizes as well as subsample sizes. The tendency is toward moderate over-rejection for very small samples. Accuracy improves quite rapidly with sample size for SSD tests and is impressive for most subsample sizes and moderate sample sizes. Bootstrap method does quite well in this 'friendly' least favorable case.

The second design is quite instructive. While the overall results are similar to the previous case, the differences reflect the fact that there is no FSD ranking, (or equality) and only a mild degree of Second Order Dominance. For moderate to reasonable sample sizes the tendency is to slightly under-reject FSD. This tendency is reduced by increasing the size of the subsamples. The results for SSD, confirm the theoretical consistency properties of our tests. The theoretical power properties of the subsampling test are evidenced.

Results for design 2c are quite conclusive. For moderate to large sample sizes, FSD is powerfully rejected, while SSD is not. Very small samples are seen to be dangerous in cases where CDFs cross (no FSD) and the degree of SSD is moderate. A comparison with the last design (case 2d) is quite instructive. Here there is no FSD or SSD and the test is quite capable of producing the correct inference.

In terms of a comparison with the tests investigated in Barrett and Donald (1999), we seem to do better in some cases and worse in others. Generally speaking their performance is better under the null hypothesis and ours is better under the alternatives. There is evidence that the subsampling tests are more powerful for SSD hypotheses than the bootstrap

5.1.3 Klecan, McFadden, and McFadden (1991): Multivariate Normal Processes

The previous designs had independent prospects and i.i.d observations. In this section we investigate the three different exchangeable multinormal processes of Klecan et al. (1991),

$$X_{jt} = (1 - \lambda) \left[\alpha_j + \beta_j \left(\sqrt{\rho} Z_{0t} + \sqrt{1 - \rho} Z_{jt} \right) \right] + \lambda X_{j,t-1}, \tag{19}$$

where (Z_{0t}, Z_{1t}, Z_{2t}) are i.i.d. standard normal random variables, mutually independent. The parameters $\lambda = \rho = 0.1$ determine the mutual correlation of X_{1t} and X_{2t} and their autocorrelation. The parameters α_j, β_j are actually the mean and standard deviation of the marginal distributions of X_{1t} and X_{2t}. This scheme produces autocorrelated and mutually dependent prospects consistent with the assumptions of Klecan et al., *but only as far as the cross-sectional dependence.* Again, these designs slightly favor their test assumptions. The marginals and the true values of the statistics are:

Design	X_1	X_2	d^*	s^*
3a	$N(0,1)$	$N(-1,16)$	0.1981	$0.0000(SSD)$
3b	$N(0,16)$	$N(1,16)$	$0.0000(FSD)$	$0.0000(SSD)$
3c	$N(0,1)$	$N(1,16)$	0.1981	0.5967

The results are given in Tables 3F and S, cases a-c. Design 3a is in the alternative for FSD, and in the null for SSD. Again we note that we need large samples and subsample sizes to infer this low degree of SSD, but have very good power in rejecting FSD (especially for large number of subsamples even in very small samples of 50). Design 3b is rather strongly in the null. These designs correspond exactly to experiments 1,2, and 3 in Table 2 of Klecan et al (1991).

Small sample sizes lead to over estimation of size but, again, the larger number of subsamples do better in these situations. Interestingly, the number and size of subsamples do not appear conse-quential for moderate to large samples. Otherwise the theoretical power and consistency properties are strongly confirmed. The final design 3c is clearly in the alternative for both FSD and SSD. Our procedures show their expected power in rejecting dominance. For very small samples (50), again larger number of subsamples do uniformly much better than otherwise (the subsample size seems not as important), but minvol method is inferior for size calculations. The subsampling tests are generally more powerful than the bootstrap for SSD than FSD cases.

5.1.4 Style Analysis

As a brief example of the residual-based testing, here we investigate a test of stochastic dominance of different residuals from a style regression based on the Klecan et al. designs of the previous section. Return-based style analysis [originally proposed in Sharpe (1992)] is a popular practitioner tool to study fund managers' performance. The style regression for the returns R_t of a given fund is

$$R_t = \alpha + \sum_{k=1}^{K} \beta_k F_{kt} + \varepsilon_t, \tag{20}$$

where F_{kt} is the (observed) return of the some asset class, for $k = 1, \ldots, K$, the $\beta'_k s$ are the factor loadings, while ε_t is an idiosyncratic disturbance term that contains the part of the fund's performance not explained by the factors. The disturbance term $u_t = \alpha + \varepsilon_t$ represents the own choice of the fund manager and is called the selectivity of the fund. It is of interest to compare the u_t from different funds and to rank them according to some criterion. For example it is common practice to interpret the α of each fund as a measure of its success in selection. Given the considerable evidence on non-normality of stock returns, relying purely on a location measure to evaluate performance may not be

appropriate, see Ho (2003) for a discussion. One could also compare the marginal distributions in a test of the stochastic dominance of one fund over another.

We let $F_t = Z_{0t}/(1-\lambda L)$, where L is the lag operator, be a single observed factor and let $R_{jt} = X_{jt}$ be the return on asset j, where X_{jt} are those generated in designs 3a-c. We have

$$R_{jt} = \alpha_j + \gamma_j F_t + \varepsilon_{jt}, \text{ where } \gamma_j = \beta_j \sqrt{\rho} \text{ and } \varepsilon_{jt} = \frac{\beta_j(1-\lambda)\sqrt{1-\rho}}{1-\lambda L} Z_{jt}.$$

The simulations compute a test of whether $u_{1t} = \alpha_1 + \varepsilon_{1t}$ dominates $u_{2t} = \alpha_2 + \varepsilon_{2t}$ based on the dataset $\{R_{1t}, R_{2t}, F_t, t = 1, \ldots, T\}$. This involves estimating the parameters (α_j, γ_j) by least squares and obtaining the residuals and applying our subsampling method. The marginals of u_{jt} and the true values of the statistics are given below

Design	u_1	u_2	d^*	s^*
3d	$N(0, 0.7364)$	$N(-1, 11.7818)$	0.1936	$0.0000(SSD)$
3e	$N(0, 11.7818)$	$N(1, 11.7818)$	$0.0000(FSD)$	$0.0000(SSD)$
3f	$N(0, 0.7364))$	$N(1, 11.7818)$	0.1930	0.6024

The results are given in Tables 3RS and 3RF. There is a slight deterioration in performance due to estimating the parameters, but otherwise all methods work well as before.

5.2 Application: Daily Stock Index Returns

Finally, we applied our tests to a dataset of daily returns on the Dow Jones Industrials and the S&P500 stock returns from 8/24/88 to 8/22/00, a total of 3131 observations. The means are 0.00055 and 0.00068 respectively, while the standard deviations are 0.00908 and 0.0223 respectively; the series are certainly mutually dependent and dependent over time. Figure 1 plots the c.d.f.'s and integrated c.d.f. [denoted s.d.f.] of the two series. This shows that the two c.d.f.'s cross, but the integrated c.d.f. of the Dow Jones index dominates that of the S&P500 index over this time period. Thus we expect to have only SSD or higher ranking.

CDF's of DJ_Ind and S&P500   SDF's of DJ_Ind and S&P500
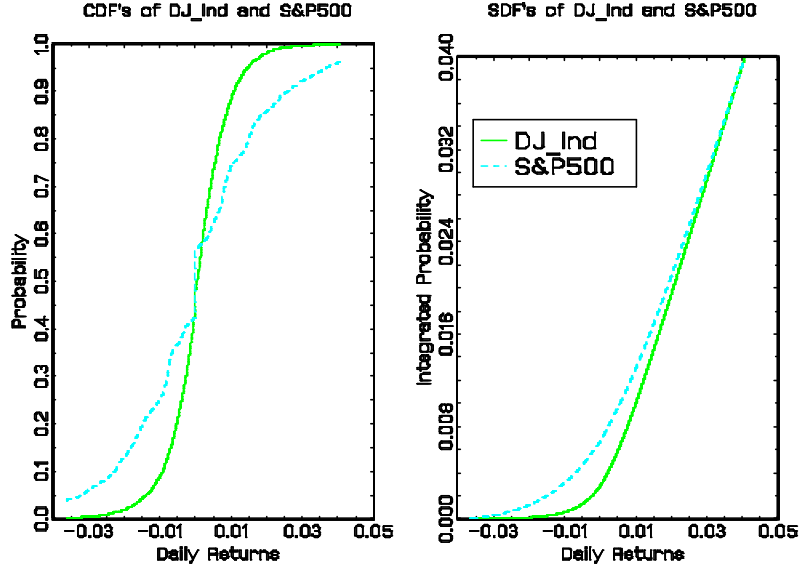
In Figure 2 we plot the surface $\int_y^x \left[F_{1N}(t) - F_{2N}(t) \right] dt$ against x, y on a grid of $x > 0, y < 0$. This surface is also everywhere positive, consistent with the hypothesis that the S&P500 index prospect dominates the Dow Jones index.

Prospect Dominance Function of DJ_Ind minus S&P500
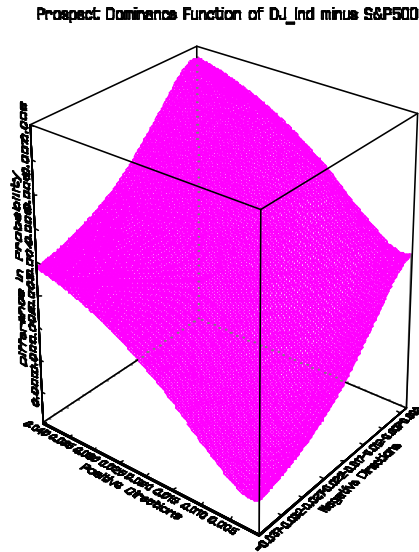
In Figure 3 we plot the p-value of our tests of the null hypotheses $d^* \leq 0$, $s^* \leq 0$, and $p^* \leq 0$ against subsample size. The results suggest strongly that the evidence is against $d^* \leq 0$ but in favour of $s^* \leq 0$ and $p^* \leq 0$.[13] Any of the automatic methods described in 4.1.2 would yield the

---

[13] In the test of prospect dominance we subtracted off the risk free rate measured by one month t-bill rates.

same conclusion. For comparison, the recentered bootstrap p-values are 0.1448, 0.9999, and 0.9999 respectively.
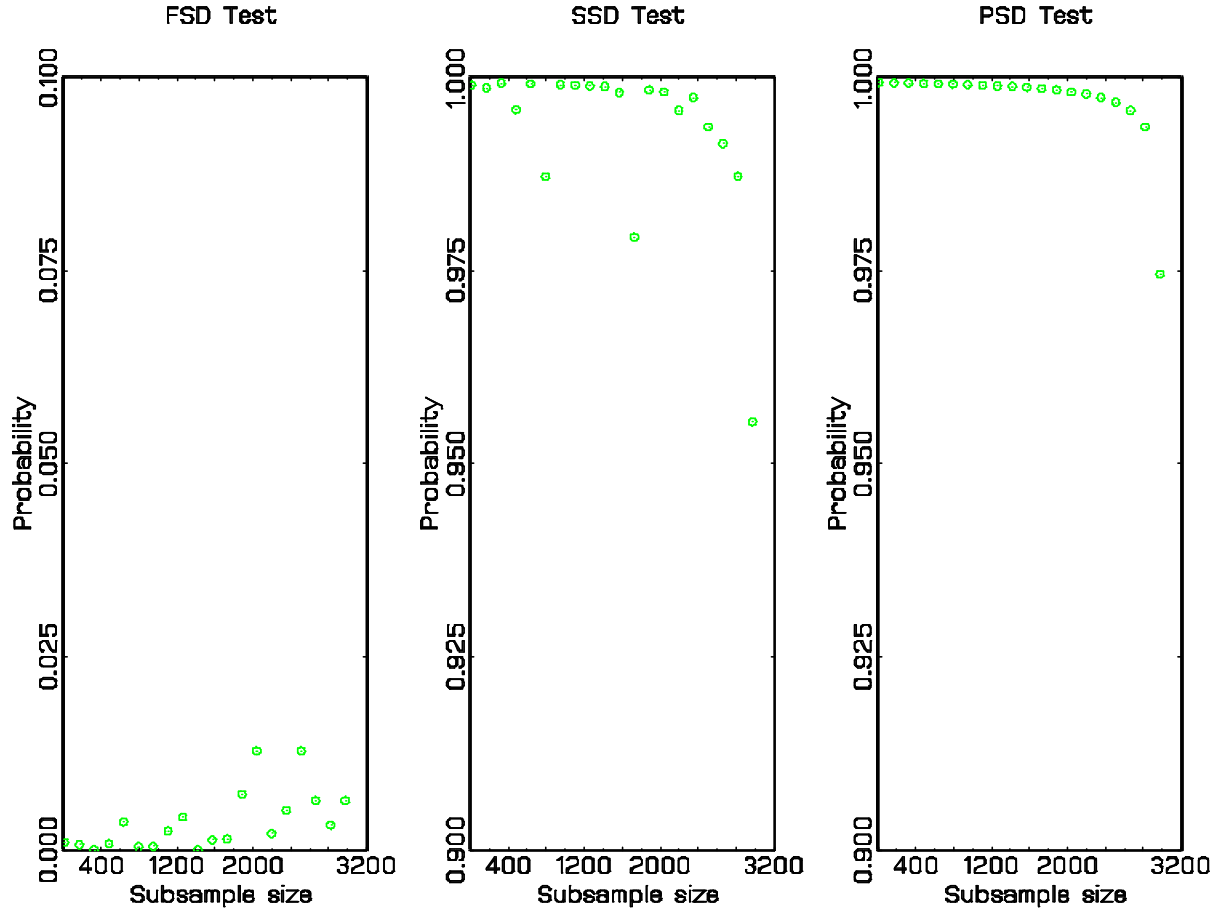
This is a rather striking result. The ranking of these return series depends on whether the prospect theory or the usual risk measures are favoured. The S&P500 index apparently offers investors better lottery opportunities on the upside, but better insurance features on the downside.

6 Concluding Remarks

We have obtained the asymptotic distribution of well known tests for FSD and SSD and demonstrated their consistency in a very general setting that allows generic dependence of prospects and non i.i.d observations. The availability of this technique for empirical situations in which ranking is done conditional on desirable controls is of consequence for widespread use of uniform ranking in empirical finance and welfare.

The performance of the subsampling technique is rather good in the cases we considered when the sample size is at least 500. We also find that the subsample method has both theoretically and empirically better power than a recentred bootstrap [except when sample sizes are very small, in which case the comparison seems to go the other way].

A   Appendix

Below we sketch the proof of Theorems in the main text only for the test D_N. The corresponding proofs for the other tests S_N and D_N are omitted for brevity and are available in our working paper version. We let C_j for some integer $j \geq 1$ denote a generic constant. (It is not meant to be equal in any two places it appears.) Let $\|Z\|_q$ denote the L^q norm $(E|Z|^q)^{1/q}$ for a random variable Z. The following lemma holds for all $k = 1, ..., K$:

Lemma 1 *Suppose Assumption 1 holds. Then, for each $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\overline{\lim_{N \to \infty}} \left\| \sup_{\rho_d^*((x_1,\theta_1),(x_2,\theta_2)) < \delta} \left| \nu_{kN}^d(x_1, \theta_1) - \nu_{kN}^d(x_2, \theta_2) \right| \right\|_q < \varepsilon, \tag{A.1}$$

where

$$\rho_d^* ((x_1, \theta_1), (x_2, \theta_2)) = \left\{ E \left[1(X_{ki}(\theta_1) \leq x_1) - 1(X_{ki}(\theta_2) \leq x_2) \right]^2 \right\}^{1/2}. \tag{A.2}$$

PROOF OF LEMMA 1. The result follows from Theorem 2.2 of Andrews and Pollard (1994) with $Q = q$ and $\gamma = 1$ if we verify the mixing and bracketing conditions in the theorem. The mixing condition is implied by Assumption 1(i). The bracketing condition also holds by the following argument: Let $\mathcal{F}_d^k = \{ 1(X_{ki}(\theta) \leq x) : (x, \theta) \in \mathcal{X} \times \Theta_k \}$. Then, F_d^k is a class of uniformly bounded functions satisfying the L^2-continuity condition, because we have

$$\sup_{i \geq 1} E \sup_{\substack{(x',\theta') \in \mathcal{X} \times \Theta_k: \\ |x'-x| \leq r_1, \|\theta'-\theta\| \leq r_2, \sqrt{r_1^2+r_2^2} \leq r}} \left| 1(X_{ki}(\theta') \leq x') - 1(X_{ki}(\theta) \leq x) \right|^2$$

$$= E \sup_{\substack{(x',\theta') \in \mathcal{X} \times \Theta_k: \\ |x'-x| \leq r_1, \|\theta'-\theta\| \leq r_2, \sqrt{r_1^2+r_2^2} \leq r}} \left| 1(X_{ki} \leq Z_{ki}'(\theta' - \theta_0) + x') - 1(X_{ki} \leq Z_{ki}'(\theta - \theta_0) + x) \right|^2$$

$$\leq E 1 \left(|X_{ki} - Z_{ki}'(\theta - \theta_0) - x| \leq \|Z_{ki}\| r_1 + r_2 \right)$$

$$\leq C_1 \left(E \|Z_{ki}\| r_1 + r_2 \right) \leq C_2 r,$$

where the second inequality holds by Assumption 1(iii) and $C_2 = \sqrt{2} C_1 (E \|Z_{ki}\| \vee 1)$ is finite by Assumption 1(ii). Now the desired bracketing condition holds because the L^2-continuity condition implies that the bracketing number satisfies $N(\varepsilon, \mathcal{F}_d^k) \leq C_3 (1/\varepsilon)^{L_k+1}$, see Andrews and Pollard (1994, p.121).   ∎

Lemma 2 *Suppose Assumptions 1-3 hold. Then, we have* $\forall k = 1, \ldots, K$,

$$\sup_{x \in \mathcal{X}} \left| \nu_{kN}^d(x, \widehat{\theta}_k) - \nu_{kN}^d(x, \theta_{k0}) \right| \xrightarrow{p} 0. \tag{A.3}$$

PROOF OF LEMMA 2. We first verify part (a). Consider the pseudometric (A.2). We have

$$
\begin{aligned}
&\sup_{x \in \mathcal{X}} \rho_d^* \left(\left(x, \widehat{\theta}_k \right), (x, \theta_{k0}) \right)^2 \\
&= \sup_{x \in \mathcal{X}} \iint \left[1 \left(\widetilde{x} \le x + z'(\widehat{\theta}_k - \theta_{k0}) \right) - 1 \left(\widetilde{x} \le x \right) \right]^2 dH_k(\widetilde{x}|z) dP_k(z) \\
&\le \sup_{x \in \mathcal{X}} \iint 1 \left(x - \left| z'(\widehat{\theta}_k - \theta_{k0}) \right| \le \widetilde{x} \le x + \left| z'(\widehat{\theta}_k - \theta_{k0}) \right| \right) dH_k(\widetilde{x}|z) dP_k(z) \\
&\le C_1 \left\| \widehat{\theta}_k - \theta_{k0} \right\| E \left\| Z_{ki} \right\| \xrightarrow{p} 0,
\end{aligned}
\tag{A.4}
$$

where $P_k(\cdot)$ denotes the distribution function of Z_{ki} and the second inequality holds by Assumption 1(iii) and a one-term Taylor expansion, and the last convergence to zero holds by Assumptions 1(ii) and 2. Now, this result and the stochastic equicontinuity result (A.1) yield the desired result (A.3) using a standard argument. ∎

Lemma 3 *Suppose Assumptions 1-3 hold. Then, we have* $\forall k = 1, \ldots, K$,

$$\sqrt{N} \sup_{x \in \mathcal{X}} \left\| F_k(x, \widehat{\theta}_k) - F_k(x, \theta_{k0}) - \Delta_{k0}'(x) \Gamma_{k0} \overline{\psi}_{kN}(\theta_{k0}) \right\| = o_p(1).$$

PROOF OF LEMMA 3. The proof is standard and follows from a mean value expansion and several applications of triangle inequality. See our website for details. ∎

Lemma 4 *Suppose Assumptions 1-3 hold. Then, we have*

$$\left(\ \nu_{kN}^d(\cdot, \theta_{k0}) - \nu_{lN}^d(\cdot, \theta_{l0}) \quad \sqrt{N} \overline{\psi}_{kN}'(\theta_{k0}) \quad \sqrt{N} \overline{\psi}_{lN}'(\theta_{l0}) \ \right)' \Rightarrow \left(\ \widetilde{d}_{kl}(\cdot) \quad \nu_{k0}' \quad \nu_{l0}' \ \right)'$$

$\forall \ k, l = 1, \ldots, K$ *and the sample paths of* $\widetilde{d}_{kl}(\cdot)$ *are uniformly continuous with respect to pseudometric* ρ_d *on* X *with probability one, where*

$$\rho_d(x_1, x_2) = \left\{ E \left[(1(X_{ki} \le x_1) - 1(X_{li} \le x_1)) - (1(X_{ki} \le x_2) - 1(X_{li} \le x_2)) \right]^2 \right\}^{1/2}.$$

PROOF OF LEMMA 4. By Theorem 10.2 of Pollard (1990), the result of Lemma 4 holds if we have (i) total boundedness of pseudometric space (X, ρ_d) (ii) stochastic equicontinuity of $\{ \nu_{kN}^d(\cdot, \theta_{k0}) - \nu_{lN}^d(\cdot, \theta_{l0}) : N \ge 1 \}$ and (iii) finite dimensional (fidi) convergence. Conditions (i) and (ii) follow from Lemma 1. We now verify condition (iii). We need to show that $(\nu_{kN}^d(x_1, \theta_{k0}) - \nu_{lN}^d(x_1, \theta_{l0}), \ldots, \nu_{kN}^d(x_J, \theta_{k0}) - \nu_{lN}^d(x_J, \theta_{l0}),$ $\sqrt{N} \overline{\psi}_{kN}(\theta_{k0})', \ \sqrt{N} \overline{\psi}_{lN}(\theta_{l0})')'$ converges in distribution to $\left(\widetilde{d}_{kl}(x_1), \ldots, \widetilde{d}_{kl}(x_J), \nu_{k0}', \ \nu_{l0}' \right)' \ \forall x_j \ \in \ X$,

$\forall j \leq J, \forall J \geq 1$. This result holds by the Cramer-Wold device and a CLT for bounded random variables (e.g., Hall and Heyde (1980, Corollary 5.1, p.132)) because the underlying random sequence $\{X_{ki} : i = 1, \ldots, n\}$ is strictly stationary and α- mixing with the mixing coefficients satisfying $\sum_{m=1}^{\infty} \alpha(m) < \infty$ by Assumption 1 and we have $|1(X_{ki} \leq x) - 1(X_{li} \leq x)| \leq 2 < \infty$. ∎

PROOF OF THEOREM 1. Suppose that $d^* = 0$. Then, there exists a pair (k, l) that satisfies $\sup_{x \in \mathcal{X}} [F_k(x) - F_l(x)] = 0$. For such pair, we have $F_k(x) \leq F_l(x)$ for all $x \in \mathcal{X}$ but $F_k(x) = F_l(x)$ for $x \in B_{kl}^d (\subset \mathcal{X})$. We first verify that

$$
\begin{aligned}
\widehat{D}_{kl} &\equiv \sup_{x \in \mathcal{X}} \sqrt{N} \left[\overline{F}_{kN}(x, \widehat{\theta}_k) - \overline{F}_{lN}(x, \widehat{\theta}_l) \right] \\
&\Rightarrow \sup_{x \in \mathcal{B}_{kl}^d} \left[\widetilde{d}_{kl}(\cdot) + \Delta_{k0}(\cdot)' \Gamma_{k0} \nu_{k0} - \Delta_{l0}(\cdot)' \Gamma_{l0} \nu_{l0} \right] \\
&= \sup_{x \in \mathcal{B}_{kl}^d} d_{kl}(x), \text{ say.}
\end{aligned}
\tag{A.5}
$$

Note that Lemmas 2 and 3 imply

$$
\begin{aligned}
\widehat{D}_{kl}(x) &\equiv \sqrt{N} \left[\overline{F}_{kN}(x, \widehat{\theta}_k) - \overline{F}_{lN}(x, \widehat{\theta}_l) \right] \\
&= \nu_{kN}^d(x, \widehat{\theta}_k) - \nu_{lN}^d(x, \widehat{\theta}_l) + \sqrt{N} \left[F_k(x, \widehat{\theta}_k) - F_l(x, \widehat{\theta}_l) \right] \\
&= \overline{D}_{kl}(x) + o_p(1) \text{ uniformly in } x \in \mathcal{X},
\end{aligned}
$$

where

$$
\overline{D}_{kl}(x) = D_{kl}^0(x) + D_{kl}^1(x)
\tag{A.6}
$$

$$
\begin{aligned}
D_{kl}^0(x) &= \nu_{kN}^d(x, \theta_{k0}) - \nu_{lN}^d(x, \theta_{l0}) \\
&\quad + \Delta_{k0}(x) \Gamma_{k0} \sqrt{N} \overline{\psi}_{kN}(\theta_{k0}) - \Delta_{l0}(x) \Gamma_{l0} \sqrt{N} \overline{\psi}_{lN}(\theta_{l0})
\end{aligned}
$$

$$
D_{kl}^1(x) = \sqrt{N} \left[F_k(x) - F_l(x) \right].
\tag{A.7}
$$

To show (A.5), we need to verify

$$
\sup_{x \in \mathcal{X}} \overline{D}_{kl}(x) \Rightarrow \sup_{x \in \mathcal{B}_{kl}^d} d_{kl}(x).
\tag{A.8}
$$

Note that

$$
\sup_{x \in \mathcal{B}_{kl}^d} D_{kl}^0(x) \Rightarrow \sup_{x \in \mathcal{B}_{kl}^d} d_{kl}(x)
\tag{A.9}
$$

by Lemma 4 and continuous mapping theorem. Note also that $\overline{D}_{kl}(x) = D_{kl}^0(x)$ for $x \in B_{kl}^d$. Given $\varepsilon > 0$, this implies that

$$
P \left(\sup_{x \in \mathcal{X}} \overline{D}_{kl}(x) \leq \varepsilon \right) \leq P \left(\sup_{x \in \mathcal{B}_{kl}^d} D_{kl}^0(x) \leq \varepsilon \right).
\tag{A.10}
$$

29

On the other hand, Lemma 4 and Assumptions 1(i), 2(ii) and 3(iii) imply that given λ and $\gamma > 0$, there exists $\delta > 0$ such that

$$P \left(\sup_{\substack{\rho(x,y)<\delta \\ y\in\mathcal{B}_{kl}^d}} \left| D_{kl}^0(x) - D_{kl}^0(y) \right| > \lambda \right) < \gamma \tag{A.11}$$

and

$$\sup_{x\in\mathcal{X}} \left| D_{kl}^0(x) \right| = O_p(1). \tag{A.12}$$

Using the results (A.11) and (A.12) and arguments similar to those in the proof of Theorem 6 of Klecan et. al. (1991, p.15), we can verify that

$$P \left(\sup_{x\in\mathcal{B}_{kl}^d} D_{kl}^0(x) \leq \varepsilon \right) \leq P \left(\sup_{x\in\mathcal{X}} \overline{D}_{kl}(x) \leq \varepsilon + \lambda \right) + 2\gamma \tag{A.13}$$

for N sufficiently large. Taking λ and γ small and using (A.9), (A.10) and (A.13) now establish the desired result (A.8) and hence (A.5). Now the desired result of Theorem 1 follows by continous mapping theorem because the terms $\widehat{D}_{ij}$ with (i,j) satifying $\sup_{x\in\mathcal{X}} \left[F_k(x) - F_l(x) \right] > 0$ will diverge to infinity and hence will not affect the limit distribution of D_N.

Next suppose $d^* < 0$. In this case, the set B_{kl}^d is an empty set and hence $F_k(x) < F_l(x) \ \forall x \in X$ for some (k,l). Then, $\sup_{x\in\mathcal{X}} \overline{D}_{kl}(x)$ defined in (A.6) will be dominated by the term $D_{kl}^1(x)$ which diverges to minus infinity for any $x \in X$ as required. Therefore, in this case D_N will also diverge to minus infinity. ∎

PROOF OF THEOREM 2. Let the asymptotic null distribution of D_N be given by $G(w)$. This distribution is absolutely continuous because it is a functional of a Gaussian process whose covariance function is nonsingular, see Lifshits (1982). Therefore, Theorem 2 holds if we establish

$$\widehat{G}_{N,\widehat{b}_N}(w) \xrightarrow{p} G(w) \ \ \forall w \in \mathbb{R}. \tag{A.14}$$

Let $G_b(w) = P \left(\sqrt{b} d_{N,b,i} \leq w \right) = P \left(\sqrt{b} d_b(W_1, \ldots, W_b) \leq w \right)$. Note that $\sup_{l_N \leq b \leq u_N} |G_b(w) - G(w)| \to$ 0, since $b \geq l_N \to \infty$. Therefore, to establish (A.14), it suffices to verify

$$\sup_{l_N \leq b \leq u_N} \left| \widehat{G}_{N,b}(w) - G_b(w) \right| \xrightarrow{p} 0 \ \ \forall w \in \mathbb{R}, \tag{A.15}$$

since then we have $P \left(\left| \widehat{G}_{N,\widehat{b}_N}(w) - G(w) \right| > \varepsilon \right) \to 0 \ \forall \varepsilon > 0$ by triangle inequality and Assumption 4.

We now verify (A.15). For any $\varepsilon > 0$ and integer $q \in (1, (N - u_N + 1)/2)$, we have

$$P\left(\sup_{l_N \leq b \leq u_N} \left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right)$$

$$\leq \sum_{b=l_N}^{u_N} P\left(\left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right)$$

$$\leq u_N \sup_{l_N \leq b \leq u_N} P\left(\left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right)$$

$$\leq u_N \left\{4\exp\left(-\frac{\varepsilon^2}{8}q\right) + 22\left(1 + \frac{4}{\varepsilon}\right)^{1/2} q\alpha\left(\left[\frac{N - u_N + 1}{2q}\right]\right)\right\}, \tag{A.16}$$

where the last inequality follows from Bosq (1998, Theorem 1.3). Take $q = [((N - u_N + 1)/2)^\gamma]$, where $\gamma = (A-1)/(A+1)$ with A satisfying Assumption 1(i). Then, the right hand side of (A.16) is bounded by $u_N\left\{O\left(\exp\left(-(N - u_N + 1)\right)\right) + O\left((N - u_N + 1)^{-1}\right)\right\}$ which converges to zero by Assumption 4. This proves (A.15) and hence part (a) of Theorem 2. Given this result, part (b) of Theorem 2 holds since we have

$$P\left(D_N > g_{N,\widehat{b}_N}(1-\alpha)\right) = P\left(D_N > g(1-\alpha) + o_p(1)\right) \to \alpha \text{ as } N \to \infty.$$

∎

PROOF OF THEOREM 3. The proof is similar to the proof of Theorem 2.6.1 of Politis et.al. (1999). ∎

PROOF OF THEOREM 4. The proof is similar to that of Theorem 1. Consider Lemmas 1-4 with $\nu_{kN}^d(x, \theta)$ now defined by

$$\nu_{kN}^d(x, \theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[1\left(X_{ki}(\theta) \leq x\right) - F_{kN}(x, \theta)\right] \text{ for } k = 1, ..., K. \tag{A.17}$$

Then, by contiguity, the results of Lemmas 2 and 3 hold under the local alternatives. This result and Assumption 2-lc imply that

$$\sqrt{N}\left[\overline{F}_{kN}(x, \widehat{\theta}_k) - \overline{F}_{lN}(x, \widehat{\theta}_l)\right]$$
$$= \nu_{kN}^d(x, \theta_{k0}) - \nu_{lN}^d(x, \theta_{l0})$$
$$+ \Delta_{k0}(x)\Gamma_{k0}\sqrt{N}\left(\overline{\psi}_{kN}(\theta_{k0}) - E\overline{\psi}_{kN}(\theta_{k0})\right) - \Delta_{l0}(x)\Gamma_{l0}\sqrt{N}\left(\overline{\psi}_{lN}(\theta_{l0}) - E\overline{\psi}_{lN}(\theta_{l0})\right)$$
$$+ \mu_{kl}(x) + o_p(1) \text{ uniformly in } x \in \mathcal{X},$$

Therefore, it suffices to show that Lemma 4 holds under the local alternatives. This follows by a slight modification of the proof of Lemma 4 and using the CLT of Herrndorf (1984) for α-mixing arrays to verify the condition (iii) (fidi convergence) of Theorem 10.2. of Pollard (1990). ∎

PROOF OF COROLLARY 5. We know that $g_{N,\widehat{b}_N}(1-\alpha) \xrightarrow{p} g(1-\alpha)$ under the null hypothesis. By contiguity, we have $g_{N,\widehat{b}_N}(1-\alpha) \xrightarrow{p} g(1-\alpha)$ under the local alternatives. The results of Corollary 5 now follows immediately from Theorem 4. ∎

REFERENCES

Abadie, A., (2001), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." Harvard University. Forthcoming in *Journal of the American Statistical Association*.

Anderson, G.J. (1996), "Nonparametric tests of stochastic dominance in income distributions," *Econometrica* 64, 1183-1193.

Andrews, D.W.K., (1997), "A conditional Kolmogorov test," *Econometrica* 65, 1097-1128.

Andrews, D. W. K. and D. Pollard (1994), "An introduction to functional central limit theorems for dependent stochastic processes," *International Statistical Review* 62, 119-132.

Barrett, G. and S. Donald (1999), "Consistent Tests for Stochastic Dominance, unpublished manuscript, Dept. of Economics, Univ. of New South Wales.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993), *Efficient and adaptive estimation for semiparametric models*, The Johns Hopkins Univ. Press, Baltimore and London.

Bishop, J.A., J.P.Formby, and P.D. Thisle (1992), "Convergence of the South and non-South income distributions, 1969-1979", *American Economic Review* 82, 262-272.

Bosq, D. (1998), *Nonparametric statistics for stochastic processes: estimation and prediction*, Lecture Notes in Statistics 110, 2nd edition, Springer, New York.

Carlstein, E. (1986), "The use of subseries methods for estimating the variance of a general statistic from a stationary time series," *Annals of Statistics* 14, 1171-1179.

Chernozhukov, V. (2002), "Inference on Quantile Regression Process, an Alternative," Working paper, MIT.

Crawford, I. (1999), "Nonparametric tests of Stochastic Dominance in Bivariate Distributions with an Application to UK data," Institute for Fiscal Studies, WP 28/99.

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

Davidson, R. and J-Y Duclos (1997), "Statistical inference for the measurement of the incidence of taxes and transfers", *Econometrica* 52, 761-76.

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

Delgado, M., J.M. Rodriguez-Poo and M. Wolf (2001), "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters* 73, 241-250.

Hadar, J. and W.R. Russell (1969), "Rules for ordering uncertain prospects, *American Economic Review* 59, 25-34.

Hall, P. and Heyde, C. C. (1980), *Martingale limit theory and its application*, Academic Press, New York.

Hansen, B.E. (1996a), "Stochastic equicontinuity for unbounded dependent heterogeneous arrays," *Econometric Theory* 12, 347-359.

Hansen, B.E. (1996b), "Inference when a nuisance parameter is not identified under the null Hypothesis," *Econometrica* 64, 413-430.

Hansen, P.R. (2001), "An unbiased and powerful test for superior predictive ability," Working paper, Brown University.

Härdle, W., J. Horowitz and J.-P. Kreiss (2001), "Bootstrap methods for time series," Working paper, Humboldt Universität zu Berlin.

Herrndorf, N. (1984), "An invariance principle for weakly dependent sequences of random variables," *Annals of Probability* 12, 141-153.

Ho, K.Y. (2003): "Long-run stock price performance after IPO's: what do tests for stochastic dominance tell us?" *Applied Economics Letters* 10, 15-19.

Horowitz, J.L. (2000), "The Bootstrap," Forthcoming in *The Handbook of Econometrics*, volume 5.

Kahneman, D., and A. Tversky (1979), "Prospect Theory of Decisions Under Risk," *Econometrica* 47, 263-291.

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

Klecan, L., R. McFadden, and D. McFadden (1991), "A robust test for stochastic dominance," Working paper, Dept. of Economics, MIT.

Künsch, H. R. (1989), "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics* 17, 1217-1241.

Lehmann, E.L. (1986), *Testing statistical hypotheses*, 2nd Edtion, J. Wiley and Sons, New York.

Levy, H., and Z. Wiener (1998), "Stochastic Dominance and Prospect Dominance with Subjective Weighting Functions" *Journal of Risk and Uncertainty*, 147-163.

Lifshits, M. A. (1982), "On the absolute continuity of distributions of functionals of random processes," *Theory of Probability and Its Applications* 27, 600-607.

Maasoumi, E. (2001), "Parametric and nonparametric tests of limited domain and ordered hypotheses in economics", chapter 25, in B. Baltagi (Ed.) *A Companion to Econometric Theory,* Basil Blackwell.

Maasoumi, E. and A. Heshmati (2000), "Stochastic dominance amongst Swedish income distributions", 19-3, *Econometric Reviews*.

Maasoumi, E. and D. Millimet (2002), "Robust Inference Concerning Recent Trends in U.S. Environmental Quality", Department of Economics, SMU.

McFadden, D. (1989), "Testing for stochastic dominance," in Part II of T. Fomby and T.K. Seo (eds.) *Studies in the Economics of Uncertainty* (in honor of J. Hadar), Springer-Verlag.

Otten, R. & D., Barns, (2001), "Statistical tests for return-based style analysis", Working paper, Maastricht University.

Pollard, D. (1990), *Empirical processes: theory and applications*, CBMS Conference Series in Probability and Statistics, Vol. 2. Institute of Mathematical Statistics, Hayward.

Politis, D.N. and J. P. Romano (1993), "The stationary bootstrap," *Journal of the American Statistical Association* 89, 1303-1313.

Politis, D. N. and J. P. Romano (1994), "Large sample confidence regions based on subsamples under minimal assumptions." *Annals of Statistics* 22, 2031-2050.

Politis, D. N., J. P. Romano and M. Wolf (1999), *Subsampling*, Springer-Verlag, New York.

Post, T. (2002): "Empirical Tests for Stochastic Dominance Efficiency," Forthcoming in *Journal of Finance.*

Sharpe, William F., (1992), "Asset allocation: management style and performance measurement", The Journal of Portfolio Management, Winter 1992, pp. 7-19.

Shorrocks, A.F.(1983), "Ranking income distributions," *Economica* 50, 3-17.

Shorrocks A., and J. Foster (1987), "Transfer sensitive inequality measures," *Review of Economic Studies* 54, 485-497.

Tse, Y.K. and X.B. Zhang (2000), "A Monte carlo Investigation of Some Tests for Stochastic Dominance", unpublished manuscript, National University of Singapore.

Tversky, A. and D. Kahneman (1992), "Advances in Prospect Theory:Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297-323.

Whang, Y.-J (2001): "Consistent Specification Testing for Conditional Moment Restrictions," *Economics Letters* 71, 299-306.

Whitmore,G.A. and M.C. Findley (1978), *Stochastic Dominance: An approach to decision making under risk*, Heath, Lexington :Mars.

Xu, K., G. Fisher, and D. Wilson (1995), "New distribution-free tests for stochastic dominance," Working paper No. 95-02, February, Dept. of Economics, Dalhousie University, Halifax, Nova Scotia.

B Table Information

The numbers are rejection frequencies for the test of First Order Stochastic Dominance (FSD) and Second Order Stochastic Dominance (SSD) with critical values computed by the automatic methods described in section 4.1.2 for the 5%, 10%, and 20% null rejection probabilities. The rejection frequencies are computed from 1,000 replications.

The design parameters (column 1) are given in the tables in the text. Below, Table jF/S corresponds to design j=1,2,3: Table 'jF' corresponds to the test of FSD, while Table 'jS' corresponds to the test of SSD. Table 3RF corresponds to the test of FSD on residuals from designs 3, while Table 3RS corresponds to the test of SSD on residuals from designs 3.

The automatic subsample method is implemented on a grid of 20 subsample sizes: for $n = 50$, the subsample sizes are $\{20, 21, \dots, 40\}$; for $n = 500$ the subsample sizes are $\{50, 65, \dots, 350\}$; for $n = 1000$ the subsample sizes are $\{100, 120, \dots, 500\}$.

Mean and median refer to the critical values defined in (16) and (17); minvol is the minimum volatility method described in that section. The columns labelled Boot refer to the *recentered* full sample bootstrap described in 4.2 with 200 repetitions.

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.1140	0.1630	0.2610	0.1370	0.1790	0.2700	0.1850	0.2000	0.2990	0.0630	0.1060	0.1990
1a	500	0.0590	0.1080	0.1990	0.0570	0.1070	0.1980	0.1100	0.1610	0.2630	0.0560	0.1040	0.1800
$d^* = 0$	1000	0.0460	0.0950	0.1850	0.0500	0.0930	0.1860	0.0710	0.1280	0.2060	0.0490	0.1010	0.2070
	50	0.1030	0.1390	0.2470	0.1180	0.1520	0.2470	0.1600	0.1840	0.2740	0.0550	0.0860	0.1850
1b	500	0.0540	0.1010	0.2150	0.0620	0.1070	0.2170	0.1010	0.1620	0.2690	0.0510	0.0830	0.1900
$d^* = 0$	1000	0.0480	0.0830	0.1690	0.0480	0.0840	0.1710	0.0720	0.1200	0.2090	0.0590	0.0960	0.1900
	50	0.3610	0.4410	0.5760	0.3640	0.4450	0.5650	0.4120	0.4270	0.5740	0.6850	0.7440	0.8160
1c	500	0.9500	0.9530	0.9590	0.9420	0.9500	0.9570	0.8980	0.9150	0.9450	0.9830	0.9830	0.9830
$d^* > 0$	1000	0.9600	0.9610	0.9700	0.9600	0.9620	0.9700	0.9580	0.9620	0.9860	0.9950	0.9950	0.9950
	50	0.3730	0.4350	0.5430	0.3680	0.4270	0.5410	0.4100	0.4290	0.5600	0.6840	0.7660	0.8340
1d	500	0.9650	0.9680	0.9720	0.9590	0.9640	0.9720	0.8860	0.9130	0.9490	0.9840	0.9840	0.9870
$d^* > 0$	1000	0.9580	0.9590	0.9680	0.9570	0.9590	0.9700	0.9520	0.9600	0.9850	0.9940	0.9940	0.9940
	50	0.3790	0.4560	0.5890	0.3800	0.4520	0.5810	0.4180	0.4340	0.5880	0.6560	0.7230	0.8000
1e	500	0.9640	0.9720	0.9810	0.9590	0.9660	0.9770	0.8880	0.9150	0.9500	0.9920	0.9920	0.9920
$d^* > 0$	1000	0.9530	0.9560	0.9660	0.9530	0.9560	0.9650	0.9480	0.9560	0.9770	0.9920	0.9920	0.9920

Table1F: Test of FSD

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.1010	0.1500	0.2440	0.1280	0.1850	0.2810	0.2110	0.2640	0.3510	0.0660	0.1160	0.2350
1a	500	0.0490	0.0920	0.1830	0.0540	0.0990	0.1850	0.1020	0.1560	0.2780	0.0550	0.1210	0.2270
$s^* = 0$	1000	0.0540	0.0980	0.1760	0.0580	0.1010	0.1790	0.0660	0.1130	0.2180	0.0500	0.1000	0.2050
	50	0.0760	0.1150	0.2150	0.1050	0.1430	0.2480	0.1710	0.2020	0.3090	0.0610	0.1040	0.1980
1b	500	0.0660	0.1270	0.2160	0.0690	0.1280	0.2210	0.1020	0.1630	0.2960	0.0600	0.1220	0.2130
$s^* = 0$	1000	0.0680	0.1230	0.2300	0.0690	0.1260	0.2360	0.0960	0.1600	0.2770	0.0500	0.0950	0.1910
	50	0.2390	0.3320	0.4960	0.2470	0.3380	0.4970	0.3240	0.3620	0.5300	0.3360	0.4850	0.6980
1c	500	0.9060	0.9420	0.9580	0.8930	0.9250	0.9500	0.8290	0.8550	0.9150	0.4510	0.6530	0.9120
$s^* > 0$	1000	0.9570	0.9610	0.9700	0.9560	0.9610	0.9700	0.9510	0.9570	0.9880	0.5450	0.7910	0.9610
	50	0.2230	0.3000	0.4400	0.2340	0.3100	0.4280	0.3120	0.3460	0.4820	0.3290	0.4860	0.7100
1d	500	0.9070	0.9360	0.9700	0.8850	0.9160	0.9550	0.8290	0.8630	0.9160	0.4230	0.6610	0.9050
$s^* > 0$	1000	0.9570	0.9580	0.9680	0.9550	0.9580	0.9700	0.9390	0.9560	0.9820	0.5240	0.7700	0.9450
	50	0.2090	0.3060	0.4510	0.2290	0.3090	0.4490	0.3070	0.3460	0.5050	0.2990	0.4410	0.6660
1e	500	0.8970	0.9410	0.9740	0.8760	0.9230	0.9620	0.8210	0.8660	0.9060	0.4240	0.6520	0.8970
$s^* > 0$	1000	0.9490	0.9520	0.9650	0.9490	0.9510	0.9650	0.9380	0.9500	0.9770	0.4840	0.7440	0.9390

Table1S: Test of SSD

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.1110	0.1520	0.2500	0.1330	0.1630	0.2610	0.1720	0.1900	0.2890	0.0540	0.0940	0.1860
2a	500	0.0470	0.0860	0.1850	0.0490	0.0880	0.1890	0.0940	0.1450	0.2440	0.0550	0.1040	0.2030
$d^* = 0$	1000	0.0680	0.1150	0.2020	0.0710	0.1160	0.2030	0.0890	0.1360	0.2310	0.0440	0.1020	0.2120
	50	0.0790	0.1090	0.1880	0.0860	0.1200	0.1950	0.1220	0.1310	0.2210	0.0720	0.1160	0.1840
2b	500	0.0120	0.0390	0.1080	0.0170	0.0490	0.1200	0.0750	0.1570	0.2630	0.0260	0.0490	0.0820
$d^* = 0$	1000	0.0210	0.0440	0.1350	0.0280	0.0540	0.1510	0.0710	0.1210	0.2570	0.0180	0.0330	0.0730
	50	0.2960	0.3640	0.4970	0.2990	0.3650	0.5010	0.3800	0.3880	0.5230	0.4530	0.5210	0.6420
2c	500	0.9650	0.9860	0.9960	0.9460	0.9740	0.9910	0.8990	0.9400	0.9640	1.0000	1.0000	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	0.9990	1.0000	1.0000	0.9940	0.9950	1.0000	1.0000	1.0000	1.0000
	50	0.2640	0.3280	0.4400	0.2730	0.3330	0.4490	0.2550	0.2860	0.4250	0.1730	0.2330	0.3850
2d	500	0.9550	0.9800	0.9960	0.9360	0.9640	0.9880	0.9220	0.9280	0.9130	0.9880	0.9950	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	0.9980	1.0000	1.0000	0.9970	0.9980	0.9990	1.0000	1.0000	1.0000

Table2F: Test of FSD

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.0680	0.1130	0.2070	0.0980	0.1490	0.2440	0.1980	0.2330	0.3410	0.0560	0.1050	0.2100
2a	500	0.0560	0.1040	0.1850	0.0620	0.1080	0.2010	0.1190	0.1700	0.3070	0.0460	0.1010	0.2090
$s^* = 0$	1000	0.0620	0.1070	0.1950	0.0660	0.1100	0.2000	0.0840	0.1410	0.2320	0.0650	0.1260	0.2210
	50	0.0580	0.0960	0.2000	0.0940	0.1340	0.2450	0.1600	0.2000	0.2850	0.0780	0.1320	0.2210
2b	500	0.0010	0.0120	0.0800	0.0070	0.0230	0.1100	0.0860	0.1430	0.2260	0.0060	0.0200	0.0820
$s^* = 0$	1000	0.0040	0.0230	0.1120	0.0100	0.0380	0.1350	0.0500	0.0920	0.2100	0.0010	0.0070	0.0620
	50	0.0010	0.0550	0.1320	0.1650	0.2060	0.2930	0.0410	0.0360	0.0290	0.0060	0.0070	0.0090
2c	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$s^* = 0$	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.1680	0.2480	0.3880	0.1870	0.2540	0.3860	0.2300	0.2570	0.4090	0.0270	0.0960	0.3480
2d	500	0.9100	0.9580	0.9930	0.8890	0.9360	0.9720	0.8350	0.8780	0.9210	0.3320	0.7180	0.9790
$s^* > 0$	1000	0.9990	1.0000	1.0000	0.9980	1.0000	1.0000	0.9920	0.9920	0.9990	0.8600	0.9910	1.0000

Table2S: Test of SSD

		Mean			Median			MinVol			Boot		
Design	n	5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.6120	0.6830	0.7860	0.6040	0.6690	0.7680	0.5640	0.5720	0.6900	0.9590	0.9750	0.9870
3a	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9930	0.9950	1.0000	1.0000	1.0000	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	50	0.0580	0.0830	0.1540	0.0650	0.0910	0.1600	0.1040	0.1110	0.1800	0.0250	0.0490	0.0860
3b	500	0.0000	0.0000	0.0040	0.0020	0.0030	0.0160	0.0440	0.0680	0.0950	0.0000	0.0000	0.0000
$d^* < 0$	1000	0.0000	0.0000	0.0020	0.0010	0.0040	0.0130	0.0190	0.0300	0.0640	0.0000	0.0000	0.0000
	50	0.6010	0.6510	0.7770	0.5860	0.6460	0.7660	0.5480	0.5560	0.6800	0.9490	0.9830	0.9890
3c	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9960	0.9970	1.0000	1.0000	1.0000	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table3F: Test of FSD

		Mean			Median			MinVol			Boot		
Design	n	5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.0010	0.0530	0.1250	0.1600	0.1970	0.2820	0.0710	0.0530	0.0460	0.0210	0.0260	0.0320
3a	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$s^* = 0$	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.0430	0.0870	0.1710	0.1210	0.1750	0.2750	0.1990	0.2060	0.2480	0.0440	0.0760	0.1510
3b	500	0.0000	0.0010	0.0240	0.0060	0.0150	0.0500	0.1180	0.1380	0.1560	0.0000	0.0010	0.0070
$s^* = 0$	1000	0.0000	0.0030	0.0180	0.0030	0.0090	0.0230	0.0110	0.0220	0.0660	0.0000	0.0000	0.0000
	50	0.5250	0.5850	0.6990	0.5180	0.5730	0.6980	0.5330	0.5790	0.6810	0.9340	0.9380	0.9490
3c	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9890	0.9900	0.9960	1.0000	1.0000	1.0000
$s^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table3S: Test of SSD

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.5220	0.5880	0.7220	0.5160	0.5780	0.6980	0.5060	0.5200	0.6380	0.9270	0.9540	0.9750
3d	500	1.0000	1.0000	1.0000	0.9980	1.0000	1.0000	0.9940	0.9960	1.0000	1.0000	1.0000	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	50	0.0860	0.1220	0.2000	0.0980	0.1260	0.2000	0.1580	0.1800	0.2480	0.0100	0.0280	0.0670
3e	500	0.0000	0.0000	0.0040	0.0000	0.0020	0.0180	0.0280	0.0380	0.0840	0.0000	0.0000	0.0000
$d^* = 0$	1000	0.0000	0.0000	0.0000	0.0020	0.0020	0.0080	0.0060	0.0160	0.0560	0.0000	0.0000	0.0000
	50	0.5320	0.6240	0.7320	0.5360	0.6220	0.7180	0.5320	0.5400	0.6780	0.9440	0.9690	0.9850
3f	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9880	0.9920	0.9960	1.0000	1.0000	1.0000
$d^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table3RF: Test of FSD on residuals

Design	n	Mean			Median			MinVol			Boot		
		5%	10%	20%	5%	10%	20%	5%	10%	20%	5%	10%	20%
	50	0.0060	0.0080	0.0120	0.0060	0.0140	0.0320	0.0120	0.0140	0.0120	0.0250	0.0270	0.0280
3d	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$s^* = 0$	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.0680	0.0880	0.1620	0.0700	0.1060	0.1760	0.1280	0.1540	0.2100	0.0310	0.0620	0.1400
3e	500	0.0000	0.0040	0.0420	0.0060	0.0180	0.0500	0.0240	0.0400	0.1000	0.0010	0.0010	0.0040
$s^* = 0$	1000	0.0000	0.0000	0.0100	0.0020	0.0060	0.0320	0.0120	0.0180	0.0600	0.0000	0.0000	0.0000
	50	0.5320	0.5960	0.6960	0.5220	0.5920	0.6880	0.5340	0.5640	0.6540	0.9190	0.9260	0.9330
3f	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9860	0.9920	0.9980	1.0000	1.0000	1.0000
$s^* > 0$	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table3RS: Test of SSD on residuals

C   Some Supporting Material

C.1   Design

We show the cdf's and integrated cdf's for the various designs used in our simulations.

Design 1a cdf

Design 1a sdf

Design 1b cdf

Design 1b sdf

Design 1c cdf

Design 1c sdf

Design 1d cdf

Design 1d sdf

Design 1e cdf

Design 1e sdf

Design 2a cdf

Design 2a sdf

Design 2b cdf

Design 2b sdf

Design 2c cdf

Design 2c sdf

Design 2d cdf

Design 2d sdf

Design 3a cdf

Design 3a sdf

Design 3b cdf

Design 3b sdf

Design 3c cdf

Design 3c sdf

44

Design 3d cdf

Design 3d sdf

Design 3e cdf

Design 3e sdf

Design 3f cdf

Design 3f sdf

C.2   Fixed Subsample Size Results

We show some results from the subsample based test when a fixed subsample is used. The figure below shows the rejection frequencies of the FSD test on designs 1a-1e [row 1 is design 1a, etc] as a function of subsample size.
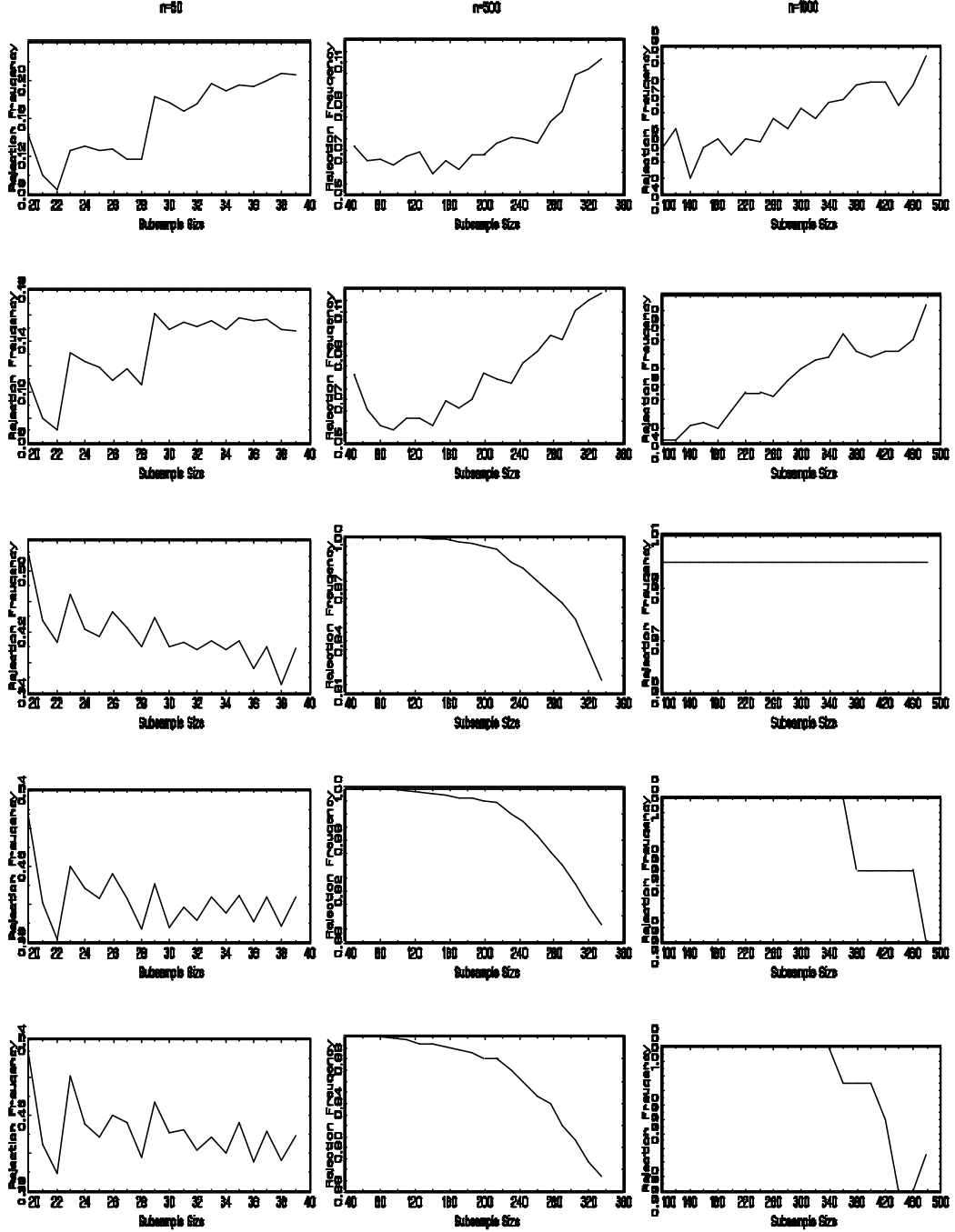
C.2.1 Multivariate Simulations

In this simulation we generate data from $N(0, I_k)$, where the dimensionality $k = 2, 3, \ldots, 10$. We use the automatic subsample rules applied in the bivariate case and sample sizes $n = 50, 500, 1000$. Note that the Monte Carlo standard errors are approximately 0.013 when the target is 0.20. The null hypothesis is true throughout here. The results are shown in Tables 4a,b. The results suggest that the performance does not deteriorate too rapidly with dimensionality. In fact, in many cases size is better with larger dimension.

n	k	Mean			Median			MinVol		
		5%	10%	20%	5%	10%	20%	5%	10%	20%
	2	0.2180	0.2743	0.3603	0.2221	0.2702	0.3726	0.2221	0.2835	0.3552
	3	0.2006	0.2456	0.2979	0.2057	0.2508	0.3050	0.2016	0.2385	0.3040
	4	0.2170	0.2764	0.3582	0.2139	0.2764	0.3327	0.2282	0.2794	0.3552
	5	0.2037	0.2242	0.2958	0.1883	0.2016	0.2170	0.2006	0.2098	0.2518
50	6	0.2190	0.2733	0.4094	0.1474	0.1607	0.2518	0.1730	0.2006	0.2815
	7	0.2221	0.3173	0.3910	0.0952	0.1484	0.2805	0.1894	0.2416	0.3163
	8	0.2825	0.3009	0.3183	0.1433	0.2190	0.2897	0.2180	0.2702	0.3071
	9	0.2405	0.2467	0.2477	0.1627	0.2047	0.2436	0.2016	0.2242	0.2405
	10	0.1914	0.1914	0.1986	0.1627	0.1760	0.1883	0.1791	0.1863	0.1914
	2	0.1218	0.1914	0.3009	0.1157	0.1924	0.2948	0.1157	0.1965	0.3142
	3	0.1484	0.1945	0.2784	0.1505	0.2027	0.2774	0.2027	0.2702	0.3306
	4	0.1157	0.1607	0.2528	0.1146	0.1576	0.2569	0.1228	0.1627	0.2856
	5	0.1290	0.1801	0.2866	0.1320	0.1740	0.2835	0.2078	0.2671	0.3429
500	6	0.1290	0.1945	0.2661	0.1402	0.1965	0.2651	0.1576	0.2006	0.2856
	7	0.0952	0.1392	0.2313	0.0880	0.1412	0.2375	0.0962	0.1576	0.2958
	8	0.0962	0.1423	0.2395	0.0921	0.1464	0.2426	0.1351	0.1842	0.2886
	9	0.0890	0.1586	0.2467	0.0942	0.1668	0.2733	0.1228	0.1965	0.2968
	10	0.1085	0.1771	0.2907	0.1116	0.1863	0.3009	0.1853	0.2242	0.3163
	2	0.1116	0.1484	0.2446	0.1085	0.1484	0.2497	0.1249	0.1996	0.2917
	3	0.0850	0.1249	0.2272	0.0942	0.1228	0.2068	0.0921	0.1341	0.2170
	4	0.0880	0.1187	0.2272	0.0942	0.1218	0.2293	0.1310	0.1760	0.2702
	5	0.1024	0.1423	0.2282	0.0993	0.1412	0.2293	0.1331	0.1853	0.2661
1000	6	0.1003	0.1392	0.2323	0.1013	0.1361	0.2344	0.0972	0.1453	0.2446
	7	0.0686	0.1290	0.1996	0.0676	0.1259	0.2160	0.0716	0.1136	0.2313
	8	0.0993	0.1658	0.2487	0.1075	0.1750	0.2538	0.1044	0.1924	0.2743
	9	0.0819	0.1238	0.2344	0.0870	0.1331	0.2303	0.0921	0.1720	0.2733
	10	0.0890	0.1464	0.2764	0.0860	0.1535	0.2641	0.1638	0.2119	0.3040

Table4a: Test of FSD

n	k	Mean			Median			MinVol		
		5%	10%	20%	5%	10%	20%	5%	10%	20%
	2	0.1648	0.2119	0.2979	0.1955	0.2446	0.3347	0.2395	0.2692	0.3460
	3	0.1238	0.1546	0.1975	0.1730	0.1996	0.2712	0.2661	0.2497	0.2426
	4	0.1024	0.1208	0.1679	0.1024	0.1208	0.2047	0.0911	0.0798	0.0737
	5	0.0962	0.1167	0.1525	0.0635	0.1024	0.2027	0.0235	0.0235	0.0256
50	6	0.0942	0.1075	0.1494	0.0747	0.1177	0.2252	0.0051	0.0051	0.0051
	7	0.0757	0.0921	0.1310	0.0798	0.1341	0.2477	0.0031	0.0031	0.0031
	8	0.0645	0.0768	0.1249	0.1075	0.1597	0.2753	0.0000	0.0000	0.0000
	9	0.0522	0.0614	0.1167	0.1269	0.1812	0.3019	0.0000	0.0000	0.0000
	10	0.0450	0.0737	0.1290	0.1228	0.1720	0.2989	0.0000	0.0000	0.0000
	2	0.1249	0.1668	0.2569	0.1208	0.1658	0.2631	0.1157	0.1801	0.2651
	3	0.0890	0.1453	0.2426	0.1249	0.1668	0.2651	0.2272	0.2764	0.3449
	4	0.0481	0.0890	0.1617	0.0768	0.1341	0.1883	0.1720	0.1873	0.1955
	5	0.0665	0.1075	0.1464	0.0768	0.0901	0.0962	0.0901	0.0993	0.0993
500	6	0.0665	0.0890	0.1290	0.0389	0.0399	0.0399	0.0399	0.0399	0.0399
	7	0.0450	0.0706	0.1054	0.0113	0.0113	0.0113	0.0113	0.0113	0.0113
	8	0.0706	0.0880	0.1075	0.0092	0.0092	0.0092	0.0092	0.0092	0.0092
	9	0.0829	0.0850	0.1044	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061
	10	0.0522	0.0645	0.0809	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.1095	0.1443	0.2190	0.1126	0.1515	0.2272	0.1341	0.1801	0.2528
	3	0.0665	0.1003	0.1791	0.0645	0.1136	0.2139	0.0819	0.1341	0.2590
	4	0.0553	0.1105	0.1812	0.0706	0.1259	0.1934	0.1198	0.1668	0.2170
	5	0.0645	0.0962	0.1249	0.0706	0.1013	0.1198	0.1013	0.1198	0.1198
1000	6	0.0420	0.0553	0.0645	0.0563	0.0563	0.0563	0.0563	0.0563	0.0563
	7	0.0174	0.0215	0.0307	0.0184	0.0184	0.0184	0.0184	0.0184	0.0184
	8	0.0041	0.0102	0.0246	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041
	9	0.0031	0.0082	0.0307	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031
	10	0.0031	0.0061	0.0205	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031

Table 4b: Test of SSD