Consistent Testing for Stochastic Dominance under General Sampling Schemes*

Oliver Linton[†]
London School of Economics

Esfandiar Maasoumi[‡]
Southern Methodist University

Yoon-Jae Whang[§]
Korea University

December 9, 2003 (First Version April 2001)

Abstract

We propose a procedure for estimating the critical values of the extended Kolmogorov-Smirnov tests of Stochastic Dominance of arbitrary order 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 based on subsampling and we show that the resulting tests are consistent and powerful against some $N^{-1/2}$ local alternatives. We also propose some heuristic methods for selecting subsample size and demonstrate in simulations that they perform reasonably. We describe an alternative method for obtaining critical values based on recentering the test statistic and using full sample bootstrap methods. We compare the two methods in theory and in practice.

*We would like to thank three referees, Joel Horowitz, Hashem Pesaran, Bernard Salanié, Steve Satchell, and Michael Wolf for helpful comments.

[†]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: `lintono@lse.ac.uk`. My research was supported by the Economic and Social Research Council of the United Kingdom and the Cowles Foundation.

[‡]Department of Economics, Southern Methodist University, Dallas, Texas, 75275-0496. Email: maasoumi@mail.smu.edu.

[§]Department of Economics, Korea University, Seoul 136-701, Korea. Email Address : `whang@korea.ac.kr` I would like to thank the hospitality of the Cowles Foundation during my visits. This work was supported by the Korea Research Foundation.

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 based on specific utility (loss) functions. The latter rankings are obtained with 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. In contrast, *uniform* order relations such as Stochastic Dominance (SD) rankings can produce "majority" assessments based on the expected utility paradigm and its mathematical regularity conditions. These relations are defined over relatively large classes of utility functions.

In this paper we propose resampling procedures for estimating the critical values of a suitably extended Kolmogorov-Smirnov test for Stochastic Dominance (SD) amongst K-competing states. Alternative implementations of this test have been examined by several authors including McFadden (1989), Klecan, McFadden, and McFadden (1991), and Barrett and Donald (2003).

Econometric tests for the existence of SD orders involve composite hypotheses on inequality restrictions. These restrictions may be equivalently formulated in terms of distribution function distances, their quantiles, or other conditional moments. 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 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) surveys 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, as we do.

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 the exchangeability assumption.[1] Barrett and Donald (2003) propose an alternative simulation

[1] 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.

method based on an idea of Hansen (1996b) for deriving critical values, also in the case where the prospects are mutually independent, and the data are i.i.d. These tests are consistent against all alternatives at least under the stated sampling assumptions.

We propose to estimate the critical values using the subsampling method proposed in Politis and Romano (1994). We also investigate a more standard full-sample bootstrap applied to a recentered test statistic. We prove that our *subsampling* test is consistent against all (nonparametric) alternatives. Because choice of subsample size may be important in practice, our main results are proven for random subsamples - this is the first result of its kind that we are aware of for subsampling. We give three practical methods of selecting subsample size.

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 non i.i.d. Accommodating generic dependence between the variables which are to be ranked is especially necessary in many 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 the latter assumptions appear to be patently false in many empirical settings. Some program evaluation settings rely on appropriately randomized treatment and/or randomized assignment and may be free of this "dependence problem".

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 Maasoumi and Millimet (2003) 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 (based on the actual observations) with tests of dominance amongst the residual distributions.[2] 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. Similarly, 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

[2] 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.

residuals from a linear regression.

Also, given the recent credible challenges to the standard risk aversion and expected utility paradigm, we propose a test of 'Prospect Stochastic Dominance', and propose consistent critical values using subsampling.

Finally, we also describe a full sample bootstrap procedure and make a comparison between this method and the subsampling procedure. The methods relying on standard bootstrap [as well as the Barrett and Donald (2003) simulation methods] typically try to mimic the asymptotic null distributions in the *least favorable case.* This is a subset of the boundary of the null where the marginal distribution functions are equal. However, the *boundary* of stochastic dominance is composite and hence the tests based on the approximation of the least favorable case are not asymptotically similar on this boundary. 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.

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 (2003), and the multivariate normal with exchangeable and correlated prospects as in Klecan et al. (1991). Optimal choice of the subsample size is rather like choosing the bandwidth in nonparametric estimation. 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.

In section 2 we discuss the various concepts of stochastic dominance, while in section 3 we introduce our test statistics. In section 4 we give its asymptotic null distribution, while in section 5 we define our subsampling procedure and obtain its asymptotic properties. In section 6 we describe a full sample bootstrap approach to obtaining critical values and compare the theoretical properties of the two resampling methods. In section 7 we report the results of some simulations and present an application. Proofs are contained in the appendix.

2 Stochastic Dominance

The following 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). Let $F_1(x)$ 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.

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

Weak orders of SD obtain by eliminating the requirement of strict inequality at some point. When dominance is not present, any strong ordering by specific *indices* that correspond to specific utility functions in $\mathcal{U}_1$ and $\mathcal{U}_2$, will not enjoy general acceptance.

Our methods are applicable to higher orders of dominance. 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 any SD order and tests. Define $D_k^{(s)}(x) = \int_{-\infty}^{x} D_k^{(s-1)}(t)dt$, $k = 1, 2$, where $D_k^{(1)}(x) = F_k(x)$. We say that X_1 Stochastically Dominates X_2 at order s, if $D_1^{(s)}(x) \leq D_2^{(s)}(x)$ for all x with strict inequality for some x.

In this paper we shall also consider the concept of prospect stochastic dominance (PSD). 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.[3] 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

[3] 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.

losses but risk aversion for gains. This leads naturally to the definition [c.f. Levy and Wiener (1998, Theorem 4)]

Definition 3 *X_1 Prospect Stochastic Dominates X_2, denoted $X_1 \succeq_{PSD} X_2$, if and only if either*

(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).

Now consider the second component of prospect theory, (b), the transformation of probabilities. 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 certain transforming of the probabilities.

Finally, Levy and Levy (2002) also discuss the concept of Markowitz Stochastic Dominance (MSD). In this case individuals rank outcomes according to the expected value of reverse S-shaped utility functions $u \in \mathcal{U}_M \subseteq \mathcal{U}_1$ for which $u''(x) \geq 0$ for all $x > 0$ but $u''(x) \leq 0$ for all $x < 0$. These properties represent risk seeking for gains but risk aversion for losses. Levy and Levy (2002, p1339) show that $X_1 \succeq_{MSD} X_2$ when $\left(\int_{-\infty}^{y} + \int_{x}^{\infty}\right) F_1(t)dt \leq \left(\int_{-\infty}^{y} + \int_{x}^{\infty}\right) F_2(t)dt$ for all pairs (x, y) with $x > 0$ and $y < 0$ with strict inequality for some (x, y). As Levy and Levy (2002, p1340) discuss, MSD is not exactly the opposite of PSD. However, when the outcomes have a common mean they are opposites: if $X_1 \succeq_{PSD} X_2$, then $X_2 \succeq_{MSD} X_1$ in that case.

3 The Test Statistics

Suppose 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 "residual" dominance, we suppose that $\{X_{ki} : i = 1, \ldots, N\}$ might depend on an unknown finite dimensional parameter $\theta_{k0} \in \Theta_k \subset \mathbb{R}^{L_k}$:

$$X_{ki} = Y_{ki} - Z_{ki}^{\top}\theta_{k0}, \tag{1}$$

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

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

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 $X_{ki}(\theta) = Y_{ki} - Z_{ki}^{\top}\theta$, $X_{ki} = X_{ki}(\theta_{k0})$, 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

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

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 $(X_1, \ldots, X_k)^\top$. Define

$$D_k^{(1)}(x, \theta) = F_k(x, \theta) \text{ and} \tag{3}$$
$$D_k^{(s)}(x, \theta) = \int_{-\infty}^{x} D_k^{(s-1)}(t, \theta) dt \text{ for } s = 2, 3, \ldots. \tag{4}$$

With some abuse of notation, for $s \geq 1$, let $D_{kl}^{(s)}(x, \theta) = D_k^{(s)}(x, \theta) - D_l^{(s)}(x, \theta)$, $D_k^{(s)}(x) = D_k^{(s)}(x, \theta_{k0})$, and $D_{kl}^{(s)}(x) = D_{kl}^{(s)}(x, \theta_{k0})$. Now, for a given integer $s \geq 1$, define the following functionals of the distribution functions:

$$d_s^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} \left[D_{kl}^{(s)}(x)\right] \tag{5}$$
$$p^* = \min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} \left[D_{kl}^{(2)}(x) - D_{kl}^{(2)}(y)\right], \tag{6}$$

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:

$$\mathbf{H}_0^d : d_s^* \leq 0 \; vs. \; \mathbf{H}_1^d : d_s^* > 0, \tag{7}$$

$$\mathbf{H}_0^p : p^* \leq 0 \; vs. \; \mathbf{H}_1^p : p^* > 0. \tag{8}$$

The null hypothesis $\mathbf{H}_0^d$ implies that the prospects in $\mathcal{A}$ are not s–th degree stochastically maximal, i.e., there exists at least one prospect in $\mathcal{A}$ which s–th degree stochastically dominates the others. Likewise for the prospect stochastic dominance test. It is also sometimes of interest to test the related hypothesis that a particular outcome k dominates all other outcomes: in this case, we merely replace 'the minimum over all k, l with $k \neq l$' in (5) and (6) by 'the minimum over all l with $k \neq l$'.

The test statistics we consider are based on the empirical analogues of (5) - (6). That is, they are defined to be:

$$D_N^{(s)} = \min_{k \neq l} \sup_{x \in \mathcal{X}} \sqrt{N} \left[\overline{D}_{kl}^{(s)}(x, \widehat{\theta}_k) \right]$$
$$P_N = \min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} \sqrt{N} \left[\overline{D}_{kl}^{(2)}(x, \widehat{\theta}_k) - \overline{D}_{kl}^{(2)}(y, \widehat{\theta}_k) \right],$$

where:

$$\begin{aligned} \overline{D}_k^{(1)}(x, \theta) &= \overline{F}_{kN}(x, \theta) \\ \overline{D}_k^{(s)}(x, \theta) &= \int_{-\infty}^{x} \overline{D}_k^{(s-1)}(t, \theta) dt \text{ for } s \geq 2 \\ \overline{D}_{kl}^{(s)}(x, \theta) &= \overline{D}_k^{(s)}(x, \theta) - \overline{D}_l^{(s)}(x, \theta) \text{ for } s \geq 1. \end{aligned}$$

For $s = 1$ and 2, we note that $D_N^{(s)}$ is the same as the Klecan et al. (1991)'s test statistic, except that we have allowed the prospects to have been estimated from the data and stochastic dominance of any pre-specified order.

We next discuss the issue of how to compute $D_N^{(s)}$ 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 $D_N^{(1)}$ can be (exactly) replaced by a maximum taken over all the distinct points in the combined sample. Regarding the computation of $D_N^{(s)}$ for $s \geq 2$, Klecan et al. (1991) propose a recursive algorithm for exact computation of $D_N^{(2)}$ and Barrett and Donald (2003) propose an extension to $D_N^{(3)}$, see also Davidson and Duclos (1999). Integrating by parts, we have

$$D_k^{(s)}(x) = \frac{1}{(s-1)!} \int_{-\infty}^{x} (x - t)^{s-1} dF_k(t),$$

which holds for all x provided $E \left\| X_k \right\|^{s-1} < \infty$ for $s \geq 1$. Therefore, it can be computed by its empirical analogue

$$\overline{D}_k^{(s)}(x, \theta) = \frac{1}{N(s-1)!} \sum_{i=1}^{N} (x - X_{ki}(\theta))^{s-1} 1 \left(X_{ki}(\theta) \leq x \right)$$

for $s \geq 1$. To reduce the computation time, it may be preferable to compute approximations to the suprema in $D_N^{(s)}$ and 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

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

where $\mathcal{X}_J \subset \mathbb{R}_+$. Theoretically, provided the set of evaluation points becomes dense in the joint support, the distribution theory is unaffected by using this approximation.

4 Asymptotic Null Distributions

4.1 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 $k = 1, \ldots, K$ and for all sequence of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$, $\sup_{x \in \mathcal{X}} \sup_{\theta : \|\theta - \theta_{k0}\| \leq \xi_N} ||(\partial/\partial\theta) D_k^{(s)}(x, \theta) - \Delta_{k0}^{(s)}(x)|| \to 0$, where $\Delta_{k0}^{(s)}(x) = (\partial/\partial\theta) D_k^{(s)}(x, \theta_{k0})$; (iii) $\sup_{x \in \mathcal{X}} ||\Delta_{k0}^{(s)}(x)|| < \infty$ for $k = 1, \ldots, K$.

For the tests $D_N^{(s)}$ (for $s \geq 2$) and P_N we need the following modification of Assumption 1:

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 $A > \max\{(s-1)rq/(r-q), 1 + 2/\delta\}$ 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) $E\,\|Z_{ki}\|^{(s-1)r} < \infty$ for $k = 1, \ldots, K$, for all $i \geq 1$.

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 $D_N^{(s)}$ for $s \geq 2$ (or P_N). Without the estimated parameters, weaker conditions on the dependence can be assumed.

2. Assumption 3 shows that, for higher order stochastic dominance, we need less smoothness of the distribution functions. Intuitively, this is true because integration is a smoothing operation in general.

3. When there are no estimated parameters, Assumptions 2 and 3 and the moment conditions in Assumptions 1(ii) and 1*(ii) are redundant.

4.2 The Null Distributions

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

To help understanding of the reader, we first introduce a heuristic argument for the test $D_N^{(1)}$ in the simplest setting, in which 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 $d_1^* = 0$. 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 $D_N^{(1)} = \min\{\sup_{x\in\mathcal{X}}[\overline{A}_N(x)], \sup_{x\in\mathcal{X}}[-\overline{A}_N(x)]\}$ 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 $D_N^{(1)}$ will diverge to minus infinity.

We now turn to our general setting and make the above heuristic statement more rigorous. For a given integer $s \geq 1$, define the empirical processes in x (and y) and θ to be:

$$\begin{aligned} \nu_{kN}^{(s)}(x,\theta) &= \sqrt{N}\left[\overline{D}_k^{(s)}(x,\theta) - D_k^{(s)}(x,\theta)\right] \\ \nu_{kN}^{p}(x,y,\theta) &= \nu_{kN}^{(2)}(x,\theta) - \nu_{kN}^{(2)}(y,\theta). \end{aligned} \tag{9}$$

Let $(\ \widetilde{d}_{kl}^{(s)}(\cdot) \quad \nu_{k0}^\top \quad \nu_{l0}^\top\ )^\top$ be a mean zero Gaussian process with covariance function given by

$$C_{kl}^{(s)}(x_1,x_2) = \lim_{N\to\infty} E \begin{pmatrix} \nu_{kN}^{(s)}(x_1,\theta_{k0}) - \nu_{lN}^{(s)}(x_1,\theta_{l0}) \\ \sqrt{N}\psi_{kN}(\theta_{k0}) \\ \sqrt{N}\psi_{lN}(\theta_{l0}) \end{pmatrix} \begin{pmatrix} \nu_{kN}^{(s)}(x_2,\theta_{k0}) - \nu_{lN}^{(s)}(x_2,\theta_{l0}) \\ \sqrt{N}\psi_{kN}(\theta_{k0}) \\ \sqrt{N}\psi_{lN}(\theta_{l0}) \end{pmatrix}^\top, \tag{10}$$

where $\overline{\psi}_{kN}(\theta_{k0}) = (1/N)\sum_{i=1}^{N}\psi_k(Y_{ki}, Z_{ki}, \theta_{k0})$ for all $1 \leq k \leq K$. We analogously define $(\ \widetilde{p}_{kl}(\cdot,\cdot)\quad \nu_{k0}^{\top}\quad \nu_{l0}^{\top}\ )^{\top}$ to be a mean zero Gaussian process with covariance function $C_{kl}^{p}(x_1, y_1, x_2, y_2)$, which is equal to (10) except that $\nu_{jN}^{(s)}(x_i, \theta_{j0})$ is replaced by $\nu_{jN}^{p}(x_i, y_i, \theta_{j0})$ for $j = k, l$ and $i = 1, 2$. The limiting null distributions of our test statistics are given in the following theorem.

Theorem 1. (a) *Suppose that Assumptions 1-3 hold when $s = 1$ and Assumptions 1*, 2 and 3 hold when $s \geq 2$. Then, under the null $\mathbf{H}_0^d$, we have*

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

where $\mathcal{I}^{(s)} = \{(k,l)|k \neq l,\ \sup_{x\in\mathcal{X}}[D_k^{(s)}(x) - D_l^{(s)}(x)] = 0\}$ and $\mathcal{B}_{kl}^{(s)} = \{x \in \mathcal{X} : D_k^{(s)}(x) = D_l^{(s)}(x)\}$.

(b) Suppose that Assumptions 1, 2 and 3 hold with $s = 2$. Then, under the null $\mathbf{H}_0^p$, we have*

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

where $\mathcal{I}^{p} = \{(k,l)|k \neq l,\ \sup_{x,-y\in\mathcal{X}_+}[D_{kl}^{(2)}(x) - D_{kl}^{(2)}(y)] = 0\}$, $\mathcal{B}_{kl}^{p} = \{(x,y) : x \in \mathcal{X}_+, -y \in \mathcal{X}_+$ *and* $D_{kl}^{(2)}(x) = D_{kl}^{(2)}(y)\}$, and $\Xi_{k0}(x,y) = \Delta_{k0}^{(2)}(x) - \Delta_{k0}^{(2)}(y)$.

Theorem 1 shows that the asymptotic null distribution of $D_N^{(s)}$ (P_N) is non-degenerate at the boundary $d_s^* = 0$ ($p^* = 0$) of the null hypothesis and depends on the "true" parameters $\{\theta_{k0} : k = 1, \ldots, K\}$ and the full joint distribution function F of $\{X_{ki} : k = 1, \ldots, K\}$. The latter implies that the asymptotic critical values for $D_N^{(s)}$ and P_N can not be tabulated once and for all. However, we define below various procedures to estimate them from the data.

5 Critical Values by Subsampling

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.

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 first define the subsampling procedure. Let $W_i = \{(Y_{ki}, Z_{ki}) : k = 1, \ldots, K\}$ for $i = 1, \ldots, N$ and T_N denote our test statistic $D_N^{(s)}$ or P_N. With some abuse of notation, the test statistic T_N can be re-written as a function of the data $\{W_i : i = 1, \ldots, N\}$:

$$T_N = \sqrt{N} t_N(W_1, \ldots, W_N),$$

where $t_N(W_1, \ldots, W_N)$ is given by $\min_{k \neq l} \sup_{x \in \mathcal{X}} [\overline{D}_{kl}^{(s)}(x, \widehat{\theta}_k)]$ for $T_N = D_N^{(s)}$ and $\min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} [\overline{D}_{kl}^{(2)}(x, \widehat{\theta}_k) - \overline{D}_{kl}^{(2)}(y, \widehat{\theta}_k)]$ for $T_N = P_N$. Let

$$G_N(w) = P\left(\sqrt{N} t_N(W_1, \ldots, W_N) \leq w\right) \tag{11}$$

denote the distribution function of T_N. Let $t_{N,b,i}$ be equal to the statistic t_b evaluated at the subsample $\{W_i, \ldots, W_{i+b-1}\}$ of size b, i.e.,

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

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 T_N using the distribution of the values of $t_{N,b,i}$ computed over $N - b + 1$ different subsamples of size b. That is, we approximate the sampling distribution G_N of T_N by

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

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 $T_N > g_{N,b}(1 - \alpha)$. 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.[4]

[4] In Barrett and Donald (2003) the outcomes have different sample sizes, say M and N with $M < N$ ($K = 2$), and the data were mutually independent and independent over i. In that case, the subsampling algorithm can be applied to the two separate series with subsample sizes $\widehat{b}_M$ and $\widehat{b}_N$ respectively. Alternatively, one can use subsample $\widehat{b}_M$ for the vector sample of size M, and then apply the subsampling algorithm with $\widehat{b}_{N-M}$ to the incremental scalar sample. This latter approach works also when there is dependence amongst the outcomes in the common sample.

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 2-4 hold. In addition, suppose that Assumption 1 (when $s = 1$) or Assumption 1*(when $s \geq 2$) holds if $T_N = D_N^{(s)}$, and Assumption 1* with $s = 2$ holds if $T_N = P_N$. Then, under the null hypothesis $\mathbf{H}_0^d$ ($\mathbf{H}_0^p$), we have that*

$$(a)\ g_{N,\widehat{b}_N}(1-\alpha) \xrightarrow{p} \begin{cases} g(1-\alpha) & \text{if } d_s^* = 0\ (p^* = 0) \\ -\infty & \text{if } d_s^* < 0\ (p^* < 0) \end{cases}$$

$$(b)\ P[T_N > g_{N,\widehat{b}_N}(1-\alpha)] \to \begin{cases} \alpha & \text{if } d_s^* = 0\ (p^* = 0) \\ 0 & \text{if } d_s^* < 0\ (p^* < 0) \end{cases}$$

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

Remarks.

1. When $d_s^* = 0$ (or $p^* = 0$), Theorem 2 shows that the level α critical value $g_{N,\widehat{b}_N}(1-\alpha)$ obtained from the subsampling distribution of the test statistic $T_{\widehat{b}_N}$ converges to the critical value $g(1-\alpha)$ from the limit distribution of T_N. This suggests that, at the *boundary* of the null hypothesis, the asymptotic significance level of the test T_N using the subsample critical value is α as desired.

2. When $d_s^* < 0$ (or $p^* < 0$), i.e., in the *interior* of the null hypothesis, Theorem 2 suggests that the type I error of the test T_N (based on the subsample critical value $g_{N,\widehat{b}_N}(1-\alpha)$) is asymptotically zero. This holds because both T_N and $g_{N,\widehat{b}_N}(1-\alpha)$ diverge to $-\infty$ as $N \to \infty$ in this case, but the rate of divergence is faster for T_N than for $g_{N,\widehat{b}_N}(1-\alpha)$ as long as $\widehat{b}_N$ goes to infinity at a slower rate than N, i.e., when Assumption 4 holds, see proof of Theorem 2 for details.

5.1 Asymptotic Power Properties

In this section, we investigate power properties of our tests. To help the reader understand why the subsampling test has non-trivial power against fixed and local alternatives, we first discuss a simple testing problem: Let $\{X_1, \ldots, X_N\}$ be a random sample from $N(\mu, 1)$ and the null and alternative hypotheses of interest are given by $\mathbf{H}_0 : \mu = 0$ and $\mathbf{H}_1 : \mu > 0$ respectively. Consider the t-test

statistic $T_N = \sqrt{N}\overline{X}_N$, which satisfies $T_N \Rightarrow N(0,1)$ as $N \to \infty$ under $\mathbf{H}_0$. Let $g_{N,b}(1-\alpha)$ be the subsample critical value, i.e., the $(1-\alpha)$-th quantile of the subsampling distribution of $T_b = \sqrt{b}\overline{X}_b$, where b denotes the subsample size (that satisfies $b \to \infty$ and $b/N \to 0$ as $N \to \infty$). Note that $T_b \Rightarrow N(0,1)$ as $b \to \infty$ under $\mathbf{H}_0$. Clearly, the test (that rejects $\mathbf{H}_0$ if $T_N > g_{N,b}(1-\alpha)$) has asymptotically correct size α. Now, suppose that the alternative hypothesis $\mathbf{H}_1$ is true. Then, both T_N and $g_{N,b}(1-\alpha)$ diverge (in probability) to ∞ but the latter diverges at a slower rate than the former, so that the test would reject $\mathbf{H}_0$ with high probability for N large. More specifically, note that, under $\mathbf{H}_1$, both X_N and $\overline{X}_b$ converges (in probability) to μ (>0) as $N, b \to \infty$ and hence

$$\begin{aligned} P\left(T_N > g_{N,b}(1-\alpha)\right) &= P\left(\sqrt{N/b}\overline{X}_N > g_{N,b}(1-\alpha)/\sqrt{b}\right) \\ &= P\left(\sqrt{N/b}\mu > \mu\right) + o(1) \to 1, \end{aligned}$$

where the latter convergence holds since $\underline{\lim}_{N\to\infty}(N/b) > 1$. This establishes that the subsampling test is consistent against $\mathbf{H}_1$. On the other hand, consider a sequence of local alternatives $\mathbf{H}_a : \mu(= \mu_N) = \delta/\sqrt{N}$, where $\delta > 0$. Under $\mathbf{H}_a$, we have $T_N \Rightarrow N(\delta, 1)$, while $T_b = \sqrt{b}(\overline{X}_b - \mu_N) + (b/N)^{1/2}\delta \Rightarrow N(0,1)$ since $b/N \to 0$. This implies that

$$P\left(T_N > g_{N,b}(1-\alpha)\right) \to P\left(N(\delta,1) > z_{1-\alpha}\right) > \alpha,$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-th quantile of the standard normal distribution. This establishes that the test has the same first order non-trivial local power as the test based on the normal critical values, and is asymptotically locally unbiased as desired.

We now come back to our tests of stochastic dominance. We first establish that the test $D_N^{(s)}$ (P_N) is consistent against the fixed alternative hypothesis $\mathbf{H}_1^d$ $(\mathbf{H}_1^p)$. As in the previous section, we shall let T_N denote the test statistic $D_N^{(s)}$ (or P_N).

Theorem 3. *Suppose that the assumptions in Theorem 2 hold. Then, under the alternative hypothesis* $\mathbf{H}_1^d$ (or $\mathbf{H}_1^p$), *we have*

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

Next, we determine the power of the test T_N against a sequence of contiguous alternatives converging to the boundary $d_s^* = 0$ (or $p^* = 0$) of the null hypothesis at the rate $N^{-1/2}$. To this end, we allow the distribution functions to depend on N under the local alternatives and denote $F_{kN}(x,\theta) = P(X_i(\theta) \le x)$. Also, define

$$\begin{aligned} D_{kN}^{(1)}(x,\theta) &= F_{kN}(x,\theta) \text{ and} \\ D_{kN}^{(s)}(x,\theta) &= \int_{-\infty}^{x} D_{kN}^{(s-1)}(t,\theta)dt \text{ for } s = 2,3,\ldots. \end{aligned}$$

and let $F_{k,N}(x,\theta_{k0}) = F_{kN}(x)$. For the test $T_N = D_N^{(s)}$, we consider the following sequence of local alternatives:

$$D_{kN}^{(s)}(x) = D_k^{(s)}(x) + \frac{\delta_k^{(s)}(x)}{\sqrt{N}} \text{ for } 1 \leq k \leq K; N \geq 1, \tag{12}$$

where $\delta_k^{(s)}(\cdot)$ is a real function satisfying $\min_{(k,l)\in\mathcal{I}^{(s)}} \sup_{x\in\mathcal{B}_{kl}^{(s)}} [\delta_k^{(s)}(x) - \delta_l^{(s)}(x)] > 0$ and $D_k^{(s)}(\cdot)$ satisfies $d_s^* = 0$, see Theorem 1 for the definition of $\mathcal{I}^{(s)}$ and $\mathcal{B}_{kl}^{(s)}$. (The latter restriction ensures that the alternative functionals (12) shrink to the functionals satisfying the *null* restriction asymptotically.) Likewise, for the test $T_N = P_N$, the local alternatives are defined to be

$$D_{kN}^{(2)}(x) = D_k^{(2)}(x) + \frac{\delta_k^{p}(x)}{\sqrt{N}} \text{ for } 1 \leq k \leq K; N \geq 1, \tag{13}$$

where $\delta_k^p(\cdot)$ is a real function with $\min_{(k,l)\in\mathcal{I}^p} \sup_{(x,y)\in\mathcal{B}_{kl}^p} [(\delta_k^p(x) - \delta_l^p(x)) - (\delta_k^p(y) - \delta_l^p(y))] > 0$ and $D_k^{(2)}(\cdot)$ satisfies $p^* = 0$.

To analyze the asymptotic behavior of the test under local alternatives, we need to modify the assumptions in Section 4.1. 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) $\sup_{N\geq 1} E\left\|Z_{ki}\right\|^2 < \infty$ 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) $\sup_{N\geq 1} E\left\|\psi_k(Y_{ki}, Z_{ki}, \theta_{k0})\right\|^{2+\delta} < \infty$ 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 sequences of positive constants $\{\xi_N : N \geq 1\}$ such that $\xi_N \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\theta:\|\theta-\theta_{k0}\|\leq\xi_N} ||\ (\partial/\partial\theta)D_{kN}^{(s)}(x, \theta) - \Delta_{k0}^{(s)}(x)|| \to 0$ for $k = 1, \ldots, K$, where $\Delta_{k0}^{(s)}(x)$ is as in Assumption 3(ii); (iii) $\sup_{x\in\mathcal{X}} ||\Delta_{k0}^{(s)}(x)|| < \infty$ for $k = 1, \ldots, K$.

For the tests $D_N^{(s)}$ (for $s \geq 2$) and P_N we need to modify Assumption 1-lc:

Assumption 1*-lc: (i) $\{(Y_{Nki}, Z_{Nki}) =: (Y_{ki}, Z_{ki}) : i = 1, \ldots, N\}$ is a strictly stationary and α-mixing array with $\alpha(m) = O(m^{-A})$ for some $A > \max\{(s-1)rq/(r-q),\ 1 + 2/\delta\}$ 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) $\sup_{N\geq 1} E\left\|Z_{ki}\right\|^{(s-1)r} < \infty$ for $k = 1, \ldots, K$, for all $i \geq 1$.

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. Now, the asymptotic distributions of $D_N^{(s)}$ and P_N under the local alternatives are given in the following theorem:

Theorem 4. *(a) Suppose that Assumptions 1-lc, 2-lc and 3-lc hold when $s = 1$ and Assumptions 1*-lc, 2-lc and 3-lc hold when $s \geq 2$. Then, under the local alternatives, we have*

$$D_N^{(s)} \Rightarrow L_D^{(s)}, \quad \textit{where:}$$

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

$\mathcal{I}^{(s)}$, $\mathcal{B}_{kl}^{(s)}$ and $\Delta_{k0}^{(s)}(x)$ are defined as in Theorem 1 and $(\ \widetilde{d}_{kl}^{(s)}(\cdot),\ \ \nu_{k0}^\top,\ \ \nu_{l0}^\top\ )^\top$ is the Gaussian process defined in Section 4.2.

(b) Suppose that Assumptions 1-lc, 2-lc and 3-lc hold with $s = 2$. Then, under the local alternatives, we have*

$$P_N \Rightarrow L_P, \quad \textit{where:}$$

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

$\mathcal{I}^p$, $\mathcal{B}_{kl}^p$ and $\Xi_{k0}(x,y)$ are defined as in Theorem 1 and $(\ \widetilde{p}_{kl}(\cdot,\cdot),\ \ \nu_{k0}^\top,\ \ \nu_{l0}^\top\ )^\top$ is the Gaussian process defined in Section 4.2.

Theorem 4 implies that asymptotic local power of our tests based on the subsample critical value is given by the following Corollary:

Corollary 5. *Suppose that the Assumptions in Theorem 4 hold. Then, under the local alternatives, we have*

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

as $N \to \infty$, where $T_N = D_N^{(s)}$ (or P_N), $L = L_D^{(s)}$ (or L_P), and $g_{N,\widehat{b}_N}(1-\alpha)$ and $g(1-\alpha)$ are defined as in Theorem 2.

Remarks.

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

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

under the local alternatives. When $T_N = D_N^{(s)}$, for example, this follows because we have

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

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

by Anderson's lemma (e.g., see Bickel et al. (1993, p.466)) and the left-hand-sides of (14) and (15) are equal by Theorem 4. A similar result also applies to the test $T_N = P_N$ using an analogous argument.

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

5.2 Choice of Subsample Size

In practice, the choice of b is important but rather difficult. Subsample size can be interpreted as a sort of smoothing parameter, like a bandwidth in kernel regression, except that it is integer valued and increases with sample size. Although a lot has been written on bandwidth choice for density and regression estimation by kernel methods, most of that theory is not relevant here for two reasons. First, b does not affect the first order distribution of the test under either null or local alternatives, and so its effect is second order. Second, in the testing application the objectives of size and power are often in conflict: tests that have good size [i.e., null rejection frequency close to the limiting value] tend to have poor power and vice versa. Fan and Linton (2003) have characterized this trade-off explicitly using higher order expansions in the context of a test of a parametric null regression against nonparametric alternatives. Unless the practitioner is willing to specify a preference function over the two conflicting objectives there is no unique best amount of smoothing. Nevertheless, there have been various proposals in the literature. Politis, Romano, and Wolf (1999) discuss various methods for selecting subsample size in the context of estimation and testing problems. 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.

We propose a number of practical criteria for choosing b and investigate below how well they do in practice. Let

$$B_N = \{b_{N1} < b_{N2} < \cdots < b_{Nr_N};\ b_j \text{ integers less than } N\}$$

be a set of candidate subsample sizes. We can allow B_N to be a very large set, including almost all, but not all, of $\{1,\ldots,N\}$. Specifically, we suppose that $b_{N1} \xrightarrow{p} \infty$ and $b_{Nr_N}/N \xrightarrow{p} 0$. Therefore,

r_N, the total number of elements of B_N, is allowed to increase to infinity at any rate slower than N. For example, we could take B_N to include all integers between $\log\log N$ and $N/\log\log N$, which is a very wide range [in practice, one may want to consider a coarser grid to save on computational time]. For comparison, many results in the literature consider such sets with more limited range: Horowitz and Spokoiny (2001) also in a testing situation consider a set of bandwidths $H_n = [H_{\min}, H_{\max}]$ with $H_{\min} \geq n^{-\gamma}$ with $0 < \gamma < 1/3$ [for the scalar case]; see also Härdle and Marron (1985).

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:

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

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

and reject when $T_N > \overline{g}_N(1-\alpha)$ in the first case and reject when $T_N > g_N^{Med}(1-\alpha)$ in the second case.[5] 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.

6 Critical Values by Recentered Bootstrap

We next define an alternative to our subsampling procedure based on full-sample bootstrap applied to a recentered test statistic. When the data are mutually dependent but independent over time, the following bootstrap procedure provides consistent critical values. Let $\widehat{\varepsilon}_{ki} = Y_{ki} - \widehat{\mu}_{k0} - Z_{ki}^{\top}\widehat{\theta}_k$

[5] 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 $b \in B_N$.

denote the residual computed using the original sample $\mathcal{W} = \{W_i : i = 1, \ldots, N\}$, where $W_i = \{(Y_{ki}, Z_{ki}) : k = 1, \ldots, K\}$. Let $\varepsilon_i^* \equiv (\varepsilon_{1i}^*, \ldots, \varepsilon_{Ki}^*)'$ for $i = 1, \ldots, N$ be the bootstrap residual vector drawn randomly *with replacement* from the empirical *joint* distribution of centered residual vectors $\{\widehat{\varepsilon}_i^c = (\widehat{\varepsilon}_{1i} - \overline{\widehat{\varepsilon}}_1, \ldots, \widehat{\varepsilon}_{Ki} - \overline{\widehat{\varepsilon}}_K)' : i = 1, \ldots, N\}$, where $\overline{\widehat{\varepsilon}}_k = \sum_{i=1}^N \widehat{\varepsilon}_{ki}/N$.[6] Drawing ε_i^* as a vector will enable the bootstrap sample to preserve the general mutual dependence that may exist in the original sample. Then compute $Y_{ki}^* = \widehat{\mu}_{k0} + Z_{ki}^\top \widehat{\theta}_k + \varepsilon_{ki}^*$. Using the bootstrap sample $\mathcal{W}^* = \{W_i^* : i = 1, \ldots, N\}$, 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.[7] Define $X_{ki}^*(\theta) = Y_{ki}^* - Z_{ki}^\top \theta$ and let

$$\begin{aligned}
\overline{D}_k^{*(1)}(x, \theta) &= \overline{F}_{kN}^*(x, \theta) = \frac{1}{N} \sum_{i=1}^N 1\left(X_{ki}^*(\theta) \leq x\right) \\
\overline{D}_k^{*(s)}(x, \theta) &= \int_{-\infty}^x \overline{D}_k^{*(s-1)}(t, \theta) dt \text{ for } s \geq 2 \\
\overline{D}_{kl}^{*(s)}(x, \theta) &= \overline{D}_k^{*(s)}(x, \theta) - \overline{D}_l^{*(s)}(x, \theta) \text{ for } s \geq 1.
\end{aligned}$$

for $k, l = 1, \ldots, K$. Define the recentered statistics

$$\begin{aligned}
\overline{D}_{kc}^{*(s)}(x) &= \overline{D}_k^{*(s)}(x, \widehat{\theta}_k^*) - E^* \overline{D}_k^{*(s)}(x, \widehat{\theta}_k) \\
\overline{D}_{klc}^{*(s)}(x) &= \overline{D}_{kc}^{*(s)}(x) - \overline{D}_{lc}^{*(s)}(x).
\end{aligned} \tag{18}$$

(Here and in the following discussion, $E^*(\cdot)$ denotes the expectation relative to the distribution of the bootstrap sample $\mathcal{W}^*$ conditional on the original sample $\mathcal{W}$.) In the case of the independent bootstrap sampling as above, we have $E^* \overline{D}_k^{*(s)}(x, \widehat{\theta}_k) = \overline{D}_k^{(s)}(x, \widehat{\theta}_k)$. The centred *bootstrap test statistics* are defined by

$$D_N^{*(s)} = \min_{k \neq l} \sup_{x \in \mathcal{X}} \sqrt{N} \left[\overline{D}_{klc}^{*(s)}(x)\right] \tag{19}$$

$$P_N^* = \min_{k \neq l} \sup_{x, -y \in \mathcal{X}_+} \sqrt{N} \left[\overline{D}_{klc}^{*(2)}(x) - \overline{D}_{klc}^{*(2)}(y)\right], \tag{20}$$

We then compute the *bootstrap distribution* of $T_N = D_N^{*(s)}$ (or P_N^*) conditional on the original sample and take the critical value from this distribution. That is, we approximate the sampling distribution

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

[7] 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 $\{(X_{1i}, \ldots, X_{Ki}) : i = 1, \ldots, N\}$.

H_N of T_N by

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

where $T^*_{N,i}$ is the value of $T_N = D^{*(s)}_N$ (or P^*_N) computed from the i-th bootstrap sample and M is the number of the bootstrap samples. Let $H_N(1-\alpha)$ denote the $(1-\alpha)$-th sample quantile of $\widehat{H}_N(\cdot)$, 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 $T_N > H_N(1-\alpha)$.

This procedure can be shown to satisfy Theorems 2 and 3 in this special case of i.i.d. sampling. We investigate the finite sample behaviour below. As we argue below, the recentering in D^*_N is crucial and is used to impose the restriction (21). 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).[8]

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 bootstrapped are the centered residuals $\{\widehat{\varepsilon}^c_i : i = 1, \ldots, N\}$. Let L denote the length of the blocks satisfying $L \propto N^\gamma$ for some $0 \leq \gamma \leq 1$. With non-overlapping blocks, block 1 is observations $\{\widehat{\varepsilon}^c_j : j = 1, \ldots, L\}$, block 2 is observations $\{\widehat{\varepsilon}^c_{L+j} : j = 1, \ldots, L\}$, and so forth. There are B different blocks, where $BL = N$. With overlapping blocks, block 1 is observations $\{\widehat{\varepsilon}^c_j : j = 1, \ldots, L\}$, block 2 is observations $\{\widehat{\varepsilon}^c_{1+j} : j = 1, \ldots, L\}$, 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. In the case of non-overlapping bootstrap, the recentering (18) may be done with $E^*\overline{D}^{*(s)}_k(x, \widehat{\theta}_k) = \overline{D}^{(s)}_k(x, \widehat{\theta}_k)$ as in the independent sampling case. However, when the overlapping block bootstrap is used, we need to recenter the statistic with

$$E^*\overline{D}^{*(s)}_k(x, \widehat{\theta}_k) = \overline{D}^{(s)}_{k,OB}(x, \widehat{\theta}_k), \text{ where:}$$

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

$$
\begin{aligned}
\overline{D}_{k,OB}^{(1)}(x,\theta) &= \frac{1}{N}\sum_{i=1}^{N} w(i,L,N)1\left(X_{ki}(\theta)\leq x\right), \\
\overline{D}_{k,OB}^{(s)}(x,\theta) &= \int_{-\infty}^{x} \overline{D}_{k,OB}^{(s-1)}(t,\theta)dt \text{ for } s\geq 2, \\
w(i,L,N) &= \begin{cases} i/L & \text{if } i\in[1,L-1] \\ 1 & \text{if } i\in[L,N-L+1] \\ (N-i+1)/L & \text{if } i\in[N-L+2,N]. \end{cases}
\end{aligned}
$$

The remaining steps are the same as in the independent case described above.[9]

6.1 Comparison between Subsampling and recentered Bootstrap

In contrast to subsampling, in the full sample recentered bootstrap one has to impose the null hypothesis in the resampling schemes. This is also true of the multiplier simulation procedure of Barrett and Donald (2003). The usual practice in the literature has been to impose the *least favorable case* where

$$F_1(x) = \cdots = F_K(x) \text{ for all } x\in\mathcal{X}. \tag{21}$$

This is easy to apply when the prospects are mutually independent and independent over time and there are no estimated parameters - you just pool the data into a common distribution and draw with replacement from that. In an innovative paper, Klecan et al. (1991) 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.[10] The recentering of the test statistic that we have made imposes (21) implicitly, thereby avoiding having to impose these nasty restrictions in the resampling algorithm. Of course, the boundary between the null and alternative hypothesis is a large and complicated set, while (21) is a much smaller and simpler set. We show below that imposing the least favorable case can have negative consequences.

In general, it reasonable to expect that the full sample method such as the bootstrap approach may be more efficient than the subsampling approach provided the former works, see e.g., Hall and

[9] 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. If the stationary bootstrap is used, the recentering is defined with $E^*\overline{D}_k^{*(s)}(x,\widehat{\theta}_k) = \overline{D}_k^{(s)}(x,\widehat{\theta}_k)$.

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

Jing (1996). However, we shall show that the bootstrap might not be very satisfactory in some situations in our testing problem. For this purpose, we first need to characterize the asymptotic distribution of the bootstrap test statistic conditional on the original sample. Instead of providing more technical details, we just briefly sketch the main ideas to derive the latter distribution. Consider the statistic $D_N^{*(s)}$ in (19) with $E^*\overline{D}_k^{*(s)}(x,\theta) = \overline{D}_k^{(s)}(x,\theta)$. By rearranging terms, we have

$$\begin{aligned}\sqrt{N}\overline{D}_{klc}^{*(s)}(x) &= \left[\nu_{kN}^{*(s)}(x,\widehat{\theta}_k^*) - \nu_{lN}^{*(s)}(x,\widehat{\theta}_k^*)\right] \\ &\quad + \left[\nu_{kN}^{(s)}(x,\widehat{\theta}_k^*) - \nu_{kN}^{(s)}(x,\widehat{\theta}_k)\right] - \left[\nu_{lN}^{(s)}(x,\widehat{\theta}_l^*) - \nu_{lN}^{(s)}(x,\widehat{\theta}_l)\right] \\ &\quad + \sqrt{N}\left[D_k^{(s)}(x,\widehat{\theta}_k^*) - D_k^{(s)}(x,\widehat{\theta}_k)\right] - \sqrt{N}\left[D_l^{(s)}(x,\widehat{\theta}_l^*) - D_l^{(s)}(x,\widehat{\theta}_l)\right],\end{aligned} \tag{22}$$

where $\nu_{kN}^{*(s)}(x,\theta) = \sqrt{N}[\overline{D}_k^{*(s)}(x,\theta) - \overline{D}_k^{(s)}(x,\theta)]$ denotes an empirical process based on the bootstrap sample $\mathcal{W}^*$ and $\nu_{kN}^{(s)}(x,\theta)$ is as defined in (9). Under suitable regularity conditions (and using the stochastic equicontinuity arguments as in Lemmas 1-2 below and consistency of $\widehat{\theta}_k^*$ for $\widehat{\theta}_k$), we may show that: Uniformly in $x \in \mathcal{X}$,

$$\begin{aligned}\sqrt{N}\overline{D}_{klc}^{*(s)}(x) &= \nu_{kN}^{*(s)}(x,\widehat{\theta}_k) - \nu_{lN}^{*(s)}(x,\widehat{\theta}_k) \\ &\quad + \Delta_{k0}^{(s)}(x)^\top \Gamma_{k0}\sqrt{N}\overline{\psi}_{kN}^*(\widehat{\theta}_k) - \Delta_{l0}^{(s)}(x)^\top \Gamma_{l0}\sqrt{N}\overline{\psi}_{lN}^*(\widehat{\theta}_l) + o_p(1)\end{aligned} \tag{23}$$

conditional on $\mathcal{W}$ with probability one, where $\overline{\psi}_{kN}^*(\theta) = (1/N)\sum_{i=1}^N \psi_k(Y_{ki}^*, Z_{ki}^*, \theta)$. We note that the recentering (18) is crucial because, without recentering, we would have an additional random term in (23) that may diverge asymptotically conditional on the original sample.[11] Using an argument analogous to the proof of Theorem 1, the representation (23) implies that $D_N^{*(s)}$ has the asymptotic distribution (conditional on $\mathcal{W}$ with probability one) given by

$$L_D^B \equiv \min_{k\neq l}\sup_{x\in\mathcal{X}}\left[\widetilde{d}_{kl}^{(s)}(x) + \Delta_{k0}^{(s)}(x)^\top\Gamma_{k0}\nu_{k0} - \Delta_{l0}^{(s)}(x)^\top\Gamma_{l0}\nu_{l0}\right] \tag{24}$$

Similarly, P_N^* has the asymptotic bootstrap distribution:

$$L_P^B \equiv \min_{k\neq l}\sup_{x,-y\in\mathcal{X}_+}\left[\widetilde{p}_{kl}(x,y) + \Xi_{k0}(x,y)^\top\Gamma_{k0}\nu_{k0} - \Xi_{l0}(x,y)^\top\Gamma_{l0}\nu_{l0}\right]. \tag{25}$$

Compare the distribution L_D^B (L_P^B) with the asymptotic null distribution given in Theorem 1(a) (1(b)). It is easy to see that the two distributions are equal when the distributions $F_k(\cdot)$ for $k = 1,\ldots,K$ are equal, i.e., when the least favorable case (21) is true. However, our test statistic $D_N^{(s)}$ (P_N) has a non-degenerate limit distribution everywhere on the boundary "$d_s^* = 0$" ("$p^* = 0$") of our null hypothesis $\mathbf{H}_0^d$ ($\mathbf{H}_0^p$). Note that "$d_s^* = 0$" (or "$p^* = 0$") is in fact a *composite* hypothesis and includes the least

[11] Essentially, the recentering has the effect of annihilating a term corresponding to the term (A.8) in the expansion (A.6) in Appendix.

favorable case (21) as a special case. Therefore, when (21) fails to hold but $d_s^* = 0$ (or $p^* = 0$") is true,[12] then the test based on the bootstrap critical value would not have asymptotic size α. In fact, this is true with any test which implicitly imposes the restriction (21), e.g., simulation-based tests or subsampling-based test using recentered statistics.[13] This implies that the bootstrap-based test (as well as the latter tests) 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 also Hansen (2001) for a similar result in a different context. Against some alternatives that are far from the boundary, however, the bootstrap test might be more powerful than the subsampling test because the former uses the full sample information.

7 Numerical Results

7.1 Simulations

We examine 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 (2003), 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 guard against the criticism of rigging of the performance evaluation; we may also compare our procedures with those of the authors' in regard to size and power, although this is not our main purpose. 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.

[12] 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.

[13] The recentred subsample method, like our uncentred subsample method, works under quite general sampling schemes. In some cases, the former might have be more powerful than the latter in small samples, because critical values from recentred statistics are generally $O_p(1)$, while those from uncentred statistics diverge at $b^{1/2}$ rate. However, the cost of recentring in our testing problem is that it makes the test not asymptotically similar on the boundary, as is true with the recentred bootstrap test.

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, 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 5. 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. We also report results for the subampling method with recentering as proposed in Chernozhukov (2002), and the uncentered full-sample bootstrap.

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 consistently works slightly better under the null hypothesis, while the subsample method frequently works better under the alternative. In 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 and significant relative to the simulation error of 0.0069 even in the larger sample. This is consistent with our theory. However, we note that in the smallest sample size, the recentered bootstrap does much better for all designs, and seems to perform adequately in many cases. The recentered subsampling method seems to have much worse size but usually has better power properties for small sample sizes. As expected the uncentered bootstrap performs terribly, almost never rejecting under either null or alternative hypotheses for the nominal 5% tests. Recentering seems essential for the full sample bootstrap but not so for the subsampling method. Regarding the automatic subsample methods, the mean critical value method seems to have the best overall performance. In comparison with the methods used by the other authors, again the result is a split decision. However, it is worth reminding the reader that these designs, especially the first two settings, favor the alternative methods which are designed for i.i.d. observations on independent or exchangeable prospects.

7.1.1 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) 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 Davidson and Duclos 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: $F^{-1}(v) = [(1-v)^{-\frac{1}{\beta}} - 1]^{\frac{1}{\alpha}}, \quad 0 \leq v < 1$. We investigated the five different Burr designs of Tse and Zhang (2000), which are given below along with the population values of d_1^*, d_2^* :

Design	X_1	X_2	d_1^*	d_2^*
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. 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 "least favorable" case of equality of the two distributions. The estimated CDFs "kiss" at 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

comparison, the Davidson and Duclos and Anderson tests reported in Tse and Zhang (2000) tend to under-reject, although not substantially: for example with $N = 1000$, their implementations vary in rejection frequency from 3.08% to 4.47%.

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 recentered 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.

7.1.2 Barrett and Donald (2003): 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 (2003) in a Monte Carlo study of the Klecan et al. (1991) tests along with some of its competitors. Let,

$$X_k = \exp(\mu_k + \sigma_k Z_k),$$

where Z_k are standard normal and mutually independent.

Design	X_1	X_2	d_1^*	d_2^*
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 Barrett and Donald (2003). The results shown in Tables 2F and 2S, cases a-d correspond exactly to cases 1,2,3, and 4 of Barrett and Donald (2003).

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 (2003), we seem to do better in some cases and worse in others. For example, in Table 2F, 2a the corresponding number for their implementation is 0.031 for n=50 and 0.044 for n=500. In Table 2S, 2a, for n=50 they have 0.032 and for n=500 they have 0.044. In Table 2S,2c they report zeros for n=50 and n=500. Generally speaking their performance is better under the null hypothesis and ours is better under the alternatives.

We note that the comparison of the automatic selection methods is similar to the previous example. There is evidence that the subsampling tests are more powerful for SSD hypotheses than the bootstrap.

7.1.3 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),

$$X_{ki} = (1-\lambda)\left[\alpha_k + \beta_k\left(\sqrt{\rho}Z_{0i} + \sqrt{1-\rho}Z_{ki}\right)\right] + \lambda X_{k,i-1}, \tag{26}$$

where (Z_{0i}, Z_{1i}, Z_{2i}) are i.i.d. standard normal random variables, mutually independent. The parameters $\lambda = \rho = 0.1$ determine the mutual correlation of X_{1i} and X_{2i} and their autocorrelation. The parameters α_k, β_k are actually the mean and standard deviation of the marginal distributions of X_{1i} and X_{2i}. This scheme produces autocorrelated and mutually dependent prospects consistent with the assumptions of Klecan et al. (1991), *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_1^*	d_2^*
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 consequential 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.

In terms of a comparison with the tests investigated in Klecan et al. (1991), we seem to do better in some cases and worse in others. In Table 3F, 3a the corresponding number for their implementation is 0.096, in 3b, it is 0.060, in 3c it is 0.95, all for n=50. In Table 3S, 3a, they have 0.020, in 3b they have 0.060, and in 3c they have 0.950, all for n=500.

7.1.4 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. (1991) 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_i of a given fund is

$$R_i = \alpha + \sum_{j=1}^{J} \beta_j F_{ji} + \varepsilon_i, \tag{27}$$

where F_{ji} is the (observed) return of the some asset class, for $j = 1, \ldots, J$, the β_j's are the factor loadings, while ε_i is an idiosyncratic disturbance term that contains the part of the fund's performance not explained by the factors. The disturbance term $u_i = \alpha + \varepsilon_i$ represents the own choice of the fund manager and is called the selectivity of the fund. It is of interest to compare the u_i 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 $F_i = Z_{0i}/(1-\lambda L)$, where L is the lag operator, be a single observed factor and let $R_{ki} = X_{ki}$ be the return on asset k, where X_{ki} are those generated in designs 3a-c. We have

$$R_{ki} = \alpha_k + \gamma_k F_i + \varepsilon_{ki}, \text{ where } \gamma_k = \beta_k\sqrt{\rho} \text{ and } \varepsilon_{ki} = \frac{\beta_k(1-\lambda)\sqrt{1-\rho}}{1-\lambda L} Z_{ki}.$$

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

Design	u_1	u_2	d_1^*	d_2^*
3Rd	$N(0, 0.7364)$	$N(-1, 11.7818)$	0.1936	$0.0000(SSD)$
3Re	$N(0, 11.7818)$	$N(1, 11.7818)$	$0.0000(FSD)$	$0.0000(SSD)$
3Rf	$N(0, 0.7364))$	$N(1, 11.7818)$	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.

7.2 Application: Daily Stock Index Returns

The SD concepts have been applied extensively in the finance literature, see for recent work Post (2002) and Abhyankar and Ho (2003). But almost all the SD testing in this area is either informal, or assumes that one or both distributions are *a priori known*. In this section, we apply 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 of these series are 0.00055 and 0.00068 respectively, while the standard deviations are 0.00908 and 0.0223, yielding Sharpe ratios of 6.1% and 3.1% respectively. The series are mutually dependent and dependent over time, as permitted by our assumptions but not by say Barrett and Donald (2003). Figure 1 plots the c.d.f.'s and integrated c.d.f. [denoted s.d.f.] of the two series over the central part of the pooled distribution. This shows that the two c.d.f.'s cross, but the s.d.f. of the Dow Jones index dominates that of the S&P500 index over this time period and this range of values.

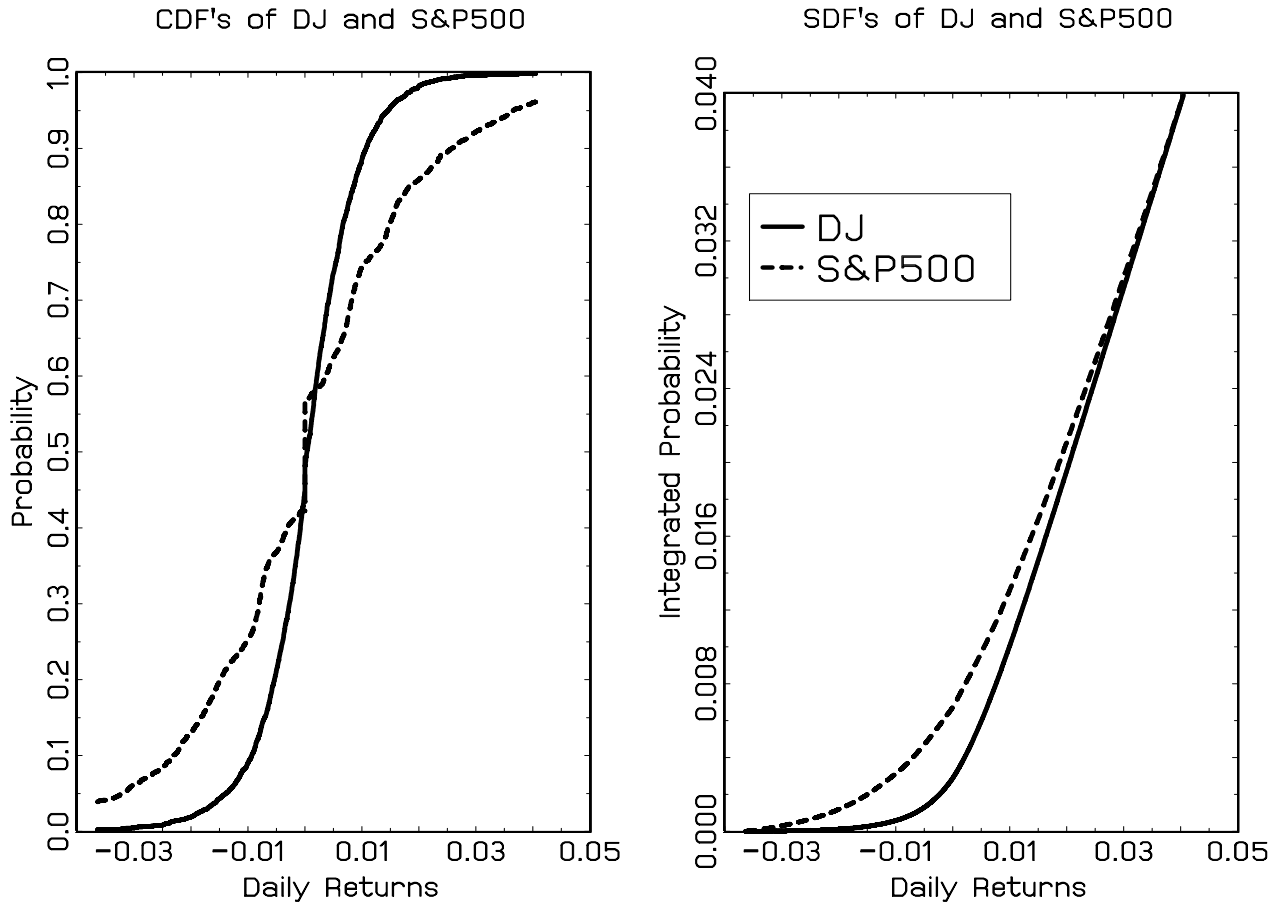

Figure 1ab

In Figure 2 we plot the surface $\int_y^x [F_{1N}(t) - F_{2N}(t)]\, dt$ 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. This can be confirmed again by looking at Figure 1a. A sufficient condition for a random variable X to prospect dominate Y is that: $\max\{X, 0\}$ first order dominates $\max\{Y, 0\}$ [risk aversion on the positive side] and that $\min\{Y, 0\}$ first order dominates $\min\{X, 0\}$ [risk seeking on the negative side]. This seems to be the case in Figure 1a.

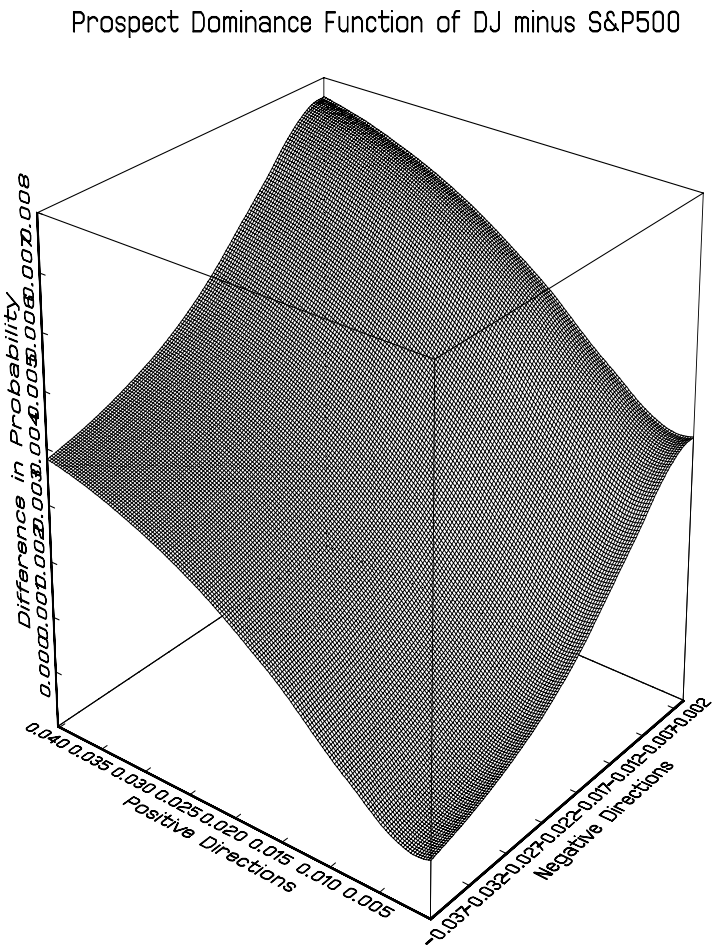

Figure 2. Shows $\int_y^x [F_{1N}(t) - F_{2N}(t)]\, dt$ against $x > 0, y < 0$.

In Figure 3 we plot the p-value of our tests of the null hypotheses $d_1^* \leq 0$, $d_2^* \leq 0$, and $p^* \leq 0$ against subsample size [we are testing maximality here because a priori it is not obvious, which if any dominance relation should exist between the series]. The results suggest strongly that the evidence is against $d_1^* \leq 0$ but in favour of $d_2^* \leq 0$ and $p^* \leq 0$.[14] Any of the automatic methods described in 5.2 would yield the same conclusion. For comparison, the recentered bootstrap p-values are 0.1448, 0.9999, and 0.9999 respectively.

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

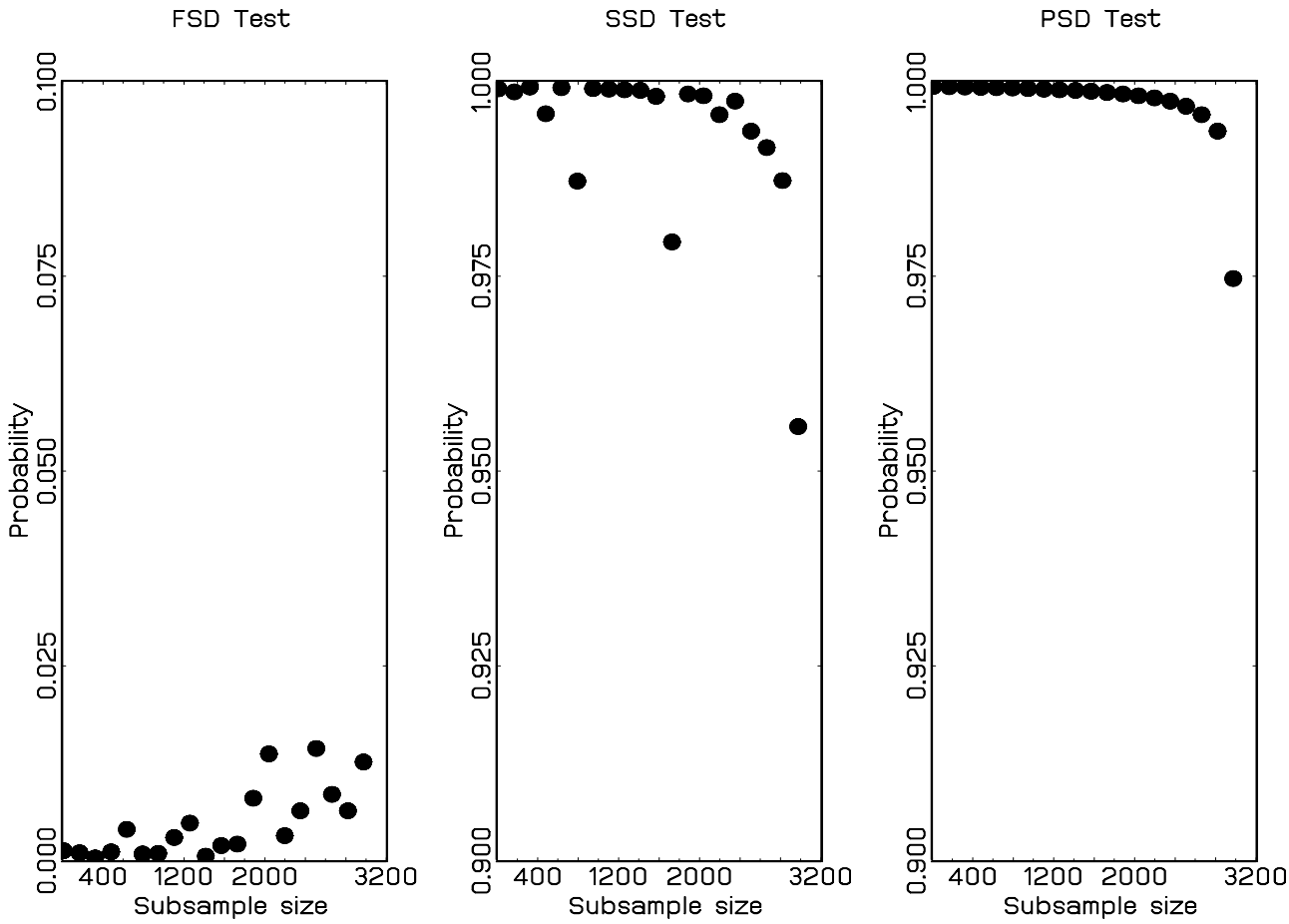

Figure 3.

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. Although we do not report it here, the ranking according to the MSD criterion agrees with that of SSD and Sharpe ratios. We refer the reader to Levy and Levy (2002) for further discussion of the comparison between MSD and PSD criteria and their appropriateness for individual behaviour.

8 Concluding Remarks

We have obtained the asymptotic distribution of well known tests for stochastic dominance of various type 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 gave theoretical reasons why the subsample method might have better power than a recentered bootstrap against some alternatives close to the null hypothesis, although there may be reasons why for alternatives far from the null hypothesis, the recentered bootstrap would have better power. Certainly, the full sample bootstrap works much better with small sample sizes and should be preferred in such cases.

We have chosen, like Klecan et al. (1991), to base our tests on the supremum criterion. However, as in testing other hypotheses about c.d.f.'s there are many other criteria that could be used, see Shorack and Wellner (1986) for a discussion. For example, the weighted supremum test based on $d_s^* = \min_{k \neq l} \sup_{x \in \mathcal{X}} [w(x) D_{kl}^{(s)}(x)]$ for some non-negative weighting function $w(x)$ is one possibility. Also, one can take distances other than supremum like $d_s^* = \min_{k \neq l} \int [\max\{0, D_{kl}^{(s)}(x)\}]^p w(x) dx$ for positive p and non-negative weighting function $w(x)$; with $p = 2$ it amounts to a sort of one-sided Cramér von Mises test. This class of criteria have been used recently in Hall and Van Keilegom (2003) in another context. The main results should carry over to these situations. Unfortunately, it is unlikely that any of these tests has a compelling advantage over any other. Although there has been much recent work on the many different notions of efficiency for these sorts of tests [excellently summarized in Nikitin (1995)], the main finding is that which one is best depends on the alternative being tested.

A Appendix

Below we sketch the proof of Theorems in the main text only for the test $D_N^{(1)}$. The corresponding proofs for the tests $D_N^{(s)}$ for $s \geq 2$ and P_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 $k = 1, \ldots, K$:

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

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

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, $\mathcal{F}_d^k$ is a class of uniformly bounded functions satisfying the L^2-continuity condition, because we have

$$\begin{aligned}
&\sup_{i\geq 1} E \sup_{\substack{(x',\theta')\in\mathcal{X}\times\Theta_k: \\ |x'-x|\leq r_1, \|\theta'-\theta\|\leq r_2, \sqrt{r_1^2+r_2^2}\leq r}} \left|1\left(X_{ki}(\theta') \leq x'\right) - 1\left(X_{ki}(\theta) \leq x\right)\right|^2 \\
&= E \sup_{\substack{(x',\theta')\in\mathcal{X}\times\Theta_k: \\ |x'-x|\leq r_1, \|\theta'-\theta\|\leq r_2, \sqrt{r_1^2+r_2^2}\leq r}} \left|1\left(X_{ki} \leq Z_{ki}^\top(\theta'-\theta_0) + x'\right) - 1\left(X_{ki} \leq Z_{ki}^\top(\theta-\theta_0) + x\right)\right|^2 \\
&\leq E1\left(\left|X_{ki} - Z_{ki}^\top(\theta-\theta_0) - x\right| \leq \|Z_{ki}\| r_1 + r_2\right) \\
&\leq C_1\left(E\|Z_{ki}\| r_1 + r_2\right) \leq C_2 r,
\end{aligned}$$

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

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

$$\sup_{x\in\mathcal{X}} \left|\nu_{kN}^{(1)}(x, \widehat{\theta}_k) - \nu_{kN}^{(1)}(x, \theta_{k0})\right| \xrightarrow{p} 0. \tag{A.3}$$

Proof of Lemma 2. Consider the pseudometric (A.2). We have

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

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* $\forall k = 1, \ldots, K$,

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

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*

$$\left(\nu_{kN}^{(1)}(\cdot,\theta_{k0}) - \nu_{lN}^{(1)}(\cdot,\theta_{l0}), \quad \sqrt{N}\psi_{kN}^{\top}(\theta_{k0}), \quad \sqrt{N}\psi_{lN}^{\top}(\theta_{l0}) \right)^{\top} \Rightarrow \left(\widetilde{d}_{kl}^{(1)}(\cdot), \quad \nu_{k0}^{\top}, \quad \nu_{l0}^{\top} \right)^{\top}$$

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

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

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 $\{\nu_{kN}^{(1)}(\cdot,\theta_{k0}) - \nu_{lN}^{(1)}(\cdot,\theta_{l0}) : N\ge 1\}$ and (iii) finite dimensional (fidi) convergence. Conditions (i) and (ii) follow from Lemma 1. We now verify condition (iii). We need to show that $(\nu_{kN}^{(1)}(x_1,\theta_{k0}) - \nu_{lN}^{(1)}(x_1,\theta_{l0}),\ldots,\nu_{kN}^{(1)}(x_J,\theta_{k0}) - \nu_{lN}^{(1)}(x_J,\theta_{l0}),$ $\sqrt{N}\psi_{kN}(\theta_{k0})^{\top},\ \sqrt{N}\psi_{lN}(\theta_{l0})^{\top})^{\top}$ converges in distribution to $\left(\widetilde{d}_{kl}^{(1)}(x_1),\ldots,\widetilde{d}_{kl}^{(1)}(x_J),\nu_{k0}^{\top},\ \nu_{l0}^{\top}\right)^{\top}$ $\forall x_j\in X$, $\forall j\le J, \forall J\ge 1$. 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 $|1(X_{ki}\le x) - 1(X_{li}\le x)| \le 2 < \infty$. ■

Proof of Theorem 1. Suppose that $d_1^* = 0$. Then, there exists a pair (k,l) that satisfies $\sup_{x\in\mathcal{X}}\left[F_k(x) - F_l(x)\right] = 0$. For such pair, we have $F_k(x)\le F_l(x)$ for all $x\in\mathcal{X}$ but $F_k(x) = F_l(x)$ for $x\in B_{kl}^{(1)}(\subset\mathcal{X})$. We first verify that

$$\begin{aligned}
\widehat{D}_{kl}^{(1)} &\equiv \sup_{x\in\mathcal{X}}\sqrt{N}\left[\overline{F}_{kN}(x,\widehat{\theta}_k) - \overline{F}_{lN}(x,\widehat{\theta}_l)\right] \\
&\Rightarrow \sup_{x\in\mathcal{B}_{kl}^{(1)}}\left[\widetilde{d}_{kl}^{(1)}(x) + \Delta_{k0}^{(1)}(x)^{\top}\Gamma_{k0}\nu_{k0} - \Delta_{l0}^{(1)}(x)^{\top}\Gamma_{l0}\nu_{l0}\right] \qquad \text{(A.5)} \\
&\equiv \sup_{x\in\mathcal{B}_{kl}^{(1)}} d_{kl}^{(1)}(x), \quad \text{say.}
\end{aligned}$$

Note that Lemmas 2 and 3 imply

$$\begin{aligned}
\widehat{D}_{kl}^{(1)}(x) &\equiv \sqrt{N}\left[\overline{F}_{kN}(x,\widehat{\theta}_k) - \overline{F}_{lN}(x,\widehat{\theta}_l)\right] \\
&= \nu_{kN}^{(1)}(x,\widehat{\theta}_k) - \nu_{lN}^{(1)}(x,\widehat{\theta}_l) + \sqrt{N}\left[F_k(x,\widehat{\theta}_k) - F_l(x,\widehat{\theta}_l)\right] \\
&= A_{kl}^{(1)}(x) + o_p(1) \text{ uniformly in } x\in\mathcal{X}, \qquad \text{(A.6)}
\end{aligned}$$

where

$$A_{kl}^{(1)}(x) = A_{kl}^{a}(x) + A_{kl}^{b}(x) \tag{A.7}$$

$$\begin{aligned} A_{kl}^{a}(x) &= \nu_{kN}^{(1)}(x, \theta_{k0}) - \nu_{lN}^{(1)}(x, \theta_{l0}) \\ &\quad + \Delta_{k0}^{(1)}(x)^{\top}\Gamma_{k0}\sqrt{N}\overline{\psi}_{kN}(\theta_{k0}) - \Delta_{l0}^{(1)}(x)^{\top}\Gamma_{l0}\sqrt{N}\overline{\psi}_{lN}(\theta_{l0}) \\ A_{kl}^{b}(x) &= \sqrt{N}\left[F_k(x) - F_l(x)\right]. \end{aligned} \tag{A.8}$$

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

$$\sup_{x\in\mathcal{X}} A_{kl}^{(1)}(x) \Rightarrow \sup_{x\in\mathcal{B}_{kl}^{(1)}} d_{kl}^{(1)}(x). \tag{A.9}$$

Note that

$$\sup_{x\in\mathcal{B}_{kl}^{(1)}} A_{kl}^{a}(x) \Rightarrow \sup_{x\in\mathcal{B}_{kl}^{(1)}} d_{kl}^{(1)}(x) \tag{A.10}$$

by Lemma 4 and continuous mapping theorem. Note also that $A_{kl}^{(1)}(x) = A_{kl}^{a}(x)$ for $x \in B_{kl}^{(1)}$. Given $\varepsilon > 0$, this implies that

$$P\left(\sup_{x\in\mathcal{X}} A_{kl}^{(1)}(x) \le \varepsilon\right) \le P\left(\sup_{x\in\mathcal{B}_{kl}^{(1)}} A_{kl}^{a}(x) \le \varepsilon\right). \tag{A.11}$$

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

$$P\left(\sup_{\substack{\rho_d(x,y)<\delta \\ y\in\mathcal{B}_{kl}^{(1)}}} |A_{kl}^{a}(x) - A_{kl}^{a}(y)| > \lambda\right) < \gamma \tag{A.12}$$

and

$$\sup_{x\in\mathcal{X}} |A_{kl}^{a}(x)| = O_p(1). \tag{A.13}$$

Using the results (A.12) and (A.13) and arguments similar to those in the proof of Theorem 6 of Klecan et. al. (1991, p.15), we can verify that

$$P\left(\sup_{x\in\mathcal{B}_{kl}^{(1)}} A_{kl}^{a}(x) \le \varepsilon\right) \le P\left(\sup_{x\in\mathcal{X}} A_{kl}^{(1)}(x) \le \varepsilon + \lambda\right) + 2\gamma \tag{A.14}$$

for N sufficiently large. Taking λ and γ small and using (A.10), (A.11) and (A.14) now establish the desired result (A.9) and hence (A.5). Now the desired result of Theorem 1 follows by continuous mapping theorem because the terms $\widehat{D}_{ij}^{(1)}$ with (i,j) satisfying $\sup_{x\in\mathcal{X}}\left[F_k(x) - F_l(x)\right] > 0$ will diverge to infinity and hence will not affect the limit distribution of $D_N^{(1)}$.

Next suppose $d_1^* < 0$. In this case, the set $\mathcal{B}_{kl}^{(1)}$ is an empty set and hence $F_k(x) < F_l(x)\ \forall x \in \mathcal{X}$ for some (k,l). Then, $\sup_{x\in\mathcal{X}} A_{kl}^{(1)}(x)$ defined in (A.7) will be dominated by the term $A_{kl}^{b}(x)$ which diverges to minus infinity for any $x \in \mathcal{X}$ as required. Therefore, in this case $D_N^{(1)}$ will also diverge to minus infinity. ■

Proof of Theorem 2. Consider first the case $d_1^* = 0$. Let the asymptotic null distribution of $T_N = D_N^{(1)}$ in this case 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(a) holds if we establish

$$\widehat{G}_{N,\widehat{b}_N}(w) \xrightarrow{p} G(w) \quad \forall w \in \mathbb{R}. \tag{A.15}$$

Let $G_b(w) = P\left(\sqrt{b}t_{N,b,i} \leq w\right) = P\left(\sqrt{b}t_b(W_1, \ldots, W_b) \leq w\right)$. 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.15), it suffices to verify

$$\sup_{l_N \leq b \leq u_N} \left|\widehat{G}_{N,b}(w) - G_b(w)\right| \xrightarrow{p} 0 \quad \forall w \in \mathbb{R}, \tag{A.16}$$

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

$$\begin{aligned}
& P\left(\sup_{l_N \leq b \leq u_N} \left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right) \\
\leq & \sum_{b=l_N}^{u_N} P\left(\left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right) \\
\leq & \ u_N \sup_{l_N \leq b \leq u_N} P\left(\left|\widehat{G}_{N,b}(w) - G(w)\right| > \varepsilon\right) \\
\leq & \ u_N \left\{4\exp\left(-\frac{\varepsilon^2}{8}q\right) + 22\left(1 + \frac{4}{\varepsilon}\right)^{1/2} q\alpha\left(\left[\frac{N - u_N + 1}{2q}\right]\right)\right\},
\end{aligned} \tag{A.17}$$

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.17) 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.16) and hence part (a) of Theorem 2. Given this result, part (b) of Theorem 2 also holds since we have

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

Next, suppose $d_1^* < 0$. Then, part (a) immediately follows from the proof of Theorem 1. To verify part (b), let

$$\begin{aligned}
\widehat{G}_{N,b}^0(w) &= (N - b + 1)^{-1} \sum_{i=1}^{N-b+1} 1\left(t_{N,b,i} \leq w\right) = \widehat{G}_{N,b}\left(\sqrt{b}w\right) \\
G_b^0(w) &= P\left(t_b(W_1, \ldots, W_b) \leq w\right).
\end{aligned}$$

By an argument analogous to those used to verify (A.16), we have $\sup_{l_N \leq b \leq u_N} \left| \widehat{G}^0_{N,b}(w) - G^0_b(w) \right| \xrightarrow{p} 0$. Since $t_b(W_1, \ldots, W_b) \xrightarrow{p} d_1^*$, this implies that

$$g^0_{N,\widehat{b}_N}(1-\alpha) = \inf \left\{ w : \widehat{G}^0_{N,\widehat{b}_N}(w) \geq 1-\alpha \right\} \xrightarrow{p} d_1^* < 0.$$

Therefore, we have

$$\begin{aligned} P\left(T_N > g_{N,\widehat{b}_N}(1-\alpha)\right) &= P\left(\sqrt{N} t_N(W_1, \ldots, W_N) > \sqrt{\widehat{b}_N} g^0_{N,\widehat{b}_N}(1-\alpha)\right) \\ &\leq P\left(\sqrt{\frac{N}{u_N}} t_N(W_1, \ldots, W_N) > d_1^*\right) + o(1) \to 0, \end{aligned}$$

using the result $\underline{\lim}_{N\to\infty} \left(\frac{N}{u_N}\right) > 1$ and $t_N(W_1, \ldots, W_N) \xrightarrow{p} d_1^* < 0$. This establishes Theorem 2. ■

Proof of Theorem 3. The proof is similar to the proof of Theorem 2 under $d_s^* < 0$. ■

Proof of Theorem 4. The proof is similar to that of Theorem 1. Consider Lemmas 1-4 with $\nu^{(1)}_{kN}(x, \theta)$ now defined by

$$\nu^{(1)}_{kN}(x, \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[1\left(X_{ki}(\theta) \leq x\right) - F_{kN}(x, \theta)\right] \text{ for } k = 1, \ldots, K. \tag{A.18}$$

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

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

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*.

Abhyankar, A., and K-Y. Ho (2003), "Exploring long-run abnormal performance using stochastic dominance criteria: Additional evidence from IPO's. Manuscript available from http://www.cc.ncu.edu.tw/~kengyuho/researchpaper.htm.

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 (2003), "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.

Fan, Y., and O.B. Linton (2003), "Some higher order theory for a consistent nonparametric model specification test," *The Journal of Statistical Planning and Inference* 109, 1-2, 125-154.

Härdle, W., and J.S. Marron (1985), "Optimal bandwidth selection in nonparametric regression function estimation," *The Annals of Statistics* 13, 1465-81.

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.

Hall, P. and I. Van Keilegom (2003), "Testing for monotone increasing hazard rate," Discussion paprer 0228, Institut de Statistique, UCL.

Hall, P. and B. -Y. Jing (1996), "On sample reuse methods for dependent data," *Journal of the Royal Statistical Society*, Series B, 58, 727-737.

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.

Horowitz, J.L., and V. Spokoiny (2001), "An adaptive, rate-optimal test of a parametric mean regression model against a nonparametric alternative," *Econometrica*, 69, 599-632.

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.

Levy, M., and H. Levy (2002), "Prospect Theory: Much Ado About Nothing?" *Management Science*, 48, 1334-1349.

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 (2003), "Robust Inference Concerning Recent Trends in U.S. Environmental Quality", forthcoming, *Journal of Applied Econometrics.*

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.

Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, Cambridge.

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.

Shorack, G.R., and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics.* John Wiley and Sons, New York.

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", Forthcoming in Journal of Statistical Computation and Simulation. Available at http://www.mysmu.edu/faculty/yktse/yktsehp.htm.

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.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.1140	0.4100	0.1370	0.4140	0.1850	0.4700	0.0000	0.0630
1a, $d_1^* = 0$	500	0.0590	0.2360	0.0570	0.2420	0.1100	0.2960	0.0000	0.0560
	1000	0.0460	0.1830	0.0500	0.1870	0.0710	0.2170	0.0000	0.0490
	50	0.1030	0.3710	0.1180	0.3720	0.1600	0.4140	0.0000	0.0550
1b, $d_1^* = 0$	500	0.0540	0.2440	0.0620	0.2580	0.1010	0.2880	0.0000	0.0510
	1000	0.0480	0.1620	0.0480	0.1590	0.0720	0.1970	0.0000	0.0590
	50	0.3610	0.8310	0.3640	0.8380	0.4120	0.8460	0.0000	0.6850
1c, $d_1^* > 0$	500	0.9500	0.9620	0.9420	0.9620	0.8980	0.9620	0.0000	0.9830
	1000	0.9600	0.9630	0.9600	0.9630	0.9580	0.9630	0.0000	0.9950
	50	0.3730	0.8170	0.3680	0.8190	0.4100	0.8280	0.0000	0.6840
1d, $d_1^* > 0$	500	0.9650	0.9710	0.9590	0.9710	0.8860	0.9720	0.0000	0.9840
	1000	0.9580	0.9680	0.9570	0.9690	0.9520	0.9670	0.0000	0.9940
	50	0.3790	0.8190	0.3800	0.8160	0.4180	0.8450	0.0000	0.6560
1e, $d_1^* > 0$	500	0.9640	0.9820	0.9590	0.9820	0.8880	0.9820	0.0000	0.9920
	1000	0.9530	0.9610	0.9530	0.9610	0.9480	0.9610	0.0000	0.9920

Table1F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 1 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.1010	0.2480	0.1280	0.2590	0.2110	0.3150	0.0000	0.0660
1a, $d_2^* = 0$	500	0.0490	0.1290	0.0540	0.1340	0.1020	0.1800	0.0000	0.0550
	1000	0.0540	0.1010	0.0580	0.1030	0.0660	0.1360	0.0000	0.0500
	50	0.0760	0.2010	0.1050	0.2200	0.1710	0.2840	0.0000	0.0610
1b, $d_2^* = 0$	500	0.0660	0.1480	0.0690	0.1550	0.1020	0.1840	0.0000	0.0600
	1000	0.0680	0.1390	0.0690	0.1430	0.0960	0.1590	0.0000	0.0500
	50	0.2390	0.6880	0.2470	0.6900	0.3240	0.7480	0.0000	0.3360
1c, $d_2^* = 0$	500	0.9060	0.7340	0.8930	0.7280	0.8290	0.7370	0.0000	0.4510
	1000	0.9570	0.7410	0.9560	0.7370	0.9510	0.7710	0.0000	0.5450
	50	0.2230	0.6410	0.2340	0.6450	0.3120	0.7090	0.0000	0.3290
1d, $d_2^* > 0$	500	0.9070	0.7390	0.8850	0.7370	0.8290	0.7480	0.0000	0.4230
	1000	0.9570	0.7270	0.9550	0.7320	0.9520	0.7350	0.0000	0.5240
	50	0.2090	0.6480	0.2290	0.6470	0.3070	0.7110	0.0000	0.2990
1e, $d_2^* > 0$	500	0.8970	0.7230	0.8760	0.7210	0.8210	0.7520	0.0000	0.4240
	1000	0.9490	0.7120	0.9490	0.7060	0.9380	0.7359	0.0000	0.4840

Table1S. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 1 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.1110	0.4120	0.1330	0.4110	0.1720	0.4680	0.0000	0.0540
2a, $d_1^* = 0$	500	0.0470	0.2440	0.0490	0.2560	0.0940	0.2850	0.0000	0.0550
	1000	0.0680	0.1800	0.0710	0.1900	0.0890	0.2290	0.0000	0.0440
	50	0.0790	0.2950	0.0860	0.2970	0.1220	0.3530	0.0000	0.0720
2b, $d_1^* = 0$	500	0.0120	0.1120	0.0170	0.1170	0.0750	0.1580	0.0000	0.0260
	1000	0.0210	0.0970	0.0280	0.0960	0.0710	0.1060	0.0000	0.0180
	50	0.2960	0.8330	0.2990	0.8290	0.3800	0.8280	0.0000	0.4530
2c, $d_1^* > 0$	500	0.9650	1.0000	0.9460	1.0000	0.8990	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	0.9990	1.0000	0.9940	1.0000	0.0000	1.0000
	50	0.2640	0.5360	0.2730	0.5330	0.2550	0.5620	0.0000	0.1730
2d, $d_1^* > 0$	500	0.9550	1.0000	0.9360	0.9980	0.9220	0.9990	0.0000	0.9880
	1000	1.0000	1.0000	0.9980	1.0000	0.9970	1.0000	0.0000	1.0000

Table2F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 2 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.0680	0.2000	0.0980	0.2060	0.1980	0.2940	0.0000	0.0560
2a, $d_2^* = 0$	500	0.0560	0.1340	0.0620	0.1400	0.1190	0.1600	0.0000	0.0460
	1000	0.0620	0.1220	0.0660	0.1230	0.0840	0.1300	0.0000	0.0650
	50	0.0580	0.2020	0.0940	0.2150	0.1600	0.2850	0.0000	0.0780
2b, $d_2^* = 0$	500	0.0010	0.0480	0.0070	0.0480	0.0860	0.0850	0.0000	0.0060
	1000	0.0040	0.0300	0.0100	0.0340	0.0500	0.0560	0.0000	0.0100
	50	0.0010	0.0110	0.1650	0.0110	0.0410	0.0100	0.0000	0.0060
2c, $d_2^* = 0$	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.1680	0.3440	0.1870	0.3560	0.2300	0.4590	0.0000	0.0270
2d, $d_2^* > 0$	500	0.9100	0.7560	0.8890	0.7700	0.8350	0.7600	0.0000	0.3320
	1000	0.9990	0.9370	0.9980	0.9400	0.9920	0.9130	0.0000	0.8600

Table2S. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 2 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.6120	0.9960	0.6040	0.9950	0.5640	0.9970	0.0000	0.9590
3a, $d_1^* > 0$	500	1.0000	1.0000	1.0000	1.0000	0.9930	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000	0.0000	1.0000
	50	0.0580	0.2310	0.0650	0.2240	0.1040	0.2680	0.0000	0.0250
3b, $d_1^* < 0$	500	0.0000	0.0010	0.0020	0.0010	0.0440	0.0040	0.0000	0.0000
	1000	0.0000	0.0000	0.0010	0.0000	0.0190	0.0000	0.0000	0.0000
	50	0.6010	0.9980	0.5860	0.9980	0.5480	0.9950	0.0000	0.9490
3c, $d_1^* > 0$	500	1.0000	1.0000	1.0000	1.0000	0.9960	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000

Table3F. Rejection frequencies for the test of First Order Stochastic Dominance for Design 3 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.0010	0.0180	0.1600	0.0190	0.0710	0.0220	0.0000	0.0210
3a, $d_2^* = 0$	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.0430	0.1530	0.1210	0.1600	0.1990	0.2050	0.0000	0.0440
3b, $d_2^* = 0$	500	0.0000	0.0050	0.0060	0.0060	0.0060	0.0180	0.0180	0.0000
	1000	0.0000	0.0000	0.0030	0.0000	0.0110	0.0000	0.0000	0.0000
	50	0.5250	0.9260	0.5180	0.9260	0.5330	0.9300	0.0000	0.9340
3c, $d_2^* > 0$	500	1.0000	1.0000	1.0000	1.0000	0.9890	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000

Table3S. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 3 with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.5220	0.9860	0.5160	0.9860	0.5060	0.9870	0.0000	0.9270
3Rd, $d_1^* > 0$	500	1.0000	1.0000	0.9980	1.0000	0.9940	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000	0.0000	1.0000
	50	0.0860	0.1390	0.0980	0.1410	0.1580	0.1680	0.0000	0.0100
3Re, $d_1^* = 0$	500	0.0000	0.0030	0.0000	0.0020	0.0280	0.0040	0.0000	0.0000
	1000	0.0000	0.0000	0.0020	0.0000	0.0060	0.0000	0.0000	0.0000
	50	0.5320	0.9910	0.5360	0.9910	0.5320	0.9880	0.0000	0.9440
3Rf, $d_1^* > 0$	500	1.0000	1.0000	1.0000	1.0000	0.9880	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000

Table3RF. Rejection frequencies for the test of First Order Stochastic Dominance for Design 3R with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.

		Subsample						Bootstrap	
		Mean		Median		MinVol			
Design	n	uncent	recent	uncent	recent	uncent	recent	uncent	recent
	50	0.0060	0.0310	0.0600	0.0320	0.0120	0.0310	0.0000	0.0250
3Rd, $d_2^* = 0$	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50	0.0680	0.1240	0.0700	0.1310	0.1280	0.1820	0.0000	0.0310
3Re, $d_2^* = 0$	500	0.0000	0.0040	0.0060	0.0040	0.0240	0.0140	0.0180	0.0010
	1000	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000
	50	0.5320	0.9570	0.5220	0.9570	0.5340	0.9570	0.0000	0.9190
3Rf, $d_2^* > 0$	500	1.0000	1.0000	1.0000	1.0000	0.9860	1.0000	0.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000

Table3RS. Rejection frequencies for the test of Second Order Stochastic Dominance for Design 3R with critical values computed by the automatic methods [Mean, median, and Minvol] described in section 5.2 for the 5% null rejection probabilities. Recent refers to the recentered subsampling or full sample bootstrap method, while uncent refers to the uncentered subsampling or full sample bootstrap method.