AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 2/96

Nonparametric Inference for Second Order Stochastic Dominance

by

Friedrich Schmid*
Mark Trede

February 1996

JEL articles classification: C12, C13, C14, C15, D81

Keywords: second order stochastic dominance, nonparametric inference, permutation tests, Monte Carlo methods

Abstract: This paper deals with nonparametric inference for second order stochastic dominance of two random variables. If their distribution functions are unknown they have to be inferred from observed realizations. Thus, any results on stochastic dominance are influenced by sampling errors. We establish two methods to take the sampling error into account. The first one is based on the asymptotic normality of point estimators, while the second one, relying on resampling techniques, can also cope with small sample sizes. Both methods are used to develop statistical tests for second order stochastic dominance. We argue, however, that tests based on resampling techniques are more useful in practical applications. Their power in small samples is estimated by Monte Carlo simulations for a couple of alternative distributions. We further show that these tests can also be used for testing for first order stochastic dominance, often having a higher power than tests specifically designed for first order stochastic dominance such as the Kolmogorov-Smirnov test or the Wilcoxon-Mann-Whitney test. The results of this paper are relevant in various fields such as finance, life testing and decision under risk.

*Seminar für Wirtschafts– und Sozialstatistik, Universität zu Köln, Albertus–Magnus–Platz, 50923 Köln, Deutschland, Tel: +49-221-470 2813, Fax: +49-221-470 5074, e–mail: schmid@wiso.uni-koeln.de

1 Introduction

Stochastic dominance plays an important role in reliability, life testing and various branches of economics such as finance and decision under risk. In those fields both first order and second order stochastic dominance are of interest. A real-valued random quantity X dominates (or is equivalent to) a random quantity Y in the sense of first order stochastic dominance if $E(u(X)) \geq E(u(Y))$ for every *nondecreasing* function u where E denotes expectation. X dominates (or is equivalent to) Y in the sense of second order stochastic dominance if $E(u(X)) \geq E(u(Y))$ for every *nondecreasing and concave* function u. In an economic context u denotes a utility function and an agent having a nondecreasing and concave utility function is called risk avers. For a recent survey on stochastic dominance and its applications see e.g. Levy (1992).

If the distribution functions F and G of X and Y, respectively, are known it is straightforward to investigate analytically whether stochastic dominance occurs or not. In general, however, the distribution functions are unknown and have to be inferred from observed realizations of X and Y, be it in a parametric or nonparametric setting. Hence statistical procedures concerning first and second order stochastic dominance are called for.

Developing statistical tests in a parametric setting is standard; in the nonparametric framework there are some well-known tests of first order stochastic dominance such as the Kolmogorov-Smirnov test. However, the literature on nonparametric testing of second order dominance is fairly sparse. Deshpande and Singh (1985) suggest an asymptotic test in the framework of the one sample problem (i.e., G is assumed to be known and observations are from X whose distribution function is unknown). Contributions in the framework of the two sample problem (both F and G are unknown) are Eubank, Schechtman and Yitzhaki (1993), Kaur, Rao and Singh (1994) and Xu, Fisher and Willson (1994). These authors present large sample tests which are based on the asymptotic distribution of various test criteria. A different approach was taken by McFadden (1989) and Klecan, McFadden and McFadden (1991) who suggest tests based on the bootstrap principle which are applicable even for small samples.

This paper is organized as follows. Section 2 introduces the notation and states some basic relations which are used throughout the paper. Section 3 is concerned with point estimation. Section 4 is concerned with nonparametric confidence estimation. We describe a method for constructing simultaneous confidence intervals which is based on the asymptotic normality of point estimators. As a method for obtaining confidence bands we suggest a bootstrap procedure which can easily be carried out. Section 5 is the core part of this paper being concerned with testing for second order stochastic dominance. Tests based on the asymptotic normality of estimators (from section 4) are briefly described. However, we recommend tests based on the permutation principle which are easy to perform. It is shown by simulations that the latter tests keep the prescribed level α even at very low sample sizes. For selected families of alternatives the power of the permutation tests is investigated. We show that tests for second order stochastic dominance can also be used for testing first order stochastic dominance. This may even result in an increase of power. Section 6 summarizes our findings and discusses issues relevant for practical applications — including the effect of possible dependencies.

2 Definitions and Assumptions

Let $\mathcal{F}$ denote the set of continuous distribution functions F on the real line and let $\mathcal{F}_k \subset \mathcal{F}$ denote the subset of distribution functions having a finite k-th moment. For $F \in \mathcal{F}_1$ let

$$\mu_F := \int_{-\infty}^{\infty} x dF(x)$$

denote the mean. The existence of the mean μ_F of F implies $\lim_{x\to-\infty} xF(x) = 0$ and $\lim_{x\to\infty} x(1 - F(x)) = 0$, see Serfling (1980).

For $F \in \mathcal{F}_1$ define

$$\begin{aligned} I\!\!F(t) &:= \int_{-\infty}^{t} F(x)dx \\ &= \int_{-\infty}^{t} (t-x)dF(x) \\ &= E((t-X)\mathbf{1}_{\{t-X\}}). \end{aligned}$$

$I\!\!F$ is continuous, nondecreasing and convex and we have $\lim_{t\to-\infty} I\!\!F(t) = 0$ and $\lim_{t\to\infty} I\!\!F(t) = \infty$. $I\!\!F$ is not a proper distribution function but it might be called a second order distribution function.

For distribution functions $F, G \in \mathcal{F}_1$ let

$$I\!\!D(t) := I\!\!F(t) - I\!\!G(t)$$

for $t \in \mathbb{R}$. Obviously,

$$\lim_{t \to -\infty} \mathbb{D}(t) = 0.$$

Further

$$\begin{aligned}
\lim_{t \to \infty} \mathbb{D}(t) &= \lim_{t \to \infty} \left(\int_{-\infty}^{t} (t - x) dF(x) - \int_{-\infty}^{t} (t - x) dG(x) \right) \\
&= \lim_{t \to \infty} t(F(t) - G(t)) - \int_{-\infty}^{t} x dF(x) + \int_{-\infty}^{t} x dG(x) \\
&= \mu_G - \mu_F.
\end{aligned}$$

It can be shown that

$$|\mathbb{D}(t)| \leq C \qquad t \in \mathbb{R}$$

for $F, G \in \mathcal{F}_1$ where C may depend on F and G.

Let $\mathcal{U}_1$ denote the set of nondecreasing functions $u : \mathbb{R} \longrightarrow \mathbb{R}$. The following equivalence concerning first order stochastic dominance is well known (see e.g. Levy (1992))

(i) $E(u(X)) \geq E(u(Y))$ for $u \in \mathcal{U}_1$ whenever the expectations exist and are finite.

(ii) $F(x) \leq G(x)$ for $x \in \mathbb{R}$.

An analogous relation holds for second order stochastic dominance. Let $\mathcal{U}_2$ denote the set of nondecreasing and concave functions and $F, G \in \mathcal{F}_1$, then the following statements are equivalent

(i) $E(u(X)) \geq E(u(Y))$ for every $u \in \mathcal{U}_2$.

(ii) $\mathbb{F}(t) \leq \mathbb{G}(t)$ for $t \in \mathbb{R}$.

Obviously first order stochastic dominance implies second order stochastic dominance. Further, it is easy to see that first order stochastic dominance is invariant with respect to strictly increasing and continuous transformations of X and Y, while second order stochastic dominance is invariant only with respect to the much smaller group of affine transformations $x \mapsto ax + b$ where $a > 0, b \in \mathbb{R}$.

The statistical inference for second order stochastic dominance in this paper refers to the function $t \mapsto \mathbb{D}(t)$ and is performed within the framework of the two sample problem for independent observations (see section 6 for a discussion of more general dependence structures). Therefore, for $F, G \in \mathcal{F}_1$ let $(X_i)_{i\in\mathbb{N}}$ and $(Y_j)_{j\in\mathbb{N}}$ denote two sequences of i.i.d. random variables defined on an appropriate probability space where $X_i \sim F$ and $Y_j \sim G$. Further, let (X_i) and (Y_j) be independent. Statistical inference is based on finite samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$.

3 Nonparametric Point Estimation

Let

$$\begin{aligned}\hat{F}_n(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{x-X_i\geq 0\}}\\ \hat{G}_m(y) &= \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}_{\{y-Y_j\geq 0\}}\end{aligned}$$

denote the empirical distribution function based on the sample $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$. Let

$$\begin{aligned}\hat{\mathbb{F}}_n(t) &= \int_{-\infty}^{t} (t-x)d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}(t-X_i)\mathbf{1}_{\{t-X_i\geq 0\}}\\ \hat{\mathbb{G}}_m(t) &= \int_{-\infty}^{t} (x-y)d\hat{G}_m(y) = \frac{1}{m}\sum_{j=1}^{m}(t-Y_j)\mathbf{1}_{\{t-Y_j\geq 0\}}.\end{aligned}$$

The natural point estimator for $\mathbb{D}(t)$ is

$$\hat{\mathbb{D}}_{n,m}(t) := \hat{\mathbb{F}}_n(t) - \hat{\mathbb{G}}_m(t).$$

It is easy to see that $t \mapsto \mathbb{D}_{n,m}(t)$ is continuous and piecewise linear between $Z_{(i)}$ and $Z_{(i+1)}$ where $Z_{(i)}$ is the i-th order statistic of the combined sample $(Z_1, \ldots, Z_{n+m}) = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$.

Further, for n, m fixed,

$$\begin{aligned}\lim_{t\to-\infty} \hat{\mathbb{D}}_{n,m}(t) &= 0\\ \lim_{t\to\infty} \hat{\mathbb{D}}_{n,m}(t) &= \bar{Y} - \bar{X} = \frac{1}{m}\sum_{j=1}^{m} Y_j - \frac{1}{n}\sum_{i=1}^{n} X_i.\end{aligned}$$

Letting $n, m \to \infty$ we obtain the following

Proposition 1 *For $F, G \in \mathcal{F}_1$ and fixed $t \in \mathbb{R}$ we have*

$$P\left(\lim_{\min(n,m)\to\infty} \hat{\mathbb{D}}_{n,m}(t) = \mathbb{D}(t)\right) = 1.$$

The proof of proposition 1 follows from the strong law of large numbers. One can further show that the convergence stated in proposition 1 is uniform on compact intervals $[a, b]$. We therefore have

Proposition 2 *For $F, G \in \mathcal{F}_1$ and $-\infty < a < b < \infty$ we have*

$$P\left(\lim_{\min(n,m)\to\infty} \sup_{t\in[a,b]} |\hat{\mathbb{D}}_{n,m}(t) - \mathbb{D}(t)| = 0\right) = 1.$$

The proof of proposition 2 can be carried out using the technique which is also applied for the proof of the Glivenko-Cantelli theorem.

4 Nonparametric Confidence Estimation

In order to assess the reliability of the point estimation $\hat{\mathbb{D}}_{n,m}(t)$ one has to compute confidence intervals. There are two different ways to obtain the range in which the true curve $\mathbb{D}(t)$ is likely to be. First, one might compute confidence intervals at a single specified point t_0 or at K prescribed points $t_1, \ldots, t_K$. To do so, the distributional properties of $\hat{\mathbb{D}}_{n,m}(t)$ have to be used. This approach is followed in section 4.1. Second, and more interestingly, a confidence band for the entire function $t \mapsto \mathbb{D}(t)$ can be derived. This will be done by resampling techniques (section 4.2).

4.1 Simultaneous Confidence Intervals

To obtain simultaneous confidence intervals one has to establish the asymptotic joint normality of $(\hat{\mathbb{D}}_{n,m}(t_1), \ldots, \hat{\mathbb{D}}_{n,m}(t_K))'$. It is easy to see that for $F, G \in \mathcal{F}_1$ we have

$$E(\hat{\mathbb{D}}_{n,m}(t)) = \mathbb{D}(t) = \mathbb{F}(t) - \mathbb{G}(t).$$

For $F, G \in \mathcal{F}_2$ we have

$$\begin{aligned}\mathrm{Var}(\hat{\mathbb{D}}_{n,m}(t)) &= \frac{1}{n}\mathrm{Var}((t - X)\mathbf{1}_{\{t-X\geq 0\}}) \\ &\quad + \frac{1}{m}\mathrm{Var}((t - Y)\mathbf{1}_{\{t-Y\geq 0\}}).\end{aligned}$$

For K fixed points $t_1 < \ldots < t_K$ one can apply a Multivariate Central Limit Theorem to $(\hat{I\!\!D}_{n,m}(t_1), \ldots, \hat{I\!\!D}_{n,m}(t_K))'$ and obtain

$$\begin{bmatrix} \hat{I\!\!D}_{n,m}(t_1) \\ \vdots \\ \hat{I\!\!D}_{n,m}(t_K) \end{bmatrix} \overset{appr.}{\sim} N(\mu(t_1, \ldots, t_K), \Sigma(t_1, \ldots, t_K))$$

where

$$\mu(t_1, \ldots, t_K) = (I\!\!D(t_1), \ldots, I\!\!D(t_K))'$$

and the element Σ_{kl} of $\Sigma = \Sigma(t_1, \ldots, t_K)$ is

$$\begin{aligned} \Sigma_{kl} &= \frac{\mathrm{Cov}((t_k - X)\mathbf{1}_{\{t_k - X \geq 0\}}, (t_l - X)\mathbf{1}_{\{t_l - X \geq 0\}})}{n} \\ &\quad + \frac{\mathrm{Cov}((t_k - Y)\mathbf{1}_{\{t_k - Y \geq 0\}}, (t_l - Y)\mathbf{1}_{\{t_l - Y \geq 0\}})}{m}. \end{aligned}$$

Σ_{kl} can be estimated by

$$\begin{aligned} \hat{\Sigma}_{kl} &= \frac{1}{n}\Bigg[\frac{1}{n}\sum_{i=1}^{n}(t_k - X_i)(t_l - X_i)\mathbf{1}_{\{t_k - X_i \geq 0\}}\mathbf{1}_{\{t_l - X_i \geq 0\}} \\ &\quad - \left(\frac{1}{n}\sum_{i=1}^{n}(t_k - X_i)\mathbf{1}_{\{t_k - X_i \geq 0\}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}(t_l - X_i)\mathbf{1}_{\{t_l - X_i \geq 0\}}\right)\Bigg] \\ &+ \frac{1}{m}\Bigg[\frac{1}{m}\sum_{j=1}^{m}(t_k - Y_j)(t_l - Y_j)\mathbf{1}_{\{t_k - Y_j \geq 0\}}\mathbf{1}_{\{t_l - Y_j \geq 0\}} \\ &\quad - \left(\frac{1}{m}\sum_{j=1}^{m}(t_k - Y_j)\mathbf{1}_{\{t_k - Y_j \geq 0\}}\right)\left(\frac{1}{m}\sum_{j=1}^{m}(t_l - Y_j)\mathbf{1}_{\{t_l - Y_j \geq 0\}}\right)\Bigg]. \end{aligned}$$

The simplest method to derive simultaneous confidence intervals is to apply Bonferroni's approach. Let c denote the $(1 - \alpha/(2K))$-quantile of the standard normal distribution. Then

$$P(\hat{I\!\!D}_{n,m}(t_i) - c\sqrt{\Sigma_{ii}} \leq I\!\!D(t_i) \leq \hat{I\!\!D}_{n,m}(t_i) + c\sqrt{\Sigma_{ii}}) = 1 - \frac{\alpha}{K}$$

for $i = 1, \ldots, K$. Applying Bonferroni's inequality gives

$$P(\hat{I\!\!D}_{n,m}(t_i) - c\sqrt{\Sigma_{ii}} \leq I\!\!D(t_i) \leq \hat{I\!\!D}_{n,m}(t_i) + c\sqrt{\Sigma_{ii}} \quad \text{for } i = 1, \ldots, K) \geq 1 - \alpha$$

and therefore a set of simultaneous $(1 - \alpha)$ confidence intervals. In practical applications Σ_{ii} has to be substituted by $\hat{\Sigma}_{ii}$.

The Bonferroni simultaneous confidence intervals are usually unnecessarily wide since they do not take into account the correlation between the $\hat{I\!D}_{n,m}(t_i), i = 1, \ldots, K$.

The same is true if we use Sidak's inequality

$$P(\hat{I\!D}_{n,m}(t_i) - \tilde{c}\sqrt{\Sigma_{ii}} \leq I\!D(t_i) \leq \hat{I\!D}_{n,m}(t_i) + \tilde{c}\sqrt{\Sigma_{ii}} \quad \text{for } i = 1, \ldots, K)$$

$$\geq \prod_{i=1}^{K} P(\hat{I\!D}_{n,m}(t_i) - \tilde{c}\sqrt{\Sigma_{ii}} \leq I\!D(t_i) \leq \hat{I\!D}_{n,m}(t_i) + \tilde{c}\sqrt{\Sigma_{ii}}) \tag{1}$$

which is valid for arbitrarily correlated normal variables. Setting

$$\tilde{c} = \Phi^{-1}\left(\frac{1 + (1-\alpha)^{\frac{1}{K}}}{2}\right) = \Phi^{-1}\left(1 - \frac{1 - (1-\alpha)^{\frac{1}{K}}}{2}\right)$$

the righthand side of (1) becomes

$$\prod_{i=1}^{K} (1-\alpha)^{\frac{1}{K}} = 1 - \alpha.$$

There is practically no difference between Bonferroni's and Sidak's intervals since for usual values of α and K

$$c = \Phi^{-1}\left(1 - \frac{\alpha}{2K}\right) \approx \Phi^{-1}\left(\frac{1 + (1-\alpha)^{\frac{1}{K}}}{2}\right) = \tilde{c}.$$

4.2 Confidence Bands

For practical purposes derivation of a confidence band for $t \mapsto I\!D(t)$ is more important than simultaneous confidence intervals at isolated points $t_1, t_2, \ldots, t_K$. If we could derive the distribution of

$$\sup_{t \in I\!R} |I\!D(t) - \hat{I\!D}_{n,m}(t)|$$

and the corresponding quantiles a $(1-\alpha)$ confidence band for $t \mapsto I\!D(t)$ would be given by

$$t \mapsto \hat{I\!D}_{n,m}(t) \pm c_{n,m}$$

where $c_{n,m}$ is the $(1-\alpha)$-quantile of $\sup_{t \in I\!R} |I\!D(t) - \hat{I\!D}(t)_{n,m}|$ because

$$P(\hat{I\!D}_{n,m}(t) - c \leq I\!D(t) \leq \hat{I\!D}_{n,m}(t) + c \text{ for every } t \in I\!R)$$

$$= P(\sup_{t \in I\!R} |I\!D(t) - \hat{I\!D}_{n,m}(t)| \leq c) \geq 1 - \alpha.$$

It is, however, impossible to derive this distribution analytically for arbitrary $F, G \in \mathcal{F}_1$. A possible remedy is to derive the distribution and the $(1-\alpha)$-quantile approximately by bootstrapping. We therefore obtain a bootstrap approximation $c^*_{n,m,B}$ which depends on n and m and the number of bootstrap replications B.

Figures 1 and 2 depict the point estimates $\hat{I\!D}_{n,m}(t)$ and their 95 %-confidence bands of two particular examples. Figure 1 is constructed by drawing two samples, one from $X \sim N(0,1)$ and one from $Y \sim N(0, 1/25)$. These samples have been used both for calculating $\hat{I\!D}_{n,m}(t)$ and for deriving the 0.975-quantile $c^*_{25,25,500}$ with $B = 500$ bootstrap replications. Literally, one may only conclude that judging solely from the data the true but unknown curve $I\!D(t)$ is likely to lie entirely inside the confidence bands. Therefore, since $\hat{I\!D}_{n,m}(t) - c^*_{25,25,500}$ crosses the horizontal axis from below, the confidence bands give evidence that either Y stochastically dominates X (which is actually correct) or that there is no stochastic dominance at all. Of course, other realizations of the samples result in different curves, and it is not always the case that the confidence bands intersect the horizontal axis.

Figure 2 is built on two samples from the same distribution, namely the standard normal $N(0,1)$. This time the confidence bands enclose the $I\!D(t) = 0$ line as well. Looking at these confidence bands not much can be said. There might be stochastic dominance in either direction or none at all.

Figure 1: $\hat{I\!D}_{n,m}(t)$ with bootstrapped 95 %-confidence band ($F(x) = \Phi(x)$, $G(x) = \Phi(5x)$, $n = m = 25$, $B = 500$)

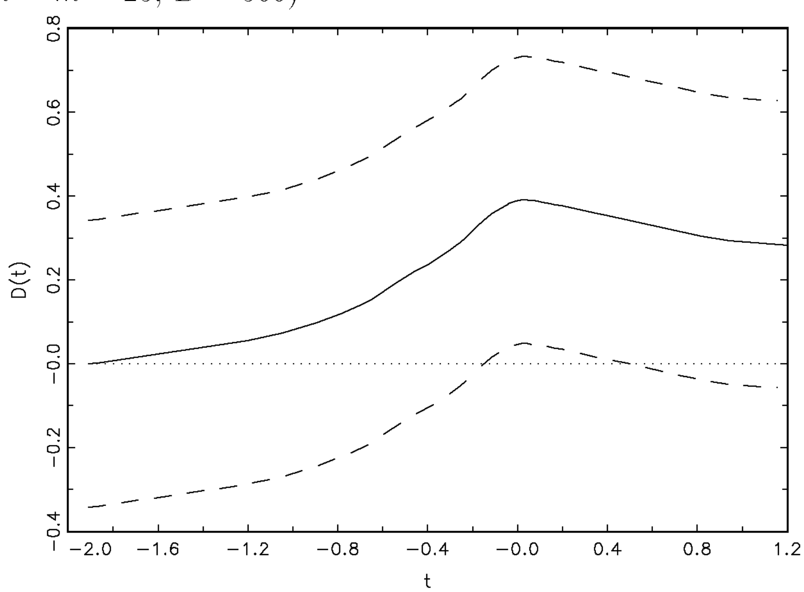
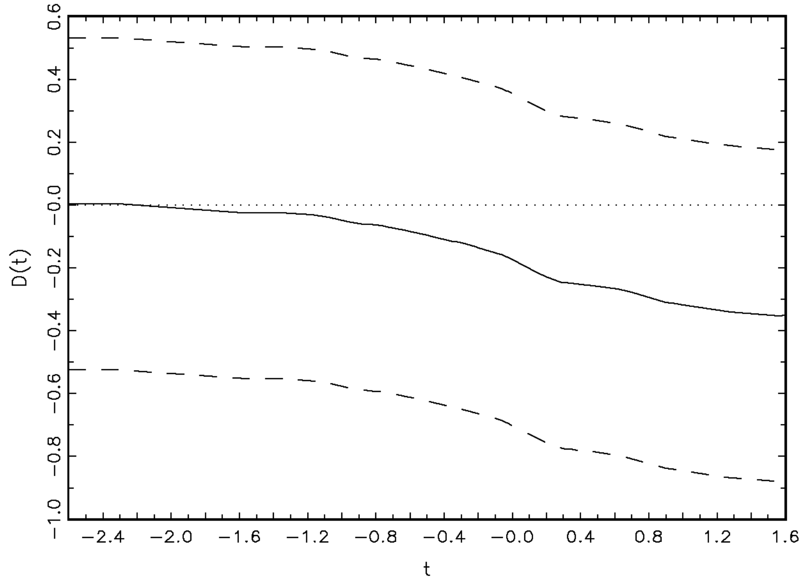


Figure 2: $\hat{I\!\!D}_{n,m}(t)$ with bootstrapped 95 %-confidence band ($F(x) = G(x) = \Phi(x)$, $n = m = 25,\ B = 500$)



5 Nonparametric Testing

This section is devoted to testing for stochastic dominance, i.e., a null hypothesis is formulated and we are looking for a test (given by a test statistic and a rejection region) which keeps a prescribed level α for the error probability of the first kind.

The null hypothesis we are dealing with is one-sided:

$$H_0\colon\ I\!\!D(t) \leq 0 \text{ for } t \in I\!\!R$$
$$H_1\colon\ \text{not } H_0$$

H_0 implies $I\!\!D(t) = I\!\!F(t) - G\!\!\!\!G(t) \leq 0$ for $t \in I\!\!R$, i.e., X second order stochastically dominates (or is equivalent to) Y. H_1 implies that either Y second order dominates X or that there is no dominance at all.

5.1 Tests Based on the Asymptotic Normality of $\hat{I\!\!D}_{n,m}$

For K prescribed points $t_1 < t_2 < \ldots < t_K$ the random vector

$$\hat{I\!\!D}_{n,m} = (\hat{I\!\!D}_{n,m}(t_1), \ldots, \hat{I\!\!D}_{n,m}(t_K))'$$

is asymptotically normally distributed with expectation $\mathbb{D}$ and covariance matrix Σ (see section 3). It is, therefore, natural to replace the testing problem

$$H_0\colon\ \mathbb{D}(t) \le 0 \text{ for } t \in \mathbb{R}$$
$$H_1\colon\ \text{not } H_0$$

by

$$H_0^*\colon\ \mathbb{D}(t_i) \le 0 \text{ for } i = 1, \ldots, K$$
$$H_1^*\colon\ \text{not } H_0^*$$

and to reject H_0^* if

$$\frac{\hat{\mathbb{D}}_{n,m}(t_i)}{\sqrt{\Sigma_{ii}}} > c$$

for at least one $i \in \{1, \ldots, K\}$.

The error probability of the first kind is given by

$$\begin{aligned}
\sup_{H_0} P(H_0 \text{ rejected}||H_0 \text{ true}) &= P(H_0 \text{ rejected}||F = G) \\
&= P\left(\frac{\hat{\mathbb{D}}_{n,m}(t_i)}{\sqrt{\Sigma_{ii}}} > c \text{ for at least one } i||F = G\right) \\
&\le \sum_{i=1}^{K} P\left(\frac{\hat{\mathbb{D}}_{n,m}(t_i)}{\sqrt{\Sigma_{ii}}} > c||F = G\right).
\end{aligned}$$

Choosing $c = \Phi^{-1}(1 - \alpha/K)$ the error probability of the first kind is $\le \alpha$, in fact it will be much less than α in most cases.

Another method for determining the critical level c is the following.

$$\begin{aligned}
P(H_0 \text{ rejected}||F = G) &= P\left(\frac{\hat{\mathbb{D}}_{n,m}(t_i)}{\sqrt{\Sigma_{ii}}} > c \text{ for at least one } i||F = G\right) \\
&= 1 - P\left(\frac{\hat{\mathbb{D}}_{n,m}(t_i)}{\sqrt{\Sigma_{ii}}} \le c \text{ for } i = 1, \ldots, K||F = G\right) \\
&\approx 1 - \Phi(c, c, \ldots, c||\Lambda)
\end{aligned}$$

where $\Phi(x_1, \ldots, x_K||\Lambda)$ denotes the joint distribution function of K correlated $N(0,1)$ variables with correlation matrix $\Lambda = (\Lambda_{ij})$ where

$$\Lambda_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}}\sqrt{\Sigma_{jj}}}.$$

If we can solve

$$1-\alpha = \Phi(c,\ldots,c||\Lambda)$$

with respect to c we arrive at a test whose error probability of the first kind is approximately α. The solution of this equation is, however, extremely difficult for larger values of K. Besides we have to replace the Σ_{ij} by their estimates $\hat{\Sigma}_{ij}$ which will introduce further sampling errors. We therefore do not recommend this approach to testing for practical applications.

5.2 Permutation Tests

The permutation principle for testing dates back to Fisher (1935). An application oriented exposition is given e.g. in Efron and Tibshirani (1993) and Good (1993). It can easily be applied to the one-sided testing problem

$$\begin{aligned} H_0&: \ \mathbb{D}(t) \le 0,\ t \in \mathbb{R} \\ H_1&: \ \text{not } H_0 \end{aligned}$$

under study. As test statistics we consider the sup-statistic

$$\begin{aligned} T_1 &= \sup_{t\in\mathbb{R}} \hat{\mathbb{D}}_{n,m}(t) \\ &= \sup_{t\in\mathbb{R}} (\hat{\mathbb{F}}_n(t) - \hat{\mathbb{G}}_m(t)) \\ &= \max_{i=1,\ldots,n+m} \hat{\mathbb{D}}_{n,m}(z_{(i)}) \end{aligned}$$

where $z_{(i)}$ is the i-th order statistic of the pooled sample $(z_1,\ldots,z_{n+m}) = (x_1,\ldots,x_n, y_1,\ldots,y_m)$. T_1 can be viewed as the second order analog of the one-sided Kolmogorov-Smirnov statistic.

Another suitable statistic is the integral statistic

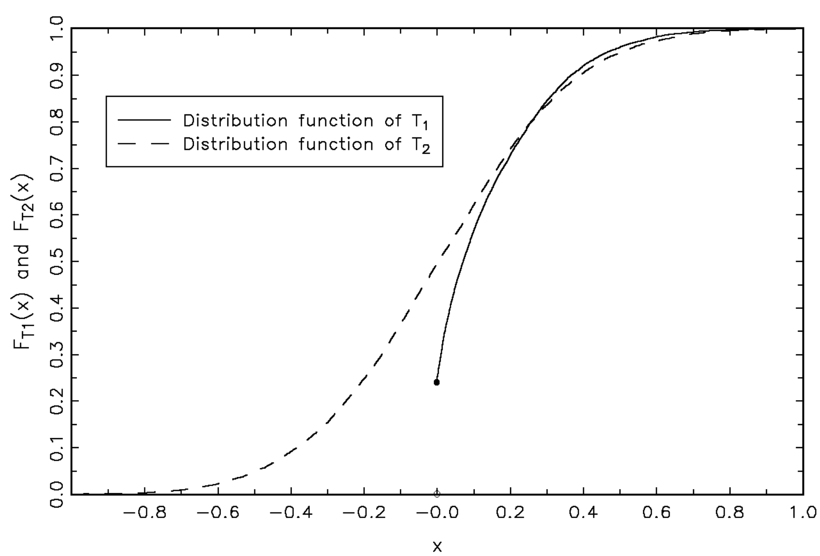
$$\begin{aligned} T_2 &= \int_{-\infty}^{\infty} (\hat{\mathbb{F}}_n(t) - \hat{\mathbb{G}}_m(t)) d(\hat{F}_n(t) + \hat{G}_m(t)) \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{n}\sum_{i=1}^{n} (t-x_i)\mathbf{1}_{\{t-x_i\ge 0\}} - \frac{1}{m}\sum_{j=1}^{m} (t-y_j)\mathbf{1}_{\{t-y_j\ge 0\}} \right) d(\hat{F}_n(t) + \hat{G}_m(t)) \\ &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{n} (x_k-x_i)\mathbf{1}_{\{x_k-x_i\ge 0\}} - \frac{1}{m^2}\sum_{k=1}^{m}\sum_{j=1}^{m} (y_k-y_j)\mathbf{1}_{\{y_k-y_j\ge 0\}} \\ &\quad + \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} (y_k-x_i)\mathbf{1}_{\{y_j-x_i\ge 0\}} - \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} (x_k-y_j)\mathbf{1}_{\{x_i-y_j\ge 0\}}. \end{aligned}$$

T_2 can be viewed as the second order analog of the Wilcoxon statistic.

H_0 is rejected, of course, if the test statistics are too large. It should be stressed, however, that the distributions of T_1 and T_2 depend on F and G even if $F = G$. It is therefore not possible to prepare a table with critical values for T_1 and T_2.

Figure 3 displays the distribution functions of T_1 and T_2 for the particular case where $n = m = 25$ and $F = G = \Phi$. The distribution of T_1 has a point mass on 0. This is due to the fact that $\hat{I\!D}_{n,m}(t) = 0$ for $t \leq z_{(1)}$. The remaining part of the distribution is continuous. The distribution of T_2 is of continuous type for $F, G \in \mathcal{F}_1$ and $F = G$.

Figure 3: Distribution functions of T_1 and T_2 for $F = G = \Phi$ and $n = m = 25$



Application of the permutation principle results in a test which keeps exactly the prescribed level α (even for very small values of n and m) if we draw all of the $\binom{n+m}{n}$ different subsets of order n from the vector $z = (z_{(1)}, \ldots, z_{(n+m)})$ and randomize. Randomization, however, is unusual in practical applications and selection of all subsets of z becomes quickly intractable. Therefore, the determination of the number B of subsets to be taken becomes a crucial point for the applicability and validity of the method. A small Monte Carlo experiment sheds some light on this issue.

In what follows H_0 is rejected if $T > c$ where c is determined by $c = T_{(B(1-\alpha))}$ where $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(B)}$ is the ordered sequence of values of the test statistic for B randomly drawn subsets of size n of z. Table 1 shows error probabilities of the first kind — obtained for $B = 500$ in the case of $n = m = 10, 15, 20, 25, 30$ and

$\alpha = 0.01, 0.05, 0.1$ with $F = G = \Phi$. Table 2 is related to $F = G = U$ for $U(x) = x$ for $0 \leq x \leq 1$, i.e., U is the distribution function of a uniform distribution on $[0, 1]$.

Table 1: Error probabilities of the first kind of the test statistics T_1 and T_2 under $F = G = \Phi$

	Test statistic T_1			Test statistic T_2		
$n = m$	0.1	0.05	0.01	0.1	0.05	0.01
10	0.1024	0.0516	0.0126	0.1008	0.0504	0.0124
15	0.1028	0.0514	0.0120	0.1070	0.0510	0.0116
20	0.0994	0.0542	0.0134	0.0992	0.0536	0.0132
25	0.1018	0.0466	0.0106	0.0986	0.0480	0.0100
30	0.1038	0.0522	0.0146	0.0932	0.0536	0.0138

Table 2: Error probabilities of the first kind of the test statistics T_1 and T_2 under $F = G = U$

	Test statistic T_1			Test statistic T_2		
$n = m$	0.1	0.05	0.01	0.1	0.05	0.01
10	0.1078	0.0572	0.0130	0.1062	0.0574	0.0118
15	0.1006	0.0474	0.0094	0.1000	0.0500	0.0090
20	0.1058	0.0532	0.0124	0.1050	0.0542	0.0122
25	0.1008	0.0512	0.0124	0.1054	0.0514	0.0114
30	0.1004	0.0486	0.0120	0.1000	0.0480	0.0124

Our conclusion from the results of tables 1 and 2 is that $B = 500$ results in satisfactory agreement of prescribed and attained error probability of the first kind. A similar simulation for $B = 100$ gave a far less satisfactory picture. We are going to use $B = 500$ in what follows.

The permutation principle determines the critical value c conditioned on $F = G$. This is the boundary of the null hypothesis $H_0 : \mathbb{F}(t) - \mathbb{G}(t) \leq 0$ since $\mathbb{F}(t) - \mathbb{G}(t) = 0$ for $t \in \mathbb{R}$ if and only if $F(x) = G(x)$ for $x \in \mathbb{R}$. It should be stressed, however, that the null hypothesis contains pairs of distribution functions (F, G) where $F \neq G$. We do not yet have a formal proof that

$$\sup_{(F,G) \in H_0} P(H_0 \text{ is rejected} || (F, G)) \leq \alpha,$$

but the construction of the tests as well as our simulation results suggest that the overall level of the tests is indeed $\leq \alpha$. Nevertheless, some care concerning the overall error probability of the first kind should be taken.

5.3 Power of the Tests

Though the tests described in section 5.2 are nonparametric it is useful to investigate their power within parametric families of alternative distributions. Power can then be displayed graphically as a function of the parameter, resulting in a power curve.

We consider the following two families of alternative distributions:

- Type I:

  $F(x) = \Phi(x) \qquad x \in I\!R$

  $G_c(x) = \Phi(cx) \quad x \in I\!R$ and $0 < c < \infty$

  where Φ denotes the distribution function of the standard normal distribution. (F, G_c) belongs to H_0 for $c \leq 1$ and to H_1 for $c > 1$.

- Type II:

  $F(x) = x \qquad 0 \leq x \leq 1$

  $$G_\delta(x) = \begin{cases} (2x)^\delta/2 & 0 \leq x \leq 0.5 \\ 1-(2(1-x))^\delta/2 & 0.5 < x \leq 1 \end{cases} \quad \text{and } 0 < \delta < \infty$$

  (F, G_δ) belongs to H_0 for $\delta \leq 1$ and to H_1 for $\delta > 1$.

Note that there is no first order stochastic dominance for families I and II on the set of alternatives.

The power of the tests is determined by Monte Carlo simulation for the case $n = m = 25$ and $\alpha = 0.05$. Samples from F and G are generated and the permutation tests (with T_1 and T_2) are carried out as described above for $B = 500$. The number of Monte Carlo replications is $N = 5000$.

This procedure is performed for some ten different values of the parameters of families I and II. Fitting cubic splines to the estimated points yields the power curves shown in figures 4 and 5. The inlaid figures depict the underlying distribution functions: the one giving more weight to the tails is the null distribution while the other one is an example of the alternative distributions.

Concerning power the test based on the integral statistic T_2 obviously dominates the second order Kolmogorov test based on T_1. The error probability of the first kind of

Figure 4: Power curves of the tests based on T_1 and T_2 for alternative distributions of Type I (with $n = m = 25$ and $B = 500$)

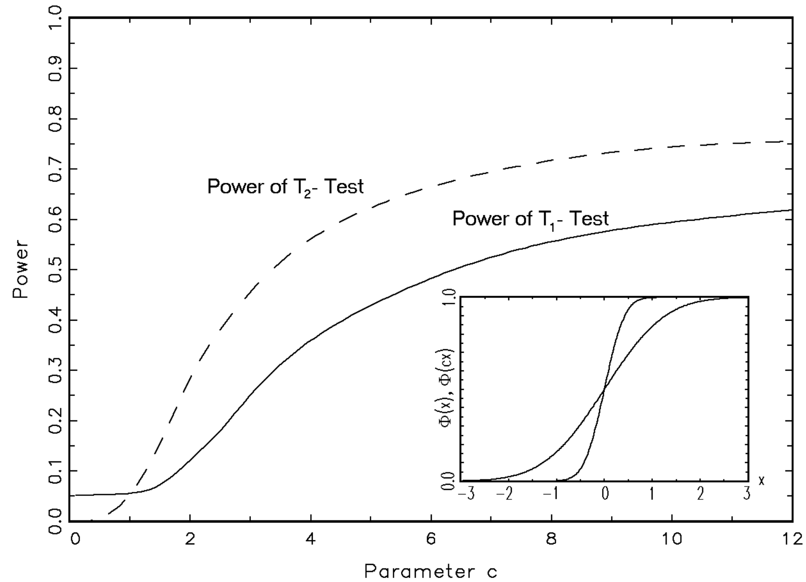
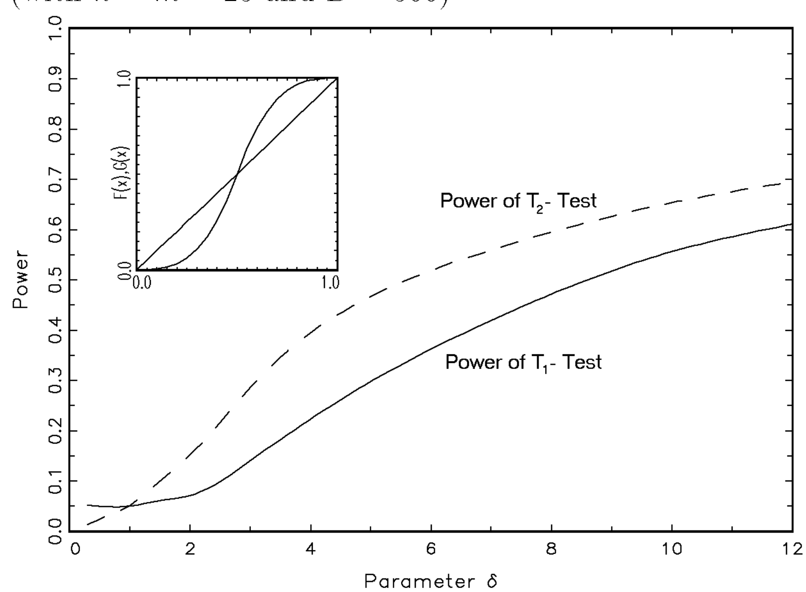


Figure 5: Power curves of the tests based on T_1 and T_2 for alternative distributions of Type II (with $n = m = 25$ and $B = 500$)



the test based on T_2 tends to zero for $c \to 0$ (Type I distributions) and $\delta \to 0$ (Type II distributions) whereas it remains approximately constant for T_1.

It should be pointed out that the tests suggested in section 5.2, though being designed for testing second order stochastic dominance, can also be applied for testing first order stochastic dominance. First order stochastic dominance implies second order stochastic dominance and if we can reject the null hypothesis of second order stochastic dominance, the null hypothesis of first order stochastic dominance must also be rejected. Therefore, instead of testing

$$H_0\colon\ F(x) \le G(x),\ x \in \mathbb{R}$$
$$H_1\colon\ \text{not } H_0$$

we may test

$$H_0^*\colon\ \mathbb{F}(t) \le \mathbb{G}(t),\ t \in \mathbb{R}$$
$$H_1^*\colon\ \text{not } H_0^*$$

with one of the tests described above and reject H_0 if H_0^* is rejected. It is interesting to note that this procedure may result in a considerable increase in power. This will be demonstrated by two examples based on the following alternatives.

- Type III:

  $F(x) = \Phi(x) \qquad x \in \mathbb{R}$

  $G_\mu(x) = \Phi(x-\mu) \quad x \in \mathbb{R}$ and $\mu \in \mathbb{R}$

  For $\mu > 0$ we have $F(x) > G_\mu(x)$ for $x \in \mathbb{R}$. Hence the null hypothesis is violated and there is first order stochastic dominance of Y over X.

- Type IV:

  $F(x) = 1 - e^{-x} \qquad x > 0$

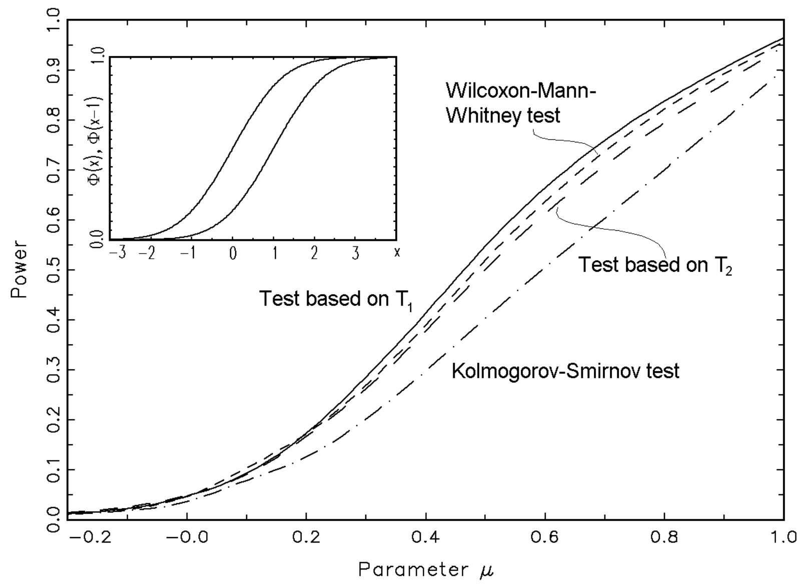
  $G_a(x) = 1 - e^{-ax}$ for $x > 0$ and $a > 0$

  For $a < 1$ we have $F(x) > G(x)$ for $x \in \mathbb{R}$ and thus again there is first order stochastic dominance of Y over X.

Figures 6 and 7 show the power curves of the second order tests and of the (first order) Kolmogorov–Smirnov test and the Wilcoxon–Mann–Whitney test. Again, the inlaid figures display the null distribution as well as an example of the alternative distributions. The distribution functions being more on the right belong to the alternative distributions.

For both types of alternatives the second order Kolmogorov test (based on T_1) dominates the familiar first order Kolmogorov test. This might be due to the fact that T_1 is smoother than the original Kolmogorov statistic. Power curves for T_2 and the Wilcoxon statistic are very close together.

Figure 6: Power curves of various tests for alternative distributions of Type III

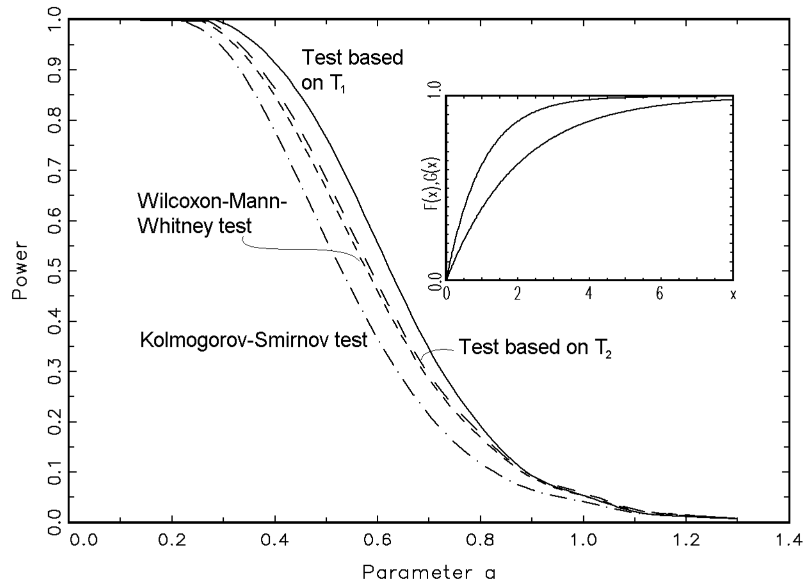


6 Concluding Remarks

This paper is concerned with nonparametric statistical inference for second order stochastic dominance. Procedures for point and confidence estimation as well as testing procedures have been proposed. Concerning tests for second order stochastic dominance we recommend the application of the permutation principle for the determination of the critical values of the test statistics.

The corresponding tests are easy to perform and are reliable even for very small sample sizes — which is an advantage over asymptotic tests suggested in 5.1 and elsewhere. The power of the two permutation tests described in this paper is compared for a number of alternative distributions but more alternatives have to be

Figure 7: Power curves of various tests for alternative distributions of Type IV



taken into account in order to identify those instances where one test is superior to the other.

Another and probably more serious question concerns the robustness of the permutation tests with respect to dependencies in the data $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ which may occur in economic applications. Though we have assumed for the ease of exposition that $Z = (Z_1, \ldots, Z_{n+m}) = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ are i.i.d. under $F = G$ the permutation principle is still applicable if $(Z_1, \ldots, Z_{n+m})$ are only exchangeable, i.e., if $(Z_1, \ldots, Z_{n+m})$ and $(Z_{\sigma(1)}, \ldots, Z_{\sigma(n+m)})$ are identically distributed for every permutation σ of the numbers $1, \ldots, n+m$. Exchangeability of Z can be interpreted in the following setting. There are (under $F = G$) i.i.d. random variables $Z' = (X_1', \ldots, X_n', Y_1', \ldots, Y_m')$ and a random variable U (being independent of Z') with

$$\begin{aligned} X_i &= X_i' + U \qquad i = 1, \ldots, n \\ Y_j &= Y_j' + U \qquad j = 1, \ldots, m. \end{aligned}$$

In an economic context X_i and Y_j could be seen as returns on two assets. It is assumed that there are individual determinants X_i' and Y_j' as well as a general additive market shock U influencing all returns in the same way. The influence of exchangeability on the power of the tests is a problem which deserves investigation.

Though exchangeability is more general than independence it often will not be the suitable model for the data. An alternative which will be reasonable in some instances is to assume that (X_i, Y_i), $i = 1, \ldots, n$ are matched pairs, i.e., (X_i, Y_i) are jointly determined and may show arbitrary dependencies. Testing with matched pairs is well developed in the framework of first order stochastic dominance. Tests for second order stochastic dominance still have to be developed but it is easy to see that the application of the permutation principle is still possible (see e.g. Good (1993)). This will be the subject of a forthcoming paper.

References

Deshpande, J.V.; Singh, H. (1985) : "Testing for Second Order Stochastic Dominance", *Communications in Statistics — Theory and Methods* 14: 887–897.

Efron, B.; Tibshirani, R.J. (1993) : *An Introduction to the Bootstrap*, New York.

Eubank, R.; Schechtman, E.; Yitzhaki, S. (1993) : "A Test for Second Order Stochastic Dominance", *Communications in Statistics — Theory and Methods* 22: 1893–1905.

Fisher, R.A. (1935) : *Design of Experiments*, Edinburgh.

Good, P. (1993) : *Permutation Tests*, Springer, New York.

Kaur, A.; Rao, B.L.S.P.; Singh, H. (1994) : "Testing for Second-order Stochastic Dominance of Two Distributions," *Econometric Theory* 10: 849–866.

Klecan, L.; McFadden, R.; McFadden, D. (1991) : A Robust Test for Stochastic Dominance, mimeographed.

Levy, H. (1992) : "Stochastic Dominance and Expected Utility: Survey and Analysis", *Management Science* 38: 555–593.

McFadden, D. (1989) : "Testing for Stochastic Dominance", in: T. Fombay, T.K. Seo (eds.), *Studies in Economics of Uncertainty*, New York.

Serfling, R.J. (1980) : *Approximation Theorems of Mathematical Statistics*, New York.

Xu, K.; Fisher, G.; Willson, D. (1994) : New Distribution-Free Tests for Stochastic Dominance, mimeographed.