Tests for First-Order Stochastic Dominance

TERESA LEDWINA[1] AND GRZEGORZ WYŁUPEK[2]

[1] Institute of Mathematics, Polish Academy of Sciences, Poland
[2] Institute of Mathematics, University of Wrocław, Poland

In a recent article (Ledwina and Wyłupek, 2012a), we proposed and studied the two new tests for detecting stochastic dominance. In this article, we further discuss and investigate these tests and construct another one. Some useful theoretical considerations on the new construction are presented. Simulation results are given to demonstrate a very good performance of the three new solutions under large sample sizes and alternatives typical for income analysis. Data sets are used to illustrate the methods discussed in this article.

Keywords Data driven test; Inequality constraints; Linear rank statistic; Model selection criterion; Multiple comparison; Stochastic order

Mathematics Subject Classification Primary: 62G10; Secondary: 60E15, 62G20

Correspondence address to Grzegorz Wyłupek, Institute of Mathematics, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland; E-mail: wylupek@math.uni.wroc.pl

1. Introduction

Stochastic dominance relation among two variables plays important role in reliability, biometrics, actuarial sciences and econometrics. For some evidence and discussion see Robertson et al. (1988), Shaked and Shanthikumar (1994), Davidson and Duclos (2000), Müller and Stoyan (2002), Barrett and Donald (2003), Sen and Silvapulle (2005), Levy (2006). In this paper we restrict attention to the stochastic dominance of the first order, which is also known as stochastic order. Moreover, we focus on classical scheme and consider two independent samples with possibly different sample sizes, from two populations that have the corresponding continuous distribution functions F and G. In this setting, the recognized Kolmogorov-Smirnov test dominates many other solutions; see the simulation results in Barrett and Donald (2003) as well as in Ledwina and Wyłupek (2012a) (LW hereafter). The last mentioned paper suggested, however, the two new tests that outperform the Kolmogorov-Smirnov test under several types of alternatives and under small and moderately large sample sizes. The new tests introduced in LW are essential refinements of some procedures already investigated in econometric literature. More precisely, the first statistic, denoted by $M_{D(N)}$, can be considered as some weighted variant of the Kolmogorov-Smirnov solution. The second one, denoted by Q_T, is a flexible and relatively easy to implement replacement of the Wald statistic for the underlying testing problem with inequality restrictions. Both new solutions are consistent under weak and natural assumptions. We present these solutions in some detail in Section 2.1.

This paper has two main contributions. First, we investigate the two new tests by simulations under large sample sizes and with special emphasize on detecting lack of the order among typical income distributions. To be more specific, the first new statistic $M_{D(N)}$ is a minimum of some optimal linear rank statistics, each one related to a comparison of the two pertaining empirical distributions in one of the $D(N)$ dyadic points from (0,1). Here, similarly as in Section 5.2 of LW, we consider $D(N)$ to be the largest dimension of the partition less than or equal to N. Hence, $D(N)$ is a nondecreasing function of N and tends to infinity as N does. The second new statistic Q_T is some quadratic form of the above mentioned linear rank statistics with the dimension of the form specified by the selection rule T depending on Schwarz-type penalty. If $D(N)$ is growing then the first statistic tends slowly to $-\infty$ while the penalty in the second statistic goes to $+\infty$. Therefore, it is interesting to see how both procedures work under relatively large sample sizes, typical for example in some econometric applications. In this article we also consider another weighted variant of the two-sample one-sided Kolmogorov-Smirnov statistic, denoted by M_N^*, which was introduced in LW as some approximation of $M_{D(N)}$. Our simulations exhibit that $M_{D(N)}$ and M_N^* work nicely under large sample sizes. They have, due to the weighting built into them, equally high power in detecting differences in all parts of the underling distributions. The solution Q_T, by its construction, is more focused on differences in central part. In this respect Q_T is similar in behaviour to the Kolmogorov-Smirnov solution. However, in comparison with this standard, Q_T is much more sensitive in detecting differences in tails.

Our second contribution relies on studying some modifications of the solution Q_T to get more comprehensive procedure of that kind, working well under both small and large sample sizes. In particular, Section 2.2 introduces an interesting modification $T1$ of T, which penalizes less the successive dimensions. In our studies and experiments, this selection rule is shown to be a reasonable solution leading to a good omnibus statistic Q_{T1} with higher average power than the test statistic Q_T. More precisely, Q_{T1} works nicely under both small and large sample sizes as well as has equally high sensitivity in detecting differences in the tails and in the central part of the two distribution functions. Similarly as its forerunner Q_T, Q_{T1} leads to a consistent quadratic test. Further comments on this construction are given in Section 5.

The reminder of this article is organized as follows. In Section 2, we define the null hypothesis of the first-order stochastic dominance, introduce notations and definitions of $M_{D(N)}$, M_N^*, and Q_T. Moreover, we present there the modification $T1$ of the selection rule T and the related data driven test statistic Q_{T1}. The four tests are then compared through simulation experiments in Section 3. For completeness, we also included there the Kolmogorov-Smirnov test, the best existing standard in the area. Real data sets are studied in Section 4. Concluding remarks are given in Section 5. Appendices A, B, and C contain some complementary materials. In particular, Appendix B presents and discusses three additional selection rules and empirical powers of the pertaining data driven tests. This material illustrates how delicate question is a calibration of data driven tests. Appendix C provides some theoretical results, including consistency of the new data driven test.

2. Testing methods under consideration

2.1. Preliminaries

Let X and Y be continuous outcome variables that may, for example, represent incomes in two different years. Let F and G be the distribution functions of X and Y, respectively. Then we say that Y stochastically dominates X at first order if $F(x) \geq G(x)$ for each $x \in \mathbb{R}$. We assume that both F and G are unknown and focus on testing

$$\mathcal{H}^+ : F(x) \geq G(x) \ \text{ for each } \ x \in \mathbb{R}$$

against

$$\mathcal{A} : F(x) < G(x) \ \text{ for some } \ x \in \mathbb{R}.$$

For this purpose, consider two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ obeying distribution functions F and G, respectively. Then $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ is the pooled sample of size $N = m + n$. We assume throughout that $\eta = \lim_{N\to\infty}(m/N)$ exists and $\eta \in (0, 1)$. Denote by R_i, $i = 1, \ldots, m$ and $i = m + 1, \ldots, N$ the rank of X_i and Y_i in the pooled sample. LW argued that to construct tests for $\mathcal{H}^+$ against $\mathcal{A}$ it is reasonable to use the following rank statistics

$$L_j = \sum_{i=1}^{N} c_{Ni}\, l_j\Big(\frac{R_i - 0.5}{N}\Big), \qquad \text{where} \qquad c_{Ni} = \sqrt{\frac{mn}{N}} \begin{cases} -m^{-1} & \text{if} \quad 1 \leq i \leq m, \\ n^{-1} & \text{if} \quad m < i \leq N, \end{cases}$$

$$l_j(z) = -\sqrt{\frac{1 - a_j}{a_j}}\, \mathbb{1}(0 \leq z < a_j) + \sqrt{\frac{a_j}{1 - a_j}}\, \mathbb{1}(a_j \leq z \leq 1),$$

$\mathbb{1}(A)$ is the indicator of the set A while $a_1, a_2, \ldots$ are the successive points of the form $(2i - 1)/2^{k+1}$, $k = 0, 1, \ldots,$ $i = 1, 2, \ldots, 2^k$. More precisely, each L_j is asymptotically (locally) optimal statistic to test that $(F - G) \circ H^{-1}(a_j) \geq 0$ against $(F - G) \circ H^{-1}(a_j) < 0$, where $H = \eta F + (1 - \eta)G$. In other words, L_j serves to compare carefully the difference between F and G at the point $H^{-1}(a_j)$, i.e. at the a_j-quantile of the combined distributions. Comparisons of that kind are coherent with practice in income studies, cf. Anderson (1996), p. 1188. Significantly small negative values of L_j indicate $\mathcal{A}$. Since each l_j is a nondecreasing function then $P(L_j < c\,|F \geq G) \leq P(L_j < c\,|F = G)$, for any $c \in \mathbb{R}$ and $j = 1, 2, \ldots$. Moreover, under $F = G$, L_j's are asymptotically $N(0,1)$ while, for any fixed m and n, $P(L_j < c\,|F = G)$ can be easily simulated. Here, as in Section 5.2 of LW, we concentrate on $k \leq K = K(N)$ where $K(N)$ is the largest natural number such that $D(N) = 2^{K(N)+1} - 1 \leq N$.

2.2. Two solutions related to multiple comparisons

In view of the interpretation of L_j's and multiple testing approach, LW proposed to reject $\mathcal{H}^+$ in favour of $\mathcal{A}$ if

$$M_{D(N)} = \min_{1 \leq j \leq D(N)} L_j \tag{1}$$

is too small. Given the significance level α and total sample size N, the critical value of this test shall be denoted by $c_M(\alpha, D(N))$.

The second test rejects the null hypothesis for large values of

$$M_N^* = \inf_{Z_{(1)} \leq x \leq Z_{(N)}} \sqrt{\frac{mn}{N}} \, \frac{\{F_m(x) - G_n(x)\}}{\sqrt{H_N(x)\{1 - H_N(x)\}}}\,,$$

where F_m and G_n are the empirical distribution functions in the first and the second sample, respectively, $H_N = (m/N)F_m + (n/N)G_n$ while $Z_{(1)} \leq \ldots \leq Z_{(N)}$ are ordered observations in the pooled sample. Additionally, the respective quotient in M_N^* is defined to be 0 at $Z_{(N)}$. The statistic M_N^*, introduced in Section 3 of LW as some analogue of $M_{D(N)}$, is a two-sample counterpart of the goodness-of-fit statistic studied by Eicker (1979); also cf. Canner (1975) for very similar construction.

2.3. Quadratic test statistic Q_T

To present the third solution introduced in LW we need additional notations. Let $\mathcal{D}(N) = \{\, d = 2^{k+1} - 1 : k = 0, 1, \ldots, K(N)\,\} = \{1, 3, \ldots, D(N)\}$, $L_j^+ = \max\{\,-L_j, 0\,\}$ and for $d \in \mathcal{D}(N)$ define

$$\boldsymbol{L}_d^+ = (\,L_1^+, \ldots, L_d^+\,) \qquad \text{and} \qquad Q_d = [\,\boldsymbol{L}_d^+\,][\,\boldsymbol{L}_d^+\,]', \tag{2}$$

where the prime denotes transposition. So, given N, for the successive $d \in \mathcal{D}(N)$ we consider increasingly finer partitions of $[0,1]$ and the related sums of squares, Q_d, of the negative L_j's amongst $L_1, \ldots, L_d$. Large values of Q_d indicate $\mathcal{A}$. A choice of d, which is decisive to the distribution of Q_d, was done in LW by the selection rule T, which was defined as follows

$$T = T0 = \min\{\, d : Q_d - d\pi^{(0)}(\alpha, N) \geq Q_j - j\pi^{(0)}(\alpha, N),\; d, j \in \mathcal{D}(N)\,\}, \tag{3}$$

where the penalty $\pi^{(0)}(\alpha, N)$ for the parameter d was given by

$$\pi^{(0)}(\alpha, N) = \begin{cases} p^{(0)}(\alpha, N) & \text{if} \quad M_{D(N)} \geq -\sqrt{t(\alpha)\log N}, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

while $p^{(0)}(\alpha, N) = \log N$ and, by (1), $M_{D(N)} = \min_{1 \leq j \leq D(N)} L_j$. Here, as before, α denotes the prescribed significance level. The penalty $\pi^{(0)}(\alpha, N)$ depends on α via the tuning parameter $t(\alpha)$.

A simple automatic rule for selecting $t(\alpha)$ was described in Section 5.1 of LW. For illustration, for sample sizes and the level $\alpha = 0.01$ considered in Table 3 of LW the rule T leads to $t(\alpha) \in [2.00, 2.45]$. Note also that the penalty in (4) depends on data at hand and therefore is a random variable.

Throughout the current paper we shall call an event of the form $\{M_{D(N)} \geq c\}$, appearing in the penalty, the 'switch' while c shall be named the barrier in the 'switch'. Moreover, the quadratic form $Q_\bullet$ with the size determined by a selection rule shall be called a data driven statistic.

Simulation results in LW show that for $N \in [100, 1000]$ the tests based on $M_{D(N)}$ and Q_T work nicely. Q_T is slightly more sensitive in detecting differences between F and G in the middle range of the a_j-quantiles of H while $M_{D(N)}$ slightly better detects differences in extreme quantiles.

The form of the penalty $\pi^{(0)}(\alpha, N)$ implies that, under the null hypothesis, the concentration of T on $d = 1$ grows to 1 as $N \to \infty$ (cf. Lemma 4(i) in LW). This stabilizes the related critical values and is profitable in this sense. On the other hand, for large N the penalty $\log N$, appearing in (4) when the 'switch' does not work, is so heavy that T concentrates too often on $d = 1$ under alternatives. This, however, is not profitable to the power. Therefore, in the next section we propose a modification $T1$ of the rule T which is less restrictive under both $\mathcal{H}^+$ and $\mathcal{A}$ and still takes enough care on stability of critical values. This results in slightly higher average power of Q_{T1} in comparison to that of Q_T. To close, emphasize that the selection rule T depends on the prescribed significance level α and the data. Similar feature shall obey its modification $T1$.

2.4. Proposed modification of T

As mentioned above, a key point to the sensitivity of data driven quadratic form is careful balancing its behaviour under both $\mathcal{H}^+$ and $\mathcal{A}$. In course of the present work we constructed and investigated a few data driven statistics and propose here to consider the one defined via

$$T1 = \min\{\, d : Q_d - d\pi^{(1)}(\alpha, N) \geq Q_j - j\pi^{(1)}(\alpha, N),\ d, j \in \mathcal{D}(N)\,\}, \tag{5}$$

where the penalty $\pi^{(1)}(\alpha, N)$ for the parameter d is given by

$$\pi^{(1)}(\alpha, N) = \begin{cases} p^{(1)}(\alpha, N) & \text{if} \quad M_{D(N)} \geq c_M(0.8\alpha, D(N)), \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

while $p^{(1)}(\alpha, N)$ is the smallest positive number $p^{(1)}$ such that, given the pooled sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$, under $F = G$, the concentration of $T1$ on $d = 1$ is at least $1 - \alpha$. In symbols,

$$P(T1 = 1 | F = G) \geq 1 - \alpha. \tag{7}$$

In Appendix C we discuss the question of existence of $p^{(1)}(\alpha, N)$. Since, under $F = G$, the quadratic forms defining $T1$ are distribution free, therefore in our experiments we searched for $p^{(1)}(\alpha, N)$ by simulations, starting from some preliminary guess.

The rule $T1$ has simple interpretation. In the case when $M_{D(N)}$ rejects $\mathcal{H}^+$ on the level slightly smaller than the prescribed significance level α, the penalty is 0 and one applies the largest possible sum of squares of L_j^+'s, i.e. $Q_{D(N)}$. In the opposite case more careful selection of d in Q_d is done via some Akaike-type penalty. At first glance it would be more natural to have simply $M_{D(N)} \geq c_M(\alpha, D(N))$ in the 'switch' appearing in (6). However, in such case, there are numerical problems with solving (7) in practice, as we search for $p^{(1)}(\alpha, N)$ by simulations. Therefore, to have some flexibility, we took a slightly lower barrier. For more discussion see Appendix C.

In Appendix C we also show that the associated with $T1$ quadratic form Q_{T1} has the following useful property:

$$P\big(\, Q_{T1} > c \,|\, F \geq G \,\big) \leq P\big(\, Q_{T1} > c \,|\, F = G \,\big), \ \text{ for any } c \in \mathbb{R} \text{ and any natural numbers } m,\ n. \quad (8)$$

In particular, this property implies that, given the significance level α, it is enough to find the critical value of the data driven test by solving $P\big(\, Q_{T1} > c \,|\, F = G \,\big) \leq \alpha$, only. As in the case of (7), it can be done by simulations.

We investigated empirical behaviour of the tests based on $M_{D(N)}$, M_N^*, Q_T, and Q_{T1} under both moderate ($N = 300$) and large sample sizes ($N = 10000, N = 20000$). Results under $N = 300$ are postponed to Appendix B. In experiments we also included some alternative selection rules. To do not break the main stream of the presentation, we show and discuss these additional results in Appendix B. As mentioned earlier, they illustrate how delicate is the problem of selection of the number of components. They also demonstrate that by a careful choice of the penalty one can obtain more specialized tests focusing, for example, on more frequent detection of changes in the central part of $F - G$.

3. Simulation study under large N

As mentioned in the Introduction, the Kolmogorov-Smirnov statistic

$$KS = \sup_{x \in \mathbb{R}} \sqrt{\frac{mn}{N}} \Big\{\, G_n(x) - F_m(x) \,\Big\},$$

is a recognized standard to verify $\mathcal{H}^+$. Therefore, we shall compare here its empirical power to that of Q_T, Q_{T1}, $M_{D(N)}$, and M_N^*.

In all experiments $m = n$. Throughout the notation `nr` stands for the number of Monte Carlo (MC) runs. Table 1 gives critical values of the related tests on the level $\alpha = 0.01$, under $N = 10000$ and $N = 20000$. It is seen that the concentration of $T1$ on $d = 1$ (given in percentages) is close to 1 and smaller by less than 1% from that of $T = T0$. Finally, recall that, according to the terminology introduced in Section 2.1, the barrier c in the 'switch' $\{M_{D(N)} \geq c\}$ equals $-\sqrt{t(\alpha) \log N}$ in the case of $T0$ and $c_M(0.8\alpha, D(N))$ for $T1$. The values of this parameter are also displayed in Table 1. Note that $c_M(0.8\alpha, D(N))$ is obtained by simulation in additional preliminary MC experiment with `nr` $= 10000$.

Table 1
Simulated critical values and related parameters versus $N = m + n$, $m = n$, $\alpha = 0.01$, `nr` $= 10000$.

	$N = 10000,\ K(N) = 12,\ D(N) = 8191$					$N = 20000,\ K(N) = 13,\ D(N) = 16383$				
Statistic	KS	M_N^*	$M_{D(N)}$	Q_{T0}	Q_{T1}	KS	M_N^*	$M_{D(N)}$	Q_{T0}	Q_{T1}
Critical value	1.500	-3.662	-3.641	5.570	10.001	1.499	-3.686	-3.684	5.379	11.343
Penalty $p^{(r)}(\alpha, N)$				9.21	6.40				9.90	6.00
Barrier in the 'switch'				-4.238	-3.680				-4.103	-3.720
$\widehat{P}(Tr = 1 \vert F = G), r = 0, 1$				99.860	99.010				99.830	99.000

Since stochastic dominance is often discussed in application to income distributions, in this paper we investigated empirical powers mostly under some pairs of alternative distributions F and G, standard in the area. In particular, Gamma, log-normal, Pareto, and Singh-Maddala distributions are included. The alternatives, considered under $N = 10000$, are labeled by $\mathcal{A}_1, \ldots, \mathcal{A}_9$ and are formally described in Appendix A. In Figure 1 we present graphically the related pairs of distributions F and G along with the first 31 average values of $\hat{\gamma}_j = \sqrt{(N/mn)}L_j$ (over $\texttt{nr} = 10000$ MC runs) among $D(N) = 8191$ under consideration. The quantity $\hat{\gamma}_j$ can be interpreted as the jth empirical Fourier coefficient of the so-called contrast function in the system $\{l_j\}_{j\geq 0}$. For details see Section 2 of LW. For ease of interpretation of the empirical results also recall here that

$$L_j \approx \sqrt{\frac{mn}{N}} \frac{(F_m - G_n) \circ H_N^{-1}(a_j)}{\sqrt{a_j(1-a_j)}}$$

and significantly small negative values of L_j indicate $\mathcal{A}$. This shows that $\hat{\gamma}_j$'s are approximately weighted differences between F_m and G_n evaluated at the a_j-quantiles of the reference distribution function H_N. We restricted attention to the first 31 $\hat{\gamma}_j$'s to keep readability. The underlying distribution functions F and G are very smooth and do not cross many times. In consequence, small number of the coefficients well describes the nature of discrepancies between F and G.

To have some benchmark, the parameters of $\mathcal{A}_1, \ldots, \mathcal{A}_9$ were selected in such a way that empirical powers of $M_{D(N)}$ lay in (0.65,0.70). Figure 1 shows that several different situations are taken into account. Alternatives $\mathcal{A}_6 - \mathcal{A}_9$ represent some negative differences in one of the tails while the remaining ones refer to more centrally located negative discrepancies.

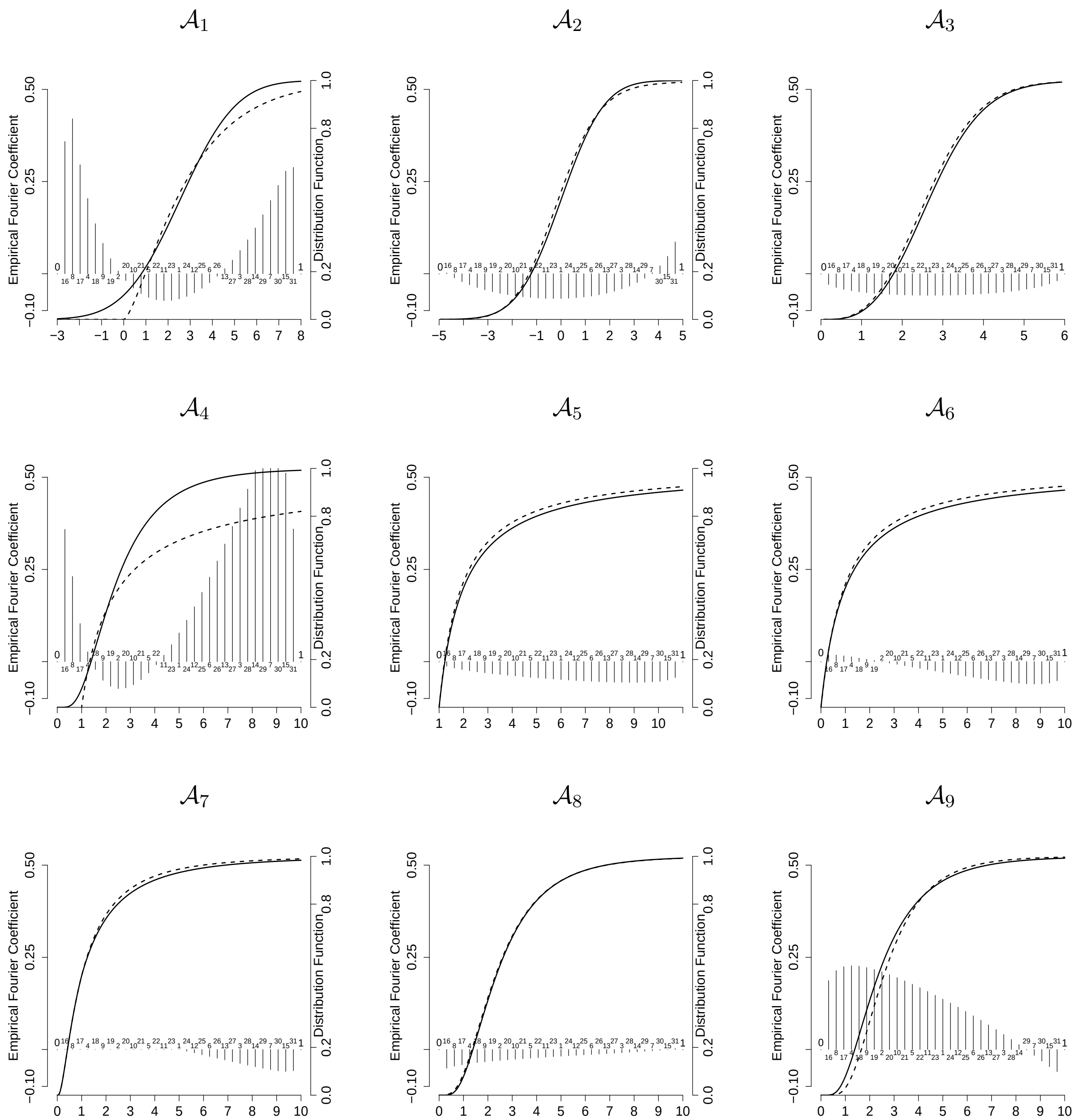

Fig. 1. F (—), G (- -), first 31 empirical Fourier coefficients $\hat{\gamma}_j$'s - vertical bars. $N = 10000$, $m = n$, $K(N) = 12$, $D(N) = 8191$, $\alpha = 0.01$, $\mathtt{nr} = 10000$.

The pertaining powers are collected in Table 2.

Table 2
Empirical powers, $N = 10000$, $m = n$, $K(N) = 12$, $D(N) = 8191$, $\alpha = 0.01$, $\mathtt{nr} = 10000$.

Test	*Alternative*									
	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9$	*average power*
KS	86.3	86.8	79.5	74.4	67.4	50.2	20.1	13.0	0.1	53.1
M_N^*	67.0	69.6	67.8	66.2	65.4	68.5	66.3	68.6	68.3	67.5
$M_{D(N)}$	67.7	70.3	68.6	66.8	66.5	69.3	67.0	69.0	68.8	68.2
Q_{T0}	85.0	84.2	74.0	42.9	61.8	46.7	38.9	40.2	40.4	57.1
Q_{T1}	69.1	72.4	72.6	65.6	67.7	67.7	65.4	67.1	67.1	68.3

It is seen that under central differences KS and Q_{T0} do very well, Q_{T1} does also very well while M_N^* and $M_{D(N)}$ are slightly weaker. Note also that the most difficult situation for the data driven statistics considered in this paper is such that $(F-G) \circ H^{-1}(a_j) > 0$ for central a_j's. Then the selection rules $T0$ and $T1$ have natural tendency to concentrate on low d's while the corresponding values of Q_d's are small. This causes some troubles with rejecting $\mathcal{H}^+$. In our simulation study the alternative $\mathcal{A}_4$ illustrates such a relatively difficult situation. In the real data analysis presented below we are also faced with similar circumstances in the case of comparison labeled by *1978 versus 1986*.

Under differences in extreme parts, KS looses its power, Q_{T0} is much better, while the three other statistics are the best and practically of the same power. The average powers collected in Table 2 reflect well the above discussed tendencies.

A similar picture follows from Table 3 in which $N = 20000$ and the parameters of $\mathcal{A}_1, \ldots, \mathcal{A}_9$ were changed to have again empirical powers of $M_{D(N)}$ in (0.65, 0.70). The related alternatives are denoted by $\mathcal{A}'_1, \ldots, \mathcal{A}'_9$. Their definitions are given in Appendix A. In view of new values of some parameters, possible displays related to $\mathcal{A}'_1, \ldots, \mathcal{A}'_9$ can be slightly different in shapes from these for $\mathcal{A}_1, \ldots, \mathcal{A}_9$ shown in Figure 1.

Table 3
Empirical powers, $N = 20000$, $m = n$, $K(N) = 13$, $D(N) = 16383$, $\alpha = 0.01$, $\texttt{nr} = 10000$.

Test	*Alternative*									
	$\mathcal{A}'_1$	$\mathcal{A}'_2$	$\mathcal{A}'_3$	$\mathcal{A}'_4$	$\mathcal{A}'_5$	$\mathcal{A}'_6$	$\mathcal{A}'_7$	$\mathcal{A}'_8$	$\mathcal{A}'_9$	*average power*
KS	87.1	86.3	81.0	73.8	69.8	48.3	20.9	10.7	0.1	53.1
M_N^*	67.5	67.1	68.0	66.1	67.2	65.2	65.9	67.6	69.0	67.1
$M_{D(N)}$	67.6	67.1	68.0	66.1	67.2	65.1	65.8	67.1	68.9	67.0
Q_{T0}	86.7	84.8	76.3	49.8	67.2	50.6	46.4	47.4	49.8	62.1
Q_{T1}	67.0	69.4	74.3	65.9	71.2	64.2	64.4	66.2	67.6	67.8

4. Real data analysis

We shall apply the discussed solutions for the income distribution comparison. The data concerns before and after tax incomes in Canada in 1978 and 1986 and comes from the Canadian Family Expenditure Survey. The two sets of data are denoted here by $Can(bt)$ and $Can(at)$, respectively. In these examples there are 8526 observations from 1978 and 9470 from 1986. These data, which were previously analyzed by Barrett and Donald (2003), were made available to us by Professor Garry F. Barrett.

The question, which was posed and investigated in the above mentioned paper, is : Do the income distributions improve over time, i.e. is the (unknown) distribution function G in 1986 less than or equal to F, corresponding to 1978, in both considered cases? This comparison is labeled in Tables V and VI of the aforementioned paper by *1986 versus 1978*. The reversed relation *1978 versus 1986* was also studied there. In this example $N = 17996$ and hence $K(N) = 13$ and $D(N) = 16383$. We start with the first comparison.

4.1. 1986 versus 1978

We indicate the structure of the data here by showing in Figure 2 the first 31 empirical Fourier coefficients $\hat{\gamma}_j$ for both sets $Can(bt)$ and $Can(at)$. Note that $\hat{\gamma}_j$'s are calculated now on the basis of the data at hand. The smallest $\hat{\gamma}_j$'s among $D(N) = 16383$ are $\hat{\gamma}_{1071} = -0.1064$ and $\hat{\gamma}_{4304} = -0.1065$, respectively. They result in $L_{1071} = -7.1269$ and $L_{4304} = -7.1336$. In view of the properties of L_j's, collected in Section 2.1, such outcomes are highly improbable when $\mathcal{H}^+$ is true. The whole data sets are summarized in Figure 2 by the respective curves of the form $\mathcal{E}_N(z) = [\{F_m - G_n\} \circ H_N^{-1}(z)]/\sqrt{z(1-z)},\ z \in (0,1)$.

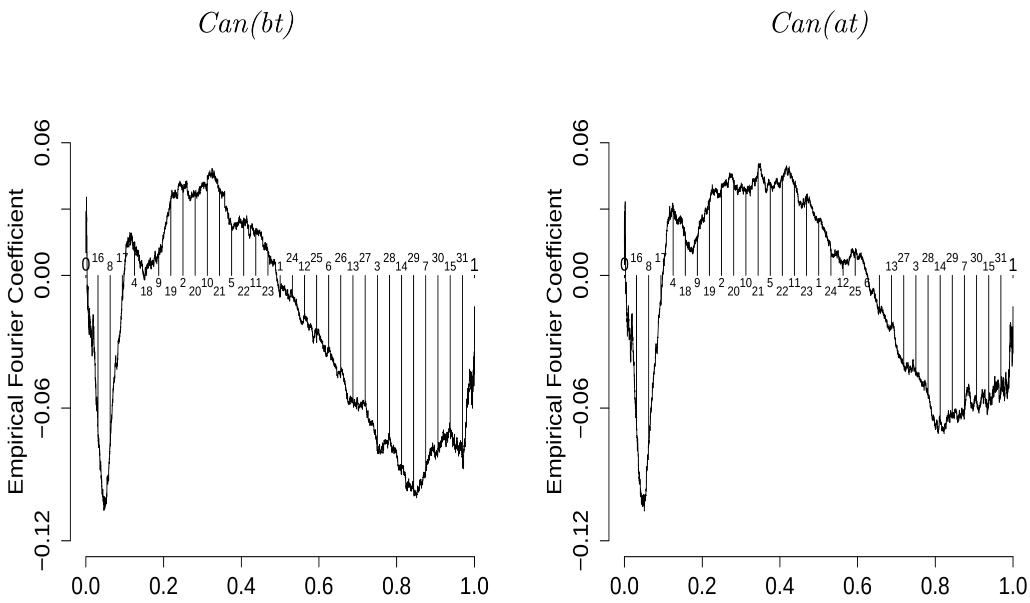

Fig. 2. First 31 empirical Fourier coefficients $\hat{\gamma}_j$'s for family income distributions comparison *1986 versus 1978* - vertical bars, the curves - respective values of $\mathcal{E}_N(z) = [\{F_m - G_n\} \circ H_N^{-1}(z)]/\sqrt{z(1-z)},\ z \in (0,1)$.

Critical values on the level $\alpha = 0.01$ and the respective parameters of the data driven solutions Q_{T0} and Q_{T1} are collected in Table 4. Table 5 shows the values of the investigated statistics and the pertaining selection rules.

Table 4

Canadian family income data, *1986 versus 1978*, $m = 9470$, $n = 8526$. Simulated critical values and related parameters of data driven tests. $\alpha = 0.01$, $\texttt{nr} = 10000$.

Statistic	*Critical value*	*Penalty* $p^{(r)}(\alpha, N)$	*Barrier in the 'switch'*
Q_{T0}	5.564	9.80	-4.021
Q_{T1}	10.790	5.50	-3.670

Table 5

Analysis of Canadian family income data, *1986 versus 1978*, $m = 9470$, $n = 8526$, $\texttt{nr} = 10000$.

			$Can(bt)$					$Can(at)$		
Significance level				$\alpha = 0.01$						$\alpha = 0.01$
Statistic	KS	M_N^*	$M_{D(N)}$	Q_{T0}	Q_{T1}	KS	M_N^*	$M_{D(N)}$	Q_{T0}	Q_{T1}
Value of statistic	2.469	-7.126	-7.128	178340.000	178340.000	1.888	-7.134	-7.134	94610.000	94610.000
p-value	0.000	0.000	0.000	—	—	0.001	0.000	0.000	—	—
Value of Tr, $r = 0, 1$	—	—	—	16383	16383	—	—	—	16383	16383

The null hypothesis is rejected by the data driven tests on the level as small as $\alpha = 0.01$ for both sets *Can(bt)* and *Can(at)*. p-values of other statistics show strong disagreement of the data with the null hypothesis as well. We are not calculating p-values of $Q_{Tr},\ r = 0, 1$, as these statistics depend on the prescribed significance level α and interpretation of such results would be unclear.

4.2. 1978 versus 1986

For comparison *1978 versus 1986* graphical representation of the data requires multiplying respective $\mathcal{E}_N$'s and all empirical Fourier coefficients in Figure 2 by -1. This, in particular, results in the case of $Can(bt)$ in $\hat{\gamma}_1 > 0$, $\hat{\gamma}_3 > 0$, and $\hat{\gamma}_5 < 0$, but small in magnitude. As mentioned in Section 3, such situation is particulary difficult to be detected by our data driven tests, especially when the remaining negative Fourier coefficients are relatively large. This is just the case. In the comparison *1978 versus 1986* the negative Fourier coefficients are larger than in the previous case and rejection of the related null hypotheses is not so easy, in general. To be specific, now the smallest empirical Fourier coefficients are $\hat{\gamma}_{2713} = -0.0484$ and $\hat{\gamma}_{11059} = -0.0506$ in the case of $Cat(bt)$ and $Can(at)$, respectively, while in the easier case, *1986 versus 1978*, they were twice smaller, approximately. This situation is reflected by p-values of KS, M_N^* and $M_{D(N)}$ which are equal to 0.009, 0.038, 0.038 in the case of $Can(bt)$ and 0.005, 0.024, 0.023 for $Can(at)$. In view of these results, we simulated critical values and related parameters of data driven tests on levels $\alpha = 0.01, \ldots, 0.05$ and collected them in Table 6.

Table 6

Canadian family income data, *1978 versus 1986*, $m = 8526$, $n = 9470$. Simulated critical values and related parameters of data driven tests versus α. $\texttt{nr} = 10000$.

Statistic $Q_{Tr}, r = 0, 1$			Q_{T0}					Q_{T1}		
α	0.01	0.02	0.03	0.04	0.05	0.01	0.02	0.03	0.04	0.05
Critical value	5.849	4.494	3.652	3.103	2.697	10.034	9.833	8.746	8.216	8.226
Penalty $p^{(r)}(\alpha, N)$	9.80	9.80	9.80	9.80	9.80	5.30	4.70	4.30	3.90	3.80
Barrier in the 'switch'	−4.081	-4.020	-4.020	- 4.081	-4.141	-3.736	-3.511	-3.376	- 3.281	- 3.209

Table 7

Analysis of Canadian family income data, *1978 versus 1986*, $m = 8526$, $n = 9470$, $\texttt{nr} = 10000$.

Statistic			Q_{T0}					Q_{T1}		
α	0.01	0.02	0.03	0.04	0.05	0.01	0.02	0.03	0.04	0.05
					$Can(bt)$					
Value of statistic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	21910.650
Value of Tr, $r = 0, 1$	1	1	1	1	1	1	1	1	1	16383
					$Can(at)$					
Value of statistic	2.651	2.651	2.651	2.651	2.651	2.651	2.651	38758.124	38758.124	38758.124
Value of Tr, $r = 0, 1$	1	1	1	1	1	1	1	16383	16383	16383

Table 7 shows values of the data driven statistics and selection rules under these α's. It is seen that in the case of $Can(bt)$ the statistic Q_{T0} accepts the null hypothesis on all the listed levels while Q_{T1} rejects $\mathcal{H}^+$ on the level 0.05. The case $Can(at)$ is slightly easier for Q_{T1} as it rejects $\mathcal{H}^+$ already on the level 0.03. Q_{T0} still accepts the null hypothesis on all the considered levels. This shows that, in the above difficult circumstances, Q_{T1} is more sensitive than Q_{T0} and comparable in strength to M_N^* and $M_{D(N)}$. As mentioned earlier, the situations under consideration resemble that under the alternative $\mathcal{A}_4$. Under such conditions the KS statistic is doing very well and it is also manifested in this example.

5. Conclusions and remarks

This article proposes a new data driven test for testing stochastic dominance. The new solution is compared with some recent constructions introduced in Ledwina and Wyłupek (2012a) and standard in the area the Kolmogorov-Smirnov test. Extensive simulations done under moderate and large sample sizes show that the new and recent solutions are competitive to the Kolmogorov-Smirnov test.

It is worth noting that two simple constructions $M_{D(N)}$ and M_N^* work very well. The constructions are based on minima of some weighted two-sample processes. The weights are such that the resulting tests have similar sensitivity in detecting differences between F and G in each point of the ranges of $(F-G)\circ H^{-1}$ and $F-G$, respectively. The construction of $M_{D(N)}$ allows for slightly more parsimonious calculations and is consistent implementation of the idea of Anderson (1996) to use intersection-union tests for inequality constraints in income comparisons.

The new data driven test based on Q_{T1} is much more complicated than M_N^* and $M_{D(N)}$ and comparable to them in average power. Other data driven constructions, more focused on detecting centrally located discrepancies are discussed as well. Though Q_{T1} is relatively complex, we find it to be interesting. First, as mentioned earlier, this construction is relatively simple to calculate and consistent substitute for the Wald-type statistic introduced in Davidson and Duclos (2000). In particular, we solved the most delicate problem related to deciding on in which and in how many points to control the difference $F-G$. In our construction it is done by some carefully designed selection rule. Since, as the sample size is growing, the selection rule is checking more and more dense sets of candidate points, the resulting test is consistent against very large set of alternatives. So, this construction solves some problem already noticed. In contrast to Q_T, the test based on Q_{T1} distributes its power more uniformly over the space of alternatives. On the other hand, one can consider a dual problem of testing of lack of stochastic dominance against presence of the dominance. This is also an important question. As shown in Ledwina and Wyłupek (2012b) and Wyłupek (2013) statistics of the Kolmogorov-Smirnov type are not adequate then while carefully elaborated data driven tests work nicely. Therefore, the new construction can also be stimulating in such a context.

Appendix A. Description of alternatives

To define alternatives $\mathcal{A}_1,\ldots,\mathcal{A}_9$ and $\mathcal{A}'_1,\ldots,\mathcal{A}'_9$ we shall use the following distributions.

$P(a)$ is an abbreviation of $Pareto(a)$, $a \geq 1$, and stands for the distribution with cdf given by $1-(1/x)^a$, $x > 1$.

$N(a,b)$, $a \in \mathbb{R}$, $b > 0$, is defined by the density $\exp\{-(x-a)^2/(2b^2)\}/\{\sqrt{2\pi}b\}$, $x \in \mathbb{R}$.

χ_k^2 stands for the chi-squared distribution with k degrees of freedom and the density $\exp\{-x/2\}x^{(k-2)/2}/\{2^{k/2}\Gamma(k/2)\}$, $x > 0$.

$LN(a,b)$, $a \in \mathbb{R}$, $b > 0$, denotes the log-normal distribution with the density $\exp\{-(\log x - a)^2/(2b^2)\}/\{\sqrt{2\pi}\, bx\}$, $x > 0$.

$LNC(a,b)$, $a > 0$, $b > 0$, is the distribution of a random variable Z defined by two cdf's of $LN(\cdot,\cdot)$ in the following way $P(\, Z \le z\,) = \{\, LN(0,a)(z)\,\}\, \mathbb{1}(z \le 1) + \{\, LN(0,b)(z)\,\}\, \mathbb{1}(z > 1)$, $z \in \mathbb{R}$.

$SM(a,b,c)$ stands for *Singh–Maddala*(a,b,c) distribution with parameters $a \ge 1$, $b \ge 1$, $c \ge 1$ which obeys cdf $1 - [1 + (x/b)^a]^{-c}$, $x > 0$, cf. Klonner (2000).

$S(a,b)$ is shortening of *Stable*(a,b), $a \in (0,2]$, $b \in [-1,1]$, which stands for $S_a(1,b,0)$ distribution according to the notation of Samorodnitsky and Taqqu (1994), cf. pp. 5, 9 and 35-38.

As in LW, each alternative $\mathcal{A}_j$ $[\mathcal{A}'_j]$ is described by the pair of distributions F and G as follows: F/G. With such notations we have

j	$\mathcal{A}_j$	$\mathcal{A}'_j$
1	$N(2.545, 2)/\chi^2_3$	$N(2.493, 2)/\chi^2_3$
2	$N(0, \sqrt{2})/S(1.875, 1)$	$N(0, \sqrt{2})/S(1.910, 1)$
3	$SM(3, 5, 5)/SM(2.927, 4.927, 5)$	$SM(3, 5, 5)/SM(2.948, 4.948, 5)$
4	$LN(0.85, 0.60)/P(0.745)$	$LN(0.85, 0.60)/P(0.716)$
5	$P(1)/P(1.073)$	$P(1)/P(1.052)$
6	$SM(1, 1, 1)/SM(1.044, 1, 1.044)$	$SM(1, 1, 1)/SM(1.030, 1, 1.030)$
7	$LNC(1, 1.080)/LN(1, 1)$	$LNC(1, 1.056)/LN(1, 1)$
8	$LN(0.85, 0.6)/0.953LN(0.85, 0.6) + 0.047LN(0.4, 0.9)$	$LN(0.85, 0.6)/0.97LN(0.85, 0.6) + 0.03LN(0.4, 0.9)$
9	$LN(0.85, 0.60)/LN(0.945, 0.505)$	$LN(0.85, 0.60)/LN(0.919, 0.531)$

Appendix B. Three additional selection rules and related results

The three variants $T2$, $T3$, and $T4$ we introduce here have the common form

$$Tr = \min\big\{\, d : Q_d - d\pi^{(r)}(\alpha, N) \ge Q_j - j\pi^{(r)}(\alpha, N),\ d, j \in \mathcal{D}(N)\,\big\}, \tag{9}$$

for $r = 2, 3, 4$. The penalties $\pi^{(r)}(\alpha, N)$ for the parameter d are as follows.
First

$$\pi^{(2)}(\alpha, N) = \begin{cases} p^{(2)}(\alpha, N) & \text{if} \quad M_{D(N)} \ge c_M(\alpha, D(N)), \\ 0 & \text{otherwise}, \end{cases} \tag{10}$$

where $p^{(2)}(\alpha, N)$ is the smallest positive number such that, under $F = G$, the concentration of $T2$ on $d = 1$ is at least $1 - \alpha - 0.01$ i.e.

$$P(T2 = 1|F = G) \ge 1 - \alpha - 0.01.$$

Second

$$\pi^{(3)}(\alpha, N) = \begin{cases} p^{(3)}(\alpha, N) & \text{if} \quad M_{D(N)} \ge c_M(\alpha, D(N)), \\ 2 & \text{otherwise}, \end{cases} \tag{11}$$

where $p^{(3)}(\alpha, N)$ is the smallest positive number such that, under $F = G$, the concentration of $T3$ on $d = 1$ is at least $1 - \alpha$ i.e.

$$P(T3 = 1|F = G) \ge 1 - \alpha. \tag{12}$$

Finally,

$$\pi^{(4)}(\alpha, N) = \begin{cases} p^{(4)}(\alpha, N) & \text{if} \quad M_{D(N)} \ge c_M(\alpha, D(N)), \\ 2 & \text{otherwise}, \end{cases}$$

where $p^{(4)}(\alpha, N)$ is defined in the spirit of Ledwina and Wyłupek (2013). To be specific, set

$$A_c = \min\{\, d : Q_d - dc \geq Q_j - jc,\ d, j \in \mathcal{D}(N) \,\}.$$

Then $p^{(4)}(\alpha, N)$ is the smallest positive number c such that, under $F = G$, the concentration of A_c on $d = 1$ is at least $1 - \alpha$ i.e.

$$P(A_c = 1|F = G) \geq 1 - \alpha.$$

Table 8

Simulated critical values and related parameters, $N = 300$, $m = n$, $K(N) = 7$, $D(N) = 255$, $\alpha = 0.01$, $\mathtt{nr} = 100000$.

Statistic	KS	M_N^*	$M_{D(N)}$	Q_{T0}	Q_{T1}	Q_{T2}	Q_{T3}	Q_{T4}
Critical value	1.501	-3.371	-3.367	6.453	10.453	551.785	9.013	28.756
Penalty $p^{(r)}(\alpha, N)$				5.70	5.20	3.70	5.50	4.20
Barrier in the 'switch'				-3.502	-3.433	-3.352	-3.352	-3.352
$\widehat{P}(Tr = 1\|F = G), r = 0, \dots, 4$				99.287	99.004	98.006	99.008	98.580

Table 8 contains critical values of the tests based on KS, M_N^*, and $M_{D(N)}$, under the level $\alpha = 0.01$ and $N = 300$, along with that corresponding to Q_{Tr}, $r = 0, \dots, 4$. The parameters of Q_{Tr} are also given there. Observe that the rule $T2$ contains natural barrier in the 'switch' but allows instead for slightly smaller concentration of $T2$ on $d = 1$ than in the case of $T1$. Such strategy results in penalty smaller than $p^{(1)}(\alpha, N)$ but causes an explosion of the pertaining critical value. The rule $T3$ also has built-in a natural barrier in the 'switch' and, to protect against possible troubles with simultaneous solving (11) and (12), the Akaike's penalty 2 imposed in the case when the 'switch' works. This construction results in relatively large penalty $p^{(3)}(\alpha, N)$ and moderately large critical value. The rule $T4$ can be considered to be intermediate between $T2$ and $T3$.

Obviously, the definitions of Tr, $r = 2, 3, 4$, also have influence into the empirical powers of the resulting data driven Q_{Tr}. Table 9 contains empirical powers under alternatives $\mathscr{A}_1 - \mathscr{A}_{12}$ defined and studied in LW. We refer to this paper for definitions, empirical Fourier coefficients and other displays. Here note only that $\mathscr{A}_1 - \mathscr{A}_3$ correspond to differences in central part of distributions F and G while $\mathscr{A}_4 - \mathscr{A}_{12}$ represent several discrepancies closer to the tails of F and G.

Table 9

Empirical powers, $N = 300$, $m = n$, $K(N) = 7$, $D(N) = 255$, $\alpha = 0.01$, $\mathtt{nr} = 5000$.

Test	*Alternative*											
	$\mathscr{A}_1$	$\mathscr{A}_2$	$\mathscr{A}_3$	$\mathscr{A}_4$	$\mathscr{A}_5$	$\mathscr{A}_6$	$\mathscr{A}_7$	$\mathscr{A}_8$	$\mathscr{A}_9$	$\mathscr{A}_{10}$	$\mathscr{A}_{11}$	$\mathscr{A}_{12}$
KS	24.8	41.7	24.3	77.5	11.6	1.3	18.0	1.7	0.6	11.9	18.6	1.6
M_N^*	29.5	32.9	15.1	90.0	50.4	79.7	36.6	50.9	65.0	40.5	78.6	63.7
$M_{D(N)}$	29.7	32.8	14.9	90.1	50.3	79.7	36.5	50.4	64.9	40.3	78.2	63.7
Q_{T0}	33.6	43.8	29.8	86.6	42.5	70.0	31.8	41.6	54.8	33.2	75.9	52.1
Q_{T1}	31.8	32.3	16.0	88.6	47.4	75.9	33.9	47.1	62.5	36.6	77.3	63.1
Q_{T2}	35.7	35.6	8.7	83.8	35.5	17.7	33.0	8.0	3.4	29.1	47.1	13.8
Q_{T3}	32.7	36.2	17.9	88.0	42.4	37.4	34.7	15.7	8.1	34.3	55.2	18.8
Q_{T4}	34.6	34.8	9.7	86.4	41.0	33.8	34.5	13.7	7.0	32.8	53.5	18.2

It is seen that Q_{T0} is better than Q_{T1} in detecting $\mathscr{A}_1 - \mathscr{A}_3$ while in the remaining situations the relation is reversed. The additional variants Q_{Tr} also work nicely for $\mathscr{A}_1 - \mathscr{A}_3$ but in some other cases have considerable break down. This discussion shows that proper tuning of the penalty is a very delicate problem. Finally, observe that M_N^* and $M_{D(N)}$ have stable overall power while KS appears to be very weak in some cases.

Some weaknesses of Q_{T2}, Q_{T3}, and Q_{T4} disappear under large N. However, these statistics still show deficiency in detecting changes in tails even for N as large as 10000. For illustration we show in Table 10 their empirical powers under alternatives $\mathcal{A}_1 - \mathcal{A}_9$; cf. Figure 1 and Table 2. In Table 11 we give simulated critical values and the corresponding parameters of these tests.

Table 10
Empirical powers of Q_{T2}, Q_{T3}, and Q_{T4}, $N = 10000$, $m = n$, $K(N) = 12$, $D(N) = 8191$, $\alpha = 0.01$, $\texttt{nr} = 10000$.

Test	*Alternative*									
	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9$	*average power*
Q_{T2}	67.8	78.3	83.6	67.0	77.7	70.9	66.4	62.2	32.8	67.4
Q_{T3}	71.1	78.6	82.6	72.7	76.5	69.2	55.7	43.2	3.4	61.4
Q_{T4}	67.0	79.2	83.9	57.4	77.6	69.0	55.3	43.1	3.4	59.5

Table 11
Simulated critical values and related parameters of Q_{T2}, Q_{T3}, and Q_{T4}, $N = 10000$, $m = n$, $K(N) = 12$, $D(N) = 8191$, $\alpha = 0.01$, $\texttt{nr} = 10000$.

Statistic	*Critical value*	*Penalty* $p^{(r)}(\alpha, N)$	*Barrier in the 'switch'*	$\widehat{P}(Tr = 1 \mid F = G)$
Q_{T2}	7751.138	3.80	-3.637	98.060
Q_{T3}	9.244	4.70	-3.637	99.020
Q_{T4}	12.922	4.30	-3.637	98.790

Appendix C. Some justifications

Below, we shall discuss some basic properties of the new solutions, which are partially based on our earlier results. For this purpose it is convenient to introduce the same notations as in Section 4 of LW.

Let $\mathcal{D} = \{ 2^{k+1} - 1 : k = 0, 1, \ldots, K \}$, where K is a given number while $D = 2^{K+1} - 1$. Moreover, $M_D = \min_{1 \le j \le D} L_j$. In this setting, consider auxiliary selection rule T^* given by

$$T^* = \min\big\{ d : Q_d - d\pi^* \ge Q_j - j\pi^*,\ d, j \in \mathcal{D} \big\} \tag{13}$$

with

$$\pi^* = \begin{cases} p^* & \text{if} \quad M_D \ge c^*, \\ 0 & \text{otherwise}, \end{cases} \tag{14}$$

where $p^* > 0$ and p^* and c^* are arbitrary otherwise.

By the same argument as used in the case of Q_T in the proof of Lemma A.1 and Corollary A.1 in LW we get

$$P\big(Q_{T^*} > c \,|\, F \ge G \big) \le P\big(Q_{T^*} > c \,|\, F = G \big), \text{ for any } c \in \mathbb{R} \text{ and any natural numbers } m,\ n.$$

Remark C.1. *Taking above $K = K(N)$ and $D = D(N)$, as defined in Section 2.1, and setting $c^* = c_M(0.8\alpha, D(N))$ the inequality (8) follows.*

Two our next conclusions shall be of asymptotic nature. Therefore, we consider $K = k(N)$ and $D = d(N)$, where, as in LW, $d(N) \to \infty$ as $N \to \infty$ and $d(N) = o(N)$. Note that $d(N)$ tends to infinity slower than $D(N)$, we took in our constructions and simulations. Our choice of $D(N)$ was primary made to have simple definitions and to avoid some discussions like for instance: which $d(N)$ of the form N^δ, $\delta \in (0,1)$, to take?

For the above defined $K = k(N)$, $D = d(N)$, $\mathcal{D} = \{\, 2^{k+1} - 1 : k = 0, 1, \ldots, k(N)\,\}$, consider T^* with $c^* = c_M^* = c_M(0.8\alpha, d(N))$ and the problem of solving

$$P(T^* = 1|F = G) \geq 1 - \alpha \tag{15}$$

with respect to p^*, cf. (13) and (14).

C.1. On solving (7) under the above conditions

Under the above notations and restrictions, solving (7) reduces to solving (15).

Proposition C.1. *Under the above assumptions, for any $\alpha \in (0,1)$ and for N large enough, there exists the smallest p^* for which (15) holds.*

Proof. Let us start with indicating a particular p^* solving (15). For this purpose set $\mathcal{S} = \{M_{d(N)} \geq c_M^*\}$ with c_M^* as above and denote by $\mathcal{S}^c$ the complement of $\mathcal{S}$. It holds that

$$P(T^* > 1|F = G) = P(\{T^* > 1\} \cap \mathcal{S}|F = G) + P(\{T^* > 1\} \cap \mathcal{S}^c|F = G) \tag{16}$$

$$\leq P(\{T^* > 1\} \cap \mathcal{S}|F = G) + 0.8\alpha. \tag{17}$$

Moreover,

$$P(\{T^* > 1\} \cap \mathcal{S}|F = G) = P\Big(\bigcup_{d \in \mathcal{D} \setminus \{1\}} \Big\{ \sum_{j=2}^{d} [L_j^+]^2 > (d-1)p^* \Big\} \cap \mathcal{S}|F = G\Big). \tag{18}$$

If $c_M^* \geq 0$ then $L_j^+ = 0$, $j = 2, \ldots, d(N)$, and (18) equals 0 for any positive p^*. In the opposite case, again the form of $\mathcal{S}$ implies that (18) equals 0 for $p^* = [c_M^*]^2$. This shows that if $p^* = [c_M^*]^2$ then

$$P(T^* = 1|F = G) \geq 1 - 0.8\alpha. \tag{19}$$

Now observe that the second component in (16) is independent from p^*. By (18), the first component of (16) increases when p^* decreases. Therefore, the left-hand side of (19) decreases when p^* decreases. Moreover, the expression in (18) is right-continuous function of p^*. Besides, for $p^* = 0$, we have

$$P(T^* > 1|F = G) = P\Big(\sum_{j=2}^{d(N)} [\max\{-L_j, 0\}]^2 > 0|F = G\Big) = P(\min_{1 \leq j \leq d(N)} L_j < 0|F = G). \tag{20}$$

However, $\min_{1 \leq j \leq d(N)} L_j \to -\infty$; cf. Appendix B.4 in Ledwina and Wyłupek (2012b). Therefore, the right-hand side of (20) tends to 1 as $N \to \infty$.

Remark C.2. Obviously, the above proof works in the case when $c_M^* = c_M(0.8\alpha, d(N))$ is replaced by $c_M^{**} = c_M(\alpha, d(N))$. In practice, however, the values of c_M^* and c_M^{**} are simulated. Therefore, in view of restricted accuracy of simulation experiments it may happen that the estimated by MC probability of $\{M_D \geq c_M^{**}\}$ shall be slightly smaller than $1-\alpha$, what makes such more natural solution difficult to apply, as then ensuring the counterpart of (7) is impossible. More precisely, the possible (additional in view of theoretical results) small fraction of outcomes for which the penalty is allowed to be 0 makes $T1$ stochastically too large.

Remark C.3. Throughout we simulate quantiles using the function `quantile` in the program R with default settings. In particular, c_M^{**} is defined as the largest value $c_M(\alpha)$ such that $P(M_D < c_M(\alpha)|F = G) \leq \alpha$. To solve a problem of possible tied observations, we extend the definition of L_j as follows $L_j = -\sqrt{n/mN}\,\sum_{i=1}^m l_j(H_N(X_i) - 1/(2N)) + \sqrt{m/nN}\,\sum_{i=1}^n l_j(H_N(Y_i) - 1/(2N))$.

C.2. Consistency of the test based on Q_{T^*}

We still require $d(N) = o(N)$ and $d(N) \to \infty$ as $N \to \infty$. In such setting LW proved consistency of the tests based on M_D and Q_T, provided that Assumption 1 in LW was satisfied. Below, we show that then the test based on Q_T^* is consistent as well. For formulation and discussion of this assumption see LW, pp. 735-736. It shows that Assumption 1 is not very restrictive one.

Proposition C.2. *If T^* is given by (13) and (14) with $c^* = c_M^* = c_M(0.8\alpha, d(N))$ then the test rejecting $\mathcal{H}^+$ in favour of $\mathcal{A}$ for large values of Q_{T^*} is consistent.*

Proof. By (3) of Lemma A.1 in LW and Lemma 5.9.1 of Lehmann and Romano (2005), $P(Q_D > c|F \geq G) \leq P(Q_D > c|F = G),\ c \in \mathbb{R}$. Besides, under $F = G,\ Q_D = O_P(D)$. Hence, under $\mathcal{H}^+$,

$$Q_{T^*} = O_P(d(N)). \tag{21}$$

As before, $\mathcal{S} = \{M_{d(N)} \geq c_M^*\}$. Let j_0 be the smallest index j such that $\gamma_j < 0$. Since on $\mathcal{S}^c$ the penalty π^* equals 0 then

$$P(T^* \geq j_0|\mathcal{A}) \geq P(\{T^* \geq j_0\} \cap \mathcal{S}^c|\mathcal{A}) \to 1 \text{ as } N \to \infty. \tag{22}$$

Let q stands for the critical value of the data driven test based on Q_{T^*}. Then, by (22), for N sufficiently large

$$P(Q_{T^*} \geq q|\mathcal{A}) \geq P(Q_{j_0} \geq q|\mathcal{A}) \geq P([L_{j_0}^+]^2 \geq q|\mathcal{A}).$$

However, L_{j_0} is, under $\mathcal{A}$, asymptotically normal with mean $\Delta_N(j_0) = O(\sqrt{N})$; cf. Appendices A and B in Ledwina and Wyłupek (2012b) for details. Hence, the consistency follows.

Acknowledgments

We are deeply grateful to Prof G. F. Barrett for the set of data as well as to Prof. E. Rafajłowicz and Dr. M. Tykierko for advice on speeding up the computer calculations. Calculations have been carried out in Wrocław Centre for Networking and Supercomputing (`http://www.wcss.wroc.pl`) under grant No. 199. Our work was supported by grant N N201 608440 from the National Science

Centre, Poland.

References

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

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

Canner, P.L. (1975). A simulation study of one- and two-sample Kolmogorov-Smirnov statistics with a particular weight function. *Journal of the American Statistical Association* 70:209-211.

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

Eicker, F. (1979). The asymptotic distribution of the suprema of the standardized empirical process. *Annals of Statistics* 7:116-138.

Klonner, S. (2000). The first-order stochastic dominance ordering of the Singh-Maddala distribution. *Economics Letters* 69:123-128.

Ledwina, T., Wyłupek, G. (2012a). Nonparametric tests for stochastic ordering. *TEST* 21:730-756.

Ledwina, T., Wyłupek, G. (2012b). Two-sample test against one-sided alternatives. *Scandinavian Journal of Statistics* 39:358-381.

Ledwina, T., Wyłupek, G. (2013). Detection of non-Gaussianity. *Submitted.*

Lehmann, E.L., Romano, J.P. (2005). *Testing Statistical Hypotheses.* New York: Springer.

Levy, H. (2006). *Stochastic Dominance: Investment Decision Making Under Uncertainty.* New York: Springer.

Müller, A., Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Chichester: Wiley.

Robertson, T., Wright, F.T., Dykstra, R.L. (1988). *Order Restricted Statistical Inference.* New York: John Wiley and Sons.

Samorodnitsky, G., Taqqu, M. (1994). *Stable non-Gaussian Random Processes.* New York: Chapman and Hall.

Sen, P.K., Silvapulle, M. (2005). *Constrained Statistical Inference.* Hoboken: Wiley.

Shaked, M., Shanthikumar, J.G. (1994). *Stochastic Orders and their Applications.* Boston: Academic Press.

Wyłupek, G. (2013). Data-driven tests for trend. *Communications in Statistics-Theory and Methods* 42:1406-1427.