AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 315–324

Permutation Tests for Univariate and Multivariate Ordered Categorical Data

Fortunato Pesarin and Luigi Salmaso
University of Padova, Italy

Abstract: In this paper, we provide solutions for univariate and multivariate testing problems with ordered categorical variables by working within the nonparametric combination of dependent permutation tests (see Pesarin, 2001). Two applications and Monte Carlo simulations for power comparisons of NPC solutions to most competitors from the literature are shown.

Keywords: Restricted Alternatives, Univariate and Multivariate Response.

1 Introduction

Problems of testing for ordered categorical variables are of great interest in many application disciplines, where a finite number of $Q \geq 1$ of such variables are observed on each individual unit. Testing of hypotheses for multivariate ordered categorical variables is known to be quite a difficult problem especially when testing for stochastic dominance, that is for a set of restricted alternatives. Stochastic dominance problems present peculiar difficulties, especially within the framework of likelihood ratio tests (see e.g. Cohen et al., 2003; Silvapulle and Sen, 2005). Several solutions have been proposed for the univariate case, most of which are based on the restricted maximum likelihood ratio test. These solutions are generally criticized, because their asymptotic null and alternative distributions are mixtures of chi-squared variables the weights of which essentially depend on underlying population distribution F and so the related degree of accuracy is difficult to assess and to characterize; thus their use when F is unknown is somewhat questionable in practice. Moreover, due to the extreme difficulty of modelling the related likelihood function (see Joe, 1997), the multivariate case is considered as almost impossible to be analyzed within the likelihood ratio approach, especially when the number Q of observed variables is larger than two and when there are more than two samples to analyze.

By working within the nonparametric combination of dependent permutation tests (NPC; see Pesarin, 2001), it is possible to find exact solutions to that kind of problems. The NPC approach works as a general methodology for most multivariate situations, as for instance in cases where sample sizes are smaller than the number of observed variables, or in some cases where there are non-ignorable missing values, or when some of the variables are categorical (ordered and nominal) and others are quantitative, and in many other complex situations. In particular, it is of interest when testing for a set of restricted alternatives in which context it shows a specific efficacy.

Section 2 examines the basic problem for stochastic dominance alternatives in the one-dimensional two-sample design and presented a brief review of NPC approach; Section 3 considers some multivariate extensions; Section 4 is devoted to the discussion of two application examples; in Section 5 there are some Monte Carlo simulations for power comparisons of NPC solutions to most competitors from the literature; Section 6 is devoted to some concluding remarks.

2 The Univariate Two-Sample Basic Problem

Let us assume that the support of a univariate non-degenerate ordered categorical variable X is partitioned into a finite number $K \geq 2$ of ordered classes $(A_1, A_2, \ldots, A_K)$, and that the data are classified according to two levels of a treatment, giving rise to a typical two-sample design. Thus, given two independent samples of respectively $n_j > 2$, $j = 1, 2$, independent and identically distributed (iid) observations, $X_j = \{X_{ji}, i = 1, \ldots, n_j\}$ say, we want to test $H_0 : \{X_1 \stackrel{d}{=} X_2\} = \{F_1(A_k) = F_2(A_k),\ k = 1, \ldots, K-1\}$, against $H_1 : \{X_1 \stackrel{d}{>} X_2\} = \{F_1(A_k) \leq F_2(A_k),\ k = 1, \ldots, K-1\}$, where at least one inequality is strict and the function $F_j(A_k) = \Pr\{X_j \leq A_k\}$ plays the role of cumulative distribution function (CDF) for X_j, $j = 1, 2$. By assuming that no reverse inequality such as $F_1(A_k) > F_2(A_k)$, $k = 1, \ldots, K-1$, is possible, the alternative can also be written in the form $H_1 : \{\cup_{k=1}^{K-1}[F_1(A_k) < F_2(A_k)]\}$. Observed data are generally organized in a $2 \times K$ contingency table such as $\{f_{jk} = \sum_{i \leq n_j} I(X_{ji} \in A_k),\ k = 1, \ldots, K,\ j = 1, 2\}$, where $I(\cdot) = 1$ if event $(\cdot)$ occurs and 0 otherwise. Symbols $N_{jk} = \sum_{s \leq k} f_{js}$, $n_j = N_{jK}$, and $f_{\cdot k} = f_{1k} + f_{2k}$ indicate cumulative and marginal frequencies, respectively.

Permutation analysis is much easier if, in place of usual contingency tables, data are unit-by-unit represented by listing the $n = n_1 + n_2$ individual records. In the $2 \times K$ design, the data set is represented by $X = \{X(i),\ i = 1, \ldots, n;\ n_1, n_2\}$, where it is intended that the first n_1 records belong to the first sample and the rest to the second. Sometimes we also use symbol X to denote the pooled data set. Thus, if $(u_1^*, \ldots, u_n^*)$ indicates a permutation of individual units $(1, \ldots, n)$, then $X^* = \{X(u_i^*),\ i = 1, \ldots, n;\ n_1, n_2\}$ indicates the corresponding permutation of the data set X. It is worth observing that in univariate two-sample designs, since they contain exactly the same amount of information on F, the marginal frequencies $\{n_1, n_2, f_{\cdot 1}, \ldots, f_{\cdot K}\}$, the pooled data set X as well as any of its permutations X^* are equivalent sets of sufficient statistics under H_0. In multivariate problems, due to the well-known difficulty of expressing marginal frequencies in easy to read ways, as a set of sufficient statistics we only consider pooled data set $X = \{X(i),\ i = 1, \ldots, n;\ n_1, n_2\}$, or any of its permutations $X = \{X(u_i^*),\ i = 1, \ldots, n;\ n_1, n_2\}$, where $X(i) = [X_1(i), \ldots, X_Q(i)]'$ is the vector of $Q \geq 1$ responses for ith individual unit. Also note that the assumed iid condition implies that in H_0 the data of two samples are exchangeable, so that the permutation testing principle can be applied. In this framework, as a solution to the two-sample one dimensional testing problem we may consider the permutation test statistic:

$$T_{AD}(X^*) = T_{AD}^* = \sum_{k=1}^{K-1} (\hat{F}_{2k}^* - \hat{F}_{1k}^*) \left[\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k}) \frac{4n_1}{n_2(n-1)} \right]^{-1/2},$$

where $\hat{F}_{jk}^* = N_{jk}^*/n_j$, $j = 1, 2$, $\bar{F}_{\cdot k} = N_{\cdot k}/n$ are permutation and marginal empirical distribution functions (EDFs) respectively, N_{1k}^*, N_{2k}^*, $k = 1, \ldots, K-1$ are permutation cumulative frequencies. Note that large values for T_{AD} are significant. Statistic T_{AD} essentially compares two EDFs and corresponds to the discrete version of a statistic following Cramér-von Mises' two-sample goodness-of-fit test statistic for stochastic dominance alternatives, adjusted according to Anderson-Darling. T_{AD} is permutationally equivalent to $\sum_k (\hat{F}_{2k}^* - \hat{F}_{1k}^*)[\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k})]^{-1/2}$ in the sense that for each data set X they induce exactly the same p-value. The p-value is defined as $\lambda_{AD} = \Pr\{T^*_{AD} \geq T^o_{AD}|X\}$, where T^o_{AD} represents the observed value of T_{AD}. Thus, according to the general testing rule, if $\lambda_{AD} \leq \alpha$ the null hypothesis is rejected at significance level $\alpha > 0$.

Testing analysis by NPC methods requires that problems can be broken down into a set of simpler sub-problems for each of which a permutation partial test is available and that these partial tests can be jointly processed. Let us observe that the kth summand in T_{AD} may be seen as a partial test for the sub-hypotheses $H_{0k} : \{F_k = G_k\}$ and $H_{1k} : \{F_k < G_k\}$, $k = 1, \ldots, K-1$. Thus the global hypotheses may equivalently be written as $H_0 : \cap_k H_{0k}$ and $H_1 : \cup_k H_{1k}$, where a suitable break down into a set of partial sub-hypotheses is emphasized. Hence, in order to obtain an overall solution one way is to properly combine all related partial results. Of course, these partial tests and associated p-values are dependent in a way that in general is extremely difficult to take into account explicitly, consequently, when considering their combination we shall take account nonparametrically of their underlying dependence relations; hence we shall work within the NPC approach. Theory and methods for this kind of solutions are fully discussed in Pesarin (2001). In the NPC approach we need combining the p-values λ_k associated with partial tests by a non-degenerate and measurable combining function ψ (for example, Fisher's: $T''_F = -\sum_k \log(\lambda_k)$ and Tippett's: $T''_T = \max_k(1-\lambda_k)$). As a further, frequently used combining function there is the so-called direct combination consisting in a function of partial test statistics instead of related p-values (for example $T''_D = \sum_k T_k$, where T_k denotes the partial test). Thus, T_{AD} may be seen as a direct nonparametric combination and so it enjoys all NPC properties.

With reference to the $2 \times K$ design, one more application of the NPC methodology is in terms of joint analysis of tests on sampling moments. To this end, let us assign ranks W to ordered classes, that is let us transform A_k into $W_k = k$, and consider the rule: Two discrete distributions defined on the same support, with a finite number K of distinct real values, are equal if and only if their first K-1 moments are equal, because their characteristic functions, as well as their probability generating functions, depend only on these few moments. Consequently, we are allowed to write global hypotheses as $H_0 : \{X_1 \overset{d}{=} X_2\} = \{\cap_{r=1}^{K-1} E(W_1^r) = E(W_2^r)\}$, $H_1 : \{X_1 \overset{d}{=} X_2\} = \{\cup_{r=1}^{K-1} E(W_1^r) > E(W_2^r)\}$.

Let us now consider for each r, $r = 1, \ldots, K-1$, the permutation partial test statistic based on comparison of the rth sampling moments or on one of its permutationally equivalent statistics, such as $T^*_{Wr} = \sum_{k \leq K} k^r f^*_{1k}/n_1$. We note that all these partial tests are exact, unbiased, and consistent, then an NPC test associated with the combining function ψ is $T''_{W\psi} = \psi(\lambda_{W1}, \ldots, \lambda_{WK-1})$; and therefore, for any ψ, combined tests $T''_{W\psi}$ are at least exact, unbiased, and consistent.

The exact determination of permutation distribution of any statistic can clearly be obtained by complete enumeration of all its permutation values. Of course, this way becomes impractical when sample sizes are not very small and for complex problems. Alternatively, it can be estimated, to the desired degree of accuracy, by a conditional Monte Carlo method consisting of a simple random sampling from the set of all permutations. This solution is especially recommended for NPC methods and in general for complex problems (see Pesarin, 2001). Main NPC routines for conditional Monte Carlo simulation are implemented in the NPC-Test® software.

3 On Unrestricted and Multivariate Extensions

In this section we consider extensions to problems such as: (a) the two-sample unidimensional design for unrestricted alternatives; (b) the multivariate extension of two-sample design for restricted alternatives and multivariate extensions of problems in (a). All NPC solutions in this section are obtained by using the direct combination of standardized partial permutation tests. Of course, instead of the direct combination, we could use any other combining function ψ.

a) In a univariate two-sample design with unrestricted alternatives the hypotheses are $H_0 : \{X_1 \stackrel{d}{=} X_2\}$ against $H_1 : \{X_1 \stackrel{d}{\neq} X_2\} = \{\cup_k [F_1(A_k) \neq F_2(A_k)]\}$. Thus, the extension of test statistic T_{AD} becomes $T^{*2}_{AD} = \sum_{k=1}^{K-1} (\hat{F}^*_{2k} - \hat{F}^*_{1k})^2 [\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k})]^{-1}$, which corresponds to a two-sample Anderson-Darling test for non-dominance alternatives, adjusted for discrete variables.

b) For an extension to the multivariate version of the two-sample design with stochastic dominance alternatives, also called component-wise stochastic dominance, let us suppose that the response is Q-dimensional $X = (X_1, \ldots, X_Q)'$ whose related numbers of ordered classes are $K = (K_1, \ldots, K_Q)'$ and that n_1 and n_2 units are independently observed from X_1 and X_2, respectively. In this framework the hypotheses under testing are then $H_0 : \{X_1 \stackrel{d}{=} X_2\} = \{\cap_{q=1}^{Q} (X_{q1} \stackrel{d}{=} X_{q2})\} = \{\cap_{q=1}^{Q} \cap_{k=1}^{K_q-1} (F_{qk} = G_{qk})\}$ and, if no reverse inequality is possible in the alternative, $H_1 : \{X_1 \stackrel{d}{>} X_2\} = \{\cup_{q=1}^{Q} (X_{q1} \stackrel{d}{>} X_{q2})\} = \{\cup_{q=1}^{Q} \cup_{k=1}^{K_q-1} (F_{qk} < G_{qk})\}$, where F_{qk} and G_{qk} play the role of CDFs for the qth variable. This problem, if not impossible, is considered as extremely difficult when approached within the restricted maximum likelihood ratio test, whereas within the NPC method its solution is straightforward. For instance, the direct combination on standardized partial tests, which shows the Anderson-Darling's structure, is:

$$T^*_{MD} = \sum_{q=1}^{Q} \sum_{k=1}^{K_q-1} (\hat{F}^*_{q2k} - \hat{F}^*_{q1k}) [\bar{F}_{q\cdot k}(1 - \bar{F}_{q\cdot k})]^{-1/2} .$$

It is worth noting that its qth component $T^*_{Dq} = \sum_{k=1}^{K_q-1} (\hat{F}^*_{q2k} - \hat{F}^*_{q1k}) [\bar{F}_{q\cdot k}(1 - \bar{F}_{q\cdot k})]^{-1/2}$ is the partial permutation test related to the qth variable, for which $\bar{F}_{q\cdot k} = (N_{q1k} + N_{q2k})/n$ is the qth pooled EDF, $q = 1, \ldots, Q$. In addition, we may consider the extension of T^*_{MD} to unrestricted or non-dominance alternatives, i.e. for, which gives:

$$T^{*2}_{MD} = \sum_{q=1}^{Q} \sum_{k=1}^{K_q-1} (\hat{F}^*_{q2k} - \hat{F}^*_{q1k})^2 [\bar{F}_{q\cdot k}(1 - \bar{F}_{q\cdot k})]^{-1} .$$

It is worth noting that T^{*2}_{MD} may be considered to correspond to well-known Hotelling's T^2 statistic for a two-sample testing for multivariate ordered categorical variables.

4 Two Examples

Data of first example are from Brunner and Munzel (2000). Data concern a trial of shoulder tip pain, and the observed variables are pain scores following laparoscopic surgery.

The pain scores could take integer values from one (low) to five (high), i.e. $1 \leq k \leq K = 5$. Treatment A (active drug) and P (placebo) were assigned randomly to 25 female patients with 14 receiving A and 11 receiving P. The pain scores recorded on the third day following surgery for treatment A are the following: 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1; for treatment P: 3, 3, 4, 3, 1, 2, 3, 1, 1, 5, 4. For this problem it is of interest to test if treatment with active drug A produces stochastically better responses than placebo P, and so the alternative hypothesis may be written as $H_1 : \{\cup_k F_{Ak} > F_{Pk}\}$. Based on $B = 100000$ conditional Monte Carlo simulations (with the software NPC-Test® 2.0), Anderson-Darling's test statistic T_{AD} on categorical classes gives a p-value of 0.00455. NPC test T''_{WT} on first four moments, after assigning ranks to categorical classes, and based on Tippett's combining function gives a p-value of 0.00457. The p-values of four partial tests respectively are: 0.00471 for T_{W1}, 0.00455 for T_{W2}, 0.01133 for T_{W3}, and 0.01939 for T_{W4}.

As a second example let us consider data from Arboretti, Pesarin, and Salmaso (2005) regarding an observational study carried out in 2004 at the University of Ferrara on the professional placing of Post-Docs. We report some result by two doctorate areas: economic-legal (EL) and scientific-technological (ST), regarding Post-Doc satisfaction on education-employment relationship including 3 variables: coherence between education and employment, use in employment of the abilities acquired during the PhD and adequacy of the PhD training for the work carried out. All variables are ordered categorical (scores from 1 to 4: not at all, not very, quite, very satisfied, i.e. $1 \leq k \leq K = 4$). Table 1 lists the marginal contingency tables for the three variables.

Table 1: Marginal contingency tables for the variables: use, adequacy and coherence

Category	Use			Adequacy			Coherence		
	EL	ST	Total	EL	ST	Total	EL	ST	Total
1	0	1	1	0	2	2	0	1	1
2	1	9	10	2	8	10	0	3	3
3	7	4	11	8	7	15	2	5	7
4	15	9	24	13	6	19	21	14	35
Total	23	23	46	23	23	46	23	23	46

It is of interest to highlight differences in the interviewed Post-Docs' multivariate and univariate satisfaction profiles. This gives rise to an unrestricted set of alternatives and so the alternative hypothesis may be written as $H_1 : \{\cup_k F_{ELk} \neq F_{STk}\}$. It is worth noting that the selection of the subjects composing the two groups has been performed by taking into account for all known confounding variables. Table 2 shows for each variable the results of Anderson-Darling's test statistic T^{*2}_{AD} on categorical classes and of the test T''_{WT} on first 3 moments combined through Tippett's combining function. The last row of Table 2 shows the global p-values related to the Anderson-Darling's test statistic T^{*2}_{MD} and to the test T'''_{WT} based on Tippett's combining function applied to the Q, ($Q = 3$), p-values of tests T''_{WT}.

Table 2: Results of multivariate permutation tests

	p-values				
	T_{W1}	T_{W2}	T_{W3}	T''_{WT}	T^{*2}_{AD}
Use	0.01015	0.01081	0.01325	0.01084	0.00503
Adequacy	0.00733	0.00746	0.00935	0.00954	0.00604
Coherence	0.01555	0.01555	0.01600	0.01600	0.01600
Overall test				0.00077	0.00053

5 Monte Carlo Power Simulations

The performance of the NPC solutions for the univariate case discussed in Section 2 with respect to two competitors proposed in the literature was studied using Monte Carlo simulations. The NPC solutions examined are the Anderson-Darling permutation test T_{AD} on categorical classes, the permutation test T''_{WF} based on Fisher's combination of sampling moments and the permutation test T''_{WT} based on Tippett's combination of sampling moments. We considered, as competitors, the Wilcoxon test with ties correction and the Brunner and Munzel rank test. Brunner and Munzel (2000) proposed a rank test for the Behrens-Fisher two-sample problem in a nonparametric model with the assumption of continuous distribution functions relaxed. For the Brunner and Munzel rank test W_N^{BF}, arbitrary distribution functions are admitted including the case where ties occur by observing ordered categorical data. The Wilcoxon rank sum test is often used in practice for ordered categorical data. This test was proposed for the two sample location problem, when, on the basis of two random samples of iid observations (one sample from the 'control' population, the second independent sample from the 'treatment' population), the aim is to investigate the presence of a treatment effect that results in a shift of location.

For the present simulation study, data were generated from ordinal distributions with four categories. By assuming the 1st sample coming from the control population and the 2nd sample coming from the treatment population, we supposed the treatment effect reduced the frequency of upper categories. The restricted alternative hypothesis is defined as $H_1 : \{X_1 \overset{d}{>} X_2\} = \{F_1(A_k) \leq F_2(A_k),\ k = 1, 2, 3\}$, where at least one inequality is strict. The associated distribution of the 2nd sample presented an absolute frequency reduction of 40% for category 4. In settings (a) and (b) reported in Table 3, this reduction entirely shifted respectively to category 3 and 1, while in setting (c) the reduction splitted over categories 2 and 3.

Table 3: Frequency distributions (%) for the data generation

Frequency distribution (%)	Category 1	2	3	4
1st sample	5	10	15	70
2nd sample (a)	5	10	55	30
2nd sample (b)	45	10	15	30
2nd sample (c)	5	30	35	30

For each configuration we performed 1000 Monte Carlo simulations and for evaluating the permutation distribution 1000 Conditional Monte Carlo iterations. For each independent pair of samples, we consider sample sizes of $n_1 = n_2 = 30$ and $n_1 = 30$, $n_2 = 20$. The results in Table 4 show a good behavior of the NPC solutions and the competitors under the null hypothesis. In particular, the simulated type-I error rates for T''_{WF} ranged for sample sizes of $n_1 = n_2 = 30$ and $n_1 = 30$, $n_2 = 20$ respectively from 0.9% to 1.2% (nominal level 1%), from 3.4% to 4.7% (nominal level 5%) and from 7.4% to 9.7% (nominal level 10%). The Wilcoxon test ranged from 1% to 1.2% (nominal level 1%), from 3.7% to 5.8% (nominal level 5%) and from 7.9% to 10.2% (nominal level 10%). The Brunner and Munzel rank test ranged from 1.1% to 1.2% (nominal level 1%), from 4.1% to 5.2% (nominal level 5%) and from 8.1% to 9.7% (nominal level 10%). The results for the setting $n_1 = 30$, $n_2 = 20$ are listed in Table 4.

Under the alternative hypothesis (see Figures 1 and 2), the Anderson-Darling test is close in power with the W_N^{BF} in configuration reported in Table 3(b), where the associated distribution of the 2nd sample differs from that of the 1st sample in the lowest and highest classes. The permutation test T''_{WF} showed a good behavior in power in all situations both for balanced and unbalanced sample sizes. For the simulation setting illustrated in Table 3(a), where the associated distribution of the 2nd sample have the absolute frequency reduction of 40% shifted from category 4 to category 3, the permutation test T''_{WF} showed the widest difference in power with respect to other tests. This difference decreased as the frequency reduction of category 4 splitted over the other categories.

Table 4: Achieved significance levels ($n_1 = 30$, $n_2 = 20$)

α nominal	T_{AD}	T''_{WF}	T''_{WT}	W	W_N^{BF}
0.01	0.012	0.012	0.013	0.012	0.012
0.025	0.034	0.029	0.031	0.031	0.029
0.05	0.056	0.047	0.055	0.058	0.052
0.1	0.100	0.097	0.097	0.102	0.097
0.2	0.178	0.176	0.184	0.190	0.186
0.3	0.291	0.273	0.284	0.297	0.292
0.4	0.392	0.359	0.370	0.376	0.375
0.5	0.481	0.463	0.479	0.484	0.484
0.6	0.569	0.558	0.560	0.574	0.574
0.7	0.664	0.650	0.651	0.661	0.661
0.8	0.763	0.757	0.760	0.776	0.770
0.9	0.889	0.889	0.882	0.905	0.899
1	1	1	1	1	1

6 Concluding Remarks

The nonparametric combination method is suitable and effective for many multivariate testing problems which, in a parametric framework, are very difficult or even impossible to solve. One major feature of the nonparametric combination of dependent tests,

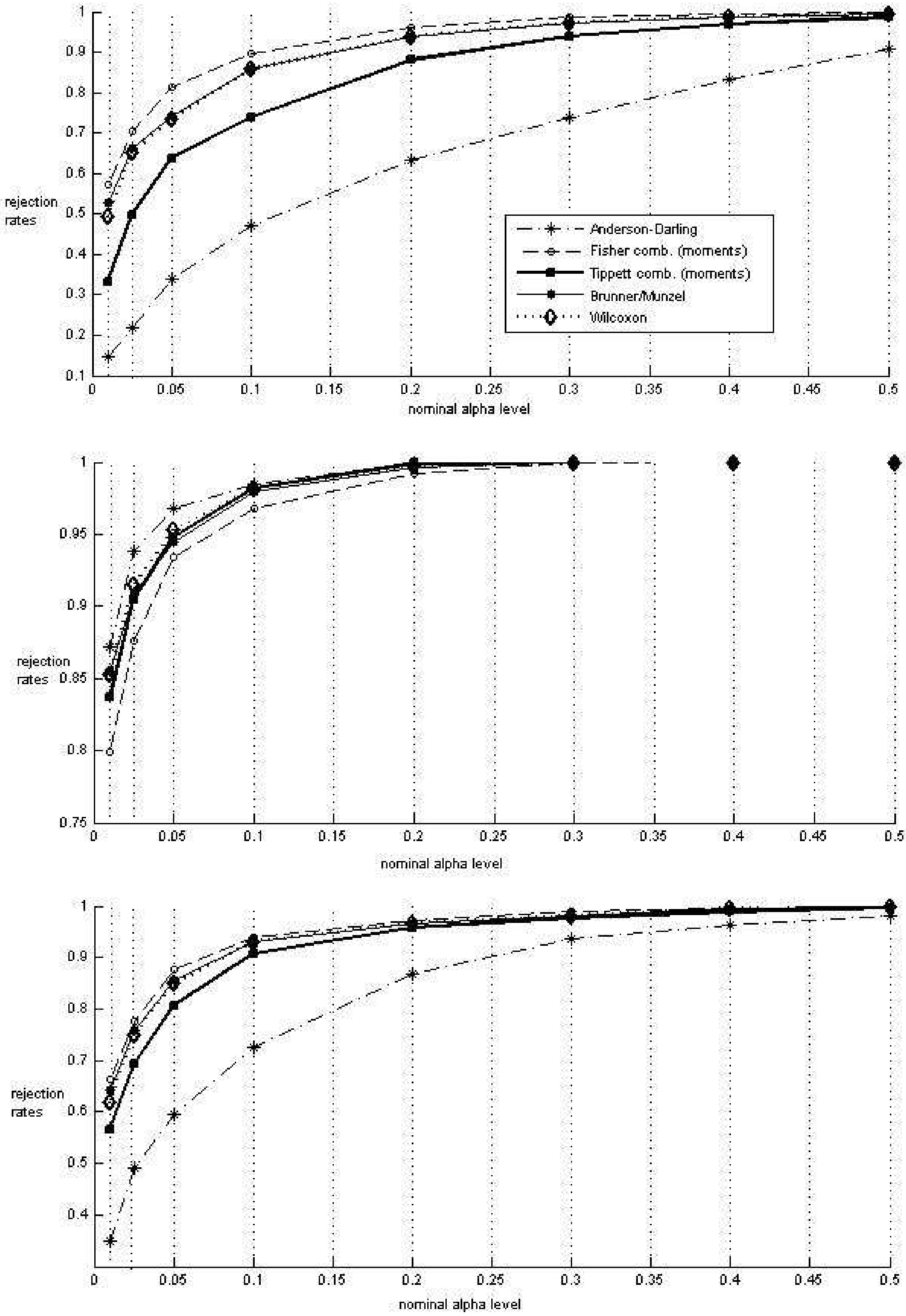

Figure 1: Empirical power for $n_1 = 30$, $n_2 = 20$ for the three distribution configurations

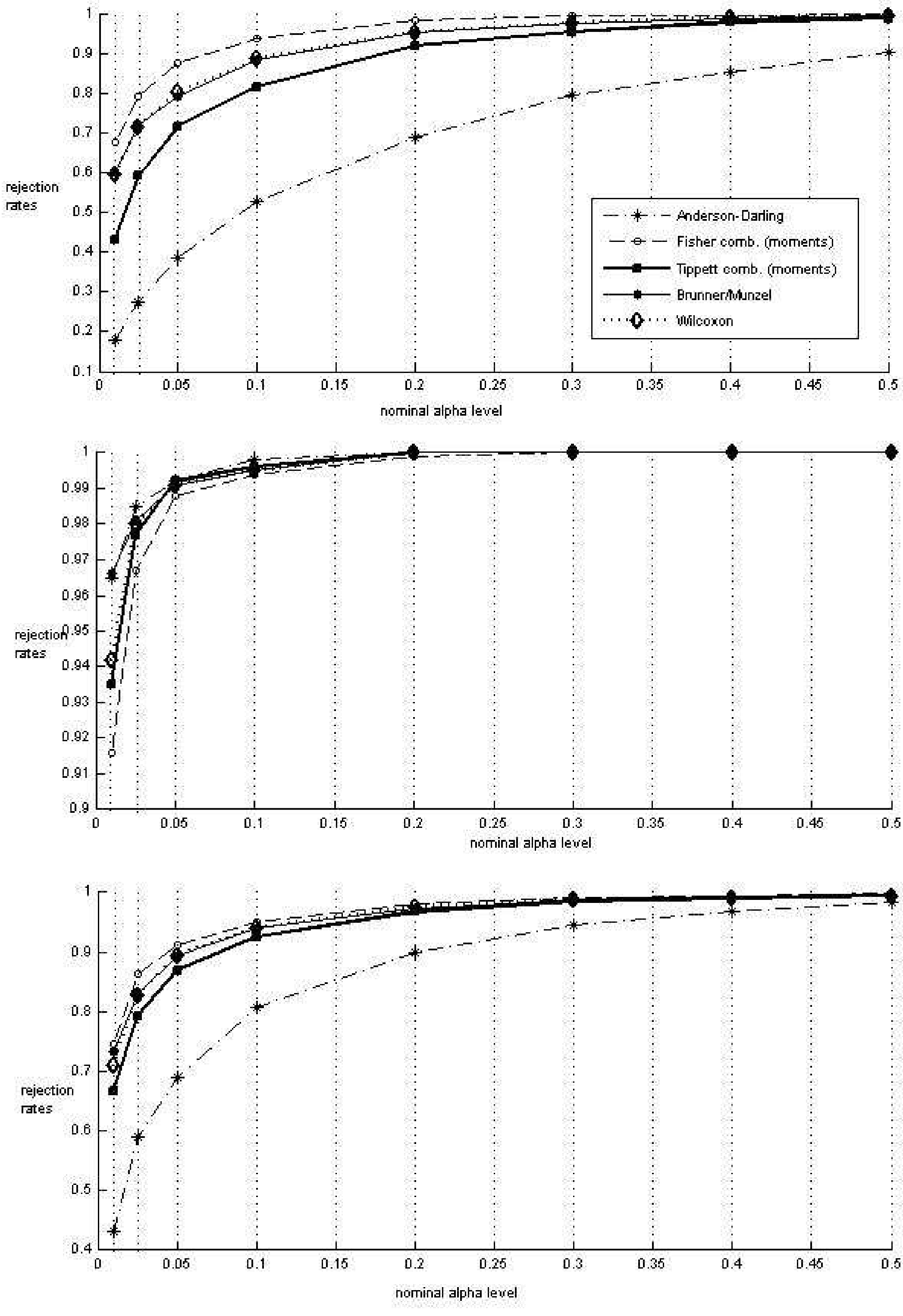

Figure 2: Empirical power for $n_1 = 30$, $n_2 = 30$ for the three distribution configurations

provided that the permutation principle applies, is that one must pay attention to a set of partial tests, each appropriate for the related sub-hypothesis, because the underlying dependence relation structure is nonparametrically captured by the combining procedure. In particular, the researcher is not explicitly required to specify the dependence structure of response variables. This aspect is of great importance especially for non-normal or categorical variables, in which dependence relations are generally too difficult to define and, even when well-defined, are hard to cope with (see Joe, 1997). The researcher is only required to make sure that all partial tests are marginally unbiased, a sufficient condition which is generally easy to check.

Monte Carlo experiments, reported in this contribution, show that the Fisher combining function have good power behavior both for balanced and unbalanced sample sizes and in some situations it is more powerful than the Brunner-Munzel test and the Wilcoxon test. Thus nonparametric combination tests are relatively efficient and much less demanding in terms of underlying assumptions compared with parametric competitors. Moreover, standard distribution-free methods based on ranks, which are generally not conditional on sufficient statistics, rarely present a better unconditional power behavior.

References

Arboretti, G. R., Pesarin, F., and Salmaso, L. (2005). Nonparametric approaches for multivariate testing with mixed variables and for ranking on ordered categorical variables with an application to the evaluation of PhD programs. In S. Sawilowsky (Ed.), *Real Data Analysis.* American Educational Research Association, Age Publishing (to appear).

Brunner, E., and Munzel, U. (2000). The nonparametric Behrens-Fisher problem: asymptotic theory and small-sample approximation. *Biometrical Journal*, *42*, 17-25.

Cohen, A., Kemperman, J. H. B., Madigan, D., and Sakrowitz, H. B. (2003). Effective directed tests for models with ordered categorical data. *Australian and New Zealand Journal of Statistics*, *45*, 285-300.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Pesarin, F. (2001). *Multivariate Permutation Test With Application to Biostatistics*. Chichester: Wiley.

Silvapulle, M. J., and Sen, P. K. (2005). *Constrained Statistical Inference, Inequality, Order, and Shape Restrictions*. New York: Wiley.

Authors' address:

Fortunato Pesarin and Luigi Salmaso
Department of Statistics
University of Padova
via C. Battisti, 241
35121 Padova, Italy
Tel. +39 049 8274143
Fax +39 049 8274170
E-mail: luigi.salmaso@unipd.it, http://www.gest.unipd.it/∼salmaso