International Journal of Statistics and Management System
Vol. 5 No. 1 (January-June, 2020)

Submitted: 13th March 2020 *Revised: 25th April 2020* *Accepted: 29th May 2020*

Nonparametric order–restricted inference for factorial and temporal data *

Xin Gao †

Abstract

Often in medical studies, the detection of biological trend underlying different treatments or varying time points are of primary interest to biologists. To assess the significance of the underlying trend, a researcher may restrict attention to an ordered alternative and thus increase the power of his test. As the crucial normality assumptions for parametric inferences are often untenable in practice, we propose a nonparametric procedure to test for completely ordered alternatives with monotone, non–monotone, or cyclical orderings. The approach consists of forming a statistic which measures the correlation between the empirical ranking of the treatments based on the data and the criterion ranking induced by the alternatives. By considering a vector of Spearman correlations on multiple subsets of treatment effects, the proposed method is extended to test for incompletely ordered alternatives with multiple sub–orderings. Using the projection technique, the variance–covariance structure associated with the test statistic is derived and consistently estimated under unbalanced factorial designs and repeated measures designs. The limiting distribution and asymptotic relative efficiency of the proposed test under Pitman alternatives are established. The application of the proposed test to assess biological trends is demonstrated through the analysis on real data sets.

*This research is supported by Natural Science and Engineering Research Council of Canada Grants.

Key words and phrases: Asymptotic relative efficiency; linear rank statistics; ordered alternatives; rank transform; repeated measures; trend.

AMS 2000 subject classifications. Primary 62G10; secondary 62K15, 62G30.

†*Mailing Address:* Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3.

E–mail: xingao@mathstat.yorku.ca.

1. Introduction

In medical studies we are often interested in a trend of a biological phenomenon. In time–course microarray experiments, gene expression levels are measured at various time points. The trend of up–regulation, down–regulation or cyclical–regulation relative to the time points are associated with particular biological functions of the genes. For example a variety of distinct cyclical patterns of gene expression levels have been observed in yeast cell cycle data which correspond to genes with different functional roles in the cell cycle development (Spellman et al., 1998). Dose–response is another well–known example of a biological trend. The changing pattern of response profile under different dose levels can provide insights into both functional and side effects of a drug candidate. The goals of the experiments above share the common nature of detecting trends in biological data.

Consider a factorial experiment with a treatment factor such as the dose level or the time effect and a clinical covariate such as age, gender or certain clinical symptom. The n^{th} observation from cell (i, j) is modelled as

$$\begin{aligned} &X_{ijn} = \theta + \alpha_i + T_j + \epsilon_{ijn}, \\ &i = 1, \ldots, I; j = 1, \ldots, J; n = 1, \ldots, N_{ij}, \end{aligned} \tag{1.1}$$

where i indexes for the covariate group, j indexes for the treatment group, n indexes for the replicate number. In testing the null hypothesis of no treatment effects, a researcher may wish to be more specific about the trend in the treatment effects and restrict attention to an ordered alternative. The following examples of ordered alternatives are commonly used to describe various trends that occur in practice.

- *Monotone increasing trend:*

$$T_1 \leq T_2 \leq \cdots \leq T_J, \tag{1.2}$$

  with at least one strict inequality. Similarly a monotone decreasing trend can be defined by reversing the directions of the inequalities.

- *Up–down trend with maximum at j:*

$$T_1 \leq T_2 \cdots \leq T_j \geq T_{j+1} \cdots \geq T_J, \tag{1.3}$$

  with at least one inequality among $T_1 \leq T_2 \cdots \leq T_j$ and one among $T_j \geq T_{j+1} \geq \cdots \geq T_J$.

- *Cyclical trend with minima at* 1, j *and* J *and maxima at* k *and* l*:*

$$T_1 \leq T_2 \cdots \leq T_k \geq T_{k+1} \geq \cdots \geq T_j \leq T_{j+1} \leq \cdots \leq T_l \geq T_{l+1} \geq \cdots \geq T_J, \quad (1.4)$$

with at least one strict inequality among each monotone sub–trend. Cyclical pattern often arises when the response oscillates.

- *Union of multiple sub–trends:*

$$T_1 \leq T_2 \cdots \leq T_k, \textit{or } T_{k+1} \geq \cdots \geq T_j, \textit{or } T_{j+1} \leq \cdots \leq T_l, \textit{or } T_{l+1} \geq \cdots \geq T_J, \quad (1.5)$$

with at least one strict inequality among each monotone sub–trend.

There has been a series of seminal work on nonparametric procedures to test for the monotone increasing or decreasing alternative. For a randomized complete block design, Jonckheere (1954) and Page (1963) proposed to measure the Kendall or Spearman correlation between each block and the vector of $(1, 2, \ldots, J)$ representing the ordering of the treatments specified by the alternative. In contrast, Hollander (1967) proposed a test statistic that is the sum of linear signed rank statistics between all pairs of treatments. The three tests presented above are restricted to randomized block design with exactly one observation in each cell. Skillings and Wolfe (1977, 1978) further developed a class of test statistics that are applicable for testing against ordered alternatives in balanced or unbalanced block designs accommodating multiple observations per cell. Kepner and Robinson (1984) proposed a composite Wilcoxon signed rank statistic for randomized complete block designs and repeated measure designs with a limitation that the number of treatments has to be four or fewer. Based on the notion of compatibility of complete ranking with incomplete ranking, Alvo and Cabilio (1995) generalized Jonckheere's and Page's test to the situation in which one or more observations are missing from one or more blocks. All the aforementioned tests have been focused on the alternatives with a *complete order*, which is defined as follows.

Definition 1.1. (Robertson, Wright, and Dykstra, 1988) An ordering on a set $\boldsymbol{X}$ is complete if

(1) It is reflexive: $x \leq x$ for all $x \in \boldsymbol{X}$.

(2) It is transitive: $x, y, z \in \boldsymbol{X}$, $x \leq y$ and $y \leq z$ imply $x \leq z$.

(3) It is antisymmetric: $x, y \in \boldsymbol{X}$, $x \leq y$ and $y \leq x$ imply $x = y$.

(4) Every two elements of $\boldsymbol{X}$ are comparable: $x, y \in X$ implies that either $x \leq y$ or $y \leq x$.

In contrast to a complete order, a *partial order* is an ordering on $\boldsymbol{X}$, which is reflexive, transitive, and antisymmetric, but there may be noncomparable elements. A *quasi–order* is reflexive, and transitive, but may not be antisymmetric and may admit noncomparable elements. Every complete order is a partial order and every partial order is a quasi–order. In practice the alternatives of interest often take the form of a partial order and allow ambiguity among certain treatments. Furthermore, some quasi–ordered alternatives may allow two distinct elements to satisfy both $x \leq y$, and $y \leq x$, and thus contain symmetric binary relationship. A sub–ordering is a special type of quasi–order such that there exists a nonempty subset $\boldsymbol{X}^* \subseteq \boldsymbol{X}$, with the ordering on $\boldsymbol{X}^*$ being complete and any binary relationship involving elements in $(\boldsymbol{X}^*)^c \cap \boldsymbol{X}$ being undefined. For instance the alternatives in expressions (1.3–1.5) specify complete orderings within each sub–trend and allow unspecified orderings across different sub–trends. These alternatives can be formulated as intersections or unions of several sub–orderings. It would be desirable to extend the proposed methodology to accommodate these alternatives with multiple sub–orderings.

Moreover, biological trends are often associated with repeated measures designs in which correlation exists among the treatment factor. Except for Kepner and Robinson's method which is restricted to the case of $J \leq 4$, there has been no nonparametric method available to deal with ordered alternatives for repeated measures designs with arbitrary finite number of treatment levels. For two–way layouts without interaction effects, the rank transform method proposed by Conover & Iman (1976) and further investigated by Hora & Conover (1984) has proven to be a valid and powerful tool for unordered alternatives regarding the main effects. The technique consists of replacing the observations by their ranks in the combined sample regardless of the row and column membership. As the limiting distribution of any rank transform statistic can be readily derived through Hájek's (1968) projection method, the rank transform technique offers a potential tool to devise new tests for ordered alternatives as well. More importantly, we shall consider the factorial design with arbitrary correlation structure on the treatment factor. The method will be flexible enough to accommodate completely or incompletely ordered alternatives with arbitrary monotone, non–monotone and cyclical orderings. It is of further interest to investigate the large–sample properties of the proposed nonparametric method and compare its asymptotic relative efficiency versus the normal theory competitor.

2. The test statistic

The essential idea behind the approaches of Jonckheere (1954) and Page (1963) for testing the monotone alternative is to measure the correlation between the empirical ranking of the treatments based on the data and the vector of $(1, 2, \ldots, J)$ induced by the alternative. To accommodate other types of ordered alternatives, it is natural to extend this idea and introduce the general definition of criterion ranking.

Definition 2.1. Consider an arbitrary alternative with complete ordering. The ordering of treatment effects $\boldsymbol{T} = (T_1, \ldots, T_J)$ is specified by assigning the relationship $R(T_j, T_{j'})$ between each ordered pair of treatments T_j and $T_{j'}$ as one of the following " $<$ ", " $\leq$ ", " $=$ ", " $>$ ", or " $\geq$ ". The criterion ranking is defined as a vector $\boldsymbol{c} = (c_1, \ldots, c_J)$, where $c_j = \sum_{j'=1}^{J} u(R(T_j, T_{j'}))$, with the indicator function $u(x) = 0$ if $x =$ " $<$ ", " $\leq$ ", or " $=$ " and $u(x) = 1$ if $x =$ " $>$ " or " $\geq$ ".

The proposed test statistic will be based on the correlation measure and thus it will be invariant under any location shift or scale shift applied to the criterion ranking. For monotone increasing alternative $T_1 \leq T_2 < \cdots \leq T_J$, the corresponding criterion ranking $\boldsymbol{c}$ plus one takes the form of $(1, 2 \ldots, J)$, which coincides with the conventional criterion ranking used for this alternative. Criterion ranking is well defined for non–monotone alternatives as well. Consider the yeast cell cycle data (Spellman et al., 1998) in which the cell alternates among the G1, S, G2, M phases. The expression level of a given gene repeats at the same phase during different cycles. It is shown that a collection of genes including *CLN1*, *CLN2*, and *CLB6* etc., attain their peak expression level in G1 phase and steadily decay in the following three phases. For these genes, it is reasonable to assume a cyclic alternative with $T_1 \geq T_2 \geq T_3 \geq T_4$, and $T_{4n+j} = T_{4n'+j}$, for $n \in \mathcal{N}$, and $j = 1, \ldots, 4$. The corresponding criterion ranking $\boldsymbol{c}$ divided by the number of cycles takes the form of $(3, 2, 1, 0, 3, 2, 1, 0, \ldots)$. In another study of MCF–7 breast cancer cell line (Lobenhofer et al. 2002), the cell samples were harvested at 1, 4, 12, 24, 36, and 48 hours after the treatment. The six time points were denoted as $T_1 - T_6$. Peddada et al. (2003) analyzed this data set and demonstrated that the expression profile of insulin induced gene I has a down–up pattern $T_1 \geq T_2 \leq T_3 = T_4 = T_5 = T_6$, with minimum achieved at 4 hours and maximum achieved at 12–48 hours. The corresponding criterion ranking takes the form $\boldsymbol{c} = (1, 0, 2, 2, 2, 2)$.

Our approach consists of obtaining the empirical ranking of the J treatments and measuring its correlation with the criterion ranking. High correlation is regarded as evidence to support the ordered alternative. To apply the rank transform technique on unbalanced designs, we introduce the definition of weighted ranks in a combined sample (Gao & Alvo, 2005). Let $\Omega = \{X_{ijn}, i = 1, .., I, j = 1, ..., J, \ n = 1, ..., n_{ij}\}$ be a collection of random variables with $N = \sum_{ij} N_{ij}$. The weighted rank of X_{ijn} within this set is

$$R^*_{ijn} = \frac{N}{IJ} \sum_{i'j'} \frac{1}{N_{i'j'}} \left(\sum_{n'} I(X_{i'j'n'} \leq X_{ijn}) \right),$$

with $I(A)$ denoting the indicator function of the event A. The weighted rank is a sum of indicator functions weighted by the reciprocal of the number of replicates in each cell to circumvent the unbalance in the design. When the N_{ij}'s are equal, the weighted rank reduces to the usual rank R_{ijn}. Let the score function ϕ be a real–valued, absolutely continuous function defined on $(0, 1)$ with bounded second derivative. Weighted rank scores are generated from ϕ as $\alpha_N(R^*_{ijn}) = \phi(R^*_{ijn}/(N+1))$. To simplify notation, let $\alpha^*_{ijn} = \alpha_N(R^*_{ijn})$.

Define $\boldsymbol{S}^*_N = (S^*_N(j), j = 1, \ldots, J)$ to be a vector of weighted linear rank statistics with components $S^*_N(j) = \sum_i \frac{N}{N_{ij}} \sum_{n=1}^{N_{ij}} \alpha^*_{ijn}$. Each $S^*_N(j)$ can be viewed as a nonparametric measurement of the effect of treatment j. Let $E_0(S^*_N(j))$ denote the expectation of $S^*_N(j)$ under the null hypothesis of no treatment effects. Let $\overline{c} = \sum_j c_j/J$ be the average criterion ranking. We measure the Spearman rank correlation between the vector of $\boldsymbol{S}^*_N$ and the criterion ranking $\boldsymbol{c} = (c_1, c_2, \ldots, c_J)$:

$$\begin{aligned} Q &= \sum_{j=1}^{J} \big(c_j - -\overline{c}\big) \left(S^*_N(j) - E_0(S^*_N(j)) \right) \\ &= \sum_{j=1}^{J} (c_j - \overline{c}) S^*_N(j) \end{aligned} \tag{2.1}$$

From this expression, one arrives at an equivalent and simper form

$$Q = \sum_i \sum_j \sum_n d_{ij} \alpha^*_{ijn}, \tag{2.2}$$

with the coefficient $d_{ij} = (c_j - \overline{c})N/N_{ij}$. Large values of Q leads to the rejection of H_0. As the Q statistic can always be formulated as a weighted linear rank statistic regardless of the underlying criterion ranking, the properties of the test statistic for different ordered alternatives can be investigated under a unified framework.

3. Limiting distributions.

3.1. Asymptotic null distribution

First we consider the properties of the proposed statistic for unbalanced designs with independent observations. Assume the error terms ϵ_{ijn} in model (1.1) are independent and identically distributed according to the absolutely continuous distribution function F. The distribution function for observations in cell (i, j) is denoted as $F_{ij}(x) = F(x-\theta-\alpha_i-T_j)$. It is also assumed that as the total sample size increases, $\lim_{N\to\infty} N_{ij}/N = \rho_{ij}$, with $0 < \rho_{ij} < 1$. From expression 2.2, it is seen that Q is a sum of correlated rank scores. For statistics not expressible as sums of independent random variables such as Q, the projection technique provides an effective way of deriving asymptotic normality. The essential idea of the technique consists in approximating the statistic Q by $\sum_{ijn} Z^*_{ijn}$, where the projections Z^*_{ijn} are independent and square–integrable random variables (Hájek, 1968).

Define the average distribution function $H(x) = 1/(IJ)\sum_{ij} F_{ij}(x)$ and the average regression constants $\bar{d} = 1/(IJ)\sum_{ij}\rho_{ij}d_{ij}$. We construct the projection of $S^*_N(j)$ onto X_{ijn} denoted by Z^*_{ijn} with the following form:

$$\begin{aligned} Z^*_{ijn} &= \frac{1}{IJn_{ij}}\sum_{i'j'n'}(d_{i'j'} - \frac{\rho_{ij}}{\rho_{i'j'}}d_{ij})\int\big(u(x-X_{ijn}) - F_{ij}(x)\big)\phi'\big(H(x)\big)dF_{i'j'}(x) \\ &= \frac{1}{IJ\rho_{ij}}\sum_{i'j'}(c_{j'} - c_j)\int_{X_{ijn}}^{\infty}\phi'\big(H(x)\big)dF_{i'j'}(x) + c, \end{aligned} \tag{3.1}$$

Where c is a generic constant. The total variance of the projection variables is denoted by $\sigma^2 = \sum_{ijn}\mathrm{var}(Z^*_{ijn})$. Also define the asymptotic mean

$$\mu^* = \sum_{ijn}\frac{c_j - \overline{c}}{\rho_{ij}}\int\phi(H(x))dF_{ij}(x).$$

In light of the result in Theorem 3 and 4 in Gao & Alvo (2005), it can be shown that $(Q-u^*)/\sigma \xrightarrow{d} N(0,1)$. This asymptotic result can be extended to a broader class of score functions including the commonly used normal score function. This class of score functions can be formulated as a difference function between two square integrable and non–decreasing functions.

Next we can establish the limiting distribution of the statistic Q under the null hypothesis.

Theorem 3.1. *Assume that* $\lim_{N\to\infty} N_{ij}/N = \rho_{ij}$, *with* $0 < \rho_{ij} < 1$. *Define the average cumulative distribution function under the null hypothesis as* $H_0(x) = 1/I \sum_i F(x-\theta-\alpha_i)$ *and* $\sigma_0^2 = \sum_{ijn} \mathrm{var}(Z_{ijn}^0)$, *with*

$$Z_{ijn}^0 = \frac{(c_j - \overline{c})}{\rho_{ij}} \phi(H_0(X_{ijn})). \tag{3.2}$$

Under the null hypothesis of no treatment effect,

$$\frac{Q}{\sigma_0} \xrightarrow{d} N(0,1), \text{ as } N \to \infty.$$

To estimate the limiting variance σ_0^2, it is worthy to note that the projection variables in equation (3.2) are simply expressed in terms of average cumulative distribution function. Therefore we may replace the cumulative distribution function by the corresponding empirical distribution to obtain the consistent variance estimator. In light of this observation, we construct variables

$$C_{ijn} = \frac{(c_j - \overline{c})}{\rho_{ij}} \phi(\hat{H}_0(X_{ijn})),$$

with $\phi(\hat{H}_0(X_{ijn})) = \alpha_{ijn}^*$.

Define $\hat{\sigma_0}^2 = \sum_{ijn}(C_{ijn} - \overline{C}_{ij.})^2$, with $\overline{C}_{ij.} = 1/N_{ij} \sum_{n=1}^{N_{ij}} C_{ijn}$. According to Glivenko–Cantelli lemma, $\max_{ijn} |\alpha_{ijn}^* - \phi(H_0(X_{ijn}))| \to 0$ almost surely as $N \to \infty$. Thus we have $|\hat{\sigma}_0^2 - \sigma_0^2| \to 0$ almost surely. As a direct application of the Slutsky's theorem, we establish the following result concerning our test statistic.

Theorem 3.2. *Under the null hypothesis of no treatment effect,*

$$\frac{Q}{\hat{\sigma_0}} \xrightarrow{d} N(0,1), \text{ as } N \to \infty.$$

3.2. Asymptotic distribution under Pitman alternatives

Next we investigate the limiting distribution of Q under the alternative situation. To construct Pitman alternatives for the ordered inference, let $\boldsymbol{T} = \beta(t_1, \ldots, t_J)$ be a vector of real numbers, where β is a common factor and the vector $(t_1, \ldots, t_J)$ induces a criterion ranking $\boldsymbol{c} = (c_1, \ldots, c_J)$. A sequence of Pitman alternatives, indexed by N, is given by $F_{ij;N}(x) = F_i(x - \frac{\beta t_j}{\sqrt{N}})$. These Pitman alternatives allow for both the ordered treatment effects and the block effects. They converge to the null hypothesis of no treatment effects.

All of the limits in this section are taken under this sequence of alternatives. Denote the average distribution function under the Pitman alternatives as $H_N(x) = 1/(IJ)\sum_{ij} F_{ij;N}(x)$. It is further assumed that as the total sample size increases, $\lim_{N\to\infty} N_{ij}/N = \rho_{ij}$, with $0 < \rho_{ij} < 1$. Denote the projection variables under the Pitman alternatives as Z^N_{ijn} and define the total variance of all the projections as $\sigma^2_N = \sum_{ijn} \mathrm{var}(Z^N_{ijn})$. Under the assumption that the score function ϕ has bounded second derivative, Z^N_{ijn} is a uniformly bounded function. We have the following convergence result:

Lemma 3.3. *Under the sequence of Pitman alternatives,* $\lim_{N\to\infty} \sigma^2_N = \sigma^2_0$.

Proof. According to equation (3.1), we have

$$\begin{aligned}\sigma^2_N &= \sum_{ijn} \mathrm{var}(Z^N_{ijn}) \\ &= \sum_{ijn} \mathrm{var}\left(\frac{1}{IJ\rho_{ij}} \sum_{i'j'} (c_{j'} - c_j) \int_{X_{ijn}}^{\infty} \phi'(H_N(x)) dF_{i'j';N(x)} + c\right)\end{aligned} \tag{3.3}$$

Under the Pitman alternatives, we have the pointwise convergence $\lim_{N\to\infty} H_N(x) = H_0(x)$, and $\lim_{N\to\infty} F_{ij;N}(x) = F_i(x)$. As $\Phi(x)$ has bounded first and second derivatives, using the generalized dominated convergence theorem (Royden, 1968, p. 232), we obtain

$$\begin{aligned}&\lim_{N\to\infty} \mathrm{var}\big(\int_{X_{ijn}}^{\infty} \phi'(H_N(x)) dF_{i'j';N}(x)\big) \\ &= \lim_{N\to\infty} \int \big[\int_y^{\infty} \phi'(H_N(x)) dF_{i'j';N}(x)\big]^2 dF_{ij;N}(y) - \big[\int \big(\int_y^{\infty} \phi'(H_N(x)) dF_{i'j';N}(x)\big) dF_{ij;N}(y)\big]^2 \\ &= \int \big[\int_y^{\infty} \phi'(H_0(x)) dF_{i'}(x)\big]^2 dF_i(y) - \big[\int \big(\int_y^{\infty} \phi'(H_0(x)) dF_{i'}(x)\big) dF_i(y)\big]^2 \\ &= \mathrm{var}\big(\int_{X_{ijn}}^{\infty} \phi'(H_0(x)) dF_{i'}(x)\big).\end{aligned} \tag{3.4}$$

It follows that $\lim_{N\to\infty} \sigma^2_N \to \sigma^2_0$. □

Define the asymptotic mean of Q under the Pitman alternative as μ_N :

$$\begin{aligned}\mu_N &= \sum_{ijn} \frac{c_j - \overline{c}}{\rho_{ij}} \int \phi(H_N(x)) dF_{ij;N}(x) \\ &= \sum_{ijn} \frac{c_j - \overline{c}}{\rho_{ij}} \int \phi\big(\frac{1}{IJ} \sum_{i'} \sum_{j'} F(x - \alpha_{i'} - \frac{\beta t_{j'}}{\sqrt{N}} + \alpha_i + \frac{\beta t_j}{\sqrt{N}})\big) dF(x).\end{aligned}$$

It follows that

$$\lim_{N\to\infty}\frac{\partial\mu_N}{\sqrt{N}\partial\beta}=\lim_{N\to\infty}\sum_{ijn}\frac{c_j-\overline{c}}{\sqrt{N}\rho_{ij}}\int\phi'(H_N(x))H'_N(x)\frac{1}{IJ}\sum_{i'}\sum_{j'}\frac{(t_j-t_{j'})}{\sqrt{N}}dF(x)$$
$$=I(\int\Phi'(H_0(x))H'_0(x)dH_0(x))\sum_j(c_j-\overline{c})(t_j-\overline{t}).$$

Using a Taylor expansion, we obtain the difference of the two asymptotic means under the alternative versus the null situation:

$$\lim_{N\to\infty}\frac{\mu_N-\mu_0}{\sqrt{N}}=\beta I\int\phi'(H_0(x))H'_0(x)dH_0(x)\sum_j(c_j-\overline{c})(t_j-\overline{t}),\tag{3.5}$$

where $\overline{t}=\sum_j t_j/J$. This expression will lead to the derivation of the noncentrality parameter of the limiting distribution under the alternatives.

Theorem 3.4. *Under the sequence of Pitman alternatives,*

$$\frac{Q^2}{\hat{\sigma}_0^2}\xrightarrow{d}\chi^2(\Delta_R,1),$$

with

$$\Delta_R=\frac{\beta^2I^2\Big(\phi'(H_0(x))H'_0(x)dH_0(x)\sum_j(c_j-\overline{c})(t_j-\overline{t})\Big)^2}{\sum_{i,j}\frac{1}{\rho_{ij}}(c_j-\overline{c})^2\operatorname{var}_{H_0}(\phi(H_0(X_{i11})))}.$$

Throughout this article, we have focused on the approach of preselecting the specific criterion ranking and developing the corresponding test statistic. This approach originally proposed by Jonckheere (1954) has been commented by Abelson and Tukey (1963) as one efficient way to utilize order information in hypothesis testing. Abelson and Tukey (1963) further pointed out that although the Jonckheere's selection of criterion ranking is arbitrary, it is reasonably wise as the criterion ranking is chosen to be highly correlated with the hypothesized ordering on the parameters. Alternatively, we may choose the centered criterion ranking $\boldsymbol{c}-\overline{c}$ from all possible contrast vectors to maximize the least value of the noncentrality parameter with the specified order restrictions on the treatment effects $\boldsymbol{T}$. It is shown from Theorem 3.4 that the noncentrality parameter of the limiting distribution depends on the criterion ranking through the term:

$$(\sum_j(c_j-\overline{c})(t_j-\overline{t}))^2/\sum_{i,j}\frac{1}{\rho_{ij}}(c_j-\overline{c})^2\operatorname{var}_{H_0}(\phi(H_0(X_{i11}))).\tag{3.6}$$

Under balanced design, this optimization problem is equivalent to obtain the "maximin-r^2" between the vectors $\boldsymbol{c}-\overline{c}$ and $\boldsymbol{T}$ under the order restriction. Abelson and Tukey (1963) have derived such optimal contrast coefficients as $c_j-\overline{c} = \{(j-1)[1-((j-1)/J)]\}^{\frac{1}{2}}-\{j(1-j/J)\}^{\frac{1}{2}}$ for a monotonic ordering. These optimal contrast coefficients are applicable here to achieve the maximum–least asymptotic power for our proposed test statistic. For non–monotonic order restrictions, Abelson and Tukey have proposed a general computational intensive algorithm to obtain the optimal contrast coefficients. Later, Schaafsma and Smid (1966) have generalized these optimum contrast coefficients to unbalanced settings. The results are also useful for our test statistic to achieve the maximum–least asymptotic power under unbalanced designs.

3.3. Asymptotic relative efficiency

To compare the proposed nonparametric method versus the parametric counterpart, we restrict our attention to the ordered alternatives with linear trend, i.e. $\boldsymbol{T} = \frac{\beta}{\sqrt{N}}(c_1, \ldots, c_J)$, where $\boldsymbol{c}$ is the criterion ranking. The sequence of Pitman alternatives, indexed by N, is given by $F_{ij;N}(x) = F_i(x - \frac{\beta c_j}{\sqrt{N}})$. The corresponding parametric approach consists of converting the problem into one of regression and testing the significance of the slope of the linear trend (Park and Lee, 2000). Under this assumption, the two–way layout model of (1.1) can be reformulated as a one–way classification with one covariate and a single slope (Searle, 1987, p171):

$$X_{ijn} = \theta + \alpha_i + \beta W_{ijn} + \epsilon_{ijn},$$

with the covariate $W_{ijn} = c_j/\sqrt{N}$.

Denote $\mathrm{var}(\epsilon_{ijn}) = s^2$. Following the least–squares approach, the estimate for the slope takes the form

$$\hat{\beta} = \frac{\sum_i \sum_j \sum_{n=1}^{N_{ij}} (X_{ijn} - \overline{X_{i..}})(W_{ijn} - \overline{W_{i..}})}{\sum_i \sum_j \sum_{n=1}^{N_{ij}} (W_{ijn} - \overline{W_{i..}})^2},$$

with $N_{i.} = \sum_{j=1}^{J} N_{ij}$, $\overline{X_{i..}} = \frac{1}{N_{i.}} \sum_{j=1}^{J} \sum_{n=1}^{N_{ij}} X_{ijn}$, $\overline{W_{i..}} = \frac{1}{N_{i.}} \sum_{j=1}^{J} \sum_{n=1}^{N_{ij}} W_{ijn}$. Furthermore, we obtain

$$\hat{s}^2 = \frac{1}{N-I-1}\Big(\sum_i \sum_j \sum_{n=1}^{N_{ij}} (X_{ijn} - \overline{X_{i..}})^2 - \hat{\beta}^2 \sum_i \sum_j \sum_{n=1}^{N_{ij}} (W_{ijn} - \overline{W_{i..}})^2\Big)$$

and

$$\hat{\text{var}}(\hat{\beta}) = \frac{\hat{s}^2}{\sum_i \sum_j \sum_{n=1}^{N_{ij}} (W_{ijn} - \overline{W_{i..}})^2}.$$

Consider the usual t statistic for testing hypothesis $H_0 : \beta = 0$,

$$t = \hat{\beta} / \sqrt{\hat{\text{var}}(\hat{\beta})}.$$

By applying the central limit theorem, we can establish the limiting distribution of t^2.

Theorem 3.5. *Under the sequence of Pitman alternatives with linear trend, the parametric statistic* $t^2 \xrightarrow{d} \chi_1^2(\Delta_P)$, *with*

$$\Delta_P = \frac{\beta^2}{s^2} \sum_i \sum_j \rho_{ij} (c_j - \sum_{j=1}^{J} \frac{\rho_{ij}}{\rho_{i.}} c_j)^2.$$

According to Rao and Mitra (1971), the asymptotic relative efficiency of the proposed rank transform method versus the parametric least squares approach for detecting linear trend is the ratio of the two noncentrality parameters:

$$\begin{aligned} ARE &= \frac{\Delta_R}{\Delta_P} \\ &= \frac{I^2 s^2 \left(\int \phi'(H_0(x)) H_0'(x) dH_0(x) \sum_j (c_j - \overline{c})^2 \right)^2}{\left(\sum_{i,j} \frac{1}{\rho_{ij}} (c_j - \overline{c})^2 \, \text{var}(\phi(H_0(X_{ij1}))) \right) \left(\sum_i \sum_j \rho_{ij} (c_j - \sum_{j=1}^{J} \frac{\rho_{ij}}{\rho_{i.}} c_j)^2 \right)} \end{aligned} \tag{3.7}$$

It is observed that the ARE value generally depends on the distribution function, the cell frequencies and the criterion ranking in a complex way. Nevertheless for balanced one–way layout or balanced two–way layout with no block effects, we obtain a simplified result without the influence of the design and the underlying criterion ranking.

Lemma 3.6. *For balanced one–way layouts or balanced two–way layouts with no block effects, the ARE of the proposed nonparametric method with linear score function versus the parametric approach to detect linear trend equals* $12s^2(\int F'(x) dF(x))^2$, *which coincides with the ARE of the Wilcoxon rank test versus the t–test for two–sample problem.*

Numerical evaluations of *ARE* under different design settings were conducted across several distributions including normal, uniform, logistic, and double–exponential. Table

9.1 provides the summarized results. Two designs were considered. The first design was balanced with three rows, six columns and equal cell frequencies. The second design was unbalanced with the same number of rows and columns and unequal relative cell frequencies $\rho_{11} = 0.038$, $\rho_{12} = 0.053$, $\rho_{13} = 0.068$, $\rho_{14} = 0.076$, $\rho_{15} = 0.053$, $\rho_{16} = 0.045$, $\rho_{21} = 0.068$, $\rho_{22} = 0.053$, $\rho_{23} = 0.045$, $\rho_{24} = 0.076$, $\rho_{25} = 0.045$, $\rho_{26} = 0.038$, $\rho_{31} = 0.061$, $\rho_{32} = 0.053$, $\rho_{33} = 0.061$, $\rho_{34} = 0.038$, $\rho_{35} = 0.076$, $\rho_{36} = 0.053$. Three different sequences of Pitman alternatives were considered. They induces three different criterion rankings: $\boldsymbol{c}_1 = (0, 1, 2, 3, 4, 5)$, $\boldsymbol{c}_2 = (0, 2, 4, 4, 2, 0)$, $\boldsymbol{c}_3 = (0, 3, 0, 3, 0, 3)$. To study the effect of block parameters on ARE value, we considered the location shifts among the blocks to be 0 or 0.25. Under balanced designs, the ARE expression in (3.7) can be simplified as $Is^2\big(\int \phi'(H_0(x))H_0'(x)dH_0(x)\big)^2 / \sum_{i=1}^{I} \mathrm{var}(\phi(H_0(X_{i11})))$, which does not depend on the criterion ranking but depends on the $F_i(x)$, consequently depends on the block shift. Therefore given the same balanced design and block shift, the ARE attains the same value across the three different alternatives. As the block shift changes from 0 to 0.25, the ARE value slightly increases. For unbalanced designs, the ARE depends on both the block shift and the criterion ranking. It is observed that given other settings fixed, the ARE value differs very slightly for monotonic ordering, up–down ordering and oscillating ordering. Similar to balanced design, the ARE value under unbalanced design also slightly increases as the block shift changes from 0 to 0.25. Regarding the underlying distributions, it is shown that for light–tailed distributions such as the normal and uniform, the efficiency of the rank transform method is slightly below that of the parametric test. For heavy–tailed distributions including logistic and double–exponential distributions, the gain in efficiency over the parametric test is substantial.

4. Extension to repeated measures designs

Next we wish to extend our method to factorial designs with repeated measures on the treatment factor. The general repeated measures design can be formulated by independent random vectors $\boldsymbol{X}_{ik} = (\boldsymbol{X}_{i1k}', \ldots, \boldsymbol{X}_{iJK}')'$, $i = 1, \ldots, I$, and $k = 1, \ldots, n_i$, when $\boldsymbol{X}_{ijk} = (X_{ijk1}, \ldots, X_{ijkm_{ijk}})'$, $j = 1, \ldots, J$ and $X_{ijks} \sim F_{ij}$, $s = 1, \ldots, m_{ijk} \geq 1$. Within each ith level of the row factor, $i = 1, \ldots, I$, there are n_i independent subjects, which are repeatedly measured under all J different treatments. There are m_{ijk} replications at each (i, j, k) factor level and subject combinations. The covariance structure within each random vector $\boldsymbol{X}_{ik}$

can be arbitrary and not necessarily of equal correlation. The null hypothesis of no treatment effect is $H_0 : F_{i1}(x) = F_{i2}(x) = \cdots = F_i(x)$. Focused on location alternatives, we assume $F_{ij}(x) = F_i(x - T_j)$. Similar to the two–way layouts with independent observations, the test statistic is constructed to be a weighted linear rank statistic. The asymptotic normality of the weighted linear rank statistic with dependent observations has been established in Domhof (2001). Let $R^*_{ijks} = \frac{N}{IJ} \sum_{i'=1}^{I} \sum_{j'=1}^{J} \frac{1}{N_{i'j'}} (\sum_{k'=1}^{n_{i'}} \sum_{s'=1}^{m_{i'j'k'}} I(X_{i'j'k's'} \le X_{ijks}))$ with $N_{i'j'} = \sum_{k'=1}^{n_{i'}} m_{i'j'k'}$. Let the rank scores be generated from ϕ as $\alpha^*_{ijks} = \phi(R^*_{ijks}/(N+1))$, with $N = \sum_{i=1}^{I} \sum_{j=1}^{J} N_{ij}$.

Similar to Section 2, we introduce the weighted linear rank statistic:

$$Q = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n_i} d_{ij} \overline{\alpha}^*_{ijk.}, \tag{4.1}$$

with $d_{ij} = N(c_j \overline{c}/n_i)$, and $\overline{\alpha}^*_{ijk.} = \sum_{s=1}^{m_{ijk}} \alpha^*_{ijks}/m_{ijk}$.

Let $H_0(x) = \frac{1}{I} \sum_{i=1}^{I} F_i(x)$ be the average univariate marginal cumulative distribution function under the null hypothesis. It is further assumed that as $N \to \infty$, $\min n_i \to \infty$, and $0 < n_i/N < 1$, and $1 \le m_{ijk} \le M \le \infty$. According to Theorem 5.6 in Domhof (2001), the limiting variance of Q can be expressed as

$$\sigma_0^2 = \sum_{i=1}^{I} \sum_{k=1}^{n_i} \mathrm{var}\big(\sum_{j=1}^{J} d_{ij} \overline{Y}_{ijk.}\big), \tag{4.2}$$

with $\overline{Y}_{ijk.} = \sum_{s=1}^{m_{ijk}} \Phi(H_0(X_{ijks}))/m_{ijk}$. This variance expression implicitly incorporates the covariances within the vectors of $\boldsymbol{X}_{ik}$. According to Theorem 5.8 in Domhof (2001), we obtain a consistent variance estimate

$$\hat{\sigma}_0^2 = \sum_{i=1}^{I} \sum_{k=1}^{n_i} \left(\sum_{j=1}^{J} d_{ij} (\overline{\alpha}^*_{ijk.} - \overline{\alpha}^*_{ij..}) \right)^2, \tag{4.3}$$

with $\overline{\alpha}^*_{ij..} = \sum_{k=1}^{n_i} \overline{\alpha}^*_{ijk.}/n_i$. Combining the results above, we are able to show that under the null hypothesis, $\frac{Q}{\hat{\sigma}_0} \xrightarrow{d} N(0,1)$. Note that the limiting distribution of the statistic remains the same as the independent case. However, the variance estimation is different due to the covariance structure of the data.

5. Extension to alternatives with multiple sub–ordering

Situations may arise that the alternative hypothesis of treatment effects only specifies the ordering within a subset of the treatments $(T_{b_1}, \ldots, T_{b_K})$, where $\boldsymbol{b} = (b_1, \ldots, b_K)$ is a subset

of $(1, \ldots, J)$. To deal with this type of sub–ordering, we can form the Spearman correlation of the empirical ranking and the criterion ranking restricted on this subset of treatments:

$$Q_b = \sum_{k=1}^{K} (c_{b_k} - \overline{c_b}) \Big(S_N^*(b_k) - E_0(S_N^*(b_k)) \Big), \tag{5.1}$$

whereas $c_{b_k} = \sum_{k'=1}^{K} u(R(T_{b_k}, T_{b_{k'}}))$ and $\overline{c_b} = 1/K \sum_{k=1}^{K} c_{b_k}$. It is shown that even though Q_b is defined on a subset of treatments, it can still be reformulated as a linear rank statistic:

$$\begin{aligned} Q_b &= \sum_{k=1}^{K} (c_{b_k} - \overline{c_b}) S_N^*(b_k) \\ &= \sum_i \sum_j \sum_n d_{ij}^b \alpha_{ijn}^*, \end{aligned} \tag{5.2}$$

with $d_{ij}^b = (c_{b_k} - \overline{c_b}) N / N_{ij}$, if $j = b_k$; and $d_{ij} = 0$, otherwise. Thus the limiting distributions of Q_b can be established following the same technique as in Sections 3.1 and 3.2. It is worthy to point out that although the correlation is restricted on treatments included in the subset, the rank scores are formed on the combined sample including the treatments not being ordered. It utilizes ranking information from all the observations and therefore it is more advantageous compared to the approach of simply ignoring the treatments without hypothesized ordering.

Furthermore, the alternative may specify the sub–orderings on a collection of L subsets of the treatments with O_l representing the sub–ordering on the lth subset. For instance, the example in expression (1.3) specifies $T_1 \leq T_2 \cdots \leq T_j \geq T_{j+1} \cdots \geq T_J$ with the inequality between T_k and $T_{k'}$ for any $1 \leq k < j$ and $j < k' \leq J$ being unspecified. Let $O_1 = T_1 \leq T_2 \cdots \leq T_j$, with the first subset $= (1, 2, \ldots, j)$, and $O_2 = T_j \geq T_{j+1} \cdots \geq T_J$ with the second subset $= (j, j+1, \ldots, J)$. Then the alternative can be formulated as $H_a : \bigcap_{l=1}^{2} O_l$. The example in expression (1.5) specifies $T_1 \leq T_2 \cdots \leq T_k, or\ T_{k+1} \geq \cdots \geq T_j, or\ T_{j+1} \leq \cdots \leq T_l, or\ T_{l+1} \geq \cdots \geq T_J$. Let $O_1 = T_1 \leq T_2 \cdots \leq T_k$, $O_2 = T_{k+1} \geq \cdots \geq T_j$, $O_3 = T_{j+1} \leq \cdots \leq T_l$, $O_4 = T_{l+1} \geq \cdots \geq T_J$. Then the alternative can be formulated as $H_a : \bigcup_{l=1}^{4} O_l$. We can obtain a vector of Spearman correlations $\boldsymbol{Q} = (Q_1, \ldots, Q_L)$ with Q_l, $1 \leq l \leq L$, defined on the l^{th} subset of the treatments. By applying the projection method as outlined in Section 3.1, we can construct the projection of Q_l onto X_{ijn} as Z_{ijn}^l. Define a $L \times L$ matrix $\boldsymbol{\Sigma} = (\sigma_{ll'})$ with $\sigma_{ll'} = \sum_i \sum_j \rho_{ij} \operatorname{cov}(Z_{ij1}^l, Z_{ij1}^{l'})$.

Theorem 5.1. *Under the null hypothesis of no treatment effect,* $\boldsymbol{Q}/\sqrt{N} \to N_L(\boldsymbol{0}, \boldsymbol{\Sigma})$.

To test for $H_a : \bigcap O_l$, we can perform an intersection test which rejects H_0 and favors H_a if $Q_l \geq Z_\alpha$ for all $1 \leq l \leq L$, with Z_α being the upper $100(1-\alpha)$ point of the standard normal distribution. As under the null hypothesis $P(\bigcap(Q_l \geq Z_\alpha)) \leq \min_l P(Q_l \geq Z_\alpha) = \alpha$, the significance level of this procedure is well controlled under the correct nominal level. Alternatively we may set $u(R(T_j, T_{j'})) = 0$ for any two elements which are unrelated and still obtain one criterion ranking for the whole set of elements. For instance, consider the simple tree ordering, $T_1 \leq T_2$, $T_1 \leq T_3$, $\ldots, T_1 \leq T_k$. The extended criterion ranking will take the form of $(0, 1, 1, \ldots, 1)$. A linear rank statistic Q_E can be constructed based on this extended criterion ranking.

To compare these two approaches, we note that the intersection test is a conservative test and its power is lower than the minimum power of the Q_l, $l = 1, \ldots, L$. In contrast, the statistic Q_E combines the strength of the correlation from each sub–ordering and is more powerful than the intersection test when $H_a : \bigcap O_l$ holds true. Nevertheless, the statistic Q_E is not specific to the alternative H_a and it can yield significant result even though the overall H_a does not hold true. As an illustration, we consider the testing of the alternative H_a specifying the simple tree order $T_1 \leq T_k$, $k = 1, \ldots, K$. First we assume that the treatment effects satisfy the simple tree order with $T = (0.25, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)$. Under this parameter setting, the Q_E is advantageous as it achieves higher power than the intersection statistic. Next we assume that the parameters completely violate the hypothesized tree order with all the directions of the inequalities being reversed : $T_1 \geq T_k$, $k = 1, \ldots, K$, and $T = (1.55, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)$. The intersection test is advantageous as it will make correct conclusion and yield insignificant result. However, the Q_E test will result in highly significant result and lead to incorrect conclusion. The false significance of the Q_E statistic is due to the fact that the pattern of part of the treatment effects $(0.9, 0.9, \ldots, 0.9)$ agrees with the counterpart of the extended criterion ranking $(1, 1, \ldots, 1)$. Therefore extra caution is needed when the Q_E test is employed to test for intersection alternatives. In the simulation section, more empirical results will be provided to compare these two methods.

Next we proceed to consider the test for $H_a : \bigcup O_l$. A quadratic form $1/N \boldsymbol{Q}' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{Q}$ can be formed as a chi–squared statistic with $\hat{\boldsymbol{\Sigma}}^{-1}$ denoting the general inverse of the estimated covariance matrix. The degree of freedom of the statistic is equal to the rank of the covariance matrix $\boldsymbol{\Sigma}$, which can be readily obtained as follows.

Lemma 5.2. *Define a $IJ \times L$ matrix $\boldsymbol{D} = (\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_L)$, with $\boldsymbol{D}_l = (d^l_{11}, d^l_{12}, \ldots, d^l_{IJ})'$.*

The rank of $\boldsymbol{\Sigma}$ *equals the rank of matrix* $\boldsymbol{D}$.

6. Simulation studies

Monte Carlo simulations were conducted to investigate the type I error rates and the power of the proposed tests for a variety of ordered alternatives under different design settings. The results were summarized based on 1,000 simulation runs. The significance level was set to be 0.05. Four different distributions were simulated: normal distribution $N(0,1)$, normal with two outliers, lognormal $(0,1)$ and Cauchy distribution $Cauchy(1)$.

First we investigated the performance of the proposed rank transform statistic Q for unbalanced factorial designs with independent observations. The simulated design has three rows and six columns with unbalanced cell sizes $N_{11} = 5$, $N_{12} = 7$, $N_{13} = 9$, $N_{14} = 10$, $N_{15} = 7$, $N_{16} = 6$, $N_{21} = 9$, $N_{22} = 7$, $N_{23} = 6$, $N_{24} = 10$, $N_{25} = 6$, $N_{26} = 5$, $N_{31} = 8$, $N_{32} = 7$, $N_{33} = 8$, $N_{34} = 5$, $N_{35} = 10$, and $N_{36} = 7$. Three different ordered alternatives are considered. Under H_{1a}, a monotone increasing alternative $\boldsymbol{T} = \{0.175, 0.350, 0.525, 0.700, 0.875, 1.050\}$ was specified. The corresponding criterion ranking was set to be $\boldsymbol{c} = (0,1,2,3,4,5)$. Under H_{2a}, an up–and–down alternative $\boldsymbol{T} = \{0.2, 0.6, 0.8, 0.8, 0.6, 0.2\}$ was specified. The corresponding criterion ranking was set to be $\boldsymbol{c} = (0,2,4,4,2,0)$. Under H_{3a}, a cyclic alternating alternative $\boldsymbol{T} = \{0.4, 1.0, 0.4, 1.0, 0.4, 1.0\}$ was specified. The corresponding criterion ranking was set to be $\boldsymbol{c} = (0,3,0,3,0,3)$. We evaluated the proposed Q statistic with two score functions–normal score and linear score, respectively. Our proposed method was compared with the least–squares approach using the criterion ranking as the covariate. The corresponding results are summarized in Table 9.2. It is demonstrated that the proposed statistic Q with normal score and linear score both attain type I error rates close to the correct nominal level with small sample size regardless of the underlying distributions. In contrast, the parametric statistic only maintains type I error rate close to nominal level under normal distribution. For the other three distributions, the type I error rates of the parametric statistic are either inflated or deflated. With regard to power, the proposed nonparametric method is slightly less powerful than the parametric method for normal distribution. For the other three distributions, the proposed method consistently outperforms the parametric method to a fairly large extent. Concerning different ordered alternatives, the proposed method achieves satisfactory power for all the three ordered alternatives considered in the simu-

lation. In terms of different score functions, the test statistic with normal score appears to have slightly better power than linear score for normal distribution. This could be due to the fact that the normal score function is the optimum score function for normally distributed observations. For other distributions, it is also observed that the normal score achieves better power than linear score in most of simulation settings. Nevertheless, the difference appears to be marginal.

Next we investigated the proposed rank transform statistic Q for ordered alternatives under repeated measures designs. Two different designs were considered. The first design has three rows and six columns with equal cell sizes $M = 20$. This design corresponds to the scenario that the number of treatments is small but the number of replications is relatively large. Three different alternatives were simulated. Under H_{1a}, we have a monotone increasing alternative $\boldsymbol{T} = (1:6) * 0.125$; Under H_{2a}, we have an up–and–down alternative $\boldsymbol{T} = (1, 3, 4, 4, 3, 1) * 0.14$; Under H_{3a}, we have a cyclic alternating alternative $\boldsymbol{T} = (2, 5, 2, 5, 2, 5) * 0.125$. The design II has three rows and 20 columns with equal cell size $M = 6$. This design corresponds to the scenario that the number of treatments is relatively large whereas the number of replications within each cell is small. For design II, three different alternatives were simulated. Under H_{1a}, we have a monotone increasing alternative $\boldsymbol{T} = (1:20) * 0.035$; Under H_{2a}, we have an up–and–down alternative $\boldsymbol{T} = (1:10, 10:1) * 0.06$; Under H_{3a}, we have a cyclic alternating alternative $\boldsymbol{T} = (1:5, 1:5, 1:5, 1:5) * 0.13$. Equal correlation structure was generated among the column factors. The summarized simulation results are provided in Table 9.3. It is demonstrated that the rank transform statistic Q echoes similar performance for the repeated measures design as that of the factorial designs with independent observations. For the first design scenario, it controls the type I error rate at the correct level, whereas for the second design it maintains the type I error rate close to but slightly above the significance level. For both designs, the proposed test achieves satisfactory power for various distributions and different ordered alternatives. Thus we can draw the conclusion that the number of replications within each cell needs to be at least greater than 8 or 10 to maintain a very good type I error rate control. Regarding the number of treatments, as long as it is a bounded finite number, its size does not affect the validity of the proposed test. This observation also agrees with the asymptotic behavior of the proposed test which is derived under the assumption that the number of treatments is a finite number where the number of replications per cell goes to infinity as the total sample size goes to infinity. Furthermore, the proposed method is

advantageous compared to the parametric method for non–normal distributions.

To investigate the proposed statistics for alternatives of multiple sub–ordering, we consider the treatment effects $\boldsymbol{T} = (T_1, \ldots, T_7)$. The statistic Q_1 is designed to test for the sub–ordering $O_1 : T_1 \leq T_2 \leq T_3$; the statistic Q_2 is designed to test for the sub–ordering $O_2 : T_3 \geq T_4 \geq T_5$; the statistic Q_3 is designed to test for the sub–ordering $O_3 : T_5 \leq T_6 \leq T_7$. The statistic W is the chi–squared statistic for $H_a : \bigcup_{l=1}^{3} O_l$. The statistic R is the intersection test statistic for $H_a : \bigcap_{l=1}^{3} O_l$. Independent observations were simulated from two different noise distributions–$N(0, 1)$ and lognormal$(0, 1)$. The simulated design has three rows and seven columns with cell sizes $N_{11} = 15$, $N_{12} = 21$, $N_{13} = 27$, $N_{14} = 30$, $N_{15} = 21$, $N_{16} = 18$, $N_{17} = 18$, $N_{21} = 27$, $N_{22} = 21$, $N_{23} = 18$, $N_{24} = 30$, $N_{25} = 18$, $N_{26} = 15$, $N_{27} = 18$, $N_{31} = 24$, $N_{32} = 21$, $N_{33} = 24$, $N_{34} = 15$, $N_{35} = 30$, $N_{36} = 21$, $N_{37} = 18$. Table 9.4 provides the summary of the performance of all the test statistics under different parameter settings. In the null situation, the treatment effects were set to be $\boldsymbol{T} = (0, 0, 0, 0, 0, 0, 0)$. The Q_1, Q_2, Q_3 statistics for each of the sub–ordering and the W statistic for the union of the orderings maintain type I error rates close to the correct nominal levels. The R statistic for the intersection test has very small type I error rate demonstrating its conservativeness. In the alternative situation I, the treatment effects were set to be $\boldsymbol{T} = (0.025, 0.35, 0.75, 0.90, 0.90, 0.90, 0.90)$, for which only the O_1 holds true among the three sub–orderings. Correspondingly the Q_1 statistic and the W statistic demonstrate the satisfactory power to detect the O_1 and the $\bigcup_{l=1}^{3} O_l$. As the other two orderings do not hold true, both Q_2 and Q_3 demonstrate very low empirical power. In the alternative situation II, the treatment effects were set to be $\boldsymbol{T} = (0.30, 0.45, 1.05, 0.60, 0.45, 0.90, 1.20)$, for which all the O_1, O_2, and O_3 hold true. The Q_1, Q_2 and Q_3 all demonstrate high power to detect each of the corresponding sub–ordering. The intersection statistic R also demonstrates satisfactory power to detect $H_a : \bigcap_{l=1}^{3} O_l$. The performance of all the test statistics remains consistent between the normal distribution and lognormal distribution. In conclusion the proposed Q_l, W and R are valid and powerful statistics to detect ordered alternatives with multiple sub–orderings.

To further compare the performance of the R statistic and the Q_E statistic based on the extended criterion ranking, we consider the testing of simple tree ordering: $H_a : \bigcap_{l=2}^{7} (T_1 \leq T_l)$. The simulated design has three rows and seven columns with cell sizes same as outlined in the paragraph above. Table 9.5 provides the summary of the performance of the two test statistics under different parameter settings. In the null situation, the treat-

ment effects were set to be $\boldsymbol{T} = (0, 0, 0, 0, 0, 0, 0)$. The Q_E statistic maintains type I error rate close to the correct nominal level. The R statistic for the intersection test has very small type I error demonstrating its conservativeness. In the alternative situation I and II, the simple tree order holds true and the treatment effects were set to be $\boldsymbol{T} = (0.25, 0.9, 0.9, 0.90, 0.90, 0.90, 0.90)$, and $\boldsymbol{T} = (0.25, 0.85, 0.95, 0.85, 0.95, 0.85, 0.95)$. Both the Q_E statistic and the R statistic demonstrate the satisfactory power to detect the simple tree order, whereas the power of Q_E is higher than that of the R statistic. In the alternative situation III and IV, the simple tree order does not hold true with each inequality's direction reversed and the treatment effects were set to be $\boldsymbol{T} = (1.55, 0.9, 0.9, 0.90, 0.90, 0.90, 0.90)$, and $\boldsymbol{T} = (1.55, 0.85, 0.95, 0.85, 0.95, 0.85, 0.95)$. The intersection statistic R yields insignificant result and leads to the correct conclusion. In contrast, the Q_E test statistic yields highly significant result and leads to the incorrect conclusion. This result demonstrates that although Q_E statistic is more powerful than the intersection statistic R under H_a, it can incorrectly yield significant result when H_a does not hold true.

7. Data analysis

To illustrate the use of the proposed rank transform test, we considered a data set which contains the data for 27 patients involved in a pilot study for a new treatment for AIDS (Thompson, 1991). For each of the 27 patients, TMHR scores were gathered three times during the study: at the beginning, after 90 days of treatment and after 180 days of treatment. The study can be viewed as a one–way layout with repeated measurements. The null hypothesis is that there is no effect over time. The previous analysis in Thompson (1991) has found that the time effect is significant. However we were further concerned if the change was in the correct direction. If the patients are improving, it is expected that TMHR scores will follow a dramatic decrease from the initial date to 90 days and then stabilize from the 90 days to the 180 days. Based on the hypothesized ordering of the time effect $T_1 > T_2 = T_3$, we set the criterion ranking to be $(3, 2, 2)$. The rank transform method yields $\chi^2_1 = 96.92$ with p–value$< 10^{-16}$, whereas the least squares method yields $t^2_{79} = 84.64$ with p–value$= 1.98 \times 10^{-14}$. Both methods yield significant results to support the alternative that the TMHR scores follow a decreasing trend. To investigate the robustness of the two methods, the data of total 81 observations was modified to contain six outliers. For the modified data, the rank transform method yields $\chi^2_1 = 17.42$ with p–value$= 3 \times 10^{-5}$,

whereas the least squares method yields $t_{79} = 0.1549$ with p–value= 0.5614. The least squares method fails to detect the decreasing trend in the modified data due to the existence of outliers.

Next we investigated the biological trend underlying the gene expression levels of the gene AA028265 from a microarray data set (Sandberg et al., 2000). The expression levels were collected from two mouse strains (129SvEv vs. C57BL/6) and six different brain regions (amygdala, cerebellum, cortex, entorhinalcortex, hippocampus and midbrain). The data is provided in Table 9.6 in the Appendix. This experiment can be viewed as a two–way layout with repeated measures on the factor of brain regions. Detecting the regional trend of gene expression levels can help elucidate the unique functions and structures of each brain region. It is observed that the expression levels of AA028265 consistently follow a monotone increasing trend in the six brain regions throughout the samples. As the gene AA028265 encodes the fibromodulin, which is a collagen binding protein, it is hypothesized that the expression level of AA028265 is correlated with the collagen composition in different brain regions. To test if the observed monotone increasing trend is significant or not, we applied our proposed method and the least squares method on this data. As there are only two biological replications for each mouse strain, we proposed to use the permutation method to evaluate the significance of the test statistics instead of the asymptotic limiting distribution. To validate the permutation approach, we assumed equal correlations among the brain regions. To perform the permutation, we randomly relabelled six region groups within each biological sample and calculate the proportion of the resulting statistic being equal to or larger than the observed statistic under the null hypothesis over 1,000 permutations. The permutation p–value of the rank transform statistic is 0.005 while as the permutation p–value of the least squares statistic is 0.018. The normal quantile plot of the 24 residuals from the least squares fit was plotted and a deviation from normality was observed in the data. This can explain why the nonparametric test yields more significant result compared to the parametric method on this data set.

8. Conclusion

Based on the general definition of criterion ranking, we have proposed a rank based method to test for ordered alternatives with monotone, non–monotone, or cyclical orderings. The approach consists of measuring the Spearman correlation between the empirical ranking of

the treatments and the criterion ranking induced by the alternative. The resulting Spearman correlation can be reformulated as a linear rank statistic regardless of the underlying criterion ranking. Therefore all the asymptotic properties of the proposed statistic can be investigated using the theoretical results pertaining to the linear rank statistics. Furthermore, we have shown that the vector of Spearman correlations defined on multiple subsets of treatment effects follows an asymptotic multivariate normal distribution. Thus our proposed method is extended to test for incompletely ordered alternatives with multiple sub–orderings. With regard to the types of designs, arbitrary unbalanced designs can be analyzed by formulating the Spearman correlation in terms of a weighted linear rank statistic. By taking into account the possible correlation structure on the treatment factor, the method is extended to handle repeated measures designs. In conclusion, the proposed nonparametric method provides a comprehensive tool to perform order–restricted inference on a variety of design settings for a wide range of ordered alternatives.

9. Appendix

Proof of Theorem 3.1. Under the null hypothesis of no treatment effects, the distribution function $F_{ij}(x)$ is reduced to $F(x-\theta-\alpha_i)$. The expression of the projection variable can be simplified as

$$\begin{aligned} Z^*_{ijn} &= \frac{1}{IJ\rho_{ij}} \sum_{i'j'} (c_{j'} - c_j) \int_{X_{ijn}}^{\infty} \phi'(H_0(x)) dF_{i'j'}(x) \\ &= \frac{1}{J\rho_{ij}} \sum_{i'j'} (c_{j'} - c_j) \int_{X_{ijn}}^{\infty} \phi'(H_0(x)) dH_0(x) \\ &= \frac{(c_j - \overline{c})}{\rho_{ij}} \phi(H_0(X_{ijn})) + c. \end{aligned} \tag{9.1}$$

The constant c can be dropped as it does not attribute to the total variance. Furthermore, the asymptotic mean can be simplified as $\mu^* = \sum_{ijn} \frac{c_j - \overline{c}}{\rho_{ij}} \int \phi(H_0(x)) dF_{ij}(x) = NI \sum_j (c_j - \overline{c}) \int \phi(H_0(x)) dH_0(x) = 0$. The limiting distribution of Q follows.

Proof of Theorem 3.4. Under the sequence of Pitman alternative, $(Q-\mu_N)/\sigma_N \to N(0,1)$. As $\lim_N \sigma_N^2 = \sigma_0^2$, it follows that $Q^2/\sigma_0^2 \to \chi^2(\Delta_R)$, with

$$\Delta_R = \lim_{N\to\infty} \frac{\mu_N^2}{\sigma_0^2} = \frac{\beta^2 I^2 \Big(\int \phi'(H_0(x)) H_0'(x) dH_0(x) \sum_j (c_j - \overline{c})(t_j - \overline{t}) \Big)^2}{\sum_{i,j} \frac{1}{\rho_{ij}} (c_j - \overline{c})^2 \operatorname{var}_{H_0}(\phi(H_0(X_{i11})))} \tag{9.2}$$

Proof of Theorem 3.5. Let $M = \sum_i \sum_j \sum_n (X_{ijn} - \overline{X_{i..}})(W_{ijn} - \overline{W_{i..}})$. It follows that $E(M) = \sum_i \sum_j \sum_{n=1}^{N_{ij}} (\theta + \alpha_i + \beta W_{ijn})(W_{ijn} - \overline{W_{i..}}) = \beta \sum_i \sum_j \rho_{ij} (c_j - \sum_j \frac{\rho_{ij}}{\rho_{i.}} c_j)^2$. Furthermore, as $\operatorname{var}(X_{ijn}) = s^2$, we have $\operatorname{var}(M) = s^2 \sum_i \sum_j \sum_n (W_{ijn} - \overline{W_{i..}})^2 = s^2 \sum_i \sum_j \rho_{ij} (c_j - \sum_j \frac{\rho_{ij}}{\rho_{i.}} c_j)^2$ According to the central limit theorem, we have

$$\frac{M - E(M)}{\sqrt{\operatorname{var}(M)}} \xrightarrow{d} N(0,1), \; as \; N \to \infty. \tag{9.3}$$

As $\hat{s}^2 \to s^2$ a.s., we have $|t^2 - M^2/var(M)| \to 0$ a.s. It follows that $t^2 \xrightarrow{d} \chi_1^2(\Delta_P)$, with

$$\Delta_P = \frac{(E(M))^2}{\operatorname{var}(M)} = \frac{\beta^2}{s^2} \sum_i \sum_j \rho_{ij} (c_j - \sum_{j=1}^{J} \frac{\rho_{ij}}{\rho_{i.}} c_j)^2. \tag{9.4}$$

Proof of Lemma 3.6. Under balanced one–way layout, the noncentrality parameter for the nonparametric method with linear score can be simplified as

$$\Delta_R = \frac{12\beta^2}{J} \sum_j (c_j - \overline{c})^2 \Big(\int H'(x) dH(x) \Big)^2, \tag{9.5}$$

whereas the noncentrality parameter for the parametric method can be simplified as

$$\Delta_P = \frac{\beta^2}{s^2 J} \sum_j (c_j - \overline{c})^2. \tag{9.6}$$

Thus $ARE = \Delta_R/\Delta_P = 12 s^2 (\int H'(x) dH(x))^2$.

Proof of Theorem 5.1. Given any vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)'$, $\boldsymbol{\lambda' Q} = \sum_{l=1}^{L} \lambda_l Q_l = \sum_{ijn} (\sum_{l=1}^{L} \lambda^l d_{ij}^l) \alpha_{ijn}^*$. Therefore, $\boldsymbol{\lambda' Q}$ can be viewed as a weighted linear rank statistic with the bounded coefficients $\sum_{l=1}^{L} \lambda^l d_{ij}^l$. The projection of $\boldsymbol{\lambda' Q}$ onto X_{ijn} is given by $V_{ijn} = \sum_{l=1}^{L} \lambda_l Z_{ijn}^l$. Thus $\lim_{N\to\infty} 1/N \operatorname{var}(\boldsymbol{\lambda' Q}) = \lim_{N\to\infty} 1/N \operatorname{var}(\sum_{ijn} \sum_l \lambda_l Z_{ijn}^l) = \boldsymbol{\lambda' \Sigma \lambda} \geq 0$.

Unless $1/\sqrt{N}\boldsymbol{\lambda}'\boldsymbol{Q}$ converges to a constant, we have the assurance that $\lim_{N\to\infty} \text{var}(\boldsymbol{\lambda}'\boldsymbol{Q}) \to \infty$. Furthermore under the null hypothesis, $E(\boldsymbol{Q}) = \boldsymbol{0}$. Thus according to Theorem 3 in Gao and Alvo (2005), we have $1/\sqrt{N}\boldsymbol{\lambda}'\boldsymbol{Q} \to N(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\lambda)$. Consequently the multivariate normality of $1/\sqrt{N}\boldsymbol{Q}$ follows.

Proof of Lemma 5.2. To find out the rank of the covariance matrix $\boldsymbol{\Sigma}$, it is instructive to find the dimension of its null space $\mathcal{N}(\boldsymbol{\Sigma})$. If $\boldsymbol{\lambda} \in \mathcal{N}(\boldsymbol{\Sigma})$, then $\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} = 0$. This implies $\text{var}[(\sum_{l=1}^{L} \lambda_l \sqrt{\rho_{ij}} \sum_{ij} Z_{ij1}^l)] = \sum_{ij} \rho_{ij} \text{var}(\sum_{l=1}^{L} \lambda_l Z_{ij1}^l) = \sum_{ij} \rho_{ij} \text{var}(\sum_{l=1}^{L} \lambda_l d_{ij}^l \phi(H_0(X_{ij1}))) = 0$. As $\text{var}(\phi(H_0(X_{ij1})) > 0$, it follows that $\sum_{l=1}^{L} \lambda_l d_{ij}^l = 0$, for all i, and j. Thus we have $\boldsymbol{D}\boldsymbol{\lambda} = 0$. This further implies that the solution space of $\boldsymbol{D}\boldsymbol{\lambda} = 0$ is equivalent to the $\mathcal{N}(\boldsymbol{\Sigma})$. Thus rank of $\boldsymbol{\Sigma}$ is equal to the rank of $\boldsymbol{D}$.

Acknowledgments: The author thanks the editor, the associate editor and two referees for their valuable comments and insightful suggestions that led to a significant improvement of the material.

REFERENCES

[1] Abelson, R. P. & Tukey, J. W. (1963). Efficient utilization of non–numerical information in quantitative analysis: General theory and the case of simple order. *Ann. Math. statist.* **34** 1347–1369.

[2] Alvo, M. & Cabilio, P. (1995). Testing ordered alternatives in the presence of incomplete data. *J. Amer. Statist. Assoc.* **90** 1015–1024.

[3] Conover, W. J. & Iman, R. L. (1976). On some alternative procedures using ranks for the analysis of experimental designs. *Commun. Statist.* **A5** 1349–1368.

[4] Domhof, S. (2001). Nichtparametrische relative Effekte. PhD Thesis, University of Göttingen. (webdoc.sub.gwdg.de/diss/2001/domhof/)

[5] Gao, X. & Alvo, M. (2005). A unified nonparametric approach for unbalanced factorial designs. *J. Amer. Statist. Assoc.* **100** 926–941.

[6] Hajék, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

[7] Hora, S. C. & Conover, W. J. (1984). The F statistic in the two–way layout with rank–score transformed data. *J. Amer. Statist. Assoc.* **79** 668–673.

[8] Hollander, M. (1967). Rank tests for randomized blocks when the alternatives have an a priori ordering. *Ann. Math. Statist.* **38** 867–877.

[9] Jonckheere, A. R. (1954). A distribution–free k–sample test against ordered alternatives *Biometrika* **41** 133–145.

[10] Kepner, J. L. & Robinson, D. H. (1984). A distribution–free rank test for ordered alternatives in randomized complete block designs. *J. Amer. Statist. Assoc.* **79** 212–217.

[11] Lobenhofer, E., Bennett, L., Cable, P., Li, L., Bushel, P. & Afshari, C. (2002). Regulation of DNA replication fork genes by 17beta–estradiol. *Molec. Endocrin.* **16** 1215–1229.

[12] Page, E. B. (1963). Ordered hypothesis for multiple treatments: A significance test for linear ranks. *J. Amer. Statist. Assoc.* **58** 216–230.

[13] Park, E. & Lee, Y. J. (2000). Non–parametric test of ordered alternatives in incomplete blocks. *Statistics in Medicine* **19** 1329–1337.

[14] Peddada, S. D., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R. & Umbach, D. M. (2003) Gene selection and clustering for time–course and dose–response microarray experiments using order–restricted inference. *Bioinformatics* **19** 834–841.

[15] Rao, C. R. & Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. New York: John Wiley.

[16] Robertson, T., Wright, F. T. & Dykstra, R. L. (1988) *Order Restricted Statistical Inference*, New York: John Wiley.

[17] Royden, H. L. (1968) *Real Analysis* (2nd ed.), New York: Macmillan.

[18] Sandberg, R., Yasuda, R., Pankratz, D.G., Carter, T.A., Del Rio, J.A., Wodicka, L., Mayford, M., Lockhart, D.J. & Barlow, C. (2000). Regional and strain–specific gene expression mapping in the adult mouse brain. *Proc. Natl. Acad. Sci. USA* **97** 11038–11043.

[19] Schaafsma, W. & Smid, L. J. (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. *Ann. Math. Statist.* **37** 1161–1172.

[20] Searle, S. R. (1987). *Linear Models for Unbalanced Data*, New York: John Wiley.

[21] Skillings, John H. & Wolfe, Douglas A. (1977). Testing for ordered alternatives by combining independent distribution–free block statistics. *Commun. Statist.–Theor. Meth.* **A6**(15) 1453–1463.

[22] Skillings, John H. & Wolfe, Douglas A. (1978). Distribution–free tests for ordered alternatives in a randomized block design. *J. Amer. Statist. Assoc.* **73** 427–431

[23] Spellman P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell.* **9** 3273–3297.

[24] Thompson, G. L. (1991). A unified approach to rank tests for multivariate and repeated measures designs. *J. Amer. Statist. Assoc.* **86** 410–419.

Table 9.1: ARE of the proposed rank transform statistic relative to the least squares statistic for testing ordered alternative with linear trend

Distribution	b	Bal Design	Unbal Design		
		$C_1 - -C_3$	C_1	C_2	C_3
$N(0,1)$	0	0.955	0.920	0.910	0.915
	0.25	0.992	0.955	0.945	0.949
$U(--1,1)$	0	1.000	0.963	0.952	0.958
	0.25	1.039	1.000	0.989	0.994
$logis(1)$	0	1.097	1.056	1.046	1.051
	0.25	1.111	1.070	1.060	1.065
$dexp(1)$	0	1.500	1.440	1.434	1.437
	0.25	1.545	1.484	1.478	1.480

Table 9.1: The table lists the ARE values for balanced and unbalanced two–way layouts. The balanced design has three rows and six columns with equal cell frequencies. The unbalanced design has three rows and six columns with unequal relative cell frequencies. The parameter b measures the location shift between blocks, i.e., $F_1(x) = F(x--b)$, $F_2(x) = F(x)$, and $F_3(x) = F(x+b)$. The ARE values were evaluated under three different sequences of Pitman alternatives which induce three different criterion rankings: $\boldsymbol{c}_1 = (0,1,2,3,4,5)$, $\boldsymbol{c}_2 = (0,2,4,4,2,0)$, $\boldsymbol{c}_3 = (0,3,0,3,0,3)$.

Table 9.2: Type I error rates and power of the proposed rank transform test vs the least squares test to detect ordered alternatives for factorial designs with independent observations

Dist	Alternative	Q*	Q	LS
N	H_{1a}	0.891 (0.046)	0.880 (0.049)	0.966 (0.056)
	H_{2a}	0.725 (0.057)	0.723 (0.063)	0.864 (0.050)
	H_{3a}	0.906 (0.054)	0.894 (0.050)	0.963 (0.053)
$Outlier$	H_{1a}	0.860 (0.042)	0.832 (0.042)	0.429 (0.079)
	H_{2a}	0.351 (0.075)	0.333 (0.074)	0.035 (0.000)
	H_{3a}	0.894 (0.047)	0.882 (0.043)	0.047 (0.027)
$LogN$	H_{1a}	0.855 (0.052)	0.810 (0.058)	0.510 (0.063)
	H_{2a}	0.659 (0.057)	0.627 (0.056)	0.393 (0.036)
	H_{3a}	0.849 (0.061)	0.818 (0.059)	0.541 (0.067)
$Cauchy$	H_{1a}	0.321 (0.047)	0.318 (0.048)	0.079 (0.040)
	H_{2a}	0.221 (0.053)	0.223 (0.055)	0.087 (0.050)
	H_{3a}	0.315 (0.046)	0.311 (0.049)	0.083 (0.025)

Table 9.2: The symbols Q^* and Q stand for the proposed rank transform statistic with normal score and linear score respectively. The symbol LS stands for the least–squares statistic. The values inside parenthesis are the type I error rates are the values outside parenthesis indicate the power. Three different ordered alternatives are considered: H_{1a}, a monotone increasing alternative $\boldsymbol{T} = \{0.175, 0.350, 0.525, 0.700, 0.875, 1.050\}$; H_{2a}, an up–and–down alternative $\boldsymbol{T} = \{0.2, 0.6, 0.8, 0.8, 0.6, 0.2\}$; H_{3a}, a cyclic alternating alternative $\boldsymbol{T} = \{0.4, 1.0, 0.4, 1.0, 0.4, 1.0\}$.

Table 9.3: Type I error rates and power of the proposed rank test for repeated measures designs

Dist	Design	Alternative	Q*	Q	LS
N	I	H_{1a}	0.962 (0.056)	0.964 (0.059)	0.991 (0.058)
	I	H_{2a}	0.877 (0.058)	0.878 (0.057)	0.951 (0.058)
	I	H_{3a}	0.916 (0.064)	0.910 (0.053)	0.969 (0.064)
	II	H_{1a}	0.967 (0.072)	0.969 (0.065)	0.994 (0.086)
	II	H_{2a}	0.868 (0.076)	0.868 (0.075)	0.953 (0.075)
	II	H_{3a}	0.908 (0.080)	0.910 (0.080)	0.975 (0.080)
$Outlier$	I	H_{1a}	0.972 (0.050)	0.971 (0.048)	0.708 (0.021)
	I	H_{2a}	0.757 (0.066)	0.740 (0.068)	0.132 (0.001)
	I	H_{3a}	0.937 (0.056)	0.930 (0.058)	0.769 (0.038)
	II	H_{1a}	0.975 (0.065)	0.970 (0.068)	0.662 (0.036)
	II	H_{2a}	0.725 (0.078)	0.700 (0.070)	0.085 (0.000)
	II	H_{3a}	0.926 (0.080)	0.928 (0.076)	0.654 (0.041)
$LogN$	I	H_{1a}	0.947 (0.064)	0.931 (0.046)	0.635 (0.060)
	I	H_{2a}	0.844 (0.048)	0.808 (0.051)	0.517 (0.058)
	I	H_{3a}	0.911 (0.053)	0.888 (0.053)	0.546 (0.045)
	II	H_{1a}	0.937 (0.063)	0.914 (0.071)	0.665 (0.080)
	II	H_{2a}	0.852 (0.068)	0.822 (0.066)	0.534 (0.069)
	II	H_{3a}	0.880 (0.085)	0.855 (0.087)	0.582 (0.090)
$Cauchy$	I	H_{1a}	0.443 (0.057)	0.446 (0.057)	0.084 (0.035)
	I	H_{2a}	0.302 (0.064)	0.301 (0.064)	0.077 (0.031)
	I	H_{3a}	0.347 (0.043)	0.351 (0.042)	0.066 (0.037)
	II	H_{1a}	0.392 (0.065)	0.386 (0.064)	0.101 (0.054)
	II	H_{2a}	0.332 (0.073)	0.331 (0.071)	0.094 (0.058)
	II	H_{3a}	0.366 (0.069)	0.361 (0.066)	0.109 (0.061)

Table 9.3: The design I has three rows and six columns with equal cell sizes $M = 20$. Three different alternatives were considered: H_{1a}, a monotone increasing alternative $\boldsymbol{T} = (1:6) * 0.125$; H_{2a}, a up–and–down alternative $\boldsymbol{T} = (1, 3, 4, 4, 3, 1) * 0.14$; H_{3a}, a cyclic alternating alternative $\boldsymbol{T} = (2, 5, 2, 5, 2, 5) * 0.125$. The design II has three rows and 20 columns with equal cell size $M = 6$. Three different alternatives were considered: H_{1a}, a monotone increasing alternative $\boldsymbol{T} = (1:20) * 0.035$; H_{2a}, an up–and–down alternative $\boldsymbol{T} = (1:10, 10:1) * 0.06$; H_{3a}, a cyclic alternating alternative $\boldsymbol{T} = (1:5, 1:5, 1:5, 1:5) * 0.13$.

Table 9.4: Performance of the proposed rank transform tests to detect alternatives with multiple sub–orderings

Parameter	Statistic	$N(0,1)$	LogN(0,1)
$\boldsymbol{T} = (0,0,0,0,0,0,0)$	Q_1	0.046	0.040
	Q_2	0.040	0.059
	Q_3	0.056	0.066
	W	0.059	0.056
	R	0.002	0.001
$\boldsymbol{T} = (0.25,0.35,0.75,0.90,0.90,0.90,0.90)$	Q_1	0.843	0.782
	Q_2	0.005	0.008
	Q_3	0.065	0.072
	W	0.916	0.817
$\boldsymbol{T} = (0.30,0.45,1.05,0.60,0.45,0.90,1.20)$	Q_1	0.991	0.982
	Q_2	0.943	0.890
	Q_3	0.994	0.980
	R	0.933	0.869

Table 9.4: The design has three rows and seven columns with unequal cell sizes. The statistics Q_1 is designed to test for the sub–ordering $O_1 : T_1 \leq T_2 \leq T_3$; the statistics Q_2 is for $O_2 : T_3 \geq T_4 \geq T_5$; statistics Q_3 is for $O_3 : T_5 \leq T_6 \leq T_7$. The statistic W is the chi–squared statistic for $H_a : \bigcup_{l=1}^{3} O_l$. The statistic R is the intersection test statistic for $H_a : \bigcap_{l=1}^{3} O_l$. In the null situation, the treatment effects $\boldsymbol{T} = (0,0,0,0,0,0,0)$. In the alternative situation I, the treatment effects $\boldsymbol{T} = (0.25,0.35,0.75,0.90,0.90,0.90,0.90)$, for which only the O_1 holds true. In the alternative situation II, the treatment effects $\boldsymbol{T} = (0.30,0.45,1.05,0.60,0.45,0.90,1.20)$, for which all the O_1, O_2, and O_3 hold true.

Table 9.5: Comparison of the intersection test with the extended criterion ranking test

Parameter	Statistic	$N(0,1)$	LogN(0,1)
$\boldsymbol{T} = (0,0,0,0,0,0,0)$	T	0.062	0.063
	R	0.001	0.001
$\boldsymbol{T} = (0.25,0.90,0.90,0.90,0.90,0.90,0.90)$	T	0.992	0.940
	R	0.826	0.692
$\boldsymbol{T} = (0.25,0.85,0.95,0.85,0.95,0.85,0.95)$	T	0.993	0.940
	R	0.803	0.665
$\boldsymbol{T} = (1.55,0.90,0.90,0.90,0.90,0.90,0.90)$	T	0.990	1.000
	R	0.000	0.000
$\boldsymbol{T} = (1.55,0.85,0.95,0.85,0.95,0.85,0.95)$	T	0.990	1.000
	R	0.000	0.000

Table 9.5: The design has three rows and seven columns with unequal cell sizes. The statistics Q_E and R are designed to test for the simple tree ordering $H_a : \bigcap_{l=2}^{7}(T_1 \leq T_l)$. In the null situation, the treatment effects $\boldsymbol{T} = (0,0,0,0,0,0,0)$. In the alternative situation I and II, the treatment effects $\boldsymbol{T} = (0.25,0.90,0.90,0.90,0.90,0.90,0.90)$, and $\boldsymbol{T} = (0.25,0.85,0.95,0.85,0.95,0.85,0.95)$, for which H_a holds true. In the alternative situation III and IV, the treatment effects $\boldsymbol{T} = (1.55,0.90,0.90,0.90,0.90,0.90,0.90)$ and $\boldsymbol{T} = (1.55,0.85,0.95,0.85,0.95,0.85,0.95)$, for which H_a does not hold true.

Table 9.6: The expression levels of Gene AA028265 in the microarray data of Sandberg et al. (2000)

Strain	Sample ID	amygdale	cere–bellum	cortex	entorhinal–cortex	hippo–campus	midbrain
129SvEv	I	–96	–67	12	30	–72	53
129SvEv	II	–72	–41	–51	–61	58	120
C57BL	III	–107	22	23	59	111	130
C57BL	IV	–144	–58	–65	–73	–23	68