Rank-Score Tests in Factorial Designs with Repeated Measures

Edgar Brunner* and Ulrich Munzel*

Universität Göttingen, Göttingen, Germany

and

Madan L. Puri†

Indiana University

Received December 19, 1996

Nonparametric factorial designs for multivariate observations are considered under the framework of general rank-score statistics. Unlike most of the literature, we do not assume the continuity of the underlying distribution functions. The models studied include general repeated measures designs, compound symmetry designs, and designs for longitudinal data. In particular, designs for ordered categorical data are included. The vectors of the multivariate observations may have different lengths. Moreover, our general framework includes missing values and singular covariance matrices which occur quite frequently in practical data analysis problems. The asymptotic properties of the proposed statistics are studied under general nonparametric hypotheses as well as under a sequence of nonparametric contiguous alternatives. $L_2$-consistent estimators for the unknown covariance matrices are given and two types of quadratic forms are considered for testing the nonparametric hypotheses. The results are applied to a two-way mixed model assuming compound symmetry and to a factorial design for longitudinal data. The main idea of the proofs is based on some moment inequalities for empirical distribution functions in mixed models. The details are provided in the Appendix.


Key words and phrases: rank tests; score functions; ties; ordered categorical data; nonparametric hypotheses; mixed model; longitudinal data.

* Supported in part by DFG, Br-655/11-1, 11-2.
† Partly supported by the Office of Naval Research Contract N00014-91-J-1020.
1. INTRODUCTION

In many biological experiments and ecological, psychological, or medical studies, the subjects are observed repeatedly under different or under the same conditions. Such designs can be described by several more or less complicated models which come under the framework of mixed models. They include growth curves, longitudinal data, or repeated measures designs where a special structure for the dependencies of the multivariate observations, e.g., the compound symmetry, may or may not be assumed. The theory for parametric as well as for semi-parametric mixed models is well developed and is excellently described in some recent textbooks (see, e.g., Davidian and Giltinan, 1995; Diggle, Liang, and Zeger, 1994; Kshirsagar and Smith, 1995; or Lindsey, 1993). For nonparametric models, however, where only the distribution functions of the observations are used to define treatment effects or to express hypotheses, the theory is less developed, particularly when several factors are present in the trial.

The first ideas to use the so-called marginal model to define treatment effects in a nonparametric mixed model date back to Hollander, Pledger, and Lin (1974) and Govindarajulu (1975) and were extended later on and studied in more detail by Brunner and Neumann (1982), Thompson (1990, 1991), and Brunner and Denker (1994). In this marginal model, a treatment effect is defined through the marginal distributions $F_j$, $j = 1, \ldots, d$ of $X_k = (X_{k1}, \ldots, X_{kd})'$ where $X_k$ is the vector of observations for subject $k$. The observations $X_{kj}$ and $X_{kj'}$ coming from different subjects $k$ and $k'$ are assumed to be independent while the observations $X_{kj}$ and $X_{kj'}$ from the same subject may be dependent. Then, a treatment effect can be defined in the same way as for the Wilcoxon–Mann–Whitney statistic or for the Kruskal–Wallis statistic with independent observations. Both statistics are based on consistent estimates of the quantities $p_j = \int H \, dF_j$, where $H = d^{-1} \sum_{j=1}^d F_j$ denotes the mean of the marginal distributions. The quantities $p_j$ are called relative treatment effects, because they measure the effect of the treatment $j$ with respect to the mean $H$ of all distributions in the experiment, and are estimated in a natural way by replacing the distributions by their empirical counterparts $\hat{F}_j$ and $\hat{H}$ respectively. It is well known that the estimator $\hat{p}_j = \int \hat{H} \, d\hat{F}_j$ can be computed from the ranks $R_{kj}$ over all (dependent and independent) observations $X_{kj}$, $k = 1, \ldots, n$, $j = 1, \ldots, d$. Thus, the concept of the marginal model leads to the problem of considering the distribution of rank statistics which use a ranking over dependent and independent observations. For historical review of rank methods for mixed models, we refer to Brunner and Puri (1996) or Akritas and Brunner (1997).

A general formulation of hypotheses in nonparametric models was suggested by Akritas and Arnold (1994) who introduced the idea to formulate...
the hypotheses for the marginal model in terms of the distribution functions and derived the relevant asymptotic distribution theory under the null hypothesis as well as under a sequence of nonparametric contiguous alternatives.

The above results, however, were obtained under the assumption of the continuity of the distribution functions which means that ties were not allowed. This is rather an unrealistic assumption for applications. Based on the ideas of Munzel (1994, 1998), Brunner, Puri and Sun (1995) used the normalized version of the distribution function

\[ F(x) = \frac{1}{2}[F^+(x) + F^-(x)] \]

and of the empirical distribution function to, derive the asymptotic results for two-sample rank statistics in mixed models including the case of arbitrary ties. Here, \( F^+(x) \) denotes the right continuous version and \( F^-(x) \) denotes the left continuous version of the distribution function of a random variable. Akritas and Brunner (1997) generalized these results to factorial designs using the concept of Akritas and Arnold (1994) to formulate nonparametric hypotheses. The advantage of this concept to formulate the hypotheses in terms of the distribution functions is that it includes models with continuous distribution functions as well as models with discrete observations (e.g. ordered categorical data). Since the procedures use rankings over all observations, they are robust to outliers as well as invariant under strictly monotone transformations of the data.

In order to make these procedures broadly applicable for the analysis of real data sets, several improvements have to be worked out. First, the case of missing values has to be studied and also the situations where singular covariance matrices appear in a natural way have to be addressed. For example, in models with ordered categorical data, singular covariance matrices appear quite often. If a treatment is very effective, then (at a certain time point) all subjects may be rated in the same category of the grading scale for this treatment while the grading scores for the subjects in the control group may be different. Consider the nice data set of the shoulder tip pain trial (see, e.g., Lumley, 1996) and assume that the observations of patient no. 113 at time point 4 and of patient no. 119 at time point 5 would be missing. Then the estimated variance for the male patients under the active treatment at time point 4 and the estimated variance for the female patients under the active treatment at time point 5 becomes 0 and hence, the estimated covariance matrix becomes singular. It cannot be regarded as reasonable to delete the two time points from the analysis of the trial in order to be able to perform the computations. Hence, also singular covariance matrices have to be included in the concept. In addition, the special model of compound symmetry where the covariance matrix has a simple structure, has to be addressed. Moreover, some alternative version of the Wald-type statistic (WTS) has to be studied since it is well known that the performance of the WTS is rather poor for small samples sizes (see
It is the aim of the present paper to address these problems from both the applied as well as the theoretical viewpoint under the framework of linear rank statistics with general scores.

**Organization of the Paper.** The paper is organized as follows: In Section 2, the notations are introduced and the relative treatment effects and the estimators are defined. These estimators are linear contrasts of rank statistics. The asymptotic properties of the estimators under the hypothesis as well as under a sequence of nonparametric alternatives is considered in Section 3. Consistent estimators for the unknown covariance matrices in the multivariate model as well as in the compound symmetry model are also derived in Section 3. Finally, the asymptotic properties of the WTS and of a proposed simple modification, which is called ANOVA-type statistic (ATS), are considered in this section. Both statistics are applied in Section 4 to different designs. Some technical Lemmas which are needed to prove the theorems in the body of the paper are provided in the Appendix.

### 2. THE GENERAL MIXED MODEL AND NONPARAMETRIC HYPOTHESES

In the general mixed model, \( r \) treatment groups (the so-called whole-plot factor) are considered where every treatment group \( i \) contains \( k = 1, ..., n_i \) independent (randomly chosen) subjects. These \( n = \sum_{i=1}^{r} n_i \) subjects are observed repeatedly under \( j = 1, ..., d \) different (fixed) situations ("treatments," the so-called sub-plot factor) with \( s = 1, ..., m_{ijk} \) replications for subject \( k \) under the treatment combination \((i, j)\). Thus, there are \( M_{ik} = \sum_{j=1}^{d} m_{ij} \) repeated measures for each subject where the subjects are repeatedly observed under the same treatment as well as under different treatments. This general mixed model can be written by independent random vectors

\[
X_{ik} = (X_{i1k}, ..., X_{ikd})', \quad \text{where} \quad i = 1, ..., r, \quad k = 1, ..., n_i, \quad j = 1, ..., d, \quad (2.1)
\]

and where \( X_{ik} \sim F_{ik}(x) = \frac{1}{2} [F_{ik}^+(x) + F_{ik}^-(x)] \), \( i = 1, ..., r, \quad j = 1, ..., d, \quad k = 1, ..., n_i, \quad s = 1, ..., m_{ijk} \) (the sign \( \sim \) means "is distributed as"). Here, \( F_{ik}^+(x) = P(X_{ik} \leq x) \) is the right continuous version and \( F_{ik}^-(x) = P(X_{ik} < x) \) is the left continuous version of the distribution function. The normalized version \( F_{ik}(x) \) (see Lang, 1993, p. 289) includes continuous as well as discontinuous distribution functions and is needed for handling ties. To derive the general results, we do not assume any particular structure for the dependencies of the components of the vectors \( X_{ik} \) except that the \( X_{ik} \) are independent,
i = 1, ... , r, k = 1, ... , n, and that the bivariate marginal distribution functions of \((X_{ikr}, X_{ik's})\) do not depend on \(k, s\) and \(s'\), i.e., \((X_{ikr}, X_{ik's}) \sim F_{ij}(x, y)\). This assumption is reasonable since the observations with \(k \neq k'\) are independent replicates and the observations with \(s \neq s'\) for the same \(j\) and \(k\) are dependent replicates of the experiment. We note that this assumption on the bivariate marginal distribution functions is needed to derive the asymptotic results under contiguous alternatives (see Theorem 3.4).

The rather general notation introduced above, covers a lot of designs which are commonly used in practice. For example, if \(r = 1\) group of \(k = 1, \ldots, n\) subjects is observed under \(j = 1, \ldots, d\) treatments, then \(F_{ij} \equiv F_{ij}^*\). This design is called “one-group repeated measures design.” If there are \(r\) groups of subjects then this is the so-called split-plot design. In the two-fold nested classification with \(i = 1, \ldots, r\) treatments, \(k = 1, \ldots, n\) independent subjects are observed under treatment \(i\) where every subject receives only one treatment but is observed repeatedly \(s = 1, \ldots, m_{ik}\) times in order to get a more accurate measurement for the variable of interest. In total, there are \(N = \sum_{i=1}^{r} \sum_{k=1}^{n} m_{ik}\) observations of \(n = \sum_{i=1}^{r} n_{i}\) independent subjects.

Higher-way layouts are easily described in this setup by splitting the indices \(i\) or \(j\) into sub-indices \(i', i'', \ldots\) or \(j', j'', \ldots\), respectively. Thus, higher-way layouts with repeated measures or longitudinal data are covered by the general model defined in (2.1). Note that in most cases with longitudinal data, \(m_{ik} = 1\) or \(m_{ik} = 0\) (if the observation is missing). The case of \(m_{ik} \geq 1\) typically occurs when some material, tissue or a set of individuals is split into several homogeneous parts and the compound symmetry model can be used as an appropriate model for this design.

Since no parameters are involved in this setup, we use the distribution functions \(F_{ij}(x)\) to describe an effect (e.g., time effect or treatment effect). To this end, we consider the so-called relative treatment effects \(p_{ij} = \int H(x) \, dF_{ij}(x)\) where \(H(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{d} \sum_{k=1}^{n} m_{ik} F_{ij}(x)\) is the average of all \(N = \sum_{i=1}^{r} \sum_{j=1}^{d} \sum_{k=1}^{n} m_{ik}\) distribution functions in the experiment. The relative effects \(p_{ij}\) may be weighted independently of \(i\) and \(j\) by a score function \(J(u) : u \in (0, 1) \rightarrow \mathbb{R}\) with bounded second derivative, i.e., \(\|J''\|_{\infty} = \sup_{0 < u < 1} |J''(u)| < \infty\) and we define the relative scored effect \(p_{ij}(J) = \int J[H(x)] \, dF_{ij}(x)\). We denote by \(p(J) = (p_{11}(J), \ldots, p_{rd}(J))'\) the vector of these relative effects which are estimated by replacing \(H(x)\) and \(F_{ij}(x)\) by their empirical counterparts. To include also the case of missing values, let

\[
\lambda_{ik} = \begin{cases} 
0, & \text{if } m_{ik} = 0, \\
1, & \text{if } m_{ik} \geq 1,
\end{cases}
\]

\(k = 1, \ldots, n_i, \quad i = 1, \ldots, r, \quad j = 1, \ldots, d\)
and let \( \hat{\lambda}_{ijk}/m_{ijk} = 0 \) if \( m_{ijk} = 0 \). Moreover, a sum is understood to be 0, if the upper summation limit \( m_{ijk} \) equals 0. Further, let \( c(x - X_{ijk}) = 0 \), if the observation \( X_{ijk} \) is missing. Then the empirical functions are defined by

\[
\hat{F}_{ij}(x) = \frac{1}{n_i} \sum_{k=1}^{n_i} \hat{\lambda}_{ijk} \hat{F}_{ijk}(x) = \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{\hat{\lambda}_{ijk}}{m_{ijk}} \sum_{s=1}^{m_{ijk}} c(x - X_{ijk}), \quad \hat{\lambda}_{ij} = \sum_{k=1}^{n_i} \hat{\lambda}_{ijk},
\]

(2.3)

\[
\hat{H}(x) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{d} \sum_{k=1}^{n_i} \sum_{s=1}^{m_{ijk}} c(x - X_{ijk}).
\]

Here, \( c(u) = \frac{1}{2} [c^+(u) + c^-(u)] \) is the normalized version of the counting functions \( c^+(u) \) and \( c^-(u) \) where \( c^+(u) = 0 \) or 1 according as \( u < \) or \( \geq 0 \) and \( c^-(u) = 0 \) or 1 according as \( u \leq \) or \( > 0 \). Note that \( \hat{F}_{ij}(x) \) is the unweighted mean of \( \hat{F}_{ijk}(x) = m_{ijk}^{-1} \sum_{s=1}^{m_{ijk}} c(x - X_{ijk}), \quad k = 1, \ldots, n_i \) where \( \hat{F}_{ijk}(x) = 0 \) if \( m_{ijk} = 0 \). The relative treatment effects \( p_{ij}(J) \) are estimated by

\[
\hat{p}_{ij}(J) = J(\hat{H}) \hat{d} \hat{F}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{\hat{\lambda}_{ijk}}{m_{ijk}} \sum_{s=1}^{m_{ijk}} J(\hat{H}(X_{ijks})) = \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{\hat{\lambda}_{ijk}}{m_{ijk}} \sum_{s=1}^{m_{ijk}} \phi_{ijks},
\]

(2.4)

where \( \phi_{ijks} = J(\hat{H}(X_{ijks})) = J[1/N(R_{ijks} - \frac{1}{2})] \) is the rank-score of \( X_{ijks} \) and \( R_{ijks} \) is the mid-rank of \( X_{ijks} \) among all dependent and independent observations. The quantities \( J(\hat{H}(X_{ijks})) \) shall be called asymptotic rank-score transform of \( X_{ijks} \). since \( E[\hat{J}(\hat{H}) - J(H)]^2 \rightarrow 0 \) under suitable conditions (see Lemma A.2 in the Appendix) and \( \hat{H}(X_{ijks}) = 1/N(R_{ijks} - \frac{1}{2}) \).

The normalized version \( \hat{H}(x) \) of the combined empirical distribution function leads automatically to a symmetric definition of \( H(x) \), since \( \hat{H}(x) \in [1/2N, 1 - (1/2N)] \) and thus, unlike in the literature so far, we shall consider \( J(\hat{H}(x)) \) instead of \( J((N/N + 1) \hat{H}(x)) \). In the sequel, we will drop the expression “normalized version” for brevity and when using the above quoted functions, the “normalized version” is understood if not stated otherwise.

In the nonparametric setup introduced above, hypotheses are formulated by the distribution functions \( F_{ij} \). Let \( F = (F_{11}, \ldots, F_{rd})^\prime \) denote the vector of the distribution functions and let \( C \) denote a contrast matrix, i.e., \( C \mathbf{1} = 0 \) where \( \mathbf{1} = (1, \ldots, 1)^\prime \) and \( \mathbf{0} = (0, \ldots, 0)^\prime \). Then a nonparametric hypothesis in its most general form is written as \( H_0^F: CF = 0 \). For mixed models, such hypotheses have been introduced by Akritas and Arnold (1994) and have
been further developed and discussed by Akritas and Brunner (1997) and by Brunner and Puri (1996). For details, we refer to these papers. Some examples for nonparametric hypotheses are given in Section 4.

3. ASYMPTOTIC RESULTS

In this section, we derive the asymptotic distribution of $\sqrt{n} C \hat{p}(J) = \sqrt{n} C(\hat{p}_{i1}(J), ..., \hat{p}_{id}(J))'$ under the hypothesis $H^F_0: CF = 0$ as well as under a sequence of contiguous alternatives. Our results extend those of Akritas and Brunner (1997) to mixed models with missing values. We also consider the cases with singular covariance matrices as well as compound symmetry models. All these problems are studied under one general framework using the linear rank statistics with general scores. Moreover, for testing the nonparametric hypothesis $H^F_0: CF = 0$, we investigate the properties of the ANOVA-type statistic $ATS$ and compare it with the Wald-type statistic $WTS$.

3.1. Consistency and Basic Asymptotic Equivalence

Note that $\hat{\lambda}_q$, is the number of subjects within treatment $i$ with no missing observations in the $j$th level of the sub-plot factor. Let

$$n_0 = \min_{1 \leq i \leq r, 1 \leq j \leq d} \hat{\lambda}_q.$$  

(3.1)

The asymptotic results are derived under the following assumptions:

Assumptions. (A1) $0 \leq m_{ik} \leq m_0 < \infty$, i.e., the number of replications of a fixed treatment combination $(i, j)$ within one subject is uniformly bounded for all subjects and treatments,

(A2) $n\hat{\lambda}_q \leq N_0 < \infty$, $i = 1, ..., r$, $j = 1, ..., d$, i.e., the ratio of the total number of subjects and the number of subjects with no missing values for treatment combination $(i, j)$ is uniformly bounded,

(A3) $\|J\|_\infty < \infty$, i.e., the score function $J(\cdot)$ has a bounded second derivative.

First, we show that $\hat{p}_q(J)$ given in (2.4), is consistent for $p_q(J) = \int J(H) dF_q$ in the sense of the following proposition.

Proposition 3.1. Let $X_{ik}$ be as defined in (2.1) and let $J(\cdot)$ denote a score function with bounded first derivative, i.e., $\|J\|_\infty < \infty$. Further let $\hat{\lambda}_{ik}$ and $n_0$ be as defined in (2.2) and (3.1), respectively. If $n_0 \to \infty$, then under the assumption (A1), $E(\hat{p}_q(J) - p_q(J))^2 \to 0$, $i = 1, ..., r$, $j = 1, ..., d$.  

Proof. Let \( \varphi(X_{ik}) = J[H(X_{ik})] \) and note that \( E[\varphi(X_{ik}) \varphi(X'_{ik'})] = 0 \) if \((i, k) \neq (i', k')\). Then by the \( c_i\)-inequality, Jensen's inequality, Lemma A.2(2) (in the Appendix) and by independence,

\[
E(\hat{p}_{ij}(J) - p_{ij}(J))^2 = E \left[ J(\hat{H}) d\hat{F}_{ij} - \int J(H) dF_{ij} \right]^2 
\]

\[
= E \left( J(\hat{H}) - J(H) \right) d\hat{F}_{ij} + \left( J(H) d(\hat{F}_{ij} - F_{ij}) \right)^2 
\]

\[
\leq 2 \sum_{k=1}^{n} \sum_{k'=1}^{n} \frac{\hat{\lambda}_{ij}}{m_{ij}} \sum_{s=1}^{m_{ij}} E[J(\hat{H}(X_{ik})) - J(H(X_{ik}))]^2 
\]

\[
+ 2 \sum_{k=1}^{n} \sum_{k'=1}^{n} \frac{\hat{\lambda}_{ij}}{m_{ij}} \sum_{s=1}^{m_{ij}} \sum_{s'=1}^{m_{ij}} E[\varphi(X_{ik}) \varphi(X'_{ik'})] 
\]

\[
\leq \frac{2m_{0} \|J'\|_2^2}{n_{0}r} + 2 \frac{\|J\|_2^2}{n_{0}} = O \left( \frac{1}{n_{0}} \right). \]
Proof. It suffices to consider the \((i, j)\)th component, which is decomposed as

\[
\sqrt{n} \left[ J(\hat{H}) d(\hat{F}_y - F_y) \right] = \sqrt{n} \left[ J(H) d(\hat{F}_y - F_y) + \sqrt{n} \left[ J(\hat{H}) - J(H) \right] d(\hat{F}_y - F_y) \right].
\]

To prove the theorem, it suffices to show that \(E(\sqrt{n} \left[ J(\hat{H}) - J(H) \right] d(\hat{F}_y - F_y))^2 \to 0\) as \(n \to \infty\).

Using Taylor’s expansion, we obtain

\[
J(\hat{H}) - J(H) = J'(H)(\hat{H} - H) + \frac{1}{2} J''(\hat{\theta}_\Delta)(\hat{H} - H)^2,
\]

where \(\hat{\theta}_\Delta\) is between \(\hat{H}\) and \(H\). Thus, \(\left[ J(\hat{H}) - J(H) \right] d(\hat{F}_y - F_y) = B_1 + B_2\), where

\[
B_1 = \int J'(H)(\hat{H} - H) d(\hat{F}_y - F_y)
\]

and

\[
B_2 = \frac{1}{2} \int J''(\hat{\theta}_\Delta)(\hat{H} - H)^2 d(\hat{F}_y - F_y).
\]

It follows from the Appendix, Lemma A.3 that \(E(\sqrt{n} B_k)^2 = O(n^{-1})\), \(k = 1, 2\) which completes the proof. \(\blacksquare\)

3.2. Asymptotic Distributions

3.2.1. Asymptotic Distribution under \(H_{\theta}^F\). To derive the asymptotic distribution of \(\sqrt{n} C_{\theta}(J)\) under the hypothesis \(H_{\theta}^F\), \(CF = 0\) as well as under the sequence of alternatives contiguous to \(H_{\theta}^F\), we need some regularity assumption for the covariance matrix of \(\sqrt{n} \vec{Y}_i(\cdot, J)\) given in (3.2). Let \(\vec{Y}_i(\cdot, J) = (\vec{Y}_{i1}(J), \ldots, \vec{Y}_{idk}(J))\), where \(\vec{Y}_{ijk}(J) = m_{ijk} \sum_{s=1}^{m_{ik}} Y_{is\cdot k}(J)\) as in (3.2) and \(\vec{Y}_{i0k}(J) = 0\) if \(m_{ijk} = 0\). Further let \(V_i = \text{Cov}(\vec{Y}_i(\cdot, J))\) and

\[
\Lambda_{ik} = n_i \cdot \text{diag} \left\{ \frac{i_{1ik}}{\nu_{1i}}, \ldots, \frac{i_{dik}}{\nu_{di}} \right\} \quad \text{and} \quad V_{i,n_i} = \text{Cov}(\sqrt{n_i} \vec{Y}_i(\cdot, J)).
\]

Then, \(n_i \vec{Y}_i(\cdot, J) = \sum_{k=1}^{n} \Lambda_{ik} \vec{Y}_{i\cdot k}(\cdot, J)\) and, by independence,

\[
V_n = \text{Cov}(\sqrt{n} \vec{Y}(\cdot, J)) = \bigoplus_{i=1}^{r} \frac{n_i}{n} V_{i,n_i} = \bigoplus_{i=1}^{r} \frac{n_i}{n} \sum_{k=1}^{n} \Lambda_{ik} \Lambda_{ik}.
\]

Assumption. \((A4)\) \(V_n \to V \neq 0\) as \(n \to \infty\), where \(V\) is a matrix of constants. (This assumption includes the case of singular covariance matrices as well.)
Theorem 3.3. Let \( \mathbf{X}_n \) be as in Proposition 3.1 and let \( \mathbf{V}_n \) be as given in (3.4). If \( n_0 \to \infty \), then under the assumptions (A1)-(A4) and under \( H_0^0 \): \( \mathbf{C} \mathbf{F} = \mathbf{0} \), the statistic \( \sqrt{n} \mathbf{C} \mathbf{p}^*(J) = \sqrt{n} \mathbf{C} \{ \mathbf{J} \mathbf{d} \mathbf{F} \} \) has asymptotically (\( n_0 \to \infty \)) as multivariate normal distribution with mean \( \mathbf{0} \) and covariance matrix \( \mathbf{C} \mathbf{V}_n \mathbf{C}^\prime \).

Proof. First, we note that under \( H_0^0 \),

\[
\sqrt{n} \mathbf{C} \{ \mathbf{J} \mathbf{d} \mathbf{F} \} = \sqrt{n} \mathbf{C} \{ \mathbf{J} \mathbf{d} \mathbf{F} \}
\]

by Theorem 3.2 where \( \mathbf{Y}_i \) is \( \mathbf{Y}_i = \{ \mathbf{Y}_{i1}, ..., \mathbf{Y}_{id} \} \). Note that the vectors \( \mathbf{Y}_i \) are independent and that \( \mathbf{Y}_i \) is the mean of independent random vectors \( \mathbf{Z}_i (J) = n_i \{ \lambda_{ij1}, ..., \lambda_{ijd} \} \). From the multivariate Central Limit Theorem, it follows that \( \mathbf{Y}_i \) has asymptotically a multivariate normal distribution with mean \( \mathbf{0} \) and covariance matrix \( \mathbf{V}_i \) since \( \mathbf{Z}_i (J) \) is uniformly bounded by the assumptions of the theorem, i.e., \( \| \mathbf{Z}_i (J) \| \leq N_0 \| J \| \sqrt{d} \). If \( \mathbf{V}_i \to \mathbf{0} \) then the asymptotic distribution is a multivariate one-point distribution which can be regarded as a degenerate normal distribution.

3.2.2. Contiguous Alternatives. To investigate the set of alternatives for which the test statistics based on \( \sqrt{n} \mathbf{C} \mathbf{p}^*(J) \) are consistent, define a sequence of alternatives for the distributions of the bivariate marginals, viz.

\[
(X_{i*k}, X_{j*k}) \sim F_{n,y'}(x, y) = F_{y'}(x, y) + \frac{1}{\sqrt{n}} \{ K_{y'}(x, y) - F_{y'}(x, y) \} \quad (3.5)
\]

which implies that

\[
X_{i*k} \sim F_{n,y}(x) = F_{y}(x) + \frac{1}{\sqrt{n}} \{ K_{y}(x) - F_{y}(x) \}, \quad i = 1, ..., r,
\]

\[
j = 1, ..., d, \quad k = 1, ..., n_i, \quad s = 1, ..., m_{jk}. \quad (3.6)
\]

Let \( \mathbf{F}_n = (F_{n,11}, ..., F_{n,rc}) = \mathbf{F} + n^{-1/2}(\mathbf{K} - \mathbf{F}) \) denote the vector of the marginal distribution functions of this sequence where \( \mathbf{F} = (F_{11}, ..., F_{rc}) \) such that \( \mathbf{CF} = \mathbf{0} \) and \( \mathbf{K} = (K_{11}, ..., K_{rc}) \) is some vector of (alternative-one-dimensional distribution functions \( \mathbf{CK} \neq \mathbf{0} \). Further, let \( v(J) = \sqrt{n} \mathbf{C} \{ \mathbf{J} \mathbf{d} \mathbf{F}_n \} \) and \( \mathbf{d} \mathbf{C} \).

Theorem 3.4. Let \( \mathbf{X}_n \) be as in Proposition 3.1 where \( X_{i*k} \sim F_{n,y}(x) \) as defined in (3.6), \( k = 1, ..., n_i, \quad s = 1, ..., m_{jk}, \quad i = 1, ..., r, \quad j = 1, ..., d \). Further-
more, let $\hat{F}_i(x)$ and $\hat{H}(x)$ denote the empirical distribution functions as defined in (2.3) and let $C$ be a suitable contrast matrix such that $CF = 0$. Let $H(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{d} \sum_{k=1}^{n_j} m_{ijk} F_{ij}(x)$ denote the weighted average distribution function of $F_{i1}, \ldots, F_{ir}$ and let $H_n(x) = \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{d} \sum_{k=1}^{n_j} m_{ijk} [K_{ij}(x) - F_{ij}(x)]$

denote the weighted average distribution function of $F_{n,11}, \ldots, F_{n,rd}$. If $n_0 \to \infty$, then under the sequence of alternatives (3.5) and under the assumptions (A1)-(A4),

1. the statistics $\hat{\psi}(J) = \sqrt{n} C \hat{p}(J) = \sqrt{n} C \{J(H) d\hat{F} + \sqrt{n} C \hat{Y}(J)\}$ are asymptotically equivalent,

2. $\hat{\psi}(J)$ has asymptotically a multivariate normal distribution with mean $\psi(J) = \{J(H) dF\}$ and covariance matrix $CV_n C$.

The proof is given in the Appendix.

3.3. Estimation of Covariance Matrices

In most practical examples, the covariance matrices $V_{i,n} = n_i^{-1} \sum_{k=1}^{n_i} \lambda_{ik}^2 V_{ik} \lambda_k^2$ defined in (3.4) are unknown and must be estimated from the data. To derive a consistent estimator for $V_{i,n}$, $i = 1, \ldots, r$, we distinguish two models. The multivariate model does not assume any special pattern for the bivariate marginal distribution functions while the compound symmetry model assumes the equality of certain marginal distribution functions under the hypothesis which is stated in details in the last part of this subsection. In what follows, we provide the estimators for $V_{i,n}$ in both models and we prove the consistency of these estimators.

3.3.1. Multivariate Model. In the multivariate model, let $v_i(j)$ denote the diagonal elements of $V_{i,n}$ and let $v_i(j, j')$ denote the off-diagonal elements. Note that by (2.2),

$$v_i(j) = n_j \sqrt{\hat{\lambda}_j \hat{\lambda}_i} \sum_{k=1}^{n_j} \hat{\lambda}_{ik} v_d(f),$$

where $v_{ik}(j) = Var(\hat{Y}_{ik}(J)),$

(3.7)
Let \( \phi_{ijk} = F[1/N(R_{ijk} - \frac{1}{2})] \) denote the rank scores defined in (2.4) and let \( \bar{\phi}_{ijk} = \frac{1}{m_{ijk}} \sum_{k=1}^{m_{ijk}} \phi_{ijk} \) denote the unweighted means of the \( \phi_{ijk} \) where \( \phi_{ijk} = 0 \) if \( m_{ijk} = 0 \). Then we define the estimator \( \hat{V}_{i,n} \) with diagonal elements \( \hat{v}_i(j) \) and off-diagonal elements \( \hat{v}_i(j, j') \) given by

\[
\hat{v}_i(j) = \frac{n_i}{\lambda_{ij}} \sum_{k=1}^{n_i} \lambda_{ijk} \left( \phi_{ijk} - \bar{\phi}_{ijk} \right)^2,
\]
(3.9)

\[
\hat{v}_i(j, j') = \frac{n_i}{\lambda_{ij}(\lambda_{ij} - 1)} \sum_{k=1}^{n_i} \lambda_{ijk} \left( \phi_{ijk} - \bar{\phi}_{ijk} \right)(\phi_{ijk} - \bar{\phi}_{ijk}),
\]
(3.10)

where \( A_{ij, j'} = \sum_{k=1}^{m_{ijk}} \lambda_{ijk} \lambda_{ijk} \). We note that \( \hat{V}_{i,n} \) may not be positive semi-definite (p.s.d.) if the number of missing values is large. In this case, a p.s.d. estimator for \( \hat{V}_{i,n} \) can be obtained by the general method given by Stanish, Gillings, and Koch (1978).

**Theorem 3.5.** Let \( \hat{v}_i(j) \) and \( \hat{v}_i(j, j') \) be as given in (3.9) and (3.10), respectively. If \( n_0 \to \infty \), then under the assumptions (A1)-(A4), \( E[\hat{v}_i(j) - v_i(j)]^2 \to 0 \) and \( E[\hat{v}_i(j, j') - v_i(j, j')]^2 \to 0 \).

The proof is given in the Appendix.

### 3.3.2. Compound Symmetry Model

To state the results for the compound symmetry model, we need some further notations and we assume that \( m_{ijk} \geq 1 \), for simplicity. We note however, that balanced incomplete designs can also be treated within this framework; but for brevity, we do not consider this case here.

In the compound symmetry model, it is assumed that under \( H_0^F: F_{11} = \cdots = F_{dr} \),

1. The bivariate marginal distribution functions of \( (X_{ijk}, X_{ijs}) \) do not depend on \( j, j', k, s \) or \( s' \), i.e., \( (X_{ijk}, X_{ijs}) \sim F_{ij}(x, y) = F_{ij}^*(x, y) \), \( j \neq j' = 1, \ldots, d \).

2. The bivariate marginal distribution functions of \( (X_{ijk}, X_{ijs}) \), \( s \neq s' = 1, \ldots, m_{ijk} \) do not depend on \( j \), i.e. \( (X_{ijk}, X_{ijs}) \sim F_{ij}^{**}(x, y) = F_{ij}^{**}(x, y) \), \( j = 1, \ldots, d \).
Thus, under $H_0^p$, the variances and the covariances are given by

$$\sigma_i^2 = \sigma_i^2 = \text{Var}(Y_{ikr}(J)), \quad j = 1, \ldots, d,$$

$$c_{ij}^* = \text{Cov}(Y_{ikr}(J), Y_{ijr}(J)), \quad j \neq j' = 1, \ldots, d, \quad k = 1, \ldots, n_i,$$

$$s = 1, \ldots, m_{ik}, \quad s' = 1, \ldots, m_{ij},$$

$$c_{ij}^{**} = \text{Cov}(Y_{ijr}(J), Y_{ikr}(J)), \quad j = 1, \ldots, d, \quad k = 1, \ldots, n_i, \quad s \neq s' = 1, \ldots, m_{ik}.$$

We note that $\sigma_i^2 < \infty$, since $|Y_{ikr}(J)| = |H(Y_{ikr})| \leq \|H\|_{\infty}$, which follows immediately from the Assumption (A3) in Subsection 3.1. Thus, the fatness of the tails of $F_{ijr}(x)$ does not have any impact on the existence of these variances.

Let $Y_{ik}(J) = (Y_{i1k}, \ldots, Y_{i1k}, \ldots, Y_{i1k}, \ldots, Y_{i1k})'$. Then it follows that

$$\text{Cov}(Y_{ik}(J)) = \Sigma_{ik} = \sum_{j=1}^{d} \left( (\sigma_j^2 - c_{ij}^{**}) I_{mk} + (c_{ij}^{**} - c_{ij}^*) J_{mk} \right) + c_{ij}^* J_{mk},$$

$$V_{ik} = \text{Cov}(Y_{ik}(J)) = \sum_{j=1}^{d} \left( \frac{1}{m_{ik}} I_{mk} \right) \Sigma_{ik} \left( \frac{1}{m_{ik}} I_{mk} \right) = \sum_{j=1}^{d} \left( \sigma_j^2 - c_{ij}^* \right) \frac{1}{m_{ik}} + (c_{ij}^{**} - c_{ij}^*) J_{mk} + c_{ij}^* J_{dk},$$

$$V_{ik} = \frac{1}{n_i} \sum_{k=1}^{n_i} V_{ik} = \text{diag}(\tau_{11}, \ldots, \tau_{dd}) + c_{ij}^* J_{dd} = D_{ikn} + c_{ij}^* J_{dd},$$

where

$$\tau_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} \tau_{jk} \quad \text{and} \quad \tau_{jk} = \frac{1}{m_{ij}} (\sigma_j^2 + (m_{ij} - 1) c_{ij}^*) - c_{ij}^*. \quad (3.11)$$

Compound symmetry is only assumed for hypotheses regarding the sub-plot factor, i.e. for hypotheses which can be written as $H_0^p(I \otimes C_d) F = 0$ where $C_d$ is a suitable contrast matrix for the sub-plot factor. Thus we need only to estimate $D_{ikn}$ since $(I \otimes C_d) V_{ik}(I \otimes C_d) = \sum_{i=1}^{d} C_d D_{ikn} C_d'$.

**Theorem 3.6.** Let $D_{ikn} = \text{diag}(\tau_{11}, \ldots, \tau_{dd})$ where $\tau_{ij}$ is given in (3.11). Let $D_i$ denote the matrix corresponding to $D_{ikn}$ where $\tau_{ij}$ is replaced by $\hat{\tau}_{ij} = n_i^{-1} \sum_{k=1}^{n_i} \hat{\tau}_{jk}$ and $\hat{\tau}_{jk}$ is defined below. Furthermore, let $\theta_{ij}$ denote the rank-scores as defined in (2.4) and let $\phi_{ik} = d^{-1} \sum_{j=1}^{d} \theta_{jk}$. Then in the compound symmetry model, under the assumptions (A1)-(A3) and under $H_0^p: F_{11} = \cdots = F_{dd}$, $||D_i - D_{ikn}||_2^2 \to 0$ as $\min n_i \to \infty$. 


If \( m_{ijk} \) is not equivalent to a constant, then

\[
\hat{t}_{jk} = \frac{1}{M_{jk} - d} \left[ \frac{1}{d m_{ijk}} \sum_{i=1}^{d} \frac{1}{m_{ijk}} \right] S_{ik1}^{2} + \frac{1}{d - 1} S_{ik2}^{2},
\]

(3.12)

where \( S_{ik1}^{2} = \sum_{j=1}^{m_{ijk}} (\phi_{ijk} - \bar{\phi}_{ik})^{2} \), \( S_{ik2}^{2} = \sum_{j=1}^{d} (\phi_{ijk} - \bar{\phi}_{i,k})^{2} \), and \( M_{jk} = \sum_{j=1}^{m_{ijk}} \).

If \( m_{ijk} \equiv m \), then \( D_{ij} = \hat{t}_{ij} I_{d} \) where

\[
\hat{t}_{ij} = \frac{1}{n_{i}(d-1)} \sum_{k=1}^{n_{i}} \sum_{j=1}^{d} (\phi_{ijk} - \bar{\phi}_{j,k})^{2}.
\]

(3.13)

The proof is given in the Appendix.

Remark. Note that the results of the Theorems 3.5 and 3.6 may not be sufficient to derive the asymptotic distributions of the statistics (to be defined in the next subsection) when \( V_{i,n} \) or \( D_{i,n} \) are replaced by \( \hat{V}_{i,n} \) or \( \hat{D}_{i,n} \), respectively. Further assumptions on the covariance matrices will be discussed for the different statistics separately.

3.4. Statistics

To test the nonparametric hypothesis \( H_{CF}^{*} : \text{CF} = 0 \), we consider two statistics, namely the rank versions of the WTS and of the ATS. Other statistics which are commonly used in multivariate analysis are not discussed here since they require the equality of the covariance matrices. In a nonparametric setup, however, this assumption is only justified in a few special cases. Note that in general any assumed homoscedasticity of the parent distribution functions is not transferred to the asymptotic rank-score transform \( Y_{ijk}(J) = J[H(X_{ijk})] \) because \( H(\cdot) \) is a non-linear transformation.

3.4.1. Wald-Type Statistics (WTS). Let \( \hat{V}_{i,n} \) denote the \( L_{2} \)-consistent estimator of the covariance matrices \( V_{i,n} \), \( i = 1, \ldots, r \) as given in Theorem 3.5. Then

\[
\hat{V}_{n} = \bigoplus_{i=1}^{r} \frac{n_{i}}{n_{i}} \hat{V}_{i,n}
\]

(3.14)

is an \( L_{2} \)-consistent estimator of the covariance matrix \( V_{n} \) given in (3.4) in the sense of Theorem 3.5. Let \( [\cdot]^{+} \) denote a symmetric reflexive \( g \)-inverse of a matrix. Then we define the rank version of the WTS for testing \( H_{CF}^{*} : \text{CF} = 0 \) by

\[
Q_{n}^{W}(C) = n \hat{\psi}(J) C [\hat{V}_{n} C^{T}]^{+} C \hat{\psi}(J).
\]

(3.15)

The asymptotic distribution of \( Q_{n}^{W}(C) \) is given in the next theorem.
Theorem 3.7. Let $\hat{V}_n$ and $V_n$ be as defined in (3.14) and (3.4), respectively. Assume that $V_n \rightarrow V \neq 0$ as $n_0 \rightarrow \infty$ such that $\text{rank}(CV_n) = \text{rank}(CV)$. Then, under the assumptions (A1)-(A4) and

(1) under the sequence of alternatives (3.5) contiguous to $H_0^F$: $\mathbf{CF} = 0$,

$$Q_n(C) \overset{d}{\rightarrow} Z \sim \chi_i^2(\lambda),$$

where $\chi_i^2(\lambda)$ is the chi-square distribution with degrees of freedom $f = \text{rank}(CV)$ and noncentrality parameter $\lambda = \nu(J)[CVC^*]^+ \nu(J)$ where $\nu(J) = C[J(H) dK]$.

(2) under $H_0^F$: $\mathbf{CF} = 0$, $Q_n(C) \overset{d}{\rightarrow} Z \sim \chi_i^2(0)$, with $f = \text{rank}(CV)$.

Proof. First note that $\sqrt{n} \hat{p}(J)$ has asymptotically a multivariate normal distribution with mean $\nu(J) = \{J(H) dK\}$ and covariance matrix $CVC^*$ by Theorem 3.4 and the assumption (A4). Let $Q_n(C) = n\hat{p}(J)C[CVC^*]^+C\hat{p}(J)$. Then it follows from Theorem 9.2.3 in Rao and Mitra (1971) that $Q_n(C)$ has asymptotically a noncentral $\chi_i^2(\lambda)$-distribution with $f = \text{rank}(CV)$ degrees of freedom and noncentrality parameter $\lambda = \nu(J)[CVC^*]^+ \nu(J)$.

Next, note that $\|C\hat{V}_nC^* - CV_nC^*\|_F \rightarrow 0$ by Theorem 3.5 and that $CVC^* - CV_nC^* \rightarrow 0$ by assumption. Thus, it follows that $(CVC^*)^+ = (CVC^*)^+ \rightarrow 0$ because $\text{rank}(CV_n) = \text{rank}(CV)$ by assumption. Hence, $Q_n(C) - Q_n(C) \rightarrow 0$ in $L_2$ and the result stated in (1) follows. Statement (2) follows immediately from (1) because $\nu(J) = 0$ under $H_0^F$: $\mathbf{CF} = 0$.

3.4.2. ANOVA-Type Statistic (ATS). Let $M = C[C^*]^*(-C$ where $[\cdot]^-$ denotes some generalized inverse of A matrix. Then, we define the rank version of the ATS by

$$Q_n(C) = n\hat{p}(J)M\hat{p}(J). \quad (3.16)$$

Note that $M$ is a projection matrix and that $MF = 0 \iff CF = 0$ because $C[C^*]^-$ is a generalized inverse of $C$. Thus, it is also reasonable to use $Q_n(C)$ as a test statistic for testing the hypothesis $H_0^F$: $\mathbf{CF} = 0$. The asymptotic distribution of $Q_n(C)$ is given in the next theorem.

Theorem 3.8. Let $M = C[C^*]^-$ and let $V_n$ and $\hat{V}_n$ be as in (3.4) and (3.14) respectively. Let $\nu(J) = \{J(H) dK\}$ be as defined in Theorem 3.4. Under the assumptions (A1)-(A4), if $n_0 \rightarrow \infty$, then

(1) under the sequence of alternatives (3.5) contiguous to $H_0^F$: $\mathbf{CF} = 0$ the statistic $Q_n(C)$ given in (3.16) has asymptotically the same distribution as $\sum_{i=1}^d \sum_{j=1}^d \hat{\lambda}_{n,i}^2 Z_{n,i,j}^2$ where $\hat{\lambda}_{n,i,j}$ are the characteristic roots of $MV_n$.
and the $Z_{n,i,j}^{2}$ are independent noncentral $\chi^2$-distributed random variables where $\sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j}Z_{n,i,j}^{2} = \nu'(J)MV(J)$.

(2) under $H_{0}^{F}$: $CF = 0$, the statistic $Q_{n}(C)$ given in (3.16) has asymptotically the same distribution as $\sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j}Z_{n,i,j}^{2}$ where the $\lambda_{n,i,j}$ are the characteristic roots of $MV_{n}$ and the $Z_{n,i,j}^{2}$ are independent central $\chi^2$-distributed random variables.

Proof. The proof follows from Theorem 3.4 and well known theorems on the distribution of quadratic forms (see, e.g., Mathai and Provost, 1992, Chap. 4).

The distribution of $\sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j}Z_{n,i,j}^{2}$ can be approximated by a scaled $\chi^2$-distribution.

**Approximation Procedure by a Central $F(f, \infty)$-Distribution.**

(1) Assume that tr($MV_{n}$) $\geq t_{0} > 0$. Then, under $H_{0}^{F}$, the first two moments of the asymptotic distribution of $Q_{n}(C)$/tr($MV_{n}$) and of the $F(f, \infty)$-distribution coincide for $f = [\text{tr}(MV_{n})]^{2}/\text{tr}(MV_{n}MV_{n})$.

(2) The unknown traces tr($MV_{n}$) and tr($MV_{n}MV_{n}$) can be estimated consistently by replacing $V_{n}$ with $\hat{V}_{n}$ given in Theorems 3.5 and 3.6, respectively. This finally leads to the statistic

\[ F_{n}(C) = \frac{1}{\text{tr}(MV_{n})} Q_{n}(C) \approx F(\hat{f}, \infty), \]

where

\[ \hat{f} = \frac{[\text{tr}(\hat{MV}_{n})]^{2}}{\text{tr}(MV_{n}MV_{n})}. \]  

(Here and in the sequel, the sign $\approx$ means “approximately distributed as.”)

**Derivation.** From Theorem 3.8, $Q_{n}(C)$ has asymptotically the same distribution as $\sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j}Z_{n,i,j}^{2}$. We want to approximate the distribution of $Q_{n}(C)$ by $g \cdot Z$ where $Z$ is a random variable with a central $\chi^2$-distribution and $g$ is a constant such that the first two moments coincide. Under $H_{0}^{F}$, we obtain

\[ E(Q_{n}(C)) = \sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j} = \text{tr}(MV_{n}) = E(g \cdot Z) = gf, \]

\[ \text{Var}(Q_{n}(C)) = 2 \cdot \sum_{i=1}^{r} \sum_{j=1}^{d} \lambda_{n,i,j}^{2} = 2 \cdot \text{tr}(MV_{n}MV_{n}) = \text{Var}(g \cdot Z) = 2g^{2}f. \]
Finally, note that \( \text{tr}(\mathbf{MV}_n) \geq t_0 > 0 \) implies that \( \text{tr}(\mathbf{MV}_n \mathbf{MV}_n^\top) \geq t_1 > 0 \). Thus, \( Q_{\alpha}^2(\mathbf{C})(gf) \sim F(f, \infty) \) and the result follows from Theorem 3.5 and Theorem 3.6, respectively.

Remark. The approximation procedure goes back to Box (1954) and turns out to be fairly good for independent observations (see Brunner, Dette, and Munk, 1997). For repeated measures, \( f \) may be biased for small samples.

3.4.3. Comparison of the Two Statistics and Consistency of the Test. The main advantage of the WTS \( Q_{\alpha}^W(\mathbf{C}) \) is that its asymptotic distribution under \( H_0^f \) is a known distribution function, namely a \( \chi^2 \)-distribution. The general drawback of \( Q_{\alpha}^W(\mathbf{C}) \) is that it converges extremely slowly to its asymptotic distribution resulting in rather liberal decisions for small or moderate sample sizes. Moreover, the restrictive assumption that \( \mathbf{V}_n \rightarrow \mathbf{V} \) such that \( \text{rank}(\mathbf{CV}_n) = \text{rank}(\mathbf{CV}) \) cannot be checked. In the case of a singular covariance matrix, the set of alternatives which can be detected by \( Q_{\alpha}^W(\mathbf{C}) \) depends on the structure of the covariance matrix and on the hypothesis under consideration. Thus, there exists a set of fixed alternatives which may not be detected by \( Q_{\alpha}^W(\mathbf{C}) \). This set is given by

\[
\mathbf{p} = [\mathbf{I} - \mathbf{VC} \mathbf{C}^\top] \mathbf{z}, \quad (3.18)
\]

where \( \mathbf{z} \) is an arbitrary vector such that

\[
\mathbf{Cp} = [\mathbf{I} - \mathbf{VC} \mathbf{C}^\top \mathbf{C}] \mathbf{z} \neq \mathbf{0}. \quad (3.19)
\]

To see this, note that \( \lambda = \mathbf{p}^\top \mathbf{C}[\mathbf{C} \mathbf{V} \mathbf{C}^\top] \mathbf{p} = 0 \Leftrightarrow \mathbf{C}[\mathbf{C} \mathbf{V} \mathbf{C}^\top] \mathbf{p} = 0 \) since \( [\mathbf{C} \mathbf{V} \mathbf{C}^\top] \mathbf{p} \) is positive semidefinite and symmetric. Furthermore observe that \( \mathbf{V} \) is a \( g \)-inverse if \( \mathbf{C}[\mathbf{C} \mathbf{V} \mathbf{C}^\top] \mathbf{C} \). Thus, the solution space of the homogeneous system of linear equations \( \mathbf{C}[\mathbf{C} \mathbf{V} \mathbf{C}^\top] \mathbf{p} = 0 \) is given by (3.18) and the restriction (3.19) follows if the test is required to be consistent for alternatives of the form \( \mathbf{Cp} \neq \mathbf{0} \). From this, it is easy to see that a test based on \( Q_{\alpha}^W(\mathbf{C}) \) is consistent for all alternatives \( \mathbf{Cp} \neq \mathbf{0} \) if, e.g., either \( \mathbf{V} \) is non-singular and \( \mathbf{C} \) is of full row rank or if \( \text{rank}(\mathbf{CV}_n) = \text{rank}(\mathbf{C}) \). Note, however, that these simple conditions are only sufficient.

The ATS \( F_{\alpha}(\mathbf{C}) \) has the main disadvantage that its asymptotic distribution under \( H_0^f \) contains unknown quantities, namely the characteristic roots of \( \mathbf{MV}_n \) which are unknown in general and must be estimated. We suggest to use the Box-approximation which is known to be fairly accurate. However, note that even in the asymptotic case, the \( \chi^2/f \)-distribution is an approximation of the true distribution of the statistic under \( H_0^f \). One advantage is that it is neither necessary to assume the convergence of the covariance matrix \( \mathbf{V}_n \) to a constant matrix \( \mathbf{V} \) nor that the rank of \( \mathbf{CV}_n \) is
preserved in the limit $CV$. The only additional assumption to (A1)-(A4) which is needed for the ATS is that $\text{tr}(MV_n) \neq 0$ which means that—regarding the hypothesis of interest—there is at least some variation among the observations of the experiment. This is close to a trivial assumption. The set of alternatives detected by a test based on $F_n(C)$ is given by $v(J) = C p_F(J) = C \{J(H) \ dK \neq 0$ since

$$E(Q_n^d(C)) = \sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_{n,i,j}(1 + \mu_{n,i,j}) = \text{tr}(MV_n) + v(J) \ Mv(J)$$

under the sequence of alternatives (3.5) contiguous to $H_0^*; CF = 0$. This set does not depend on the rank of the covariance matrix—unlike as for the WTS. The main advantage of $F_n(C)$ is that the approximation by the $\chi^2$-distribution works also fairly well for rather small sample sizes (for details, see, e.g., Brunner and Langer, 1999). We recommend the use of the ATS for small and moderate sample sizes and for the case where the estimated covariance matrix is singular or close to being singular which may happen frequently with ordered categorical data.

In the next section, the general results given here will be applied to different two- and three-way layouts.

4. APPLICATIONS

Numerous examples for the application of the WTS in the non-parametric multivariate model (Wilcoxon-scores, no missing values, non-singular covariance matrices) have been given by Akritas and Brunner (1997). Since they did not consider the compound symmetry model, neither the ATS nor missing values, we now give two examples which are concerned with these situations. The application of the WTS is considered for a two-way layout with random interaction where compound symmetry is assumed. The ATS is applied in a three-way layout with two fixed factors with missing values in a multivariate model.

4.1. Two-way Layout with Random Interaction, Compound Symmetry

First, we apply the results of the previous section to a cross-classified design with one random factor ($k = 1, \ldots, n \to \infty$ levels) and with one fixed factor ($j = 1, \ldots, d$ levels). For every subject $k$ and treatment $j$, the number of replications $m_{kj}$ is assumed to be uniformly bounded, i.e. $m_{kj} \leq m_0 < \infty$.

Let $X_k = (X_{k1}, \ldots, X_{kd}) = (X_{k1}, \ldots, X_{kd})$, $k = 1, \ldots, n$ be independent random vectors where $X_{kj} = (X_{kj1}, \ldots, X_{kjem})$ denotes the vector of observations for subject $k$ and where $X_{kj} \sim F_j(x)$, $k = 1, \ldots, n$, $j = 1, \ldots, d$. 
It is reasonable to assume compound symmetry if subjects are split into homogeneous parts and if there is no treatment effect. This means that the parts of each subject are “interchangeable” under the hypothesis. However under a treatment effect, the compound symmetry structure may not be preserved. The property of interchangeable parts of the subjects is reflected by the interchangeability of the random variables $X_{ks}$ and $X_{ks'}$ for $j \neq j' = 1, \ldots, d$ and $\forall s, s'$ under the hypothesis of no treatment effect. The random variables $X_{ks}$ and $X_{ks'}$, $s, s' = 1, \ldots, m_k$, are always interchangeable since they describe replications of the same experiment under the same treatment for the same subject. Treatment effects are described by the relative treatment effects denoted by the weighted mean of the estimated relative treatment effects

\[
\hat{\beta}_j(J) = \left[ J(\hat{H}) \right]_d \hat{D}_j = \frac{1}{n} \sum_{k=1}^n \frac{1}{m_{kj}} \sum_{s=1}^{m_k} \phi_{ks} = \hat{\phi}_j.
\]

and where $\phi_{ks} = \{H(\hat{X}_{ks}) = \{1/N(R_{ks} - \frac{1}{d})\}$ and $R_{ks}$ is the rank of $X_{ks}$ among all the $N$ random variables. According to the notation in Theorem 3.6, let $D = \text{diag}\{\tau_1, \ldots, \tau_d\}$, $\tau_j = n^{-1} \sum_{k=1}^n \tau_{kj}$, and let $D_n = \text{diag}\{\hat{\tau}_1, \ldots, \hat{\tau}_d\}$, $\hat{\tau}_j = n^{-1} \sum_{k=1}^n \hat{\tau}_{kj}$ where

\[
\hat{\tau}_{kj} = \frac{1}{M_k - d} \left[ \frac{1}{m_{kj}} - \frac{1}{d} \sum_{i=1}^d \frac{1}{m_{ki}} \right] S_{k1}^2 + \frac{1}{d-1} S_{k2}^2.
\]

Then the WTS for testing $H_0^J: P_d F = 0$ is derived from (3.15) for $C = P_d$ and $V_n = D_n$. Thus, $C V_n C' = P_d \hat{D}_n P_d$. Note that $\hat{W}_n = \hat{D}_n^{-1}[I_d - J_d \hat{D}_n^{-1}/\text{tr}(\hat{D}_n^{-1})]$ is a $g$-inverse of $P - d \hat{D}_n P_d$ and note that $P_d \hat{W}_n P_d = \hat{W}_n$. Let $\hat{\phi}_{kj} = m_{kj}^{-1} \sum_{s=1}^{m_{kj}} \phi_{ks}$ denote the means of the rank-scores for subject $k$, $j = 1, \ldots, d$ and let

\[
\hat{\phi} = \frac{1}{\sum_{j=1}^d (1/\hat{\tau}_j)} \sum_{j=1}^d \frac{\hat{\phi}_{kj}}{\hat{\tau}_j}
\]

denote the weighted mean of the estimated relative treatment effects $\hat{\beta}_j(J) = \hat{\phi}_{j\cdot}$. Then, from (3.15) and Theorem 3.7,

\[
Q_n^H(P_d) = n \cdot \hat{\phi}(J) P_d [P_d \hat{D}_n P_d - P_d \hat{\beta}(J)] = n \sum_{j=1}^d \frac{1}{\hat{\tau}_j} (\hat{\phi}_{j\cdot} - \hat{\phi})^2
\]
has asymptotically a central \( \chi^2_{d-1} \) -distribution under the hypotheses \( H_0^A \) if \( D \) is of full rank and if \( D_m \rightarrow D \) such that \( \text{rank}(D_m) = \text{rank}(D) = d \). Note that in this case, the quadratic form \( Q^m_d(P_d) \) has a central \( \chi^2_d \) -distribution with \( f = \text{rank}(P_d) = d - 1 \) for any choice of the \( g \)-inverse \([P_d D_m^-1 P_d]^{-1}\).

In the case of equal cell frequencies, \( m_{ij} \equiv m \),

\[
\hat{t}_j = \frac{1}{m(d-1)} \sum_{i=1}^{n} \sum_{j=1}^{d} (\phi_{ij} - \bar{\phi}_{..})^2,
\]

and the statistic simplifies to

\[
Q^m_d(P_d) = \frac{n^2(d-1)}{\sum_{i=1}^{n} \sum_{j=1}^{d} (\phi_{ij} - \bar{\phi}_{..})^2} \sum_{j=1}^{d} (\phi_{..j} - \bar{\phi}_{..})^2,
\]

which has been given by Brunner and Neumann (1982) for the case of continuous distribution functions and for Wilcoxon-scores \( \phi_{ik} = 1/N(R_{ik} - \frac{1}{2}) \).

### 4.2. Three-way Layout, Multivariate Model

In a second example, we apply the general theory developed in Section 3 to a three-way layout with two fixed factors. Let \( X_{ik} = (X_{iak}, ..., X_{iak})' \), \( i = 1, ..., a, k = 1, ..., n_k \) be \( n = \sum_{i=1}^{n} n_k \) independent random vectors where \( X_{ijk} \sim F_{ik}(x) \), \( i = 1, ..., a, j = 1, ..., b, k = 1, ..., n_k \). This model is appropriate for a trial where \( i = 1, ..., a \) groups (whole-plot factor \( A \)) of (independent) subjects are observed repeatedly under \( j = 1, ..., d \) different situations (sub-plot factor \( B \)). The subjects are nested under the levels of factor \( A \) and are crossed with factor \( B \). In designs with longitudinal data, typically the time is the sub-plot factor \( B \). For simplicity, we apply the general results derived in Section 3 only to the case of \( m_{ik} \equiv m = 1 \) replication for each subject \( k \) and combination \((i,j)\) of the factor levels of \( A \) and \( B \). We admit however that some observations are missing (at random) where the notation introduced in (2.2) is used.

We shall give statistics for testing the nonparametric hypotheses \( H_0^A(A) \) of no group effect, \( H_0^B(B) \) of no sub-plot factor effect, and \( H_0^{AB}(AB) \) of no interaction between the groups and the sub-plot factor. Let \( P_d = I_d - (1/d) J_d \) where \( d \) is the dimension of the identity matrix \( I_d \) and of the matrix \( J_d = I_d J_d \) of 1’s. Let \( F = (F_{ij}, ..., F_{ab})' \) denote the vector of the marginal distribution functions \( F_{ij} \) of the vectors \( X_{ijk} \). Then the nonparametric hypotheses are expressed as \( H_0^A(A) : C_A F = 0, H_0^B(B) : C_B F = 0 \) and \( H_0^{AB}(AB) : C_{AB} F = 0 \), respectively, where \( C_A = P_a \otimes (1/b) I_b \), \( C_B = (1/a) I_a \otimes P_b \) and \( C_{AB} = P_a \otimes P_b \).

Let \( R_{ik} \) denote the (mid)-rank of \( X_{ijk} \) among all the \( N = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_k} \hat{t}_{ijk} \) random variables and let \( \phi_{ik} = J(1/N(R_{ik} - \frac{1}{2})) \) denote the rank-score.
of $X_{ik}$. Furthermore let $\Phi_{ik} = (\phi_{ijk}, \ldots, \phi_{iak})'$ denote the vector of the rank-scores for subject $k$ in group $i$ and let

$$
\bar{\Phi}_i = (\bar{\phi}_{i1}, \ldots, \bar{\phi}_{ia})', \quad \text{where} \quad \bar{\phi}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} \phi_{ijk}
$$

denote the mean vector of the scores in group $i$, $i = 1, \ldots, a$. Finally, let $\Phi = (\Phi_{i1}, \ldots, \Phi_{ia})'$ denote the vector of all $a \cdot b$ means of the rank-scores.

An estimator of the covariance matrix is derived from Theorem 3.5,

$$
\hat{V}_n = \bigoplus_{i=1}^{a} \frac{n_i}{n} \hat{V}_{i,n},
$$

where the elements of $\hat{V}_{i,n}$ are given in (3.9) and (3.10), respectively.

Below, the statistics for testing $H_0^F(A)$, $H_0^F(B)$ and $H_0^F(AB)$ are given. They are derived from Theorem 3.8 and from the approximation procedure. To derive the statistic for $H_0^F(A)$, let $\phi_{i1} = b^{-1} \sum_{j=1}^{b} \bar{\phi}_{ij}$, and $\hat{\sigma}_i^2 = n_i \hat{V}_{i,n} \hat{I}_b = \sum_{j=1}^{b} \hat{v}(i,j) + \sum_{j=1}^{b} \hat{\epsilon}(i,j)$. Further let

$$
\hat{\Sigma}_n = \bigoplus_{i=1}^{a} \frac{n_i}{n} \hat{\sigma}_i^2, \quad \text{and} \quad T_A = C_A [C_A C_A']^{-1} C_A = P_A \otimes \frac{1}{b} J_b.
$$

Note that $T_A$ is a projection matrix with $\text{rank}(T_A) = a - 1$ and with identical diagonal elements $(a - 1)/(ab)$. Since $\hat{\Sigma}_n$ is a diagonal matrix, an improved approximation as given in Brunner, Dette, and Munk (1997) can be used. This approximation is rather accurate for $n_i \geq 7$, $i = 1, \ldots, a$. Under $H_0^F(A)$, the distribution of

$$
F_n(T_A) = \frac{n \cdot a \cdot b}{(a - 1) \text{tr}(\hat{\Sigma}_n)} \hat{\Phi}' T_A \hat{\Phi} = \frac{a \cdot b^2}{(a - 1) \sum_{i=1}^{a} \hat{\sigma}_i^2/n_i} \sum_{i=1}^{a} (\hat{\phi}_{i1} - \bar{\phi})^2
$$

is approximately the central $F(\hat{f}_A, \hat{f}_0)$-distribution with estimated degrees of freedom

$$
\hat{f}_A = \frac{[n(a - 1) \sum_{i=1}^{a} \hat{\sigma}_i^2/n_i]^2}{a^2 \text{tr}(P_A \hat{\Sigma}_n P_A \hat{\Sigma}_n)} \quad \text{and} \quad \hat{f}_0 = \frac{[\sum_{i=1}^{a} \hat{\sigma}_i^2/n_i]^2}{\sum_{i=1}^{a} (\hat{\sigma}_i^2/n_i)^2/(n_i - 1)}.
$$

where $\text{tr}(\cdot)$ denotes the trace of a square matrix. It is easy to see that $\hat{f}_A = 1$ if $a = 2$.

To test $H_0^F(B)$: $C_B F = 0$, let $\Phi = (\Phi_{i1}, \ldots, \Phi_{ia})'$ denote the vector of all $a \cdot b$ means of the rank-scores $\Phi_{ijk} = a^{-1} \sum_{i=1}^{a} \phi_{ijk}$, for the $b$
time points. Let $\tilde{V}_B = n \sum_{i=1}^n n_i^{-1} \tilde{V}_{i,n_i}$ and let $T_B = C_B C_B^+ C_B = (\frac{1}{n}) J_u \otimes P_u$. Then, under $H_0^B(B)$, the statistic

$$F_n(T_B) = \frac{na^2}{\text{tr}(P_u \tilde{V}_B)} \phi^T P_u \phi$$

has asymptotically a central $F$-distribution with $F_n = [\text{tr}(P_u \tilde{V}_B)]^2/\text{tr}(P_u \tilde{V}_B P_u \tilde{V}_B)$.

To test $H_0^B(AB)$, let $T_{AB} = P_u \otimes P_k$. Under $H_0^B(AB)$, the statistic

$$F_n(T_{AB}) = \frac{n}{\text{tr}(T_{AB} \tilde{V}_n)} \sum_{i=1}^n \sum_{j=1}^k (\phi_{i,j} - \phi_{i,j})^2$$

has approximately a central $F$-distribution with $F_n = [\text{tr}(T_{AB} \tilde{V}_n)]^2/\text{tr}(T_{AB} \tilde{V}_n T_{AB} \tilde{V}_n)$.

**APPENDIX**

**Lemma A.1.** Let $X_{ik} \sim F_{ij}$, $i = 1, \ldots, r$, $j = 1, \ldots, d$, $k = 1, \ldots, n_i$, $s = 1, \ldots, m_{ik}$ and let $X_{ik}$ and $X_{i'k'}$ be independent for $(i, k) \neq (i', k')$. Then

1. $E[c(x - X_{ik}) - F_{ij}(x)] = 0$, $i = 1, \ldots, r$, $j = 1, \ldots, d$, $k = 1, \ldots, n_i$, $s = 1, \ldots, m_{ik}$.
2. $E[c(X_{ik} - X_{i'k'}) - F_{ij}(X_{ik})] = 0$, if $(i, k) \neq (i', k')$.
3. $\int F_{ij} F_{ij} = \frac{1}{2}$.

**Proof.** (1) By definition,

$$E[c(x - X_{ik}) - F_{ij}(x)] = P(X_{ik} < x) + \frac{1}{2} P(X_{ik} = x) - F_{ij}(x)$$

$$= \frac{1}{2} [F_{ij}(x) + F_{ij}(x)] - F_{ij}(x) = 0.$$ 

(2) This follows by noting that $E[c(X_{ik} - X_{i'k'}) - F_{ij}(X_{ik}) | X_{ik} = x] = 0$ and applying Fubini's theorem for $(i, k) \neq (i', k')$.

(3) The result follows using by parts.

To prove the asymptotic results, we first give some moment inequalities for empirical processes in the mixed model which are needed in the body of the paper.

**Lemma A.2.** Let $X_{ik}$ be as defined in (2.1) and let $J(u)$ denote a score function with bounded first derivative, i.e., $\|J^P\|_w < \infty$. Further let $\lambda_{ik}$ and $n_0$
be as defined in (2.2) and (3.1), respectively. If \( n_0 \to \infty \), then under the assumption (A1),

1. \( E[J[H(x)] - J[H(x)]]^2 \leq \frac{m_0}{m_0} \|J\|_\infty^2 \),

2. \( E[J[H(X_{j,k})] - J[H(X_{j,k})]]^2 \leq \frac{m_0}{m_0} \|J\|_\infty^2 \),

3. \( E[\hat{F}_g(x) - F_g(x)]^2 \leq \frac{1}{n_0} \),

4. \( E[\hat{F}_g(X_{j,k}) - F_g(X_{j,k})]^2 \leq \frac{1}{n_0} \),

5. \( E[\hat{H}(x) - H(x)]^4 \leq \left( \frac{m_0}{m_0} \right)^2 \),

6. \( E[\hat{H}(X_{j,k}) - H(X_{j,k})]^4 \leq \left( \frac{m_0}{m_0} \right)^2 \).

**Proof.** To prove (1), note that by the mean value theorem,

\[ E[J[H(x)] - J[H(x)]]^2 \leq \|J\|_\infty^2 \| \hat{H}(x) - H(x) \|^2. \]

Next, note that by independence,

\[
\begin{align*}
E[\hat{H}(x) - H(x)]^2 &= \frac{1}{N^2} \sum_{i=1}^{r} \sum_{i'=1}^{r} \sum_{j'=1}^{d} \sum_{k=1}^{d} \sum_{k'=1}^{d} \sum_{x=1}^{m_0} \sum_{x'=1}^{m_0} \sum_{s=1}^{m_{y,x}} \sum_{s'=1}^{m_{y,x'}} E[c(x - X_{j,k}) - F_g(x)] [c(x - X_{j',k'}) - F_g(x')] \\
&\leq \frac{1}{N^2} \sum_{i=1}^{r} \sum_{i'=1}^{r} \sum_{j'=1}^{d} \sum_{k=1}^{d} \sum_{k'=1}^{d} \sum_{x=1}^{m_0} \sum_{x'=1}^{m_0} \sum_{s=1}^{m_{y,x}} \sum_{s'=1}^{m_{y,x'}} 1 \\
&\leq \frac{m_0 d}{N} \leq \frac{m_0}{m_0} \frac{m_0}{m_0}
\end{align*}
\]

Statement (2) follows in the same way by noting that

\[
E[(X_{j,k} - X_{s,t}) - F_{g}(X_{j,k})] [c(X_{j,k} - X_{j',k'}) - F_{g}(X_{j',k'})] = 0
\]

if \((s, t) \neq (s', t')\) since either \((i, k) \neq (s, t)\) or \((i, k) \neq (s', t')\) in this case.
Statement (3) follows in the same way as statement (1), since

\[
E[\hat{F}_g(x) - F_g(x)]^2
\]

\[
= \sum_{k=1}^{n} \sum_{k'=1}^{n} \frac{\hat{\alpha}_{i} \hat{\alpha}_{i'}}{\hat{\beta}_{i} \hat{\beta}_{i'}} \frac{m_{ik} m_{i'k'}}{s} \sum_{s=1}^{m_{ik}} \sum_{s'=1}^{m_{i'k'}} E[\{c(x - X_{ik}) - F_g(x)\}] E[\{c(x - X_{i'k'}) - F_g(x)\}]
\]

\[
\leq \sum_{k=1}^{n} \sum_{k'=1}^{n} \frac{\hat{\alpha}_{i} \hat{\alpha}_{i'}}{\hat{\beta}_{i} \hat{\beta}_{i'}} \frac{m_{ik} m_{i'k'}}{s} \sum_{s=1}^{m_{ik}} E[\{c(x - X_{ik}) - F_g(x)\}]^2
\]

\[
\leq \frac{1}{\hat{\beta}_g} \leq \frac{1}{n_0}.
\]

Statement (4) follows by the same argument as used to prove (2).

To prove (5) and (6), the indices \(i\) and \(k\) are collapsed to one index \(\ell\), say, where \(\ell = 1, \ldots, n = \sum_{j=1}^{k} n_j\). Thus, the number of replications for each subject \(\ell\) under treatment \(j\) is relabeled from \(m_{ik}\) to \(m_{ij}\), \(j = 1, \ldots, d\). Similarly, the indices \(j\) and \(s\) are collapsed to one index \(t\), say, where \(t = 1, \ldots, M_t = \sum_{j=1}^{d} m_{ij}\). Thus, the random variables \(X_{ik}\) are relabeled to \(X_{it}\), \(\ell = 1, \ldots, n\), \(t = 1, \ldots, M_t\). Note that \(N = \sum_{\ell=1}^{M} n_{\ell}\) and that \(X_{it}\) and \(X_{it}\) are independent if \(\ell \neq \ell'\). Further let \(\varphi(X_{it}) = c(x - X_{it}) - F_g(x)\). Then,

\[
E[\hat{H}(x) - H(x)]^2 = \frac{1}{N^4} \sum_{\ell=1}^{n} \sum_{t=1}^{n} m_{\ell} \sum_{t'=1}^{n} m_{\ell'} \sum_{t''=1}^{n} m_{\ell''} \sum_{t'''=1}^{n} m_{\ell'''} E[\varphi(X_{\ell t}) \varphi(X_{\ell' t'}) \varphi(X_{\ell'' t''}) \varphi(X_{\ell''' t'''})]
\]

If one of the indices \(\ell, \ell', \ell'', \ell'''\) is different from the three other indices, then it follows that the expectation on the right hand side is 0 since the random variables are independent if the first indices are different. It suffices to count the number of cases where not one of the indices \(\ell, \ell', \ell'', \ell'''\) is different from the three others. This happens if the indices are either all equal or if they are pairwise equal, e.g., \(\ell = \ell' \neq \ell'' = \ell'''\). Thus,

\[
E[\hat{H}(x) - H(x)]^2 \leq \frac{1}{N^4} \sum_{\ell=1}^{n} m_{\ell} \sum_{t=1}^{n} m_{\ell'} \sum_{t''=1}^{n} m_{\ell''} \sum_{t'''=1}^{n} m_{\ell'''} E[\varphi(X_{\ell t}) \varphi(X_{\ell' t'}) \varphi(X_{\ell'' t''}) \varphi(X_{\ell''' t'''})]
\]

\[
\leq \frac{1}{N^4} \sum_{\ell=1}^{n} \left( \sum_{t=1}^{n} M_{\ell} \right)^2
\]

\[
\leq \frac{1}{N^4} \left( \max_{1 \leq \ell \leq n} M_{\ell} \sum_{t=1}^{n} M_{\ell} \right)^2 = \frac{1}{N^4} \max_{1 \leq \ell \leq n} M_{\ell}^2.
\]
Then, the assumptions (A1)(A3), that \( M_{i\ell} \leq dm_0, \ell = 1, \ldots, n \) and that \( N \geq r \) hold. Thus,
\[
E[\hat{H}(x) - H(x)]^4 \ll \frac{1}{N^3} \max_{1 \leq \ell \leq n} M_{\ell}^2 \leq \frac{m_0^2}{r n_0^2}
\]
which proves (5).

The proof for (6) follows by conditioning on \( X_{gks} \) and proceeding as in (5).

To prove Theorem 3.2, we need the following lemma.

**Lemma A.3.** Let \( B_1 = \int J(H)(\hat{H} - H) \, d\hat{F}_y - F_y \) and \( B_2 = \frac{1}{2} \int J'(\tilde{\theta}_N)(\tilde{H} - H)^2 \, d\hat{F}_y - F_y \) and consider the notations of Theorem 3.2. Then, under the assumptions (A1)–(A3), \( E(\sqrt{n} B_k)^2 = O(n_0^{-1}) \), \( k = 1, 2 \).

**Proof.** First consider \( B_1 = \int J(H)(\hat{H} - H) \, d\hat{F}_y - F_y \).

To prove Theorem 3.2, we need the following lemma.
and

\[ B_{22} = -\frac{1}{2} \int J'(\hat{\theta}_g)(\hat{f} - H)^2 \, dF_g. \]

The two terms are estimated separately.

\[
E(\sqrt{n} B_{22})^2 \lesssim \frac{n}{4} \frac{\|J'\|_2^2}{\lambda_{ij}} \sum_{k=1}^m \frac{\lambda_{jk}}{m_{jk}} \sum_{s=1}^{m_{jk}} E[\hat{H}(X_{gks}) - H(X_{gks})]^4
\]

\[
\lesssim N_0 m_0^2 \|J'\|_\infty^2 \left( \frac{1}{n_0} \right).
\]

Using Jensen’s inequality and Lemma A.2 (6). In the same way it follows that

\[
e(\sqrt{n} B_{22})^2 \lesssim \frac{n}{4} \frac{\|J'\|_2^2}{\lambda_{ij}} \int E[\hat{H}(x) - H(x)]^4 \, dF_g
\]

\[
\lesssim N_0 m_0^2 \|J'\|_\infty^2 \left( \frac{1}{n_0} \right)
\]

which completes the proof.

To prove the results for contiguous alternatives, we need the following lemma.

**Lemma A.4.** Under the assumptions (A1)–(A4),

1. \( (J[H(x)] - J[H_d(x)])^2 \lesssim (1/n) \|J'\|_\infty^2. \)
2. \( E[J[\hat{H}(x)] - J[H(x)]]^2 \lesssim (4m_0/rn_0) \|J'\|_\infty^2. \)
3. \( E[J[H(X_{gks})] - J[H_d(X_{gks})]]^2 \lesssim (1/n) \|J'\|_\infty^2. \)

**Proof.** To prove (1), we note that

\[
|H(x) - H_d(x)| \lesssim \frac{1}{\sqrt{n}} \sum_{i=1}^r \sum_{j=1}^d \sum_{k=1}^m \frac{m_{jk}}{N} |F_{gj}(x) - K_{gj}(x)| \lesssim \frac{1}{\sqrt{n}}
\]

and that

\[
J[H(x)] - J[H_d(x)] = J[H(x)] \int_{H_d(x)} J'(\theta_1) [H(x) - H_d(x)]
\]

where \( \theta_1 \) is between \( H_d(x) \) and \( H(x) \) and the result follows.
Statement (2) follows from
\[ J(\hat{H}(x)) - J(H(x)) = J'(\theta_2)(\hat{H}(x) - H(x)) \]
\[ = J'(\theta_2)(\hat{H}(x) - H_\alpha(x) + H_\alpha(x) - H(x)), \]
where \( \theta_2 \) is between \( h(x) \) and \( \hat{H}(x) \). Thus,
\[ (J(\hat{H}(x)) - J(H(x)))^2 \leq \|J'(\theta_2)\|_*^2 (2[H_\alpha(x) - H_\alpha(x)]^2 + 2[H_\alpha(x) - H(x)]^2). \]
Since \( E(\hat{H}(x)) = H_\alpha(x) \), Lemma A.2(1) is still valid if \( H \) is replaced by \( H_\alpha \).
Hence, it follows from Lemma A.2(1) that \( E(\hat{H}(x)) - H_\alpha(x))^2 \leq m_0/(r_0). \)
Together with statement (1), it follows that
\[ E(J(\hat{H}(x)) - J(H(x)))^2 \leq \frac{4m_0}{r_0} \|J'(\theta_2)\|_*^2. \]
Statement (3) follows analogously using (1).

**Proof of Theorem 3.4.** Proceeding as in the proof of Theorem 3.2, we decompose
\[ \hat{v}(J) = \sqrt{n} C \int J(\hat{H}) d\hat{F} = \sqrt{n} C \int J(H) d\hat{F} + a_1 + a_2 - a_3, \]
where
\[ a_1 = C \int [J(\hat{H}) - J(H)] dK, \]
\[ a_2 = \sqrt{n} C \int [J(\hat{H}) - J(H_\alpha)] d(\hat{F} - F_\alpha), \]
\[ a_3 = \sqrt{n} C \int [J(H) - J(H_\alpha)] d(\hat{F} - F_\alpha). \]
To prove the statement in (1), it suffices to consider the \((i, j)\)th components of \( a_j, j = 1, 2, 3 \) and it will be shown that
\[ \begin{align*}
(i) & \quad \left\| \int [J(\hat{H}) - J(H)] dK_{ij} \right\|_2^2 = O\left( \frac{1}{n_0} \right), \\
(ii) & \quad \left\| \sqrt{n} \int [J(\hat{H}) - J(H_\alpha)] d(\hat{F}_{ij} - F_{\alpha,ij}) \right\|_2^2 = O\left( \frac{1}{n_0} \right), \\
(iii) & \quad \left\| \sqrt{n} \int [J(H) - J(H_\alpha)] d(\hat{F}_{ij} - F_{\alpha,ij}) \right\|_2^2 = O\left( \frac{1}{n_0} \right).
\end{align*} \]
Statement (i) follows easily by Jensen’s inequality,

\[ E \left( \int [J(\hat{H}) - J(H)] \, dK_{ij} \right)^2 \leq \int E[J(\hat{H}) - J(H)]^2 \, dK_{ij} \leq \frac{4m_0}{m_0} \|J\|^2 \]

and using lemma A.4 (2).

To prove (ii), we note that under the sequence of alternatives (3.6),

\[ E(F_{ij}(x)) = F_n(x) \quad \text{and} \quad E(H(x)) = H_n(x) \quad (1.1) \]

and thus, Lemma A.3 still holds if \( F_{ij} \) is replaced by \( F_n(x) \) and \( H \) is replaced by \( H_n(x) \). Hence, the result follows in the same way as in the proof of theorem 3.2 using Lemma A.3.

To prove (iii), let \( \varphi(u) = J[H(u)] - J[H_n(u)] \) and consider

\[ E \left( \sqrt{n} \left[ \varphi(x) d(\hat{F}_n(x) - F_n(x)) \right] \right)^2 \]

which follows by independent, (1.1), and Lemma A.4. This completes the proof of statement (1). To prove statement (2), first note that

\[ E(\sqrt{n} C J H \, d\hat{F}) = \int J H \, d\hat{F} = \varphi(J) \quad \text{and} \quad V_n^* = \text{Cov}(\sqrt{n} \int J H \, d\hat{F}) \]

is the covariance matrix of \( \sqrt{n} \int J H \, d\hat{F} \) under the sequence of alternatives defined in (3.5). It is easily seen that \( V_n^* \) and \( V_n \) defined in (3.4) are asymptotically equivalent and the result follows from Theorem 3.3.

**Proof of Theorem 3.5.** First, we consider the diagonal elements \( v_i(j) \) and we define an “estimator” \( \hat{e}_i(j) \) with the unobservable random variables \( \hat{Y}_{ijk} \).

\[ \hat{e}_i(j) = \frac{1}{\hat{z}_{ij}} \sum_{k=1}^n \frac{n_j}{\hat{z}_{ij}} \hat{z}_{ijk} (\hat{Y}_{ijk} - \overline{Y}_{ijk})^2 \]

Recall that \( \hat{z}_{ijk}^2 = \hat{z}_{ijk} \) by definition and note that \( E[\hat{e}_i(j)] = v_i(j) \). To show the consistency of \( \hat{e}_i(j) \), consider the difference
\[
\bar{v}_{i}(j) - v_{i}(j) \\
= \frac{1}{\lambda_{ij}} \sum_{k=1}^{n} \frac{n_{i}}{n_{j}} \lambda_{ik} [ (\bar{y}_{ik}.(J) - p_{ij}(J))^2 - v_{ik}(J) ] \\
+ \frac{1}{\lambda_{ij}} \sum_{k=1}^{n} \frac{n_{i}}{n_{j}} \lambda_{ik} [ (\bar{y}_{ik}.(J) - \bar{y}_{ij}.(J))^2 - (\bar{y}_{ik}.(J) - p_{ij}(J))^2 ] \\
+ \frac{1}{\lambda_{ij}} \bar{v}_{i}(J).
\]

Thus, by independence, Jensen’s inequality, and by the assumptions (A2) and (A3),

\[
E[\bar{v}_{i}(j) - v_{i}(j)]^2 \\
\leq \frac{3}{\lambda_{ij}^2} \sum_{k=1}^{n} N_0^2 \lambda_{ik} E[ (\bar{y}_{ik}.(J) - p_{ij}(J))^2 - v_{ik}(J) ]^2 \\
+ \frac{3}{\lambda_{ij}} \sum_{k=1}^{n} N_0^2 \lambda_{ik} (4 ||J||_{\infty})^2 E[ \bar{y}_{ij}.(J) - p_{ij}(J)]^2 + O\left(\frac{1}{n_0}\right) \\
= O\left(\frac{N_0^2 ||J||_{\infty}^4}{n_0}\right).
\]

next, the unobservable random variables \(Y_{ijkl}(J) = J[H(X_{ijkl})]\) are replaced by the observable rank-scores \(\hat{\phi}_{ijkl} = J[H(X_{ijkl})] = J[1/N(R_{ijkl} - \frac{1}{2})]\) and we consider the difference \(\hat{v}_{i}(j) - \bar{v}_{i}(j)\). It follows that

\[
\left[\hat{v}_{i}(j) - \bar{v}_{i}(j)\right]^2 \\
= \left(\frac{1}{n_i} \sum_{k=1}^{n} \frac{n_i^2 \lambda_{ik}}{\lambda_{ij} \cdot \lambda_{ij} - 1}\right) \left[ \hat{\phi}_{ik}^2 - \bar{y}_{ik}^2(J) \right] \\
- \frac{n_i}{\lambda_{ij} - 1} \left[ \hat{\phi}_{ij}^2 - \bar{y}_{ij}^2(J) \right]^2 \\
\leq 8 \left(\frac{n_i^2}{\lambda_{ij}} \cdot \lambda_{ij} - 1\right)^2 \left[ \frac{n_i}{\lambda_{ij}} \sum_{k=1}^{n} \lambda_{ik} \left[ J(\hat{H}) - J[H] \right]^2 \right] ^2 d\hat{F}_{ik} \\
+ \left[ J(\hat{H}) - J[H] \right]^2 d\bar{F}_{ij} \\
\leq 32N_0^2 (N_0 + 1) ||J||_{\infty}^2 \left[ J(\hat{H}) - J[H] \right]^2 d\bar{F}_{ij}
by the $c_i$-inequality, Jensen’s inequality and the assumption (A2) and (A3). Finally, by Lemma A.2(2), it follows that

$$E[\hat{v}_i(j) - \hat{v}_i(j')]^2 \leq 32N_0^2(N_0 + 1) \|J\|_\infty^2 \|J'\|_\infty^2 \frac{m_0}{r} = O\left(\frac{1}{n_0}\right)$$

which completes the proof for the diagonal elements of $V_{i,n_i}$.

For the off-diagonal elements set

$$\bar{v}_{ij}(j', j') = \frac{\lambda_{jj'}}{\lambda_{jj'} + 1} + \sum_{k=1}^{m_i} \lambda_{jk} \bar{v}_{jk}(\bar{Y}_{jk}(J) - \bar{Y}_{jk}(J'))(\bar{Y}_{jk}(J) - \bar{Y}_{jk}(J'))$$

and note that $E[\bar{v}_{ij}(j, j')] = v_{ij}(j, j')$. Then, the consistency for the off-diagonal elements follows in a similar way as for the diagonal elements.

**Proof of Theorem 3.6.** As in the proof of Theorem 3.5, we define an “estimator” $\hat{\tau}_y$ for $\tau_y$ defined in (3.11) where the unobservable random variables $Y_{jk}(J)$ are used. Let $\bar{Y}_{i,k}(J) = d^{-1} \sum_{j=1}^d Y_{jk}(J)$ and let $S^2_{1,k} = \sum_{j=1}^d \sum_{m_i=1}^\infty (Y_{jk}(J) - \bar{Y}_{jk}(J))^2$ and $S^2_{2,k} = \sum_{j=1}^d (\bar{Y}_{jk}(J) - \bar{Y}_{jk}(J))^2$. Then it follows by Lancaster’s Theorem that $E(S^2_{1,k}) = (M_{ik} - d) \lambda$, and $E(S^2_{2,k}) = (d - 1) \beta_i + (1 - d^{-1}) \sum_{j=1}^d m_{jk} \lambda_i$ where $\lambda_i = \sigma_i^2 - c_i^* + c_i^{**}$ and $\beta_i = c_i^{**} - c_i^*$. Now let

$$c_{ik} = \frac{1}{M_{ik} - d} \left[ \frac{1}{m_{ik}} - \frac{1}{d} \sum_{j=1}^d \frac{1}{m_{ijk}} \right]$$

and let $\hat{\tau}_y = n_i^{-1} \sum_{k=1}^{n_i} \hat{\tau}_{yk}$ where

$$\hat{\tau}_{yk} = \frac{1}{d^{-1}} \frac{S^2_{2,k}}{S^2_{1,k} + \frac{1}{d^{-1}}}$$

are independent, uniformly bounded random variables, $k = 1, \ldots, n_i$, since $|c_{ik}| \leq 1$, $S^2_{1,k} \leq d m_0 \|J\|_\infty^2$ and $S^2_{2,k} \leq d \|J\|_\infty^2$. Moreover, $E(\hat{\tau}_{yk}) = \tau_{yk}$ and hence, by independence,

$$E(\hat{\tau}_y - \tau_y)^2 = \frac{1}{n_i} \sum_{k=1}^{n_i} E(\hat{\tau}_{yk} - \tau_{yk})^2 = O\left(\frac{1}{n_i}\right).$$

Finally, it is shown that $E(\hat{\tau}_y - \tau_y)^2 \to 0$ as $n_i \to \infty$. By Jensen’s inequality and by the $c_i$-inequality,
\[ E(\tilde{\epsilon}_i - \bar{\epsilon}_i)^2 \leq \frac{1}{n_i} \sum_{k=1}^{n_i} E(\tilde{\epsilon}_{ijk} - \bar{\epsilon}_{ijk})^2 \]
\[ \leq \frac{1}{n_i} \sum_{k=1}^{n_i} \left( 2\epsilon^2 g_k E[S_{i,k1}^2 - S_{i,k1}^2]^2 + \frac{2}{(d-1)^2} E[S_{i,k2}^2 - S_{i,k2}^2]^2 \right). \]

Further note that \( |Y_{ijk} + \phi_{ijk}| \leq 2\|J\|_\infty \). Thus,

\[ (S_{i,k1}^2 - S_{i,k1}^2)^2 = \left( \sum_{j=1}^{d} \left[ \sum_{x=1}^{m_{jk}} (\phi_{ijk} - Y_{ijk}^2) - m_{jk} (\phi_{ijk} - Y_{ijk}^2) \right] \right)^2 \]
\[ \leq 8d^2 \|J\|^2_\infty \left[ \sum_{j=1}^{d} \left( \sum_{x=1}^{m_{jk}} (\phi_{ijk} - Y_{ijk}^2) \right)^2 \right] \]
\[ + \left( \frac{1}{m_{jk}} \sum_{x=1}^{m_{jk}} (\phi_{ijk} - Y_{ijk}^2) \right)^2 \]
\[ \leq 16d^2 \|J\|^2_\infty \left[ \sum_{j=1}^{d} \left( \sum_{x=1}^{m_{jk}} (J[H(X_{ijk})] - J[H(X_{ijk})]) \right)^2 \right] \]

by Jensen's inequality and the \( c_r \)-inequality. Furthermore, by Lemma A.2(2),

\[ E(S_{i,k1}^2 - S_{i,k1}^2)^2 \leq \frac{16d^2 m_{jk}^2 \|J\|^2_\infty \|J'\|^2_\infty}{n_i} = O \left( \frac{1}{n_i} \right). \]

In the same way it follows that

\[ (S_{i,k1}^2 - S_{i,k1}^2)^2 \leq 16d \|J\|^2_\infty \left[ \sum_{j=1}^{d} \left( \sum_{x=1}^{m_{jk}} (J[H(X_{ijk})] - J[H(X_{ijk})]) \right)^2 \right] \]

and hence, by Lemma A.2(2),

\[ E(S_{i,k1}^2 - S_{i,k1}^2)^2 \leq \frac{16d^2 m_{jk} \|J\|^2_\infty \|J'\|^2_\infty}{n_i} = O \left( \frac{1}{n_i} \right) \]

and the result follows.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to Professor Manfred Denker for many helpful discussions and valuable comments during the preparation of the paper. Part of this paper was prepared during the first author's visit to the Department of Mathematics, Indiana University. He thanks the Department of Mathematics for its hospitality and support.
REFERENCES


