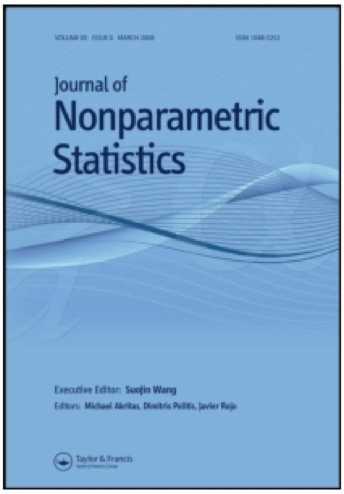

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713645758

Rank tests for anova with large number of factor levels

Haiyan Wang [a]; Michael G. Akritas [ab]
[a] Department of Statistics, Penn State University, University Park, PA, USA [b] Department of Mathematics, NTUA, Zografou, Greece

Online Publication Date: 01 June 2004

To cite this Article Wang, Haiyan and Akritas, Michael G.(2004)'Rank tests for anova with large number of factor levels',Journal of Nonparametric Statistics,16:3,563 — 589

To link to this Article: DOI: 10.1080/10485250310001624774

URL: http://dx.doi.org/10.1080/10485250310001624774

Nonparametric Statistics
Vol. 16(3–4), June–August 2004, pp. 563–589

Taylor & Francis
Taylor & Francis Group

RANK TESTS FOR ANOVA WITH LARGE NUMBER OF FACTOR LEVELS

HAIYAN WANG[a,*] and MICHAEL G. AKRITAS[a,b]

[a]*Department of Statistics, Penn State University, 325 Thomas Bldg., University Park, PA 16802, USA;* [b]*Department of Mathematics, NTUA, 15780 Zografou, Greece*

(Received 4 March 2003; Revised 14 May 2003; In final form 10 June 2003)

Recent papers (Boos, D. D and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statist. Probab. Lett.*, **23**, 183–191; Akritas, M. G. and Arnold, S. (2000). Asymptotics for ANOVA when the number of levels is large. *Journal of the American Statistical Association*, **95**, 212–226; Bathke, A. (2002). ANOVA for a large number of treatments. *Mathematical Methods of Statistics*, **11**(1), 118–132; Akritas and Papadatos 2004; Wang and Akritas 2003) have studied asymptotic properties of ANOVA F-statistics, under general distribution assumptions, when the number of levels is large. Most of these results pertain to statistics based on the original observations, which require strong moment assumptions and are sensitive to outliers. In this paper, we study the use of rank statistics as robust alternatives. Balanced and unbalanced, homoscedastic and heteroscedastic ANOVA models are considered. The main asymptotic tools are the asymptotic rank transform and Hájek's projection method. Simulation results show that the present rank statistics outperform those based on the original observations, in terms of both Type I and Type II errors.

Keywords: Rank tests; Nonparametric; Large number of factor levels; Projection method; Quadratic forms; Unbalanced designs

1 INTRODUCTION

The classical designs of one-way and crossed two-way layouts arise very commonly in scientific investigations. The classical ANOVA model assumes that the error terms ε_{ijk} are i.i.d. normal, in which case the F-statistics for testing the null hypotheses of no treatment effects or no interaction effects have certain optimality properties (cf. Arnold, 1981, Chap. 7). The study of properties of F-tests under violation of the classical assumptions of normality and homoscedasticity has a long history. See for example, Box (1954), Box and Andersen (1955), Scheffé (1959, Chap. 10) and Miller (1986, Chap. 4). However, these studies pertain only to the case when the number of treatment levels is small. In this case, Arnold (1980) showed that the classical F-test is robust to the normality assumption if in addition the sample size per treatment level tends to infinity. Portnoy (1984) considered the case where the number of treatment levels also goes to infinity with the sample size. Recently, there has been some interest in investigating the behavior of these classical tests when the number of treatment levels is large but sample size per treatment combination is limited. Motivating examples from agricultural trials are mentioned in Brownie and Boos (1994) and Wang and Akritas (2002) for one- and two-way layouts, respectively.

* Corresponding author. E-mail: hxw150@psu.edu

ISSN 1048-5252 print; ISSN 1029-0311 online
DOI: 10.1080/10485250310001624774

Boos and Brownie (1995) derived the asymptotic distribution of the classical F-statistic for balanced one-way ANOVA and balanced randomized complete block design (without interaction effects) in this setting. Akritas and Arnold (2000) studied a general class of designs, including balanced and unbalanced, fixed effect and random effect, one- and multi-factor. Independently, and using different asymptotic techniques, Bathke (2002) also generalized the results of Boos and Brownie (1995) to fixed effects balanced multi-factor designs. The aforementioned asymptotic results all pertain only to the homoscedastic case. When the sample size per treatment combination is small, the assumption of homoscedasticity is difficult to justify. Krutchkoff (1989) found that both the classical and weighted F-tests perform poorly in the heteroscedastic case with a large number of groups and few observations per group. Akritas and Papadatos (2002) and Wang and Akritas (2002) study (for the one- and two-way designs, respectively) the asymptotics for suitable test statistics in the heteroscedastic case when the number of treatment levels is large, both with small and large sample sizes per cell.

The aforementioned statistics of Akritas and Papadatos and Wang and Akritas require finite higher order moments. In addition, as is well known, test statistics based on the original observations are sensitive to outliers and can perform poorly away from the normal distribution. In this paper, we consider test statistics based on the overall ranks of the data. Rank tests were also considered in Boos and Brownie (1995) but only for the balanced homoscedastic one-way design and a balanced two-way design with no interaction. The main tools for our asymptotic development is the asymptotic rank transform (Akritas, 1990) and Hájek's projection method for quadratic forms (Akritas and Papadatos, 2002). The reported simulation studies indicate that the rank statistics have more stable Type I error rate and are much more powerful when the error distribution is away from the normal.

The rest of the paper is organized as the following. In Section 2, we present results for the one-way design. The homoscedastic and heteroscedastic cases are treated separately. Section 3 pertains to the two-way layout. For reasons explained there we consider only the heteroscedastic case. Simulation results are presented in Section 4, while the proofs are given in Section 5.

Throughout the article we use $c(x, y) = [I(x \le y) + I(x < y)]/2$, where $I(A)$ is the indicator function for the event A, and the following notation. For the one-way design, we let $\bar{X}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$, $\tilde{X}_{\cdot\cdot} = r^{-1} \sum_{i=1}^{r} \bar{X}_{i\cdot}$, $\bar{X}_{\cdot\cdot} = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij}$, where $N = \sum n_i$. For the two-way design we define $\bar{X}_{ij\cdot} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} X_{ijk}$, $\tilde{X}_{i\cdot\cdot} = c^{-1} \sum_{j=1}^{c} \bar{X}_{ij\cdot}$, $\tilde{X}_{\cdot\cdot\cdot} = (rc)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{X}_{ij\cdot}$, $\tilde{X}_{\cdot j\cdot} = r^{-1} \sum_{i=1}^{r} \bar{X}_{ij\cdot}$. Similar notation applies when X is replaced by Y or Z.

2 MAIN RESULTS ON ONE-WAY ANOVA

In one-way analysis of variance, we have independent observations $X_{ij} \sim F_i(x)$, $i = 1, \ldots, r$, $j = 1, \ldots, n_i$. In order to accommodate discrete and continuous data in the same notation, we will define all distribution function, including empirical ones, to be the average of their right and left continuous versions. We are interested in testing the hypothesis of no treatment effect in a setting where $r \to \infty$. An asymptotically equivalent form of the classical Kruskal–Wallis test statistic compares the treatment and error mean squares calculated on the ranks. Here we will consider this form and establish its asymptotic behavior when the number of treatments is large.

Let $H(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} F_i(x)$, where $N = \sum_{i=1}^{r} n_i$, and set

$$\hat{H}(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} c(X_{ij}, x),$$

where $c(x, y)$ is defined in Section 1, be its empirical version. Then $R_{ij} = 1/2 + N\hat{H}(X_{ij})$ is the (mid-)rank of X_{ij} among all N observations. Let $\text{MS}\alpha_R$ be the mean squares for treatment on mid-rank,

$$\text{MS}\alpha_R = \frac{1}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{R}_{i.} - \bar{R}_{..})^2. \tag{2.1}$$

In all that follows, $Y_{ij} = H(X_{ij}), Z_{ij} = \hat{H}(X_{ij})$, and $\sigma_i^2 = \text{Var}(Y_{ij})$. The results for homoscedastic and heteroscedastic cases are stated separately.

2.1 The Homoscedastic Case

In the homoscedastic case, the null hypothesis can be stated as $H_{0,1}(\alpha)$: $F_1 = \cdots = F_r$, which certainly implies homoscedasticity under the null hypothesis. Let

$$\text{MSE}_R = \frac{1}{N-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(R_{ij} - \bar{R}_{i.})^2 \quad \text{and} \quad F_{R,r} = \frac{\text{MS}\alpha_R}{\text{MSE}_R}$$

be the error mean squares calculated on the mid-ranks, and the corresponding F-ratio.

For convenience, we first state the two assumptions needed for the next theorem.

1. When the n_is remain fixed, assume n_i satisfy

$$B_r = \frac{1}{r}\sum_{i=1}^{r} n_i \longrightarrow b \in (1, \infty), \quad B_{1r} = \frac{1}{r}\sum_{i=1}^{r}\frac{1}{n_i} \longrightarrow b_1,$$

as $r \to \infty$.

2. When $n_i = n_i(r) \to \infty$, as $r \to \infty$. Set $n(r) = \min\{n_i(r); i = 1, \ldots, r\}$, $\kappa(r) = \max\{n_i(r); i = 1, \ldots, r\}$, and assume that

$$n(r) \longrightarrow \infty, \quad \text{and} \quad \kappa(r) - n(r) \le C(r), \quad \text{for all } r,$$

where $C(r) = o(n(r))$, as $r \to \infty$.

THEOREM 2.1 (Unbalanced homoscedastic case) *Let $H_{0,1}(\alpha)$ be satisfied.*

(a) *Under assumption* 1 *and if* $\limsup_{r\to\infty}(1/r)\sum_{i=1}^{r} n_i^{4+\delta} < \infty$, *for some* $\delta > 0$, *then*

$$r^{1/2}(F_{R,r} - 1) \longrightarrow N(0, \tau^2), \quad \text{as } r \longrightarrow \infty,$$

where, letting $\mu_4 = E[H(X_{ij}) - 1/2]^4/\sigma^4$, *with* $\sigma^2 = \text{Var}\{H(H_{1j})\}$,

$$\tau^2 = \frac{2b}{b-1} + (\mu_4 - 3)\frac{b(bb_1 - 1)}{(b-1)^2}.$$

(b) *Under assumption* 2 *listed above*,

$$\tau^{1/2}(F_{R,r} - 1) \longrightarrow N(0, 2), \quad \text{as } r \longrightarrow \infty.$$

COROLLARY 2.2 (Balanced homoscedastic case) *Let $H_{0,1}(\alpha)$ be satisfied.*

(a) *If $n \geq 2$ remains fixed, then*

$$r^{1/2}(F_{R,r} - 1) \longrightarrow N\left(0, \frac{2n}{n-1}\right), \quad \text{as } r \longrightarrow \infty.$$

(b) *If $n = n(r) \to \infty$, as $r \to \infty$, then*

$$r^{1/2}(F_{R,r} - 1) \longrightarrow N(0, 2), \quad \text{as } r \longrightarrow \infty.$$

2.2 The Heteroskedastic Case

In the heteroscedastic case, the null hypothesis that will be tested is $H_{0,2}(\alpha)$: $p_1 = \cdots = p_r$, where $p_i = E\{H(X_{ij})\}$. We remark that the p_i are called relative treatment effects, and are closely related to the d statistic used in behavioral psychology (cf. Cliff, 1993). Testing the equality of the relative treatment effects is commonly advocated in the context of rank methods (cf. Brunner, *et al.*, 2002, p. 37).

In the balanced case, the expectations of the treatment and error mean squares are equal under the null hypothesis, and we can use the same test statistic as in Section 2.1. However, this does not hold in the unbalanced heteroscedastic case. Akritas and Papadatos (2002) proposed two test statistics for dealing with heteroscedasticity in the unbalanced one-way design. The first alters MSE by considering a different weighted average of the group sample variances in such a way as to have the same expected value as MSα, and the other is a weighted Wald-type statistic. The weighted Wald-type statistic requires the sample sizes to also tend to infinity. This is supported by simulations reported in Akritas and Papadatos (2002) which suggest that the asymptotic approximation to the distribution of the statistic becomes satisfactory if the sample sizes are greater than 80. The first approach can be used with small sample sizes and has the advantage that it reduces to the usual statistic in the balanced case. Since it also has a simpler asymptotic theory we will present only a rank version of it, in Theorem 2.3. In addition we will consider, in Theorem 2.5, a different test statistic, which can be used for both small and large sample sizes.

Let

$$\text{MSE}_R^* = \frac{1}{r-1} \sum_{i=1}^{r} \left(1 - \frac{n_i}{N}\right) S_{R,i}^2, \tag{2.2}$$

where $S_{R,i}^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_{i\cdot})^2$, be the altered MSE applied on the ranks and set $F_R = \text{MS}\alpha_R / \text{MSE}_R^*$. We will consider the asymptotic distribution of $\sqrt{r}(F_R - 1)$ as r gets large.

We have the following result.

THEOREM 2.3 (Unbalanced heteroscedastic case) *Let*

$$\sigma_0^2 = \lim_{r\to\infty} \frac{2}{r} \sum_{i=1}^{r} \left(1 - \frac{n_i}{N}\right) \sigma_i^2, \quad \tau_0 = \lim_{r\to\infty} \frac{2}{r} \sum_{i=1}^{r} \left(1 - \frac{n_i}{N}\right)^2 \frac{n_i}{n_i - 1} \sigma_i^4. \tag{2.3}$$

Assume $\sigma_0^2 > 0$ and there exists $\delta > 0$ such that $r^{-1} \sum_{i=1}^{r} n_i^{2+\delta}$ is finite as $r \to \infty$. Then under $H_{0,1}(\alpha)$ or $H_{0,2}(\alpha)$,

$$\sqrt{r}(F_R^{(1)} - 1) \xrightarrow{d} N\left(0, \frac{\tau_0}{\sigma_0^4}\right).$$

The proof is given in Section 5. When the sample sizes are the same, $\text{MSE}_R^* = \text{MSE}_R$. We naturally have the following corollary.

COROLLARY 2.4 (Balanced heteroscedastic case) *Let* MSE_R *be the mean square errors calculated on the mid-ranks, and* $F_{R,r} = \text{MS}\alpha_R/\text{MSE}_R$. *Denote*

$$\tau_0 = \frac{2n}{n-1} \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \sigma_i^4 \quad \text{and} \quad \sigma_0^2 = \lim_{r\to\infty} \frac{2}{r} \sum_{i=1}^{r} \sigma_i^2.$$

Assume $\sigma_0^2 > 0$. *Then*

$$\sqrt{r}(F_{R,r} - 1) \xrightarrow{d} N\left(0, \frac{\tau_0}{\sigma_0^4}\right).$$

Note that the test statistic used in the heteroscedastic unbalanced case is valid only if the sample sizes n_i are small. This motivates us to construct another test statistic that is good for both small and large sample sizes.

Define

$$\text{MST} = \frac{1}{r-1} \sum_{i=1}^{r} (\bar{X}_{i.} - \tilde{X}_{..})^2, \quad \text{MSE}^{(2)} = \frac{1}{r} \sum_{i=1}^{r} \frac{1}{n_i} S_{X,i}^2, \text{ and } F_2 = \frac{\text{MST}}{\text{MSE}^{(2)}}, \tag{2.4}$$

where $S_{X,i}^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2$. Then under the null hypothesis of no treatment effect,

$$E(\text{MST}) = E(\text{MSE}^{(2)}) = r^{-1} \sum_{i=1}^{r} \frac{1}{n_i} \sigma_i^2.$$

It is reasonable to compare MST with $\text{MSE}^{(2)}$ for the test of no treatment effect. We will consider the rank version of above statistics.

THEOREM 2.5 (Unbalanced heteroscedastic case) *Let* $H_{0,1}(\alpha)$ *or* $H_{0,2}(\alpha)$ *be satisfied.*

(a) *If* $n_i \geq 2$ *fixed, assume the following limits exist*

$$\nu_2^2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{1}{n_i} \sigma_i^2 > 0 \quad \tau_2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{2\sigma_i^4}{n_i(n_i - 1)},$$

then $\sqrt{r}(F_{R,2} - 1) \xrightarrow{d} N(0, \tau_2/\nu_2^4)$ *as* $r \to \infty$, *where* MST_R, $\text{MSE}_R^{(2)}$, $F_{R,2}$ *are the rank version of* MST, $\text{MSE}^{(2)}$ *and* F_2 *defined in Eq.* (2.4).

(b) *If* $n_i = n_i(r) \to \infty$ *as* $r \to \infty$, *assume* $n(r) = o(r^{\delta/(4+2\delta)})$ *for some* $\delta > 0$ *and the following limits exist*

$$\nu_3^2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{n(r)}{n_i} \sigma_i^2 > 0 \quad \tau_3 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{2n^2(r)\sigma_i^4}{n_i(n_i - 1)},$$

we have $\sqrt{r}(F_{R,2} - 1) \xrightarrow{d} N(0, \tau_3/\nu_2^4)$ *as* $r \to \infty$.

COROLLARY 2.6 (Unbalanced homoscedastic case) *Let $H_{0,1}(\alpha)$ or $H_{0,2}(\alpha)$ be satisfied.*

(a) *If $n_i \geq 2$ fixed, assume the following limits exist*

$$\nu_2^2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{1}{n_i} > 0, \quad \tau_2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{2}{n_i(n_i-1)},$$

then $\sqrt{r}(F_{R,2}-1) \xrightarrow{d} N(0, \tau_2/\nu_2^4)$ as $r \to \infty$.

(b) *If $n_i = n_i(r) \to \infty$ as $r \to \infty$, assume $n(r) = o(r^{\delta/(4+2\delta)})$ for some $\delta > 0$ and the following limits exist*

$$\nu_3^2 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{n(r)}{n_i} > 0, \quad \tau_3 = \lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} \frac{2n^2(r)}{n_i(n_i-1)},$$

we have $\sqrt{r}(F_{R,2}-1) \xrightarrow{d} N(0, \tau_3/\nu_3^4)$ as $r \to \infty$.

Under the homoscedastic case with balanced sample sizes, the test statistic reduces to the classical F-statistic introduced in Section 2.1. We naturally have the following corollary.

COROLLARY 2.7 (Balanced homoscedastic case) *Let $H_{0,1}(\alpha)$ or $H_{0,2}(\alpha)$ be satisfied. $F_{R,r}$ is defined in Section 2.1*

(a) *If $n \geq 2$ fixed,*

$$\sqrt{r}(F_{R,r}-1) \xrightarrow{d} N\left(0, \frac{2n}{n-1}\right) \quad \text{as } r \to \infty.$$

(b) *If $n = n(r) \to \infty$ as $r \to \infty$, assume $n(r) = o(r^{\delta/(4+2\delta)})$ for some $\delta > 0$, then $\sqrt{r}(F_{R,r}-1) \xrightarrow{d} N(0, 2)$ as $r \to \infty$.*

Note that Theorem 2.1, Corollaries 2.2, 2.6 and 2.7 are all for homoscedastic case. However, Theorem 2.1 and Corollary 2.2 are obtained under $H_{0,1}(\alpha)$, while Corollaries 2.6 and 2.7 hold under both $H_{0,1}(\alpha)$ and $H_{0,2}(\alpha)$.

3 MAIN RESULTS ON TWO-WAY ANOVA

In two-way analysis of variance, we have independent observations $X_{ijk} \sim F_{ij}(x)$, $i = 1, \ldots, r$, $j = 1, \ldots, c$, $k = 1, \ldots, n_{ij}$. For this section we set $H(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} F_{ij}(x)$ for the average distribution function, and

$$\hat{H}(x) = N^{-1} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} c(X_{ijk}, x),$$

for its empirical version. Thus $R_{ijk} = 1/2 + N\hat{H}(X_{ijk})$ is the (mid-)rank of X_{ijk} among all observations. We also let $Y_{ijk} = H(X_{ijk})$, so the relative treatment effects are $p_{ij} = E(Y_{ijk})$. Consider the decomposition

$$p_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad i = 1, \ldots, r, \quad j = 1, \ldots, c, \quad k = 1, \ldots, n_{ij},$$

where $\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = \sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{c} \gamma_{ij} = 0$. We are interested in testing for no main effects and no interaction effects, namely $H_0(\alpha)$: all $\alpha_i = 0$, $H_0(\beta)$: all $\beta_i = 0$, and

$H_0(\gamma)$: all $\gamma_{ij} = 0$. Due to the fact that the number of column levels is fixed, testing for column effects requires different techniques and will not be presented here.

Because even under homoscedasticity the rank-transformed data can be heteroscedastic (Akritas, 1990), we will consider only the heteroscedastic case. Motivated by Wang and Akritas (2002), we will consider the test statistics defined from the following mean squares

$$\mathrm{MST}_\alpha = \frac{1}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{c}(\tilde{R}_{i..} - \tilde{R}_{...})^2$$

$$\mathrm{MST}_\gamma = \frac{1}{(r-1)(c-1)}\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{R}_{ij.} - \tilde{R}_{i..} - \tilde{R}_{.j.} + \tilde{R}_{...})^2$$

$$\mathrm{MSE}_R^{(3)} = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{S_{ij,R}^2}{n_{ij}}.$$

Thus we consider test statistics

$$F_{R,\alpha,2} = \frac{\mathrm{MST}_\alpha}{\mathrm{MSE}_R^{(3)}} \quad \text{and} \quad F_{R,\gamma,2} = \frac{\mathrm{MST}_\gamma}{\mathrm{MSE}_R^{(3)}},$$

for the hypotheses $H_0(\alpha)$ and $H_0(\gamma)$, respectively, and study the asymptotic distribution of $\sqrt{r}(F_{R,\alpha,2} - 1)$ and $\sqrt{r}(F_{R,\gamma,2} - 1)$ when r is large. Note that $F_{R,\alpha,2}$ and $F_{R,\gamma,2}$ reduce to the usual F-ratios in balanced case.

THEOREM 3.1 (Unbalanced case) *Let* $\mathrm{Var}(Y_{ijk}) = \sigma_{ij}^2$.
(a) *For* $n_{ij} \geq 2$ *fixed*, *we have*

$$\textit{under } H_0(\alpha), \quad \sqrt{r}(F_{R,\alpha,2} - 1) \xrightarrow{d} N\left(0, \frac{2(\phi^4 + \eta^4)}{c\nu_4^4}\right) \quad \textit{as } r \to \infty,$$

$$\textit{under } H_0(\gamma), \quad \sqrt{r}(F_{R,\gamma,2} - 1) \xrightarrow{d} N\left(0, \frac{2\phi^4(c-1)^2 + 2\eta^4}{c\nu_4^4}\right) \quad \textit{as } r \to \infty,$$

where

$$\nu_4^2 = \lim_{r\to\infty}\frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\sigma_{ij}^2}{n_{ij}}, \quad \phi^4 = \lim_{r\to\infty}\frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\sigma_{ij}^4}{n_{ij}(n_{ij}-1)},$$

$$\eta^4 = \lim_{r\to\infty}\frac{1}{rc}\sum_{i=1}^{r}\sum_{j_1\neq j_2}^{c}\frac{\sigma_{ij_1}^2\sigma_{ij_2}^2}{n_{ij_1}n_{ij_2}},$$

assuming the limits exist.
(b) *If* $n_{ij} = n_{ij}(r) \to \infty$, *set* $n(r) = \min\{n_{ij}(r),\ i = 1,\dots,r, j = 1,\dots,c\}$, $\kappa(r) = \max\{n_{ij}(r), i = 1,\dots,r,\ j = 1,\dots,c\}$, *and assume that*

$$n(r) \to \infty, \quad \textit{and} \quad \kappa(r) - n(r) \leq C(r), \ \textit{for all } r,$$

where $C(r) = o(n(r))$, *as* $r \to \infty$. *In addition, assume that for some* $\delta > 0$, $n(r) = o(r^{\delta/(4+2\delta)})$. *Then*

$$\text{under } H_0(\alpha), \quad \sqrt{r}(F_{R,\alpha,2} - 1) \xrightarrow{d} N\left(0, \frac{\tau_4}{\nu_5^4}\right) \quad \text{as } r \to \infty,$$

$$\text{under } H_0(\gamma), \quad \sqrt{r}(F_{R,\gamma,2} - 1) \xrightarrow{d} N\left(0, \frac{\tau_4 + \tau_5}{\nu_5^4}\right) \quad \text{as } r \to \infty,$$

where

$$\nu_5^2 = \lim_{r\to\infty} \frac{n(r)}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\sigma_{ij}^2}{n_{ij}}, \quad \tau_4 = \lim_{r\to\infty} \frac{2}{rc^2} \sum_{i=1}^{r} \left(\sum_{j=1}^{c} \sigma_{ij}^2 \right)^2,$$

$$\tau_5 = \lim_{r\to\infty} \frac{2(c-2)}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \sigma_{ij}^4,$$

assuming the limits exist.

COROLLARY 3.2 (Balanced case) *Assume* $r^{-1}c^{-1} \sum_{i=1}^{r} \sum_{j=1}^{c} \sigma_{ij}^2 \to \nu_{b,4}^2$.

(a) *For* $n \geq 2$ *fixed we have*

$$\text{under } H_0(\alpha), \quad \sqrt{r}(F_{R,\alpha,2} - 1) \xrightarrow{d} N\left(0, \frac{2(\phi_b^4 + \eta_b^4)}{c\nu_{b,4}^4}\right) \quad \text{as } r \to \infty,$$

$$\text{under } H_0(\gamma), \quad \sqrt{r}(F_{R,\gamma,2} - 1) \xrightarrow{d} N\left(0, \frac{2\phi_b^4(c-1)^2 + 2\eta_b^4}{c\nu_{b,4}^4}\right) \quad \text{as } r \to \infty,$$

where

$$\phi_b^4 = \lim_{r\to\infty} \frac{n}{rc(n-1)} \sum_{i=1}^{r} \sum_{j=1}^{c} \sigma_{ij}^4, \quad \eta_b^4 = \lim_{r\to\infty} \frac{1}{rc} \sum_{i=1}^{r} \sum_{j_1 \neq j_2}^{c} \sigma_{ij_1}^2 \sigma_{ij_2}^2,$$

assuming the limits exist.

(b) *If* $n(r) \to \infty$, *as* $r \to \infty$, *and for some* $\delta > 0$, $n(r) = o(r^{\delta/(4+2\delta)})$, *we have*

$$\text{under } H_0(\alpha), \quad \sqrt{r}(F_{R,\alpha,2} - 1) \xrightarrow{d} N\left(0, \frac{\tau_4}{\nu_{b,4}^4}\right) \quad \text{as } r \to \infty,$$

$$\text{under } H_0(\gamma), \quad \sqrt{r}(F_{R,\gamma,2} - 1) \xrightarrow{d} N\left(0, \frac{\tau_4 + \tau_5}{\nu_{b,4}^4}\right) \quad \text{as } r \to \infty,$$

where τ_4 *and* τ_5 *are defined in Theorem* 3.1.

4 SIMULATION RESULTS

The simulations reported in this section pertain only to the two-way ANOVA design. We compare the asymptotic tests based on the original observations (Wang and Akritas, 2002) with the present rank tests. Type I error-rate results are reported for both the row main effect and interaction effect, with the number of row factor levels taking values $r = 10, 15, 20, 25$. Results for the achieved power are reported only for testing for row main effects with $r = 20$. The number of column factors is $c = 2$ for all simulations. The simulations are based on

TABLE I Estimated Level for Balanced Case, $\alpha = 0.05, n = 4$.

		WA-test		*Rank test*	
Error	r	$H_0(\alpha)$	$H_0(\gamma)$	$H_0(\alpha)$	$H_0(\gamma)$
Normal(0,1)	10	0.0520	0.0465	0.0600	0.0565
	15	0.0430	0.0405	0.0505	0.0415
	20	0.0415	0.0435	0.0480	0.0460
	25	0.0345	0.0360	0.0375	0.0450
Exp(1)	10	0.0340	0.0390	0.0495	0.0590
	15	0.0255	0.0240	0.0510	0.0455
	20	0.0265	0.0280	0.0535	0.0485
	25	0.0300	0.0235	0.0485	0.0425
Lognormal(0,1)	10	0.0185	0.0185	0.0580	0.0540
	15	0.0125	0.0135	0.0420	0.0460
	20	0.0085	0.0110	0.0490	0.0455
	25	0.0070	0.0095	0.0375	0.0450
Cauchy	10	0.0040	0.0050	0.0660	0.0600
	15	0.0025	0.0015	0.0435	0.0480
	20	0.0005	0.0015	0.0550	0.0445
	25	0.0000	0.0005	0.0435	0.0440

2000 replications and use the normal, exponential, log-normal and Cauchy distributions. The elements of the asymptotic variance were estimated in the obvious way except for σ_{ij}^4 where we used an unbiased estimator as suggested in Akritas and Papadatos (2004).

Tables I and II report the achieved α-levels and power, respectively, for the balanced case with $n = 4$. Tables III and IV report the achieved α-levels and power, respectively, for the unbalanced case. When $r = 10$, the sample sizes in the unbalanced case are $n_1 = c(4, 5, 4, 6, 5, 6, 4, 5, 4, 4)$, $n_2 = c(4, 4, 4, 4, 5, 4, 4, 5, 6, 5)$; when $r = 15$, $n_1 = (5, 4, 4, 4, 4, 4, 6, 4, 4, 5, 5, 4, 4, 5, 4)$, and $n_2 = (4, 4, 4, 4, 4, 4, 7, 4, 4, 4, 4, 4, 4, 5, 4)$; when $r = 20$, $n_1 = (4, 4, 4, 4, 4, 4, 4, 4, 6, 6, 4, 4, 5, 4, 4, 4, 4, 4, 4, 5)$, and $n_2 = (4, 4, 4, 5, 4, 6, 4, 5, 4, 5, 4, 4, 4, 4, 5, 4, 4, 4, 4, 4)$; when $r = 25$, $n_1 = (6, 4, 5, 4, 5, 4, 4, 4, 5, 4, 6, 4, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4)$, and $n_2 = (6, 4, 6, 4, 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 4, 5, 4, 4, 4, 4, 4, 6, 4, 5)$. For the power comparison, only $r = 20$ is reported.

In all cases observations under the alternative were generated as $X_{ijk} = i * \tau / r +$ error, where the error term and τ are specified at the tables.

TABLE II Achieved Power for Balanced Case, $\alpha = 0.05$, $r = 20, n = 4$.

Error	τ	*WA-test*	*Rank test*
Normal(0,1)	0	0.0375	0.0460
	0.5	0.1080	0.1190
	1.0	0.4585	0.4580
	1.5	0.8950	0.8870
Exp(1)	0	0.0275	0.0505
	0.5	0.0665	0.2490
	1.0	0.4045	0.8280
	1.5	0.8395	0.9915
Lognormal(0,1)	0	0.011	0.0460
	1	0.041	0.5835
	2	0.322	0.9895
	3	0.721	1

TABLE III Estimated Level for Unbalanced Case, $\alpha = 0.05$.

		WA-test		*Rank test*	
r	*Error*	$H_0(\alpha)$	$H_0(\gamma)$	$H_0(\alpha)$	$H_0(\gamma)$
10	Normal(0,1)	0.0565	0.0545	0.0630	0.0575
	Exp(1)	0.0325	0.0325	0.0635	0.0530
	Lognormal(0,1)	0.0160	0.0260	0.0540	0.0600
	Cauchy	0.0055	0.0045	0.0505	0.0585
15	Normal(0,1)	0.0455	0.0485	0.0485	0.0495
	Exp(1)	0.0250	0.0390	0.0530	0.0585
	Lognormal(0,1)	0.0060	0.0110	0.0415	0.0570
	Cauchy	0.0000	0.0010	0.0590	0.0495
20	Normal(0,1)	0.0430	0.0405	0.0500	0.0425
	Exp(1)	0.0245	0.0270	0.0470	0.0585
	Lognormal(0,1)	0.0080	0.0120	0.0515	0.0385
	Cauchy	0.0000	0.0000	0.0450	0.0475
25	Normal(0,1)	0.0495	0.0335	0.0505	0.0405
	Exp(1)	0.0215	0.0275	0.0400	0.0415
	Lognormal(0,1)	0.0095	0.0070	0.0505	0.0405
	Cauchy	0.0005	0.0010	0.0405	0.0420

TABLE IV Achieved Power for Unbalanced Case, $\alpha = 0.05$, $r = 20$.

Error	τ	*WA-test*	*Rank test*
Normal	0	0.011	0.0460
	0.5	0.041	0.5835
	1.0	0.322	0.9895
	1.5	0.721	1
Exp(1)	0	0.0285	0.0505
	0.5	0.089	0.27
	1.0	0.457	0.8545
	1.5	0.8895	0.9955
Lognormal(0,1)	0	0.0125	0.043
	1	0.048	0.591
	2	0.3555	0.995
	3	0.751	1

From these tables it is easily seen that the Type I error rate is more stable for the rank test, and the power of the rank test is much better than that based on the original observations away from the normal distribution.

5 PROOFS

5.1 Proofs for the One-way Design

Proof of Theorem 2.1 Because $\text{MS}\alpha_R$ here is the same as in the unbalanced heteroscedastic case, several results/lemmas from the proof of Theorem 2.3 will be used.

$$\sqrt{r}(F_{R,r} - 1) = \frac{\sqrt{r}}{\text{MSE}_R}(\text{MS}\alpha_R - \text{MSE}_R).$$

The proof of $\text{MSE}_R/N^2 \xrightarrow{p} \sigma^2$ is similar to the proof of Lemma 5.2 below, and thus the proof is omitted. So we only need to consider the asymptotic distribution of $\sqrt{r}(\text{MS}\alpha_R - \text{MSE}_R)/N^2$. Let $P_{R,\alpha}$ be defined in Eq. (5.2). Then Lemma 5.3 implies that under $H_{0,1}(\alpha)$,

$$\sqrt{r}\frac{\text{MS}\alpha_R - P_{R,\alpha}}{N^2} \xrightarrow{p} 0 \quad \text{as } r \longrightarrow \infty.$$

So $\sqrt{r}(\text{MS}\alpha_R - \text{MSE}_R)/N^2$ has same asymptotic distribution as $\sqrt{r}(P_{R,\alpha} - \text{MSE}_R)/N^2$. Let MSE_Z be similarly defined as MSE_R with R_{ij} replaced by Z_{ij}. Then

$$\begin{aligned}\frac{\text{MSE}_R}{N^2} = \text{MSE}_Z &= \frac{1}{N-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Z_{ij} - p_i + p_i - \bar{Z}_{i.})^2 \\ &= \frac{1}{N-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Z_{ij} - p_i)^2 - \frac{1}{N-r}\sum_{i=1}^{r} n_i(\bar{Z}_{i.} - p_i)^2 \\ &= \frac{1}{N-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\frac{n_i - 1}{n_i}(Z_{ij} - p_i)^2 - \frac{1}{N-r}\sum_{i=1}^{r}\frac{1}{n_i}\sum_{j\neq j'}^{n_i}(Z_{ij} - p_i)(Z_{ij'} - p_i),\end{aligned}$$

and

$$\begin{aligned}\frac{P_{R,\alpha}}{N^2} = P_{Z,\alpha} &= \frac{1}{r-1}\sum_{i=1}^{r} n_i\left(1 - \frac{n_i}{N}\right)(\bar{Z}_{i.} - p_i)^2 \\ &= \frac{1}{r-1}\sum_{i=1}^{r}\frac{1}{n_i}\left(1 - \frac{n_i}{N}\right)\sum_{j\neq j'}^{n_i}(Z_{ij} - p_i)(Z_{ij'} - p_i) \\ &\quad + \frac{1}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\frac{1}{n_i}\left(1 - \frac{n_i}{N}\right)(Z_{ij} - p_i)^2.\end{aligned}$$

Hence

$$\sqrt{r}\frac{P_{R,\alpha} - \text{MSE}_R}{N^2} = \sqrt{r}(P_{Z,\alpha} - \text{MSE}_Z) = T_3(\mathbf{Z}) + T_4(\mathbf{Z}),$$

where

$$\begin{aligned}T_3(\mathbf{Z}) &= \sqrt{r}\sum_{i=1}^{r}\sum_{j\neq j'}^{n_i}\left(d_i + \frac{1}{N-r}\right)(Z_{ij} - p_i)(Z_{ij'} - p_i), \\ T_4(\mathbf{Z}) &= \sqrt{r}\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i(Z_{ij} - p_i)^2,\end{aligned} \tag{5.1}$$

with

$$d_i = \frac{N-1}{(N-r)(r-1)n_i} - \frac{1}{N(r-1)} - \frac{1}{N-r}.$$

$T_3(\mathbf{Y})$ and $T_4(\mathbf{Y})$ are similarly defined with Z_{ij} replaced by Y_{ij}. Following the same procedure as in the proof of Lemma 5.4, it can be shown that $T_3(\mathbf{Z}) - T_3(\mathbf{Y}) = o_p(1)$ as $r \to \infty$. Lemma 5.1 shows that $T_4(\mathbf{Z}) - T_4(\mathbf{Y}) = o_p(1)$. Therefore, $\sqrt{r}(P_{R,\alpha} - \text{MSE}_R)/N^2$ has same asymptotic distribution as $T_3(\mathbf{Y}) + T_4(\mathbf{Y})$. Asymptotic normality of $T_3(\mathbf{Y}) + T_4(\mathbf{Y})$ is shown in the proof of Theorem 3.2 of Akritas and Papadatos (2004), by noting that it can be written as $\sqrt{r}\mathbf{Y}'\mathbf{A}_D\mathbf{Y}$, where $\mathbf{Y} = (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{rn_r})$ and $\mathbf{A}_D$ is defined in the aforementioned reference. This completes the proof of the theorem.

LEMMA 5.1 *Let $T_4(\mathbf{Z})$ be defined in Eq.* (5.1). *Under $H_{0,1}(\alpha)$, we have $T_4(\mathbf{Z}) - T_4(\mathbf{Y}) \xrightarrow{p} 0$, as $r \to \infty$.*

Proof Let $H_{0,1}(\alpha)$ hold, and H denote the common distribution in all categories, note that $p_i = E(H(X_{i1}))$ are all equal to 0.5. Define

$$h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}) = [c(X_{i_1,j_1}, X_{i,j}) - p_{i_1}][c(X_{i_2,j_2}, X_{i,j}) - p_{i_2}] \\ - [H(X_{ij}) - p_{i_1}][H(X_{ij}) - p_{i_2}].$$

Then, under $H_{0,1}(\alpha)$,

$$T_4(\mathbf{Z}) - T_4(\mathbf{Y}) = \sqrt{r}\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i N^{-2} \sum_{i_1=1}^{r}\sum_{j_1=1}^{n_{i_1}}\sum_{i_2=1}^{r}\sum_{j_2=1}^{n_{i_2}} h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}).$$

If any two or all three pairs in $\{(i_1, j_1), (i_2, j_2), (i, j)\}$ are same, the summation is $o_p(1)$ since $h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})$ is uniformly bounded and $d_i = O(r^{-1})$. So

$$T_4(\mathbf{Z}) - T_4(\mathbf{Y}) = \sqrt{r}N^{-2} \sum_{(i_1,j_1)\neq(i_2,j_2)\neq(i,j)} d_i h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}) + o_p(1).$$

To show that the above converges to zero, we will consider its projection. Letting

$$\tilde{h}_1(x) = E(h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i_1,j_1} = x) = E[(H(X_{11}) - p_1)(c(x, X_{11}) - H(X_{11}))],$$

we have that for $(i_1, j_1) \neq (i_2, j_2) \neq (i, j)$, $E(h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i,j}) = 0$,

$$E(h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i_1,j_1}) = \tilde{h}_1(X_{i_1,j_1}), = E(h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i_2,j_2}) = \tilde{h}_1(X_{i_2,j_2}),$$

and if (i_3, j_3) is different from these three pairs, $E(h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i_3,j_3}) = 0$. So under $H_{0,1}(\alpha)$,

$$E(T_4(\mathbf{Z}) - T_4(\mathbf{Y})|X_{i_3,j_3}) = \frac{2\sqrt{r}}{N}\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i \tilde{h}_1(X_{i_3 j_3}),$$

and the projection of $T_4(\mathbf{Z}) - T_4(\mathbf{Y})$ is

$$\begin{aligned}\hat{T}_{4ZY} &= \sum_{i_3=1}^{r}\sum_{j=1}^{n_{i_3}} E(T_4(\mathbf{Z}) - T_4(\mathbf{Y})|X_{i_3 j_3}) - (N-1)E(T_4(\mathbf{Z}) - T_4(\mathbf{Y})) \\ &= \frac{2\sqrt{r}}{N}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}}\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i \tilde{h}_1(X_{i_3 j_3}) \\ &= \frac{\sqrt{r}}{N^2}\sum_{i_1=1}^{r}\sum_{j_1=1}^{n_{i_1}}\sum_{i_2=1}^{r}\sum_{j_2=1}^{n_{i_2}}\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i[\tilde{h}_1(X_{i_1 j_1}) + \tilde{h}_1(X_{i_2 j_2})].\end{aligned}$$

Thus,

$$T_4(\mathbf{Z}) - T_4(\mathbf{Y}) - \hat{T}_{4ZY} = \frac{\sqrt{r}}{N^2} \sum_{i_1=1}^{r} \sum_{j_1=1}^{n_{i_1}} \sum_{i_2=1}^{r} \sum_{j_2=1}^{n_{i_2}} \sum_{i=1}^{r} \sum_{j=1}^{n_i} d_i h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})$$
$$= \frac{\sqrt{r}}{N^2} \sum_{(i_1,j_1)\neq(i_2,j_2)\neq(i,j)} d_i h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}) + o_p(1),$$

where

$$h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}) = h(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j}) - \tilde{h}_1(X_{i_1 j_1}) - \tilde{h}_1(X_{i_2 j_2}).$$

Therefore,

$$E\left(\frac{\sqrt{r}}{N^2} \sum_{(i_1,j_1)\neq(i_2,j_2)\neq(i,j)} d_i h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})\right)^2$$
$$= \frac{r}{N^4} \sum_{(i_1,j_1\neq(i_2,j_2)\neq i,j)} \sum_{(i_3,j_3)\neq(i_4,j_4)\neq(i_5,j_5)} d_i d_{i5} E[h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})$$
$$\times h^*(X_{i_3,j_3}, X_{i_4,j_4}, X_{i_5,j_5})].$$

Note that $E(h^*(X_{i_1,j_1}, X_{i_2,j_2}, X_{i,j})|X_{i_3,j_3}) = 0$, for all indices (i_3, j_3). Thus, if the number of different pairs in $\{(i_1, j_1), (i_2, j_2), (i, j), (i_3, j_3), (i_4, j_4), (i_5, j_5)\}$ is six or five, the expectation under the summation is zero. It follows that $E(T_4(\mathbf{Z}) - T_4(\mathbf{Y}) - \tilde{T}_{4ZY})^2 = O(r^{-1})$ and thus $T_4(\mathbf{Z}) - T_4(\mathbf{Y}) - \hat{T}_{4ZY} \xrightarrow{p}$. However, $\sum_{i=1}^{r}\sum_{j=1}^{n_i} d_i = 0$ implies that $\tilde{T}_{4ZY} = 0$. So $T_4(\mathbf{Z}) - T_4(\mathbf{Y}) \xrightarrow{p} 0$.

Proof of Theorem 2.3 The proof uses three lemmas which are stated and proved after the proof of the theorem. By Lemma 5.2, $\mathrm{MSE}_R^*/N^2 \xrightarrow{p} \sigma_0^2$. So we only need to consider the asymptotic distribution of $\sqrt{r}(\mathrm{MS}\alpha_R - \mathrm{MSE}_R^*)/N^2$. Set

$$P_{R,\alpha} = (r-1)^{-1} \sum_{i=1}^{r} n_i \left(1 - \frac{n_i}{N}\right) (\bar{R}_{i.} - \mu_i)^2, \tag{5.2}$$

where $\mu_i = Np_i + 0.5$. By Lemma 5.3, we have that under $H_{0,2}(\alpha)$,

$$\sqrt{r}\left(\frac{\mathrm{MS}\alpha_R - P_{R,\alpha}}{N^2}\right) \xrightarrow{p} 0 \quad \text{as} \quad r \longrightarrow \infty.$$

So $\sqrt{r}(\mathrm{MS}\alpha_R - \mathrm{MSE}_R^*)/N^2$ has same asymptotic distribution as $\sqrt{r}(P_{R,\alpha} - \mathrm{MSE}_R^*)/N^2$. Set

$$\frac{\mathrm{MSE}_R^*}{N^2} = \mathrm{MSE}_Z^*, \quad \frac{P_{R,\alpha}}{N^2} = P_{Z,\alpha}, \quad \frac{S_{R,i}^2}{N^2} = S_{Z,i}^2,$$

where MSE_R^*, and $S_{R,i}^2$ are defined in Eq. (2.2). MSE_Z^* and $S_{z,i}^2$ are defined similarly. $P_{Z,\alpha}$ is defined similarly as $P_{R,\alpha}$.

For simplicity of notation, denote $T_P(\mathbf{Z}) = \sqrt{r}(P_{Z,\alpha} - \mathrm{MSE}_Z^*)$, and use $T_P(\mathbf{Y})$ to denote the quantity similarly defined but with Z_{ij} replaced by Y_{ij}. Lemma 5.4 shows that $T_P(\mathbf{Z}) - T_P(\mathbf{Y}) \xrightarrow{p} 0$ as $r \to \infty$. The asymptotic distribution of $T_P(\mathbf{Y})$ can be found by applying results in Akritas and Papadatos (2004) or we can verify Lyapounov's condition easily here because Y_{ij} are uniformly bounded. Indeed,

$$T_P(\mathbf{Y}) = \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)[n_i(\bar{Y}_{i.}-p_i)^2 - S_{Y,i}^2]$$
$$= \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sum_{j_1\neq j_2}^{n_i}(\bar{Y}_{ij_1}-p_i)(\bar{Y}_{ij_2}-p_i) = \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}W_i^*,$$

where $W_i^* = (1-n_i/N)\sum_{j\neq j'} e_{ij}e_{ij'}$ and $e_{ij} = Y_{ij} - p_i$. It is easy to see that W_i^*s are independent and satisfy

$$E(W_i^*) = 0 \quad \text{and} \quad \mathrm{Var}(W_i^*) = (n_i-1)^{-2}\sum_{j\neq j'}\sum_{j_1\neq j_1'} E(e_{ij}e_{ij'}e_{ij_1}e_{ij_1'}) = \frac{n_i}{n_i-1}2\sigma_i^4.$$

Because Y_{ij} are nonnegative and uniformly bounded by 1, using the inequality

$$\left|\sum_{i=1}^{m} z_i\right|^p \leq m^{p-1}\sum_{i=1}^{m}|z_i|^p, \quad m \geq 1, \ \ p \geq 1,$$

which for $p > 1$ follows from Hölder's inequality, we have

$$E(W_i^*)^{2+\delta} = \left(1-\frac{n_i}{N}\right)^{2+\delta} E[(n_i(\bar{Y}_{i.}-\mu_i)^2 - S_{Y,i}^2)^{2+\delta}]$$
$$\leq 2^{1+\delta}\left(1-\frac{n_i}{N}\right)^{2+\delta}[n_i^{2+\delta}E(\bar{Y}_{i.}-\mu_i)^{4+2\delta} + E(S_{Y,i}^{4+2\delta})]$$
$$\leq 2^{1+\delta}\left(1-\frac{n_i}{N}\right)^{2+\delta}\left(n_i^{2+\delta} + \left(\frac{n_i}{n_i-1}\right)^{2+\delta}\right) \leq 2^{3+2\delta}n_i^{2+\delta}.$$

By assumption,

$$\frac{1}{\sqrt{\sum_{i=1}^{r}\mathrm{Var}(W_i^*)}^{2+\delta}}\sum_{i=1}^{r}E(W_i^*)^{2+\delta} \leq \frac{1}{\sqrt{\sum_{i=1}^{r}2\sigma_i^4}^{2+\delta}}\sum_{i=1}^{r}(4n_i)^{2+\delta} = O(r^{-\delta/2}) \to 0,$$

as r go to infinity. So the asymptotic distribution of $T_p(\mathbf{Y})$ follows.

LEMMA 5.2 *For the one-way heteroscedastic model*,

$$\frac{\mathrm{MSE}_R^*}{N^2} \xrightarrow{p} \sigma_0^2 = \lim_{r\to\infty}\frac{1}{r}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sigma_i^2, \quad \textit{as } r \longrightarrow \infty.$$

Proof The proof will follow if we show that

$$\frac{\mathrm{MSE}_R^*}{N^2} - \mathrm{MSE}_Y^* \xrightarrow{p} 0, \quad \text{as } r \longrightarrow \infty, \tag{5.3}$$

and

$$\mathrm{MSE}_Y^* \xrightarrow{p} \sigma_0^2, \quad \text{as } r \longrightarrow \infty. \tag{5.4}$$

We have

$$
\begin{aligned}
&\frac{\mathrm{MSE}_R^*}{N^2} - \mathrm{MSE}_Y^* \\
&= \frac{1}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j=1}^{n_i}[(Z_{ij}-\bar{Z}_{i.})^2-(Y_{ij}-\bar{Y}_{i.})^2] \\
&= \frac{1}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.})^2 \\
&\quad + \frac{2}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.})(Y_{ij}-\bar{Y}_{i.}).
\end{aligned}
$$

The first summation is $O_p(N^{-1}) = o_p(1)$ since $\sup_x(\widehat{H}(x) - H(x)) = O_p(N^{-1/2})$. The second summation is bounded by

$$
\frac{1}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j=1}^{n_i}|Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.}| = O_p(N^{-1/2}) = o_p(1).
$$

So Eq. (5.3) is shown. Next, note that $0 \le S_{Y,i}^2 \le n_i/(n_i-1)$, so $\mathrm{Var}(S_{Y,i}^2) \le E(S_{Y,i}^2) \le 4$. Therefore, when $r \to \infty$,

$$
E(\mathrm{MSE}_Y^*) = \frac{1}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sigma_i^2 \longrightarrow \sigma_0^2
$$

$$
\mathrm{Var}((\mathrm{MSE}_Y^*)) = \frac{1}{(r-1)^2}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)^2\mathrm{Var}(S_{Y,i}^2) \le \frac{4r}{(r-1)^2} \longrightarrow 0.
$$

Thus (5.4) is also true and this finishes the proof.

LEMMA 5.3 *Let* $\mathrm{MS}\alpha_R$ *be defined in Eq.* (2.1) *and* $P_{R,\alpha}$ *defined in Eq.* (5.2). *Then under* $H_{0,1}(\alpha)$ *or* $H_{0,2}(\alpha)$, $\sqrt{r}(\mathrm{MS}\alpha_R - P_{R,\alpha})/N^2 \xrightarrow{p} 0$ *as* $r \to \infty$.

Proof Under $H_{0,1}(\alpha)$ or $H_{0,2}(\alpha)$, $p_i = \bar{p}.$,

$$
\begin{aligned}
\frac{\mathrm{MS}\alpha_R}{N^2} &= \frac{1}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Z}_{i.}-\bar{Z}_{..})^2 = \frac{1}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{n_i}((\bar{Z}_{i.}-p_i)-(\bar{Z}_{..}-\bar{p}_{.}))^2 \\
&= \frac{1}{r-1}\sum_{i=1}^{r}n_i\left(1-\frac{n_i}{N}\right)(\bar{Z}_{i.}-p_i)^2 - \frac{1}{N(r-1)}\sum_{i\neq i'}^{r}n_i n_i'(\bar{Z}_{i.}-p_i)(\bar{Z}_{i'.}-p_{i'}).
\end{aligned}
$$

So

$$
\begin{aligned}
\sqrt{r}\left(\frac{\mathrm{MS}\alpha_R}{N^2} - P_{R,\alpha}\right) &= -\frac{\sqrt{r}}{N(r-1)}\sum_{i\neq i'}^{r}n_i n_i'(\bar{Z}_{i.}-p_i)(\bar{Z}_{i'.}-p_{i'}) \\
&= -\frac{\sqrt{r}}{N(r-1)}\sum_{i\neq i'}^{r}\sum_{j=1}^{c}\sum_{j'=1}^{c}(Z_{ij}-p_i)(Z_{i'j'}-p_{i'}).
\end{aligned}
$$

Denote above quantity as $T(\mathbf{Z})$ and we show that $T(\mathbf{Z}) - T(\mathbf{Y}) = o_p(1)$ as $r \to \infty$, where $T(\mathbf{Y})$ is similarly defined as $T(\mathbf{Z})$ with Z_{ij} replaced by Y_{ij}.

$$\begin{aligned} T(\mathbf{Z}) &= -\frac{\sqrt{r}}{N(r-1)} \sum_{i \neq i_1}^{r} \sum_{j=1}^{n_i} \sum_{j_1=1}^{n_i} (Z_{ij} - Y_{ij} + Y_{ij} - p_i)(Z_{i_1 j_1} - Y_{i_1 j_1} + Y_{i_1 j_1} - p_{i_1}) \\ &= D_1 + D_2 + T(\mathbf{Y}), \end{aligned}$$

where

$$D_1 = -\frac{\sqrt{r}}{N(r-1)} \sum_{i \neq i_1}^{r} \sum_{j=1}^{n_i} \sum_{j_1=1}^{n_{i_1}} (Z_{ij} - Y_{ij})(Z_{i_1 j_1} - Y_{i_1 j_1})$$

$$D_2 = -\frac{2\sqrt{r}}{N(r-1)} \sum_{i \neq i_1}^{r} \sum_{j=1}^{n_i} \sum_{j_1=1}^{n_{i_1}} (Z_{ij} - Y_{ij})(Y_{i_1 j_1} - p_{i_1}).$$

Since $\sup_x (\widehat{H}(x) - H(x)) = O_p(N^{-1/2})$, we have $D_1 = O_p(r^{-1/2}) = o_p(1)$. Next,

$$\begin{aligned} \frac{N^2(r-1)^2}{4r} E(D_2^2) &= \sum_{i \neq i_1}^{r} \sum_{j=1}^{n_i} \sum_{j_1=1}^{n_{i_1}} \sum_{i' \neq i_1'}^{r} \sum_{j'=1}^{n_{i'}} \sum_{j_1'=1}^{n_{i_1'}} \\ &\quad \times E[(Z_{ij} - Y_{ij})(Y_{i_1 j_1} - p_{i_1})(Z_{i'j'} - Y_{i'j'})(Y_{i_1' j_1'} - p_{i_1'})] \\ &= \sum_{i \neq i_1}^{r} \sum_{j=1}^{n_i} \sum_{j_1=1}^{n_i} \sum_{i' \neq i_1'}^{r} \sum_{j'=1}^{n_{i'}} \sum_{j_1'=1}^{n_{i_1'}} \frac{1}{N^2} \sum_{i_2=1}^{r} \sum_{j_2=1}^{n_{i_2}} \sum_{i_3=1}^{r} \sum_{j_3=1}^{n_{i_3}} \\ &\quad \times E[(c(X_{i_2 j_2}, X_{ij}) - F_{i_2 j_2}(X_{ij}))(c(X_{i_3 j_3}, X_{i'j'}) \\ &\quad - F_{i_3 j_3}(X_{i'j'}))(Y_{i_1 j_1} - p_{i_1})(Y_{i_1' j_1'} - p_{i_1'})]. \end{aligned}$$

The expectation under the summation is zero if the number of different elements in $\{i, i_1, i', i_1', i_2, i_3\}$ is five or six or the number of different elements in $\{j, j_1, j', j_1', j_2, j_3\}$ is five or six. Also note that $c(X_{ij}, X_{i'j'})$, Y_{ij} and $F_{ij}(X)$ are all uniformly bounded by 1, so $E(D_2^2) = O(r^{-1}) \to 0$. Hence $D_2 = o_p(1)$. Therefore it remains to show that $T(\mathbf{Y}) = o_p(1)$. This can be shown easily by verifying $E(T(\mathbf{Y})) = 0$ and $\mathrm{Var}(T(\mathbf{Y})) \to 0$ due to the fact that Y_{ij}s are independent uniformly bounded random variables.

LEMMA 5.4 *Assume $r^{-1} \sum_{i=1}^{r} n_{ij}^2$ converges and $T_P(\mathbf{Z})$ and $T_P(\mathbf{Y})$ are defined in the proof of Theorem* 2.3. *We have $T_P(\mathbf{Z}) - T_P(\mathbf{Y}) = o_p(1)$ as $r \to \infty$.*

Proof

$$\begin{aligned} T_P(\mathbf{Z}) &= \frac{\sqrt{r}}{r-1} \sum_{i=1}^{r} \left(1 - \frac{n_i}{N}\right) \sum_{j_1 \neq j_2}^{n_i} (Z_{ij_1} - p_i)(Z_{ij_2} - p_i) \\ &= \frac{\sqrt{r}}{r-1} \sum_{i=1}^{r} \left(1 - \frac{n_i}{N}\right) \sum_{j_1 \neq j_2}^{n_i} (Z_{ij_1} - Y_{ij_1} + Y_{ij_1} - p_i)(Z_{ij_2} - Y_{ij_2} + Y_{ij_2} - p_i) \\ &= D_{p1} + D_{p2} + T_P(\mathbf{Y}), \end{aligned}$$

where

$$D_{p1} = \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-Y_{ij_1})(Z_{ij_2}-Y_{ij_2}),$$

$$D_{p2} = \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-Y_{ij_1})(Y_{ij_2-p_i}).$$

It remains to show that $D_{p1}\xrightarrow{p}0$ and $D_{p2}\xrightarrow{p}0$ as $r\to\infty$. $D_{p1}=O_p(r^{-1/2}N^{-1}\sum_{i=1}^{r}n_i^2)=O_p\sqrt{r}(N^{-1})=o_p(1)$ due to the fact that $\sup_x(\hat{H}(x)-H(x))=O_p(N^{-1/2})$. Next,

$$D_{p2}=\frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\left(1-\frac{n_i}{N}\right)\sum_{j_1\neq j_2}^{n_i}(Y_{ij_2}-p_i)\frac{1}{N}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}}(c(X_{i_3,j_3},X_{ij_1})-F_{i_3}(X_{ij_1}))$$

and

$$\begin{aligned}E(D_{p2}^2)=\frac{r}{(r-1)^2}&\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{j_1\neq j_2}^{n_i}\sum_{j_1'\neq j_2'}^{n_{i'}}\frac{1}{N^2}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}}\sum_{i_4=1}^{r}\sum_{j_4=1}^{n_{i_4}}\left(1-\frac{n_i}{N}\right)\left(1-\frac{n_{i'}}{N}\right)\\&\times E[(Y_{ij_2}-p_i)(Y_{i'j_2'}-p_i')(c(X_{i_3,j_3},X_{ij_1})\\&-F_{i_3}(X_{ij_1}))(c(X_{i_4,j_4},X_{i'j_1'})-F_{i_4}(X_{i'j_1'}))].\end{aligned}$$

Note that $j_1\neq j_2$, $j_1'\neq j_2'$ and

$$E[c(X_{i_1,j_1},X_{ij})-F_{i_1}(X_{ij})]=E\{E[c(X_{i_1,j_1},X_{ij})-F_{i_1}(X_{ij})|X_{ij}]\}=0.$$

So by independence, the expectation under the summation is zero if the number of different elements in $\{i,i',i_3,i_4\}$ is three or four or the number of different elements in $\{j_1,j_2,j_1',j_2',j_3,j_4\}$ is five or six. Also note that $c(X_{ij},X_{i'j'})$, Y_{ij} and $F_i(X)$ are all uniformly bounded by 1, so $E(D_2^2)=O(r^{-1}N^{-2}(\sum_{i=1}^{r}n_i^2)^2)=O(r^{-1})\to 0$. Hence $D_{p2}=o_p(1)$.

Proof of Theorem 2.5 The proof uses three lemmas which are stated and proved after the proof of the theorem. By Lemma 5.5, $n(r)\mathrm{MSE}_R^{(2)}/N^2$ converges in probability to a constant. So we only need to consider the asymptotic distribution of $\sqrt{r}(\mathrm{MST}_R-\mathrm{MSE}_R^{(2)})/N^2$ when n_i remain fixed and $n(r)\sqrt{r}(\mathrm{MST}_R-\mathrm{MSE}_R^{(2)})/N^2$ when $n_i\to\infty$ with r. Under $H_{0,1}(\alpha)$ or $H_{0,2}(\alpha)$,

$$\frac{\mathrm{MST}_R}{N^2}=\frac{1}{r-1}\sum_{i=1}^{r}(\bar{Z}_{i.}-\tilde{Z}_{..})^2=\frac{1}{r-1}\sum_{i=1}^{r}((\bar{Z}_{i.}-p_i)-(\tilde{Z}_{..}-\bar{p}))^2=P(\mathbf{Z})+D_T(\mathbf{Z}),$$

where

$$P(\mathbf{Z})=\frac{1}{r}\sum_{i=1}^{r}(\bar{Z}_{i.}-p_i)^2, \tag{5.5}$$

$$D_T(\mathbf{Z})=-\frac{1}{r(r-1)}\sum_{i\neq i'}(\bar{Z}_{i.}-p_i)(\bar{Z}_{i'.}-p_i'). \tag{5.6}$$

By Lemma 5.6, it suffices to consider the asymptotic distribution of

$$T_2(\mathbf{Z}) = \sqrt{r}\left(P(\mathbf{Z}) - \frac{\mathrm{MSE}_R^{(2)}}{N^2}\right) = \frac{1}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-p_i)(Z_{ij_2}-p_i) \quad (5.7)$$

when n_i fixed and that of $n(r)T_2(\mathbf{Z})$ when $n_i \to \infty$ with r. Let $T_2(\mathbf{Y})$ be defined similarly as $T_2(\mathbf{Z})$ with Z_{ij} replaced by Y_{ij}. By Lemma 5.7, it remains to find the asymptotic distribution of $T_2(\mathbf{Y})$ and $n(r)T_2(\mathbf{Y})$ for fixed n_i and $n_i \to \infty$ with r, respectively. Let $e_{ij} = Y_{ij} - p_i$ and $W_i = 1/n_i(n_i-1)\sum_{j_1\neq j_2}^{n_i} e_{ij_1}e_{ij_2}$, then W_is are independent with zero mean and $T_2(\mathbf{Y}) = r^{-1/2}\sum_{i=1}^{r} W_i$. We will verify Lyapounov's condition. First,

$$\begin{aligned}\mathrm{Var}(W_i) &= \frac{1}{n_i^2(n_i-1)^2}\sum_{j\neq j'}^{n_i}\sum_{j_1\neq j_1'}^{n_i} E(e_{ij}e_{ij'}e_{ij_1}e_{ij_1'})\\ &= \frac{2}{n_i^2(n_i-1)^2}\sum_{j\neq j'} E(e_{ij}^2)E(e_{ij'}^2)\\ &= \frac{2}{n_i(n_i-1)}\sigma_i^4,\end{aligned}$$

so that

$$E(n(r)T_2(\mathbf{Y})) = 0 \quad \text{and} \quad \mathrm{Var}(n(r)T_2(\mathbf{Y})) = \frac{2}{r}\sum_{i=1}^{r}\frac{n^2(r)}{n_i(n_i-1)}\sigma_i^4$$

Note that from the original definition of $T_2(\mathbf{Y})$ we can also write W_i as $(\bar{Y}_i - p_i)^2 - n_i^{-1}S_{Y,i}^2$. So by the fact that Y_{ij} are nonnegative and uniformly bounded by 1,

$$\begin{aligned}E(W_i)^{2+\delta} &= E[(\bar{Y}_{i.} - p_i)^2 - n_i^{-1}S_{Y,i}^2]^{2+\delta}\\ &\le 2^{1+\delta}[E(\bar{Y}_{i.} - p_i)^{4+2\delta} + n_i^{-2-\delta}E(S_{Y,i}^{4+2\delta})]\\ &\le 2^{1+\delta}\left[1 + \left(\frac{4}{n_i}\right)^{2+\delta}\right] \le 2^{2+\delta}.\end{aligned}$$

By assumption,

$$\begin{aligned}&\frac{1}{\sqrt{\sum_{i=1}^{r}\mathrm{Var}[n(r)W_i]}^{2+\delta}}\sum_{i=1}^{r}E[n(r)W_i]^{2+\delta}\\ &\quad\le \frac{1}{\sqrt{\sum_{i=1}^{r}2n^2(r)\sigma_i^4/n_i(n_i-1)}^{2+\delta}}\sum_{i=1}^{r}(2n(r))^{2+\delta}\\ &\quad= O(n(r)^{2+\delta}r^{-\delta/2}) \longrightarrow 0,\end{aligned}$$

as r go to infinity. So the asymptotic distribution of $n(r)T_2(\mathbf{Y})$ follows.

If n_i fixed, treat $n(r)$ as a bounded constant, we get the asymptotic distribution of $T_2(\mathbf{Y})$.

LEMMA 5.5 *Assume $r^{-1}\sum_{i=1}^{r} n(r)/n_i\sigma_i^2 \to v_3^2$, as $r \to \infty$, and let $\mathrm{MSE}_R^{(2)}$ be defined in Theorem* 2.5. *Then $\mathrm{MSE}_R^{(2)}/N^2$ converges in probability to v_2^2, if the n_i are fixed, and $n(r)\mathrm{MSE}_R^{(2)}/N^2$ converges in probability to v_3^2, if the n_i tend to infinity.*

Proof Note that $\mathrm{MSE}_R^{(2)}/N^2 = \mathrm{MSE}_Z^{(2)}$. First we will show that, in both cases,

$$n(r)(\mathrm{MSE}_Z^{(2)} - \mathrm{MSE}_Y^{(2)}) \xrightarrow{p} 0, \quad \text{as } r \longrightarrow \infty. \tag{5.8}$$

We have

$$\begin{aligned} n(r)\left(\frac{\mathrm{MSE}_R^{(2)}}{N^2} - \mathrm{MSE}_Y^{(2)}\right) &= \frac{n(r)}{r}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)} \\ &\quad \times \sum_{j=1}^{n_i}[(Z_{ij}-\bar{Z}_{i.})^2 - (Y_{ij}-\bar{Y}_{i.})^2] \\ &= \frac{n(r)}{r}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.})^2 \\ &\quad + \frac{2n(r)}{r}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.})(Y_{ij}-\bar{Y}_{i.}). \end{aligned}$$

The first summation is $O_p(N^{-1}) = o_p(1)$ since $\sup_x(\hat{H}(x) - H(x)) = O_p(N^{-1/2})$. The second summation is bounded by

$$\frac{2n(r)}{r}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j=1}^{n_i}|Z_{ij}-\bar{Z}_{i.}-Y_{ij}+\bar{Y}_{i.}| = O_p(N^{-1/2}) = o_p(1).$$

So Eq. (5.8) holds, in both cases. Next we will show

$$\mathrm{MSE}_Y^{(2)} \xrightarrow{p} \nu_2^2, \ \text{as } r \longrightarrow \infty, \quad \text{and} \quad n(r)\mathrm{MSE}_Y^{(2)} \xrightarrow{p} \nu_3^2, \ \text{if } n_i \longrightarrow \infty \ \text{as } r \longrightarrow \infty. \tag{5.9}$$

Note that $0 \le S_{Y,i}^2 \le n_i/(n_i-1)$, so $\mathrm{Var}(S_{Y,i}^2) \le E(S_{Y,i}^2)^2 \le 4$. Therefore, when $r \to \infty$,

$$E(n(r)\mathrm{MSE}_Y^{(2)}) = \frac{1}{r}\sum_{i=1}^{r}\frac{n(r)}{n_i}\sigma_i^2 \longrightarrow \nu_3^2$$

$$\mathrm{Var}(\mathrm{MSE}_Y^{(2)}) = \frac{n^2(r)}{r^2}\sum_{i=1}^{r}\frac{1}{n_i^2}\mathrm{Var}(S_{Y,i}^2) \le \frac{4}{r} \longrightarrow 0.$$

The case that n_i are fixed is similar. So (5.9) is true. Combine (5.8) and (5.9), we finish the proof of the lemma.

LEMMA 5.6 *$D_T(\mathbf{Z})$ is defined in Eq.* (5.6).

1. $\sqrt{r}D_T(\mathbf{Z}) \xrightarrow{p} 0$ *as $r \to \infty$ while n_i's remain fixed*;
2. *When $n_i = n_i(r) \to \infty$, as $r \to \infty$, assume $n^2(r) = o(r)$, then $n(r)\sqrt{r}D_T(\mathbf{Z}) \xrightarrow{p} 0$.*

Proof When $n_i = n_i(r) \to \infty$ as $r \to \infty$, we will show

$$n(r)\sqrt{r}[D_T(\mathbf{Z}) - D_T(\mathbf{Y})] \xrightarrow{p} 0 \quad \text{and} \quad n(r)\sqrt{r}D_T(\mathbf{Y}) \xrightarrow{p} 0, \tag{5.10}$$

where $D_T(\mathbf{Y})$ is similarly defined as $D_T(\mathbf{Z})$ with Z_{ij} replaced by Y_{ij}.

$$\begin{aligned} n(r)\sqrt{r}D_T(\mathbf{Z}) &= -\frac{n(r)\sqrt{r}}{r(r-1)}\sum_{i\neq i_1}^{r}\frac{1}{n_i n_{i_1}}\sum_{j=1}^{n_i}\sum_{j_1=1}^{n_{i_1}} \\ &\quad\times (Z_{ij}-Y_{ij}+Y_{ij}-p_i)(Z_{i_1j_1}-Y_{i_1j_1}+Y_{i_1j_1}-p_{i_1}) \\ &= D_{T1}+D_{T2}+n(r)\sqrt{r}D_T(\mathbf{Y}), \end{aligned}$$

where

$$D_{T1} = -\frac{n(r)}{\sqrt{r}(r-1)}\sum_{i\neq i_1}^{r}\frac{1}{n_i n_{i_1}}\sum_{j=1}^{n_i}\sum_{j_1=1}^{n_{i_1}}(Z_{ij}-Y_{ij})(Z_{i_1j_1}-Y_{i_1j_1}),$$

$$D_{T2} = -\frac{2n(r)}{\sqrt{r}(r-1)}\sum_{i\neq i_1}^{r}\frac{1}{n_i n_{i_1}}\sum_{j=1}^{n_i}\sum_{j_1=1}^{n_{i_1}}(Z_{ij}-Y_{ij})(Y_{i_1j_1}-p_{i_1}).$$

Since $\sup_x(\widehat{H}(x)-H(x)) = O_p(N^{-1/2})$, we have $D_{T1} = O_p(n(r)\sqrt{r}N^{-1}) = o_p(1)$.

$$\begin{aligned} E(D_{T2}^2) &= \frac{4n^2(r)}{r(r-1)^2}\sum_{i\neq i_1}^{r}\frac{1}{n_i n_{i_1}}\sum_{j=1}^{n_i}\sum_{j_1=1}^{n_{i_1}}\sum_{i'\neq i'_1}^{r}\frac{1}{n_{i'}n_{i'_1}}\sum_{j'=1}^{n_{i'}}\sum_{j'_1=1}^{n_{i'_1}} \\ &\quad\times E[(Z_{ij}-Y_{ij})(Y_{i_1j_1}-p_{i_1})(Z_{i'j'}-Y_{i'j'})(Y_{i'_1j'_1}-p_{i'_1})] \\ &= \frac{4n^2(r)}{r(r-1)^2}\sum_{i\neq i_1}^{r}\frac{1}{n_i n_{i_1}}\frac{1}{n_{i'}n_{i'_1}}\sum_{j=1}^{n_i}\sum_{j_1=1}^{n_i}\sum_{i'\neq i'_1}^{r}\sum_{j'=1}^{n_{i'}}\sum_{j'_1=1}^{n_{i'_1}}\frac{1}{N^2}\sum_{i_2=1}^{r}\sum_{j_2=1}^{n_{i_2}}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}} \\ &\quad\times E[(c(X_{i_2j_2},X_{ij})-F_{i_2j_2}(X_{ij}))(c(X_{i_3j_3},X_{i'j'})-F_{i_3j_3}(X_{i'j'})) \\ &\quad\times (Y_{i_1j_1}-p_{i_1})(Y_{i'_1j'_1}-p_{i'_1})]. \end{aligned}$$

The expectation under the summation is zero if the number of different elements in $\{i, i_1, i', i'_1, i_2, i_3\}$ is five or six or the number of different elements in $\{j, j_1, j', j'_1, j_2, j_3\}$ is five or six. Also note that $c(X_{ij}, X_{i'j'})$, Y_{ij} and $F_{ij}(X)$ are all uniformly bounded by 1, so $E(D_{T2}^2) = O(n^2(r)rN^{-2}) \to 0$. Hence $D_{T2} = o_p(1)$. Therefore $n(r)\sqrt{r}[D_T(\mathbf{Z}) - D_T(\mathbf{Y})] \xrightarrow{p} 0$. It remains to show that $n(r)\sqrt{r}D_T(\mathbf{Y}) = o_p(1)$. This can be shown easily because $E(D_T(\mathbf{Y})) = 0$ and

$$\begin{aligned} \mathrm{Var}(n(r)\sqrt{r}D_T(\mathbf{Y})) &= \frac{n^2(r)}{r(r-1)^2}\sum_{i\neq i'}^{r}\sum_{i_1\neq i'_1}^{r}E[(\bar{Z}_{i.}-p_i)(\bar{Z}_{i'.}-p_{i'})(\bar{Z}_{i_1.}-p_{i_1})(\bar{Z}_{i'_1.}-p_{i'_1})] \\ &= \frac{n^2(r)}{r(r-1)^2}\sum_{i\neq i'}^{r}E(\bar{Z}_{i.}-p_i)^2E(\bar{Z}_{i'.}-p_{i'})^2 \\ &= O\left(\frac{n^2(r)}{r}\right) \longrightarrow 0, \end{aligned}$$

since Z_{ij}s are independent uniformly bounded random variables.

Therefore Eq. (5.10) holds and $n(r)\sqrt{r}D_T(\mathbf{Z})\xrightarrow{p}0$ as $r \to \infty$.

When n_is are fixed, treat $n(r)$ as a bounded constant, then we have $\sqrt{r}D_T(\mathbf{Z})\xrightarrow{p}0$ as $r \to \infty$.

LEMMA 5.7 *$T_2(\mathbf{Z})$ and $T_2(\mathbf{Y})$ are defined in Eq. (5.7).*

1. *$T_2(\mathbf{Z}) - T_2(\mathbf{Y})\xrightarrow{p}0$ as $r \to \infty$ while n_i's remain fixed*;
2. *$n(r)(T_2(\mathbf{Z}) - T_2(\mathbf{Y}))\xrightarrow{p}0$ when $n_i = n_i(r) \to \infty$, as $r \to \infty$.*

Proof When $n_i = n_i(r) \to \infty$ as $r \to \infty$,

$$\begin{aligned} n(r)T_2(\mathbf{Z}) &= \frac{n(r)}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-Y_{ij_1}+Y_{ij_1}-p_i)(Z_{ij_2}-Y_{ij_2}+Y_{ij_2}-p_i) \\ &= D_{21}+D_{22}+n(r)T_2(\mathbf{Y}), \end{aligned}$$

where

$$D_{21} = \frac{n(r)}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-Y_{ij_1})(Z_{ij_2}-Y_{ij_2}),$$

$$D_{22} = \frac{n(r)}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j_1\neq j_2}^{n_i}(Z_{ij_1}-Y_{ij_1})(Z_{ij_2}-p_i).$$

It remains to show that $D_{21}\xrightarrow{p}0$ and $D_{22}\xrightarrow{p}0$ as $r \to \infty$. $D_{21} = O_p(n(r)\sqrt{r}N^{-1}) = o_p(1)$ due to the fact that $\sup_x(\widehat{H}(x) - H(x)) = O_p(N^{-1/2})$.

$$D_{22} = \frac{n(r)}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n_i(n_i-1)}\sum_{j_i\neq j_2}^{n_i}(Y_{ij_2}-p_i)\frac{1}{N}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}}(c(X_{i_3,j_3},X_{ij_1})-F_{i_3}(X_{ij_1}))$$

and

$$\begin{aligned} E(D_{22}^2) = \frac{n^2(r)}{r}&\sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{j_1\neq j_2}^{n_i}\sum_{j_1'\neq j_2'}^{n_{i'}}\frac{1}{N^2}\sum_{i_3=1}^{r}\sum_{j_3=1}^{n_{i_3}}\sum_{i_4=1}^{r}\sum_{j_4=1}^{n_{i_4}}\frac{1}{n_i(n_i-1)}\frac{1}{n_{i'}(n_{i'}-1)} \\ &\times E[(Y_{ij_2}-p_i)(Y_{i'j_2'}-p_{i'})(c(X_{i_3,j_3},X_{ij_1}) \\ &- F_{i_3}(X_{ij_1}))(c(X_{i_4,j_4},X_{i'j_1'})-F_{i_4}(X_{i'j_1'}))]. \end{aligned}$$

Note that $j_1 \neq j_2$, $j_1' \neq j_2'$ and

$$E[c(X_{i_1,j_1},X_{ij})-F_{i_1}(X_{ij})] = E\{E[c(X_{i_1,j_1},X_{ij})-F_{i_1}(X_{ij})|X_{ij}]\} = 0.$$

So by independence, the expectation under the summation is zero if the number of different elements in $\{i, i', i_3, i_4\}$ is three or four or the number of different elements in $\{j_1, j_2, j_1', j_2', j_3, j_4\}$ is five or six. Also note that $c(X_{ij}, X_{i'j'})$, Y_{ij} and $F_i(X)$ are all uniformly bounded by 1, so $E(D_{22}^2) = O(n^2(r)rN^{-2}) \to 0$. Hence $D_{22} = o_p(1)$. Therefore $n(r)(T_2(\mathbf{Z}) - T_2(\mathbf{Y})) \xrightarrow{p} 0$ when $n_i = n_i(r) \to \infty$, as $r \to \infty$.

When n_is are fixed, treat $n(r)$ as bounded constant, we get $T_2(\mathbf{Z}) - T_2(\mathbf{Y}) \xrightarrow{p} 0$.

5.2 Proofs for the Two-way Design

Proof of Theorem 3.1 By Lemma 5.8, $\mathrm{MSE}_R^{(3)}/N^2 \xrightarrow{p} \nu_4^2$ as $r \to \infty$ if n_{ij} *fixed*; $n(r)\mathrm{MSE}_R^{(3)}/N^2 \xrightarrow{p} \nu_5^2$ as $r \to \infty$ if $n(r) \to \infty$ as $r \to \infty$. So we only need to consider $\sqrt{r}(\mathrm{MST}_\alpha - \mathrm{MSE}_R^{(3)})/N^2$. Note that $R_{ijk} = 1/2 + NZ_{ijk}$, where $Z_{ijk} = \widehat{H}(X_{ijk})$. So

$$\sqrt{r}\frac{(\mathrm{MST}_\alpha - \mathrm{MSE}_R^{(3)})}{N^2} = \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{c}(\tilde{Z}_{i..} - \tilde{Z}_{...})^2 - \frac{\sqrt{r}}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{S_{ij,Z}^2}{n_{ij}}$$
$$= \frac{\sqrt{r}}{r-1}\sum_{i=1}^{r}\sum_{j=1}^{c}\left[(\tilde{Z}_{i..} - \tilde{Z}_{...})^2 - \frac{1}{c}\left(1-\frac{1}{r}\right)\frac{S_{ij,Z}^2}{n_{ij}}\right].$$

Let

$$T_\alpha(\mathbf{Z}) = (rc)^{-1/2}\sum_{i=1}^{r}\sum_{j=1}^{c}\left[(\tilde{Z}_{i..} - \tilde{Z}_{...})^2 - \frac{1}{c}\left(1-\frac{1}{r}\right)\frac{S_{ij,Z}^2}{n_{ij}}\right]. \tag{5.11}$$

Then

$$\sqrt{r}\frac{(\mathrm{MST}_\alpha - \mathrm{MSE}_R^{(3)})}{N^2} = \frac{r\sqrt{c}}{r-1}T_\alpha(\mathbf{Z}).$$

Let $p_{ij} = E(Y_{ijk})$. The projection, under $H_0(\alpha)$, of $T_\alpha(\mathbf{Z})$ onto the class of random variables of the form $\Sigma_{i=1}^r g_i(\mathbf{Z}_i)$, where $\mathbf{Z}_i = (Z_{i11}, \ldots, Z_{i1n_{i1}}, \ldots, Z_{icn_{ic}})$ and g_i are measurable with $Eg_i^2(\mathbf{Z}_i) < \infty$, is given by

$$\tilde{T}_\alpha(\mathbf{Z}) = (rc)^{-1/2}\sum_{i=1}^{r}\frac{r-1}{rc}\left[\left(\sum_{j=1}^{c}(\bar{Z}_{ij.} - p_{ij})\right)^2 - \sum_{j=1}^{c}\frac{S_{ij,Z}^2}{n_{ij}}\right]. \tag{5.12}$$

By Lemmas 5.9 and 5.10, $T_\alpha(\mathbf{Z})$ has the same asymptotic distribution as $\tilde{T}_\alpha(\mathbf{Y})$ when n_{ij} fixed, and $n(r)T_\alpha(\mathbf{Z})$ has the same asymptotic distribution as $n(r)\tilde{T}_\alpha(\mathbf{Z})$ when $n_{ij}(r) \to \infty$ as $r \to \infty$. But the asymptotic distributions of $\tilde{T}_\alpha(\mathbf{Y})$ when n_{ij} fixed and $n(r)\tilde{T}_\alpha(\mathbf{Z})$ when $n_{ij}(r) \to \infty$ as $r \to \infty$ are established in Wang and Akritas (2002). Combining these results we establish the convergences under $H_0(\alpha)$ stated in parts (a) and (b) of the theorem.

Similarly, for the test of interaction effect, let

$$T_\gamma(\mathbf{Z}) = (rc)^{-1/2}\sum_{i=1}^{r}\sum_{j=1}^{c}\left[(\bar{Z}_{ij.} - \tilde{Z}_{i..} - \tilde{Z}_{.j.} + \tilde{Z}_{...})^2 - \frac{(r-1)(c-1)}{rc}\frac{S_{ij,Z}^2}{n_{ij}}\right], \tag{5.13}$$

so

$$\sqrt{r}\frac{\mathrm{MST}_\gamma - \mathrm{MSE}_R^{(3)}}{N^2} = \frac{r\sqrt{c}}{(r-1)(c-1)}T_\gamma(\mathbf{Z}).$$

The projection, under $H_0(\gamma)$, of $T_\gamma(\mathbf{Z})$ onto the same class of random variables defined above is given by

$$\tilde{T}_\gamma(\mathbf{Z}) = \frac{(r-1)(c-1)}{(rc)^{3/2}}\sum_{i=1}^{r}\sum_{j=1}^{c}\left((\bar{Z}_{ij.} - p_{ij})^2 - \frac{S_{ij,Z}^2}{n_{ij}}\right)$$
$$- \frac{r-1}{(rc)^{3/2}}\sum_{i=1}^{r}\sum_{j_1 \neq j_2}^{c}(\bar{Z}_{ij_1} - p_{ij_1})(\bar{Z}_{ij_2} - p_{ij_2}). \tag{5.14}$$

By Lemmas 5.9 and 5.10, $T_\gamma(\mathbf{Z})$ has the same asymptotic distribution as $\tilde{T}_\gamma(\mathbf{Y})$ when n_{ij} fixed, and $n(r)T_\gamma(\mathbf{Z})$ has the same asymptotic distribution as $n(r)\tilde{T}_\gamma(\mathbf{Z})$ when $n_{ij}(r) \to \infty$ as $r \to \infty$. But the asymptotic distributions of $\tilde{T}_\gamma(\mathbf{Z})$ when n_{ij} fixed and $n(r)\tilde{T}_\gamma(\mathbf{Z})$ when $n_{ij}(r) \to \infty$ as $r \to \infty$ are established in Wang and Akritas (2002). Combining these results we establish the convergences under $H_0(\gamma)$ stated in parts (a) and (b) of the theorem. This completes the proof.

LEMMA 5.8 *$\mathrm{MSE}_R^{(3)}/N^2 \xrightarrow{p} \nu_4^2$ as $r \to \infty$ if n_{ij} fixed; $n(r)\mathrm{MSE}_R^{(3)}/N^2 \xrightarrow{p} 0\nu_5^2$ as $r \to \infty$ if $n(r) \to \infty$ as $r \to \infty$, where ν_4^2 and ν_5^2 are defined in Theorem* 3.1.

Proof: First we will show that, in both cases, $\mathrm{MSE}_R^{(3)}/N^2 - \mathrm{MSE}_Y^{(3)} = O_p(N^{-1/2}n(r)^{-1})$, where $\mathrm{MSE}_Y^{(3)}$ is defined similarly as $\mathrm{MSE}_R^{(3)}$ with R_{ijk} replaced by Y_{ijk}. We have

$$\begin{aligned}\frac{\mathrm{MSE}_R^{(3)}}{N^2} - \mathrm{MSE}_Y^{(3)} &= \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{1}{n_{ij}(n_{ij}-1)}\sum_{k=1}^{n_{ij}}[(Z_{ijk}-\bar{Z}_{ij.})^2 - (Y_{ijk}-\bar{Y}_{ij.})^2]\\ &= \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{1}{n_{ij}(n_{ij}-1)}\sum_{k=1}^{n_{ij}}(Z_{ijk}-\bar{Z}_{ij.}-Y_{ijk}+\bar{Y}_{ij.})^2\\ &\quad+\frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{1}{n_{ij}(n_{ij}-1)}\sum_{k=1}^{n_{ij}}(Z_{ijk}-\bar{Z}_{ij.}-Y_{ijk}+\bar{Y}_{ij.})(Y_{ijk}-\bar{Y}_{ij.}).\end{aligned}$$

The first summation is $O_p(N^{-1}n(r)^{-1})$ since $\sup_x|\hat{H}(x) - H(x)| = O_p(N^{-1/2})$. The second summation is bounded by

$$\frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{1}{n_{ij}(n_{ij}-1)}\sum_{k=1}^{n_{ij}}|(Z_{ijk}-\bar{Z}_{ij.})^2 - (Y_{ijk}+\bar{Y}_{ij.})| = O_p(N^{-1/2}n(r)^{-1}).$$

So

$$\frac{\mathrm{MSE}_R^{(3)}}{N^2} - \mathrm{MSE}_Y^{(3)} = O_p(N^{-1/2}n(r)^{-1}), \tag{5.15}$$

as $r \to \infty$ whether n_{ij} is fixed or not. We will be done if we show that

$$\mathrm{MSE}_Y^{(3)} \xrightarrow{p} \nu_4^2 \text{ if } n_{ij} \text{ remain fixed;} \quad n(r)\mathrm{MSE}_Y^{(3)} \xrightarrow{p} \nu_5^2 \text{ if } n(r) \longrightarrow \infty \text{ with } r, \tag{5.16}$$

as $r \to \infty$.

Denote $V_i = c^{-1}\sum_{j=1}^{c} n_{ij}^{-1}(n_{ij}-1)^{-1}\sum_{k=1}^{n_{ij}}(Y_{ijk}-\bar{Y}_{ij.})^2$, then V_is are independent random variables uniformly bounded by 1 and $\mathrm{MSE}_Y^{(3)} = r^{-1}\sum_{i=1}^{r} V_i$.

If n_{ij} remain fixed,

$$E(\mathrm{MSE}_Y^{(3)}) = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\sigma_{ij}^2}{n_{ij}} \longrightarrow \nu_4^2,$$

$$\mathrm{Var}(\mathrm{MSE}_Y^{(3)}) = \frac{1}{r^2}\sum_{i=1}^{r}\mathrm{Var}(V_i) \le \frac{1}{r^2}\sum_{i=1}^{r}E(V_i^2) \le \frac{1}{r} \longrightarrow 0.$$

If $n(r) \to \infty$ as $\alpha \to \infty$,

$$E(n(r)\mathrm{MSE}_Y^{(3)}) = \frac{n(r)}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\sigma_{ij}^2}{n_{ij}} \longrightarrow \nu_5^2,$$

$$\mathrm{Var}(n(r)\mathrm{MSE}_Y^{(3)}) = \frac{n(r)}{r^2}\sum_{i=1}^{r}\mathrm{Var}(V_i) \le \frac{n(r)}{r^2}\sum_{i=1}^{r}E(V_i^2) \le \frac{n(r)}{r} \longrightarrow 0.$$

Thus, Eq. (5.16) holds and together with Eq. (5.15) the proof of the lemma is done.

Lemma 5.9 *Let* $T_\alpha(\mathbf{Z})$, $\tilde{T}_\gamma(\mathbf{Z})$, $T_\gamma(\mathbf{Z})$ *and* $\tilde{T}_\gamma(\mathbf{Z})$ *be defined as in the proof of Theorem* 3.1. *Suppose* $1/rc\sum_{i=1}^{r}\sum_{j=1}^{c}\sigma_{ij}^2 \to \sigma^2 > 0$, *then*

(1) *If* n_{ij} *is fixed,*

$$T_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Z}) \xrightarrow{p} 0 \textit{ under } H_0(\alpha); \quad T_\gamma(\mathbf{Z}) - \tilde{T}_\gamma(\mathbf{Z}) \xrightarrow{p} 0 \textit{ under } H_0(\gamma).$$

(2) *If* $n_{ij}(r) \to \infty$, *as* $r \to \infty$, *let* $n(r) = \min\{n_{ij}(r); i = 1, \ldots, r, j = 1, \ldots, c\}$, *assume* $n(r)/r \to \infty$, *as* $r \to \infty$, *then*

$$n(r)(T_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Z})) \xrightarrow{p} 0 \textit{ under } H_0(\alpha); n(r)(T_\gamma(\mathbf{Z}) - \tilde{T}_\gamma(\mathbf{Z})) \xrightarrow{p} 0 \textit{ under } H_0(\gamma).$$

Proof It is not hard to prove that

$$\begin{aligned}
T_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Z}) &= -(rc)^{-3/2}\sum_{i\neq i'}^{r}\left(\sum_{j=1}^{c}(\bar{Z}_{ij.} - \mu_{ij})\right)\left(\sum_{j'=1}^{c}(\bar{Z}_{i'j'.} - \mu_{i'j'})\right)\\
&= -(rc)^{-3/2}\sum_{i\neq i'}^{r}\sum_{j=1}^{c}\sum_{j'=1}^{c}\\
&\quad \times (\bar{Z}_{ij.} - \bar{Y}_{ij.} + \bar{Y}_{ij.} - \mu_{ij})(\bar{Z}_{i'j'.} - \bar{Y}_{i'j'.} + \bar{Y}_{i'j'.} - \mu_{i'j'})\\
&= D_{\alpha 1} + D_{\alpha 2} + D_{\alpha 3},
\end{aligned}$$

where

$$D_{\alpha 1} = -(rc)^{-3/2}\sum_{i\neq i'}^{r}\sum_{j=1}^{c}\sum_{j'=1}^{c}(\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Z}_{i'j'.} - \bar{Y}_{i'j'.}),$$

$$D_{\alpha 2} = -2(rc)^{-3/2}\sum_{i\neq i'}^{r}\sum_{j=1}^{c}\sum_{j'=1}^{c}(\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Y}_{i'j'.} - \mu_{i'j'.}),$$

$$D_{\alpha 3} = -(rc)^{-3/2}\sum_{i\neq i'}^{r}\sum_{j=1}^{c}\sum_{j'=1}^{c}(\bar{Y}_{ij.} - \mu_{ij.})(\bar{Y}_{i'j'.} - \mu_{i'j'.}).$$

$$\begin{aligned}
T_\gamma(\mathbf{Z}) - \tilde{T}_\gamma(\mathbf{Z}) &= -\frac{c-1}{(rc)^{3/2}}\sum_{j=1}^{c}\sum_{i\neq i'}^{r}(\bar{Z}_{ij.} - \mu_{ij})(\bar{Z}_{i'j.} - \mu_{i'j}) + (rc)^{-3/2}\\
&\quad \times \sum_{i\neq i'}^{r}\sum_{j\neq j'}^{c}(\bar{Z}_{ij.} - \mu_{ij})(\bar{Z}_{i'j'.} - \mu_{i'j'})
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{c-1}{(rc)^{3/2}} \sum_{j=1}^{c} \sum_{i \neq i'}^{r} \\
&\quad \times (\bar{Z}_{ij.} - \bar{Y}_{ij.} + \bar{Y}_{ij.} - \mu_{ij})(\bar{Z}_{i'j.} - \bar{Y}_{i'j.} + \bar{Y}_{i'j.} - \mu_{i'j}) \\
&\quad + (rc)^{-3/2} \sum_{i \neq i'}^{r} \sum_{j \neq j'}^{c} (\bar{Z}_{ij.} - \bar{Y}_{ij.} + \bar{Y}_{ij.} - \mu_{ij})(\bar{Z}_{i'j'.} - \bar{Y}_{i'j'.} + \bar{Y}_{i'j'.} - \mu_{i'j'}) \\
&= D_{\gamma 1} + D_{\gamma 2} + D_{\gamma 3},
\end{aligned}
$$

where

$$
\begin{aligned}
D_{\gamma 1} &= -\frac{c-1}{(rc)^{3/2}} \sum_{j=1}^{c} \sum_{i \neq i'}^{r} (\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Z}_{i'j.} - \bar{Y}_{i'j.}) \\
&\quad + (rc)^{-3/2} \sum_{i \neq i'}^{r} \sum_{j \neq j'}^{c} (\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Z}_{i'j'.} - \bar{Y}_{i'j'.}), \\
D_{\gamma 2} &= -2\frac{c-1}{(rc)^{3/2}} \sum_{j=1}^{c} \sum_{i \neq i'}^{r} (\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Y}_{i'j.} - \mu_{i'j}) \\
&\quad + 2(rc)^{-3/2} \sum_{i \neq i'}^{r} \sum_{j \neq j'}^{c} (\bar{Z}_{ij.} - \bar{Y}_{ij.})(\bar{Z}_{i'j'.} - \mu_{i'j'}), \\
D_{\gamma 3} &= -\frac{c-1}{(rc)^{3/2}} \sum_{j=1}^{c} \sum_{i \neq i'}^{r} (\bar{Y}_{ij.} - \mu_{ij})(\bar{Y}_{i'j.} - \mu_{i'j}) \\
&\quad + (rc)^{-3/2} \sum_{i \neq i'}^{r} \sum_{j \neq j'}^{c} (\bar{Y}_{ij.} - \mu_{ij})(\bar{Y}_{i'j'.} - \mu_{i'j'}).
\end{aligned}
$$

Because $\sup_x(\widehat{H}(x) - H(x)) = O_p(N^{-1/2})$, we have

$$
D_{\alpha 1} = O_p\left(\frac{\sqrt{rc}}{N}\right); \quad D_{\gamma 1} = O_p\left(\frac{\sqrt{rc}}{N}\right).
$$

So if n_{ij} is fixed,

$$
D_{\alpha 1} = o_p(1); \quad D_{\gamma 1} = o_p(1) \text{ as } r \longrightarrow \infty.
$$

If $n_{ij}(r) \to \infty$ as $r \to \infty$,

$$
n(r)D_{\alpha 1} = o_p(1) \quad \text{and} \quad n(r)D_{\alpha 1} = o_p(1).
$$

Note that $(Y_{ijk} - \mu_{ij})$s are independent random variables with zero mean, apply Proposition 3.2 of Wang and Akritas (2002), $n(r)D_{\alpha 3} = o_p(1)$ and $n(r)D_{\gamma 3} = o_p(1)$ if $n_{ij}(r) \to \infty$ as $r \to \infty$; $D_{\alpha 3} = o_p(1)$ and $D_{\gamma 3} = o_p(1)$ if n_{ij} is fixed.

It remains to show $n(r)D_{\alpha 2} = o_p(1)$ and $n(r)D_{\gamma 2} = o_p(1)$. Since $|D_{\gamma 2}| \le (c-1)|D_{\alpha 2}|$, we only need to show

$$n(r)D_{\alpha 2} = o_p(1) \text{ if } n(r) \longrightarrow \infty \text{ as } r \longrightarrow \infty; \quad \text{and} \quad D_{\alpha 2} = o_p(1) \text{ if } n_{ij} \text{ are fixed.} \tag{5.17}$$

$$\begin{aligned} D_{\alpha 2} &= -2(rc)^{-3/2} \sum_{i \ne i'}^{r} \sum_{j=1}^{c} \sum_{j'=1}^{c} \sum_{k=1}^{n_{ij}} \sum_{k'=1}^{n_{i'j'}} \frac{1}{n_{ij} n_{i'j'}} (Z_{ijk} - Y_{ijk})(Y_{i'j'k'} - \mu_{i'j'}) \\ &= -2(rc)^{-3/2} \sum_{i \ne i'}^{r} \sum_{j=1}^{c} \sum_{j'=1}^{c} \sum_{k=1}^{n_{ij}} \sum_{k'=1}^{n_{i'j'}} \frac{1}{n_{ij} n_{i'j'} N} \sum_{i_1=1}^{r} \sum_{j_1=1}^{c} \sum_{k_1=1}^{n_{i_1 j_1}} \\ &\quad \times [c(X_{i_1 j_1 k_1}, X_{ijk}) - F_{i_1 j_1}(X_{ijk})](Y_{i'j'k'} - \mu_{i'j'}), \end{aligned}$$

where $c(x, y) = [I(x \le y) + I(x < y)]/2$. Hence,

$$\begin{aligned} E(D_{\alpha 2}^2) &= \frac{4}{(rc)^3 N^2} \sum_{i \ne i'}^{r} \sum_{j=1}^{c} \sum_{j'=1}^{c} \sum_{k=1}^{n_{ij}} \sum_{k'=1}^{n_{i'j'}} \sum_{i_1=1}^{r} \sum_{j_1=1}^{c} \sum_{k_1=1}^{n_{i_1 j_1}} \frac{1}{n_{ij} n_{i'j'}} \\ &\quad \times \sum_{i_2 \ne i_3}^{r} \sum_{j_2=1}^{c} \sum_{j_3=1}^{c} \sum_{k_2=1}^{n_{i_2 j_2}} \sum_{k_3=1}^{n_{i_3 j_3}} \sum_{i_4=1}^{r} \sum_{j_4=1}^{c} \sum_{k_4=1}^{n_{i_4 j_4}} \frac{1}{n_{i_2 j_2} n_{i_3 j_3}} \\ &\quad \times E\{[c(X_{i_1 j_1 k_1}, X_{ijk}) - F_{i_1 j_1}(X_{ijk})](Y_{i'j'k'} - p_{i'j'}) \\ &\quad \times [c(X_{i_4 j_4 k_4}, X_{i_2 j_2 k_2}) - F_{i_4 j_4}(X_{i_2 j_2 k_2})](Y_{i_3 j_3 k_3} - p_{i_3 j_3})\}, \end{aligned}$$

When five or six elements in the set $\{i, i', i_2, i_3, i_1, i_4\}$ are different, the expectation under the summation is zero. Therefore $E(D_{\alpha 2}^2) = O(r^{-1})$ and so $E(n(r)D_{\alpha 2}^2) = O(n(r)/r)$ if $n(r) \to \infty$ as $r \to \infty$.

By given condition, $n(r)/r \to \infty$ as $r \to \infty$. So Eq. (5.17) is proved.

LEMMA 5.10 *Let $\tilde{T}_\alpha(\mathbf{Z})$ and $\tilde{T}_\gamma(\mathbf{Z})$ be defined as in the proof of Theorem* 3.1. *$\tilde{T}_\alpha(\mathbf{Y})$ and $\tilde{T}_\gamma(\mathbf{Y})$ are defined by replacing Z_{ijk} with Y_{ijk} in the corresponding functions.*

(1) *If n_{ij} is fixed,*

$$\tilde{T}_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Y}) \xrightarrow{p} 0; \quad \tilde{T}_\gamma(\mathbf{Z}) - \tilde{T}_\gamma(\mathbf{Y}) \xrightarrow{p} 0.$$

(2) *If $n_{ij}(r) \to \infty$, as $r \to \infty$, let $n(r) = \min\{n_{ij}(r); i = 1, \ldots, r, j = 1, \ldots, c\}$,*

$$n(r)(\tilde{T}_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Y})) \xrightarrow{p} 0; \quad n(a)(\tilde{T}_\gamma(\mathbf{Z}) - \tilde{T}_\gamma(\mathbf{Y})) \xrightarrow{p} 0.$$

Proof It can be shown that

$$\begin{aligned} \tilde{T}_\alpha(\mathbf{Z}) &= \frac{r-1}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k_1 \ne k_2}^{n_{ij}} \frac{(Z_{ijk_1} - \mu_{ij})(Z_{ijk_2} - \mu_{ij})}{n_{ij}(n_{ij} - 1)} \\ &\quad + \frac{r-1}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j_1 \ne j_2}^{c} \sum_{k_1=1}^{n_{ij_1}} \sum_{k_2=1}^{n_{ij_2}} \frac{(Z_{ij_1 k_1} - \mu_{ij_1})(Z_{ij_2 k_2} - \mu_{ij_2})}{n_{ij_1} n_{ij_2}}, \end{aligned}$$

and $\tilde{T}_\alpha(\mathbf{Z}) - \tilde{T}_\alpha(\mathbf{Y}) = \tilde{D}_{\alpha 1} + \tilde{D}_{\alpha 2}$, where

$$\tilde{D}_{\alpha 1} = \frac{r-1}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k_1 \neq k_2}^{n_{ij}} \frac{(Z_{ijk_1} - Y_{ijk_1})(Z_{ijk_2} - Y_{ijk_2})}{n_{ij}(n_{ij}-1)}$$

$$+ \frac{r-1}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j_1 \neq j_2}^{c} \sum_{k_1=1}^{n_{ij_1}} \sum_{k_2=1}^{n_{ij_2}} \frac{(Z_{ij_1k_1} - Y_{ij_1k_1})(Z_{ij_2k_2} - Y_{ij_2k_2})}{n_{ij_1}(n_{ij_2})},$$

$$\tilde{D}_{\alpha 2} = \frac{2(r-1)}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k_1 \neq k_2}^{n_{ij}} \frac{(Z_{ijk_1} - Y_{ijk_1})(Y_{ijk_2} - \mu_{ij})}{n_{ij}(n_{ij}-1)}$$

$$+ \frac{2(r-1)}{(rc)^{3/2}} \sum_{i=1}^{r} \sum_{j_1 \neq j_2}^{c} \sum_{k_1=1}^{n_{ij_1}} \sum_{k_2=1}^{n_{ij_2}} \frac{(Z_{ij_1k_1} - Y_{ij_1k_1})(Y_{ij_2k_2} - \mu_{ij_2})}{n_{ij_1} n_{ij_2}}.$$

Follow the same procedure as for $D_{\alpha 1}$ and $D_{\alpha 2}$ in the proof of Lemma 5.9, we can show that

$$n(r)\tilde{D}_{\alpha_1} \xrightarrow{p} 0, \quad \text{if } n(r) \longrightarrow \infty \text{ with } r \quad \text{and} \quad \tilde{D}_{\alpha_1} \xrightarrow{p} 0 \text{ if } n_{ij} \text{ fixed;}$$

$$n(r)\tilde{D}_{\alpha_2} \xrightarrow{p} 0, \quad \text{if } n(r) \longrightarrow \infty \text{ with } r \quad \text{and} \quad \tilde{D}_{\alpha_2} \xrightarrow{p} 0 \text{ if } n_{ij} \text{ fixed.}$$

Then we finished the proof of the lemma.

References

Akritas, M. G. (1990). The rank transform method in some two factor designs. *J. Amer. Statist. Assoc.*, **85**, 73–78.

Akritas, M. G. and Arnold, S. (2000). Asymptotics for ANOVA when the number of levels is large. *J. Amer. Statist. Assoc.*, **95**, 212–226.

Akritas, M. G. and Papadatos, N. (2004). Heteroscedastic one-way ANOVA and lack-of-fit test. *J. Amer. Statist. Assoc.*, in press.

Arnold, S. (1980). Asymptotic validity of F tests for the ordinary linear model and the multiple correlation model. *J. Am. Statist. Assoc.*, **75**, 890–894.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

Bathke, A. (2002). ANOVA for a large number of treatments. *Mathematical Methods of Statistics*, **11**(1), 118–132.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Boos, D. D and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statist. Probab. Lett.*, **23**, 183–191.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effect of inequality of variance and correlation of errors in the two-way classification. *Ann. Math. Statist.*, **25**, 484–498.

Box, G. E. P. and Anderson, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J. R. Statist. Soc.*, **B 17**, 1–26.

Brownie, C. and Boos, D. D. (1994). Type I error robustness of ANOVA and ANOVA on ranks when the number of treatments is large. *Biometrics*, **50**, 542–549.

Brunner, E., Domhof, S. and Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* Wiley, NY.

Cliff, N. (1993). Dominance statistics: ordinal analyses to answer ordinal questions. *Psychological Bulletin*, **114**, 494–509.

Krutchkoff, R. G. (1989). Two-way fixed effects analysis of variance when the error variances may be unequal. *J. Stat. Comp. Simul.*, **32**, 177–183.

Miller, R. G., Jr. (1986). *Beyond ANOVA: Basics of Applied Statistics.* John Wiley, New York. (Reprinted Chapman & Hall, New York.)

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Annals of Statistics*, **12**, 1298–1309.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley, New York.

Wang, L. and Akritas, M. G. (2002). Two-way heteroscedastic ANOVA when the number of levels is large (Submitted).