



## NONPARAMETRIC MODELLING VIA TWO-WAY MIXED EFFECTS DESIGN

Sevgi DEMİRCİOĞLU and Bilgehan GÜVEN

Mathematics Department, Çanakkale Onsekiz Mart University, Çanakkale, Turkey.

**ABSTRACT.** The classical  $F$ -test for testing the hypothesis of no fixed main effects in a mixed effects design is valid under the assumption of normality, symmetry and variance homogeneity of the error terms assumption. We consider the two-way mixed effects design which does not require these three assumptions. A test procedure for the hypothesis of no main fixed effects is developed under this flexible model. The asymptotic distribution of the test statistic is studied for a large number of levels of the random effects.

### 1. INTRODUCTION

The assumption of independence of the random main and interaction effects is called the *symmetry assumption*. Violation of the symmetry assumption yields that the sum of squares for the main fixed effects is not distributed as a constant times the chi-squared variable. Thereby, the  $F$ -test for testing no-main fixed effects is not valid. See ([9], ch. 8).

Departure from the symmetry assumption has not been received enough attention. Studies on this topic are [2], [3], [4], [5], [6]. Studies [2]- [4] are on the unbalanced two-way mixed model while [5] and [6] studied the balanced three way mixed model with two random factors. They proposed approximate tests for the effects. Also, simulation studies given by [3] and [5] showed that the achieved sizes of the  $F$ -test are liberal when the symmetry assumption is violated.

In the unbalanced design, an exact test for the main fixed effects of a two-way mixed effects design is derived by [8]. But derivation is based on assuming that the random main and interaction effects are independent. See ([8], p.153-155).

Consider the case in which the assumption of symmetry and variance homogeneity of the error terms are violated but the assumption of normality is preserved.

---

2020 *Mathematics Subject Classification.* Primary 62G10, 62G20.

*Keywords and phrases.* Two-way mixed effects design, symmetry assumption, hypothesis testing, asymptotic normality.

✉ sevgidemircioglu@stu.comu.edu.tr-Corresponding author; bguven@comu.edu.tr

ORCID 0000-0003-3900-0713; 0000-0001-8661-1408.

For this case, Hotelling's  $T^2$  test for testing of the hypothesis of no main fixed effects is suggested when two-way mixed effects design is balanced. Also [2] and [3] extended the balanced two-way mixed effects design into the unbalanced one. Their result is that cell sizes are random as it is indicated in Section 2.

The contribution of this paper is to present a new test for testing the hypothesis of no main fixed effects for the unbalanced two-way mixed model under departure of three main assumptions: normality, variance homogeneity of the error terms and symmetry. We propose the test statistic different from studies mentioned above. To do this:

- 1) The test statistic is established by extending Hotelling's  $T^2$  test to the unbalanced two way mixed design where cell sizes are random.
- 2) We assume that the distribution of the random factors and error terms have any distribution with zero-mean. Using the asymptotic results, the asymptotic null distribution of the test statistic is established.

The rest of the article is organized as follows: Section 2 describes the nonparametric two-way mixed effects design which is mainly the two-way crossed mixed effects design without the classical assumptions. Section 3 derives the null asymptotic distribution of the Hotelling's test statistic with a large number of levels of the random effects where the design does not have fixed cell size. Section 4 presents the simulation results. Section 5 gives an illustrative example.

## 2. THE NONPARAMETRIC TWO-WAY MIXED EFFECTS DESIGN

We now describe the two-way mixed design violating normality, variance homogeneity of the error terms and symmetry. Consider a design in which the  $a$  rows accord with the fixed effects and the  $b$  columns accord with the random effects. The  $b$  levels of the random effect are considered to be obtained by simple random sampling from a population  $\mathcal{V}$ . Let  $V$  be a randomly selected element from a population  $\mathcal{V}$ . Then  $Y_{iV}$  is an observation from the cell  $(i, V)$  where  $i$  denotes the level of the fixed effect and the level of the random effects is labeled by random variable  $V$ .

We write  $Y_{iV}$  as

$$Y_{iV} = E[Y_{iV}|V] + (Y_{iV} - E[Y_{iV}|V]) = m_i(V) + e_i(V) \quad (1)$$

where  $m_i(V)$  and  $e_i(V)$  are called the random mean and error term respectively. We only assume that the distribution of  $Y_{iV}$  depends on  $(i, V)$ . Then

$$\text{Var}(m_{i_1}(V)) \neq \text{Var}(m_{i_2}(V)) \quad \text{and} \quad \text{Var}(e_{i_1}(V)) \neq \text{Var}(e_{i_2}(V)) \quad \text{for} \quad i_1 \neq i_2.$$

Further, the expression in (1) of  $Y_{iV}$  allows that

$$E[e_i(V)] = E[e_i(V)|V] = 0 \quad \text{and} \quad E[m_i(V)e_i(V)] = E[m_i(V)e_i(V)|V] = 0.$$

It follows that  $m_i(V)$  and  $e_i(V)$  are uncorrelated.

The  $m_i(V)$  is uniquely decomposed into main and interaction effects as

$$m_i(V) = \mu + \alpha_i + \beta(V) + \gamma_i(V)$$

where  $\mu$ ,  $\alpha_i$ ,  $\beta(V)$  and  $\gamma_i(V)$  are defined as

$$\mu = \frac{1}{a} \sum_{i=1}^a E[m_i(V)], \quad \alpha_i = E[m_i(V)] - \mu, \quad \beta(V) = \frac{1}{a} \sum_{i=1}^a m_i(V) - \mu,$$

$$\gamma_i(V) = m_i(V) - \mu - \alpha_i - \beta(V). \quad (2)$$

The results of Eq. (2) are:

$$\sum_{i=1}^a \alpha_i = 0, \quad E[\beta(V)] = 0, \quad \sum_{i=1}^a \gamma_i(V) = 0 \quad \text{and} \quad E[\gamma_i(V)] = 0.$$

It is evident that

1) The random effects  $\{\beta(V), \gamma_1(V), \dots, \gamma_a(V)\}$  are not independent. Thus the vector of random means  $(m_1(V), m_2(V), \dots, m_a(V))'$  has multivariate distribution.

2) The replicated observation on the cell  $(i, V)$  is denoted by  $Y_{iV_k}$  where  $k = 1, 2, \dots, n_{iV}$ . Then the cell sizes  $n_{iV}$  is random since it is a function of  $V$ .

Let  $\{V_1, V_2, \dots, V_b\}$  be simple random sampling representing  $b$  levels of the random effects drawn independently from a population  $\mathcal{V}$ . Let  $n_i(V_j)$  be the number of iid observations drawn from the cell  $(i, V_j)$ . Then iid observations and the error terms are denoted by  $Y_{iV_jk}$  and  $e_{ik}(V_j)$  where  $k = 1, 2, \dots, n_i(V_j)$

We assume that  $e_{ik}(V_j)$ ,  $k = 1, 2, \dots, n_i(V_j)$  are iid for each  $i$  and  $e_{i_1k}(V_j)$  and  $e_{i_2k}(V_j)$  are uncorrelated for each  $k$  and  $i_1 \neq i_2$ .

We write  $j$  for  $V_j$ ,  $Y_{ijk}$  for  $Y_{iV_jk}$ ,  $m_{ij}$  for  $m_i(V_j)$ ,  $\beta_j$  for  $\beta(V_j)$ ,  $\gamma_{ij}$  for  $\gamma_i(V_j)$ ,  $e_{ijk}$  for  $e_{ik}(V_j)$  and  $n_{ij}$  for  $n_i(V_j)$ . Then from (1), we have

$$E[Y_{ijk}|V_j] = m_{ij} \quad \text{and} \quad Y_{ijk} - E[Y_{ijk}|V_j] = e_{ijk}.$$

The two-way mixed model is written as

$$Y_{ijk} = m_{ij} + e_{ijk}, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, n_{ij} \quad (3)$$

where the random mean  $m_{ij}$  is defined as

$$m_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}. \quad (4)$$

with  $\sum_{i=1}^a \alpha_i = 0$  and  $\sum_{i=1}^a \gamma_{ij} = 0$  for each  $j$ . The vector random means  $(m_{1j}, m_{2j}, \dots, m_{aj})'$  for  $j = 1, 2, \dots, b$  are iid with the mean vector

$$\boldsymbol{\nu} = (\mu + \alpha_1, \mu + \alpha_2, \dots, \mu + \alpha_a)' \quad (5)$$

and the covariance matrix

$$\boldsymbol{\Sigma}_m = \{Cov(m_{ij}, m_{i'j'})\} \quad (6)$$

The vector of the error terms  $(e_{1jk}, e_{2jk}, \dots, e_{ajk})'$  are iid with zero-mean vector and the covariance matrix  $\boldsymbol{\Sigma}_e = \text{diag}(\sigma_{e_1}^2, \sigma_{e_2}^2, \dots, \sigma_{e_a}^2)$ . The elements of the vector of the error terms are uncorrelated to the elements of the random means. It is assumed that  $n_{ij}$ ,  $j = 1, 2, \dots, b$  are iid for each  $i$ . Further  $n_{ij}$  is independent of  $Y_{ijk}$ .

3. TESTING THE HYPOTHESIS OF NO MAIN FIXED EFFECTS

Consider two-way mixed design has no an empty cell such that  $P(n_{ij} \geq 1) = 1$ . Then the average of an observed cell and row are

$$\bar{Y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}, \quad \bar{Y}_{i..} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij.}$$

respectively. We have  $E[\bar{Y}_{ij.}|V_j] = m_{ij}$  and

$$Cov(E[\bar{Y}_{ij.}|V_j], E[\bar{Y}_{i'j.}|V_j]) = Cov(m_{ij}, m_{i'j}).$$

From the model (3), one can easily obtain that

$$E[\bar{Y}_{ij.}\bar{Y}_{i'j.}|V_j] = m_{ij}m_{i'j} + \delta_{ii'}\sigma_{e_i}^2/n_{ij}$$

where  $\delta_{ii'}$  denotes the kronecker delta. It follows that  $Cov(\bar{Y}_{ij.}, \bar{Y}_{i'j.}|V_j) = \delta_{ii'}\sigma_{e_i}^2/n_{ij}$ . Then

$$E[Cov(\bar{Y}_{ij.}, \bar{Y}_{i'j.}|V_j)] = \delta_{ii'}\sigma_{e_i}^2 E[n_{ij}^{-1}].$$

The unconditional covariance between  $\bar{Y}_{ij.}$  and  $\bar{Y}_{i'j.}$  is

$$\begin{aligned} Cov(\bar{Y}_{ij.}, \bar{Y}_{i'j.}) &= Cov(E[\bar{Y}_{ij.}|V_j], E[\bar{Y}_{i'j.}|V_j]) + E[Cov(\bar{Y}_{ij.}, \bar{Y}_{i'j.}|V_j)] \\ &= Cov(m_{ij}, m_{i'j}) + \delta_{ii'}\sigma_{e_i}^2 E[n_{ij}^{-1}]. \end{aligned}$$

Vectors of the observed cell averages  $(\bar{Y}_{1j.}, \bar{Y}_{2j.}, \dots, \bar{Y}_{aj.})'$ , for  $j = 1, 2, \dots, b$ , are i.i.d. as the multivariate normal distribution with mean vector  $\boldsymbol{\nu}$  and covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_m + \text{diag}(\sigma_{e_1}^2 E[n_{1j}^{-1}], \sigma_{e_2}^2 E[n_{2j}^{-1}], \dots, \sigma_{e_a}^2 E[n_{aj}^{-1}]), \tag{7}$$

where  $\boldsymbol{\nu}$  and  $\boldsymbol{\Sigma}_m$  are given in Eqs. (5) and (6) respectively.

The test statistic is based on the vector of the observed row averages

$$\bar{\mathbf{Y}} = (\bar{Y}_{1..}, \bar{Y}_{2..}, \dots, \bar{Y}_{a..})'. \tag{8}$$

The proposed test statistic for testing  $H_0: \text{all } \alpha_i = 0$  is

$$TS_A = b(\mathbf{C}\bar{\mathbf{Y}})'(\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')^{-1}\mathbf{C}\bar{\mathbf{Y}} \tag{9}$$

where  $\mathbf{C} = \{c_{ij}\}$  is an  $(a-1) \times a$  contrast matrix, i.e., its rows are orthogonal to the vector of ones and the sample covariance matrix

$$\mathbf{S} = \{S_{ii'}\} \quad \text{with} \quad S_{ii'} = \frac{\sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..})(\bar{Y}_{i'j.} - \bar{Y}_{i'..})}{b-1} \tag{10}$$

for  $i, i' = 1, 2, \dots, a$ .

**Remark 3.1** In the case that the two-way mixed effects design keeps normality of the random effects and of the error terms but violates both the homoscedasticity and symmetry assumption, the test statistic  $TS_A$  in Eq. (9) is the Hotelling's  $T^2$  test statistic distributed as  $(b-1)(a-1)/(b-a+1)F_{a-1, b-a+1}$ .

**Theorem 3.1** The asymptotic null distribution of  $TS_A$  given in Eq. (9) is the chi-squared distribution with  $a - 1$  degrees of freedom as  $b \rightarrow \infty$ .

**Proof** From central limit theorem for multivariate observations, we have

$$\sqrt{b}(\bar{\mathbf{Y}} - \boldsymbol{\nu})$$

is the multivariate normal distribution with zero-mean vector and the covariance matrix given in Eq. (7) as  $b \rightarrow \infty$  where  $\bar{\mathbf{Y}}$  is given in Eq. (8) see [1]. Let  $\mathbf{C}$  be an  $(a - 1) \times a$  contrast matrix. Then  $\mathbf{C}\boldsymbol{\nu} = \mathbf{0}$  when  $H_0 : \text{all } \alpha_i = 0$  holds. It follows that  $\sqrt{b}\mathbf{C}\bar{\mathbf{Y}} \xrightarrow{d} N_{a-1}(\mathbf{0}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$  as  $b \rightarrow \infty$ .

Observe that  $E[S_{ii'}] = \text{Cov}(m_{ij}, m_{i'j}) + \delta_{ii'}\sigma_{e_i}^2(1/b) \sum_{j=1}^b n_{ij}^{-1}$  where  $S_{ii'}$  is the  $ii'$ th entry of the sample covariance matrix  $\mathbf{S}$  given in Eq. (10). From Lemma given in Appendix, it follows that  $E[S_{ii'}] \rightarrow \text{Cov}(m_{ij}, m_{i'j}) + \delta_{ii'}\sigma_{e_i}^2 E[n_{ij}^{-1}]$  as  $b \rightarrow \infty$ . Then  $\mathbf{S}$  is a consistent estimator of  $\boldsymbol{\Sigma}$ .

Thus,  $\boldsymbol{\Sigma}$  can be replaced by  $\mathbf{S}$  where  $\mathbf{C}\mathbf{S}\mathbf{C}'$  is a positive definite matrix. Then,  $\sqrt{b}(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1/2}\mathbf{C}\bar{\mathbf{Y}} \xrightarrow{d} N_{a-1}(\mathbf{0}, \mathbf{I})$  as  $b \rightarrow \infty$ . As a result, the asymptotic null distribution of the test statistic  $TS_A$  is the chi-squared distribution with  $a - 1$  degrees of freedom as  $b \rightarrow \infty$ .

#### 4. SIMULATION STUDY

In this section, we investigate the performance and power of the proposed test by simulation study. The distribution of the cell sizes  $n_{ij}$  is considered to be a discrete uniform distribution for the unbalanced design.  $Y_{ijk}$ 's defined by Eq. (3) are generated where the number of  $Y_{ijk}$  is specified by achieved  $n_{ij}$  for a given  $(i, j)$ . Data violating the assumption of symmetry is generated by using the data generation method of Gaugler and Akritas [4]. It is as follows:

1. Set  $\mu$  and  $\alpha_i, i = 1, 2, \dots, a$  constants subject to  $\sum_{i=1}^a \alpha_i = 0$ .
2. Generate  $V_j, j = 1, 2, \dots, b$  from the standard normal distribution
3. Generate  $\lambda_i, i = 1, 2, \dots, a$  from the standard exponential distribution where  $\lambda_i$ 's are independent of  $V_j$ 's and always held constants within simulations.
4. Then  $Y_{ijk}$  is from a normal distribution with mean  $\mu + \alpha_i + V_j + V_j * (\lambda_i - \bar{\lambda})$  and variance  $\sigma_{e_i}^2$ . where  $\bar{\lambda}$  denotes the average of  $\lambda_i$ 's.

The number of levels of the random effects "b" is large. Therefore, we take  $a = b = 20$  where the number of levels of the fixed effects "a" can be taken any number greater than 2. The asymptotic null distribution of  $TS_A$  in (9) does not depend on  $\mu$ . So we take  $\mu = 0$ . For all cases and  $\sigma_{e_i}^2 = 1 \forall i$  in the homoscedastic case and  $\sigma_{e_i}^2 = 2, 25$  for  $i = 1, 2, 3, 4, 5$  and  $\sigma_{e_i}^2 = 4$  for  $i = 6, 7, 8, 9, 10$  in the heteroscedastic case. Two different values of  $\sigma_{e_i}^2$ , that we set, are enough to reflect the heteroscedastic case The distribution of the cell sizes is assumed to be identical and from a discrete uniform distribution.

The way of generating data preserves normality but violates symmetry and variance homogeneity of the error term. This fact ensures that the  $1 - \alpha$  quantile point

TABLE 1. Achieved test sizes (in  $H_0$ -row) and power (in remaining rows) for  $T^2$  and  $TS_A$  where  $P(n_{ij} = x) = 0.2$ ,  $x = 2, 3, 4, 5, 6$ ,  $i = 1, 2, \dots, 20$  and for each  $j$  if the design is unbalanced and  $P(n_{ij} = 4) = 1.0$  if the design is balanced.

$\alpha = 0.05$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.064	0.077	0.064	0.074
$H_1$	0.116	0.132	0.260	0.286
$H_2$	0.343	0.370	0.417	0.494
$H_3$	0.516	0.543	0.624	0.633
Hetero.				
$H_0$	0.064	0.075	0.079	0.096
$H_1$	0.077	0.094	0.123	0.136
$H_2$	0.127	0.152	0.292	0.315
$H_3$	0.175	0.195	0.415	0.454
$\alpha = 0.10$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.098	0.117	0.135	0.151
$H_1$	0.220	0.240	0.339	0.365
$H_2$	0.492	0.515	0.402	0.419
$H_3$	0.675	0.694	0.655	0.658
Hetero.				
$H_0$	0.117	0.133	0.133	0.153
$H_1$	0.137	0.155	0.222	0.239
$H_2$	0.223	0.238	0.367	0.383
$H_3$	0.268	0.292	0.564	0.591

can be  $(b - 1)(a - 1)/(b - a + 1)F_{a-1, b-a+1}(\alpha)$  if the test statistics has the exact distribution or  $\chi_{a-1}^2(\alpha)$  if the test statistic has the asymptotic distribution. Here,  $F_{a_1, a_2}(\alpha)$  and  $\chi_a^2(\alpha)$  are the upper quantile point of the  $F$ -distribution with respect to degrees of freedom  $a_1$  and  $a_2$  and the chi-squared distribution with degrees of freedom  $a$  respectively.

We use  $\alpha_i = 0$  for all  $i = 1, 2, \dots, 20$  for the null hypothesis. We formed three alternative hypotheses denoted by  $H_1$ ,  $H_2$  and  $H_3$ . They are:

$$\begin{aligned}
 H_1 : \alpha_1 = \dots = \alpha_{10} = 0.3, \quad \alpha_{11} = \dots, \alpha_{20} = -0.3, \\
 H_2 : \alpha_1 = \dots = \alpha_{10} = 0.5, \quad \alpha_{11} = \dots, \alpha_{20} = -0.5,
 \end{aligned}$$

$$H_3 : \alpha_1 = \dots = \alpha_{10} = 0.7, \quad \alpha_{11} = \dots, \alpha_{20} = -0.7.$$

TABLE 2. Achieved test sizes (in  $H_0$ -row) and power (in remaining rows) for  $T^2$  and  $TS_A$  where  $P(n_{ij} = x) = 0.2$ ,  $x = 7, 8, 9, 10, 11$   $i = 1, 2, \dots, 20$  and for each  $j$  if the design is unbalanced and  $P(n_{ij} = 9) = 1.0$  if the design is balanced.

$\alpha = 0.05$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.062	0.076	0.065	0.075
$H_1$	0.281	0.310	0.327	0.356
$H_2$	0.434	0.452	0.417	0.494
$H_3$	0.566	0.571	0.822	0.835
Hetero.				
$H_0$	0.058	0.073	0.079	0.095
$H_1$	0.127	0.148	0.160	0.177
$H_2$	0.325	0.356	0.484	0.510
$H_3$	0.509	0.545	0.711	0.740
$\alpha = 0.10$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.106	0.117	0.137	0.151
$H_1$	0.434	0.464	0.419	0.449
$H_2$	0.687	0.702	0.774	0.777
$H_3$	0.782	0.784	0.797	0.797
Hetero.				
$H_0$	0.109	0.127	0.132	0.154
$H_1$	0.203	0.220	0.263	0.282
$H_2$	0.461	0.488	0.545	0.574
$H_3$	0.644	0.667	0.798	0.813

Simulations are based on 1000 runs for all cases. Achieving any test size is accomplished as follows: the number of the generated test statistic from null data and exceeding the upper  $1 - \alpha$  quantile of the null distribution is divided by 1000. To obtain achieved power value of any test, the number of the generated test statistic from alternative data and exceeding the upper  $1 - \alpha$  quantile of the null distribution is divided by 1000. It should be noted that any significant difference from the presenting results is not detected when simulations are based on 10000 runs.

TABLE 3. Achieved test sizes (in  $H_0$ -row) and power (in remaining rows) for  $T^2$  and  $TS_A$  where  $P(n_{ij} = x) = 0.1$ ,  $x = 2, 3, \dots, 11$   $i = 1, 2, \dots, 20$  and for each  $j$  if the design is unbalanced and  $P(n_{ij} = 7) = 1.0$  if the design is balanced.

$\alpha = 0.05$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.063	0.075	0.065	0.075
$H_1$	0.129	0.147	0.309	0.339
$H_2$	0.441	0.462	0.506	0.526
$H_3$	0.641	0.674	0.673	0.681
Hetero.				
$H_0$	0.060	0.075	0.075	0.096
$H_1$	0.077	0.094	0.136	0.144
$H_2$	0.162	0.173	0.381	0.344
$H_3$	0.224	0.245	0.517	0.546
$\alpha = 0.10$				
	Unbalanced		Balanced	
	$T^2$	$TS_A$	$T^2$	$TS_A$
Homo.				
$H_0$	0.105	0.114	0.135	0.151
$H_1$	0.260	0.279	0.402	0.420
$H_2$	0.589	0.608	0.584	0.600
$H_3$	0.775	0.795	0.682	0.684
Hetero.				
$H_0$	0.111	0.125	0.133	0.153
$H_1$	0.146	0.161	0.239	0.256
$H_2$	0.264	0.278	0.430	0.451
$H_3$	0.317	0.339	0.659	0.681

Table 1-Table 3 address the achieved test sizes and the power values of  $T^2$  and  $TS_A$ . In all settings, the proposed test  $TS_A$  displays stronger power than Hotelling's  $T^2$  test under each of the three alternative hypotheses. Further, our simulation results verified that Hotelling's  $T^2$  test suffers from low power. It should be noted that the proposed test is slightly more liberal than Hotelling's  $T^2$  test. Additionally Hotelling's  $T^2$  test sizes achieved their nominal sizes more accurately in the unbalanced cases. Both tests have stronger power in the homoscedastic case than the heteroscedastic case.

No significant difference was observed between achieved test sizes of Hotelling's  $T^2$  and  $TS_A$  in the case of either homoscedastic or heteroscedastic. Therefore



departure from the assumption of variance homogeneity of the error terms does not effect to the proposed test.

### 5. EXAMPLE

The company selected five different transmission cases and operators to measure the diameter of hole (mm) in the rear of the transmission cases. The experiment consisted of randomly selected six operators from the workforce. Each selected operator was presented the transmission case in a completely random order. Data for this experiment is given in Table 4. ([7] p.175).

TABLE 4. Measurements of the diameter hole

Transmission Case	Operator					
	1	2	3	4	5	6
1	186	176	184	197	179	176
	188	177	185	196	176	174
			184	196	176	
2	188	181	194	195	182	173
	189	184	192	191	184	173
	195	182	188		186	
3	167	169	166	182	167	158
	166	167	166	182	166	163
		167				
4	188	181	190	188	182	176
	189	180	188	189	182	176
	192	182		190	186	174
5	169	173	184	185	174	164
	172	172	182	187	177	166
	173	175		188	176	166

We are testing the hypothesis of no difference among the transmission cases. The data summary is:

$$\bar{\mathbf{Y}} = (182.69, 185.72, 167.19, 184.55, 175.88)'$$

and

$$\mathbf{S} = \begin{bmatrix} 67.42 & 51.68 & 49.89 & 37.97 & 47.48 \\ 51.68 & 47.04 & 36.47 & 42.82 & 47.84 \\ 49.89 & 36.47 & 52.12 & 23.57 & 44.87 \\ 37.97 & 42.82 & 23.57 & 32.98 & 32.59 \\ 47.48 & 47.84 & 44.87 & 32.59 & 6.13 \end{bmatrix}.$$

The value of the test statistic  $TS_A$  in (9) is calculated where it is equal to 70.72. We have  $\chi_{4,0.005}^2 = 14.86$ . The hypothesis can be rejected.

### 6. CONCLUSIONS

We consider the unbalanced two-way mixed effects model when three main assumptions, normality, symmetry and variance homogeneity of the error term, are violated. Non symmetry causes that the sum of squares for the main fixed effects are not distributed as a constant times the chi-squared variable. As a result the  $F$ -test is not valid. Therefore, in this article, we focuss on testing the hypothesis of no main fixed effects. This article is a part of Ph. D. dissertation of the first author.

We showed that both the mean of observations and the cell sizes are random when the levels of the random effects are from a simple random sampling. The proposed test statistic  $TS_A$  is the same as the Hotelling's  $T^2$  test statistic without assuming normality. In simulation study, the proposed test  $TS_A$  displays stronger power than the Hotelling's  $T^2$  test in all cases. However, the proposed test does not allow empty cells such that  $n_{ij} \geq 1 \forall (ij)$  with probability 1. It is considered that both a test for the hypothesis of no interaction effects and empty cells are considered to be further research.

#### Appendix

**Lemma** For each  $i$ ,  $n_{ij}$ ,  $j = 1, 2, \dots, b$  are iid and take values from a finite set of distinct positive integers  $N_i$ . Then

$$p1 \lim_{n \rightarrow \infty} (1/b) \sum_{j=1}^b n_{ij}^{-1} = E[n_{ij}^{-1}] \quad i = 1, 2, \dots, a \tag{11}$$

where the symbol  $p1$  denotes convergency with probability one.

**Proof** Let  $\nu_{i\min}$  and  $\nu_{i\max}$  be the minimal and maximal elements of a finite set  $N_i$ . Then  $\nu_{i\max}^{-1} \leq n_{i1}^{-1} \leq \nu_{i\min}^{-1}$ . It follows that  $E[n_{i1}^{-1}]$  is finite. From strong law of large numbers we get Eq. (11) (see [10], p.27).

**Authors Contributions Statements** The authors jointly worked on the results and they read and approved the final manuscript.

**Declaration of Competing Interests** The authors declare that they have no competing interest.

## REFERENCES

- [1] Anderson, T. W., An Introduction to Multivariate Statistical Analysis (2d ed.), Wiley, New York, 1984.
- [2] Gaugler, T., Akritas, M. G., Testing for interaction in two-way random and mixed effects models: the fully nonparametric approach, *Biometrics*, 67 (4) (2011), 1314-1320.
- [3] Gaugler, T., Akritas, M. G., Mixed effects design: the symmetry assumption and missing data, *Journal of the American Statistical Association*, 107 (499) (2012), 1230-1238.
- [4] Gaugler, T., Akritas, M. G., Testing for main random effects in two-way random and mixed models: modifying the F statistic, *Journal of Probability and Statistics*, 2013 (2013), Article ID 708540, 11 pages.
- [5] Güven, B., A mixed model for complete three or higher way layout with two random effects factors, *Journal of Multivariate Analysis*, 139 (2015), 45-55.
- [6] Imhof, J. P., A mixed model for complete three or higher way layout with two random effects factors, *Annals of Mathematical Statistics*, 31 (4) (1960), 906-928.
- [7] Khattree, R., Rao, C. R., Handbook of Statistics 22, North-Holland, Amsterdam, 2003.
- [8] Khuri, A. I., Sinha, B. K., Statistical Test for Mixed Linear Models, Wiley, New York, 1998.
- [9] Scheffé, H., The Analysis of Variance, Wiley, New York, 1959.
- [10] Serfling, R. J., Approximation Theorems of Mathematical Statistics, Wiley, New York, 2002.