Journal of Statistics and Computer Science
Vol. 3, Issue 1, 2024, pp. 79-89
https://doi.org/10.47509/JSCS.2024.v03i01.05

A Note on Effective Transformation-based Exact F-test for Sub-Clustering Effect in Two-Fold Nested Error ANOVA Model

Yahia S. El-Horbaty[1]
Department of Mathematic, Insurance and Applied Statistics, Helwan University, Cairo, Egypt

Abstract

This article re-examines the use of exact F-tests for zero variance of the sub-clustering effect in the two-fold nested error analysis of variance model. Alternative to the classical F-test, a new outperforming version of an F-test statistic is proposed. The power of the new test is studied analytically. Using small simulation studies, the new test shows favorable performance to the classical test as well as to a simulation-based likelihood ratio test.

Keywords. Exact Tests, Variance Components, Orthogonal Transformations, ANOVA, Two-way Classification.

1. Introduction

The two-fold nested error ANOVA model (henceforth, 2FNE-model), also known as the empty three-level random intercept model (Snijders and Bosker, 2012), is a famous model used in various social science applications (Goldstein, 2011). The use of the random effects offers a wider flexibility in the 2FNE-model to accommodate for the complexity in the underlying sampling design. However, prior to proceeding through further inferential practices, testing the need for subset of those random effects may be of considerable importance. For example, in small area estimation (Rao and Molina, 2015; Pfeffermann, 2013), a great simplification is evident if one can arguably drop the sub-clustering effect in calculating the mean squared prediction error under the working small area model (Cai et al., 2020).

In this article, we are interested in testing subset of the random effects and particularly the sub-clustering effects under the 2FNE-model, for which we offer a new exact test. Focus is primarily on providing an alternative test statistic to the classical F-test, where the former is shown to possess exact finite sample properties too. Recently, Hui et al. (2019) reported that using the classical F-test may be as powerful as or even better than some recent simulation-based likelihood ratio tests (LRTs) for variance components. Stemmed from the resampling methods proposed in Ofversten (1993), we use efficient linear transformations in terms of utilizing a larger subset of the residuals from the transformed 2FNE-model than those proposed therein. Our simulation studies indicate that the proposed transformations result in higher power of the proposed tests compared to the classical test.

One of the drawbacks when the tested random effects are treated as fixed effects under the classical F-test is the power loss when clusters lack enough information to estimate such fixed effects (Scheipl et al., 2008; El-Horbaty, 2015). Nevertheless, Hui et al. (2019) advocated three advantages of this test. Namely, the computational speed, generality to broad class of models (including 2FNE-model), and its exactness. An early development of the classical F-test had been used to test the absence of random effects in linear mixed models in Wald (1941, 1947). See also Seely and El-

[1] Correspondence email: yahia_mohamed@commerce.helwan.edu.eg

Article History
Received : 04 September 2023; Revised : 30 May 2024, Accepted : 15 June 2024; Published : 30 June 2024

To cite this paper
Yahia S. El-Horbaty (2024). A Note on Effective Transformation-based Exact F-test for Sub-Clustering Effect in Two-Fold Nested Error ANOVA Model. *Journal of Statistics and Computer Science.* 3(1), 79-89.

Bassiouni (1983). Under random two-fold models with nested random effects, Khuri (1987)] considered the derivation of an exact F-test. See also Khuri and Littell (1987) for models with interaction terms. Ofversten (1993) pioneered various developments of various F-test statistics for some important unbalanced linear mixed models. See also El-Horbaty (2018, 2022b). A test for the sub-clustering effects have been taken into consideration therein. Alternatively, Zhang et al. (2016) proposed a simulation-based LRT, which avoids the problem of approximating the asymptotic properties of the LRT statistic under the balanced 2FNE-model. This work extends the original LRT statistic under models with single variance component Crainiceanu and Ruppert (2004). Asymptotic theory for the LRTs are studied in Self SG, Liang (1987) and Stram DO, Lee (1994). Permutation tests (Fitzmaurice et al. ,2007; Drikvandi et al., 2013; Samuh et al., 2012; El-Horbaty and Hanafy, 2020, 2024a; El-Horbaty, 2022a, 2023a, 2023b, 2024b) are also used to assess the need for random effects. However, the orientation of the majority of those tests have not been directed towards the 2FNE-model. Some simulation comparisons were using permutation tests were investigated in Abo-El-Hadid et al. (2021)

Compared to the test procedures in Ofversten (1993), the proposed test relies on a sequence of two transformations of the original response vector assuming normally distributed errors. The first splits off the original response vector such that two independent residual vectors can be composed, one of them includes identically and independently distributed components. The second transformation is applied to the other residual vector, where the rank of the corresponding augmented transformed design matrix is utilized to eliminate the main clustering effect before constructing the new F-statistic. The proposed statistic can be shown as a ratio of two independent quadratic forms in normally distributed variates where only one of them depends on the variance of the sub-clustering random effects.

The rest of this paper is organized as follows. Section 2 addresses the hypothesis of interest under the 2FNE-model where the classical F-test is briefly reviewed. In Section 3, the main results are summarized via a lemma and a heuristic reasoning for the new version of the F-test is provided. The empirical results using simulation experiments are given in Section 4. Section 5 concludes this work and provides some insights for future research.

2. The Classical F-test

Consider the 2FNE-model with nested random effects for representing N observations from m clusters, each composed of r_i sub-clusters of size n_{ij} for $i = 1, \dots, m$; $j = 1, \dots, r_i$. The model is given by

$$y_{ijk} = a + b_{1i} + b_{2ij} + e_{ijk} \qquad (1)$$

for $k = 1, \dots, n_{ij}$. In equation (1), y_{ijk} denotes the k^{th} observation on the response variable in the j^{th} sub-cluster from the i^{th} cluster, a denotes the overall mean, b_{1i} denotes the random effect of the i^{th} cluster, b_{2ij} denotes the random effect of the j^{th} sub-cluster from the i^{th} cluster, and e_{ijk} denotes the residual error. Clearly, the random effects b_{2ij} are nested within the random effect b_{1i} in the i^{th} group. It is conventionally assumed that the random variables b_{1i}, b_{2ij}, and e_{ijk} are mutually independent over all clusters and sub-clusters under (1), each with mean zero and respective variances σ_1^2, σ_2^2, and σ^2. Let $r = \sum_{i=1}^{m} r_i$, $n_i = \sum_{j=1}^{r_i} n_{ij}$, and $N = \sum_{i=1}^{m} n_i$.

The compact form of the 2FNE-model can be presented as

$$\boldsymbol{Y} = a\mathbf{1}_N + \boldsymbol{Z}_1\boldsymbol{b}_1 + \boldsymbol{Z}_2\boldsymbol{b}_2 + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \dots, \boldsymbol{Y}_m^T)^T$, $\mathbf{1}_N$ is an $N \times 1$ vector of ones, $\boldsymbol{Z}_1 = diag(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_m})$, $\boldsymbol{b}_1 = (b_{11}, \dots, b_{1m})^T$, $\boldsymbol{Z}_2 = diag(\boldsymbol{Z}_{21}, \dots, \boldsymbol{Z}_{2m})$, $\boldsymbol{b}_2 = (\boldsymbol{b}_{21}^T, \dots, \boldsymbol{b}_{2m}^T)^T$, and $\boldsymbol{e} = (\boldsymbol{e}_1^T, \dots, \boldsymbol{e}_m^T)^T$. For $i = 1, \dots, m$; $j = 1, \dots, r_i$, Note further that $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}^T, \dots, \boldsymbol{Y}_{ir_i}^T)^T$, $\boldsymbol{Y}_{ij} = (y_{ij1}, \dots, y_{ijn_{ij}})^T$, $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector of ones, $\boldsymbol{Z}_{2i} = diag(\mathbf{1}_{n_{i1}}, \dots, \mathbf{1}_{n_{ir_i}})$, $\mathbf{1}_{n_{ij}}$ denotes an $n_{ij} \times 1$ vector of ones, $\boldsymbol{b}_{2i} = (b_{2i1}, \dots, b_{2ir_i})^T$, $\boldsymbol{e}_i = (\boldsymbol{e}_{i1}^T, \dots, \boldsymbol{e}_{ir_i}^T)^T$, and $\boldsymbol{e}_{ij} = (e_{ij1}, \dots, e_{ijn_{ij}})^T$.

Particularly consider the following hypothesis

$$H_0: \sigma_1^2 > 0, \sigma_2^2 = 0 \quad \text{versus} \quad H_1: \sigma_1^2 > 0, \sigma_2^2 > 0, \tag{3}$$

Here, we review the construction of the classical F-test statistic as considered by [7]. See Hui et al. (2019) for an alternative derivation. However, for the sake of unifying the presentation needed to motivate the derivation of our new test statistic, we consider the construction of the test statistic as given in Ofversten (1993).

Let $rank[\mathbf{1}_N, \boldsymbol{Z}_1] = m$, $rank[\mathbf{1}_N, \boldsymbol{Z}_1, \boldsymbol{Z}_2] = \tilde{q} > m$ and note that $rank[\mathbf{1}_N, \boldsymbol{Z}_2] - rank[\mathbf{1}_N, \boldsymbol{Z}_2, \boldsymbol{Z}_1] = 0$. Further, let $\boldsymbol{C}$ be an orthogonal matrix such that $\boldsymbol{C}[\mathbf{1}_N, \boldsymbol{Z}_1, \boldsymbol{Z}_2] = [\boldsymbol{R}^T, \mathbf{0}^T]^T$, where $\boldsymbol{R}$ is of full row rank such that $rank(\boldsymbol{R}) = \tilde{q}$. The matrix $\boldsymbol{C}$ transforms model [2] as follows

$$\boldsymbol{CY} = \begin{bmatrix} t_1 \\ \boldsymbol{t}_2 \\ \boldsymbol{t}_3 \\ \boldsymbol{t}_4 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ \mathbf{0} & \boldsymbol{R}_{22} & \boldsymbol{R}_{23} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R}_{33} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a \\ \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} + \boldsymbol{Ce}, \tag{4}$$

and $\boldsymbol{R}$ can be partitioned column-wise as $[\mathbf{1}_N, \boldsymbol{Z}_1, \boldsymbol{Z}_2]$ and row-wise such that $rank[R_{11}, R_{12}, R_{13}] = 1$, $rank[\boldsymbol{R}_{22}, \boldsymbol{R}_{23}] = m - 1$, and $rank[\boldsymbol{R}_{33}] = \tilde{q} - m$. Focusing on $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$ to establish an F-test for the hypothesis in [3]. Assuming the normality of $\boldsymbol{b}_2$ and $\boldsymbol{e}$, we have

$$\boldsymbol{t}_3 \sim N[\mathbf{0}, \boldsymbol{R}_{33}\boldsymbol{R}_{33}^T\sigma_2^2 + \boldsymbol{I}_{\tilde{q}-m}\sigma^2],$$
$$\boldsymbol{t}_4 \sim N[\mathbf{0}, \boldsymbol{I}_{N-\tilde{q}}\sigma^2].$$

If the null hypothesis of (3) is true, the vectors $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$ are independent. Thus, an exact F-test for this hypothesis can be constructed as follows

$$F_0 = \frac{\boldsymbol{t}_3^T\boldsymbol{t}_3/(\tilde{q}-m)}{\boldsymbol{t}_4^T\boldsymbol{t}_4/(N-\tilde{q})} \tag{5}$$

which follows an F-distribution with degrees of freedom $\tilde{q} - m$ and $N - \tilde{q}$ respectively. We shall compare our proposals to the result in (5) as will be shown via the simulation studies later in Section 4. A heuristic justification of the powerfulness of the test statistic F_0 in (5) and an analytical derivation of its power function is given in Ofversten (1993). Reasonably large values for the test statistic should lead to a rejection of the null hypothesis in (3). Noticeably, F_0 can be derived alternatively by explicitly considering both $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ as fixed, rather than random, effects and testing only for the absence of $\boldsymbol{b}_2$. This has been the approach followed in Hui et al. (2019).

3. Main Result

We focus is on comparing the performance of the F_0 in (5) to a new proposed test statistic. Here we show that the F_0 is not the only statistic that can be derived under the null hypothesis of interest. Motivated by the transformation in (4), we propose some transformations of the observed vector $\boldsymbol{Y}$ that preserves linear combinations of all the observations in the original sample data. Note that, in Ofversten (1993), m out of N elements in the vector $\boldsymbol{CY}$ in (4) have to be sacrificed to conduct the F-test. As the statistical power of the F-test based on $\boldsymbol{CY}$ is known, at least, to be a non-decreasing function in the overall sample size, we expect that a test that utilizes a larger subset of elements in $\boldsymbol{CY}$ compared to F_0 to possess higher power. The results of the simulation experiments in the next section confirm such expectations. Lemma 1 summarizes the main result of this article and introduce the new test procedures.

Lemma 1

Under model (2), let $\boldsymbol{Q}$ be an $N \times N$ matrix such that $\boldsymbol{QY} = [\tilde{y}_1, \boldsymbol{Y}_2^T, \boldsymbol{Y}_3^T]^T$, where $\boldsymbol{QZ}_2 = [\widetilde{\boldsymbol{d}}_2, \boldsymbol{D}_2^T, \boldsymbol{0}^T]^T$ and $\boldsymbol{QQ}^T = \boldsymbol{I}_N$ with $\boldsymbol{D}_2^{*T} = [\widetilde{\boldsymbol{d}}_2, \boldsymbol{D}_2^T]^T$. Then, $\boldsymbol{D}_2^*$ is of full row rank where $rank(\boldsymbol{D}_2^*) = rank(\boldsymbol{QZ}_2) = q$. Under the null hypothesis in [3], the test statistic

$$F_1 = \frac{\boldsymbol{Y}_2^T \boldsymbol{H}\, \boldsymbol{Y}_2 / q_H}{\boldsymbol{Y}_3^T \boldsymbol{Y}_3 / (N-q)} \tag{6}$$

follows an F-distribution with respective degrees of freedom $q_H = rank(\boldsymbol{H})$ and $(N-q)$ where $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{D}_1(\boldsymbol{D}_1^T \boldsymbol{D}_1)^{-} \boldsymbol{D}_1^T$, and $\boldsymbol{D}_1$ denotes the matrix obtained from the first q rows of $\boldsymbol{QZ}_1$ with the first row being removed thereafter.

The proof of Lemma 1 is provided in the Appendix. Note also from this proof that $E_1[\boldsymbol{Y}_2^T \boldsymbol{H}\, \boldsymbol{Y}_2] > E_0[\boldsymbol{Y}_2^T \boldsymbol{H}\, \boldsymbol{Y}_2]$ where $E_0[.]$ and $E_1[.]$ denote the expectation evaluated under the null and alternative hypothesis, respectively. Since $E_1[\boldsymbol{Y}_2^T \boldsymbol{H}\, \boldsymbol{Y}_2] = trace\left(\boldsymbol{H}\left[\boldsymbol{D}_2 \boldsymbol{D}_2^T \sigma_2^2 + \boldsymbol{I}_{q-1}\sigma^2\right]\right) > trace(\boldsymbol{H}\sigma^2)$, the test statistic F_1 is expected to gain larger values as the value σ_2^2 departs from nullity. This heuristic reasoning is close in spirit to the one given in Ofversten (1993). In next section, we provide some empirical results on the performance the proposed test statistic. Again, it is not surprising that the conclusion from such results emphasizes the outperformance of F_1 over F_0 as the latter has to sacrifice m linear combinations of the response variable components from $\boldsymbol{CY}$ in (4) to compose the vectors $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$.

4. Empirical Assessments

In this section the objective is to compare the performance of the proposed test statistic F_1 (F_1-test hereafter) to the classical test statistic, F_0 (F_0-test hereafter). We also identify the situations where the proposed test is superior to the simulation-based LRT Zhang et al. (2016), eLRT hereafter, using small-scale simulation experiments.

The chosen nominal level in all simulation settings is taken to be 0.05 where we present the proportions of rejection of the null hypothesis in (3) using 1000 simulated datasets in each conducted experiment. Assume that the data generating process is defined according model (1) such that the number of main clusters $m = 10, 30$, the number of sub-clusters within each cluster is taken as $r_i = 2, 4$, each sub-cluster is of size $n_{ij} = 4, 7, 10$, and let $\sigma_1^2 = 1$. Throughout the results, we have assumed that the random effects in model (1) are normally distributed with zero mean.

We report the results using power curves, where $\sigma_2^2 = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ for all the previous simulation parameters when $r_i = 2$ as shown in Figure 1-4. Further, we enhance our reported results using Tables 1 and 2 where $r_i = 2, 4$ for limited values of σ_2^2. For both graphical and tabular summaries, we have assumed that the residual errors are either normally distributed or generated from Gamma distribution with zero mean and constant variance (i.e. $e_{ijk} = \sigma(U_k^* - 1)$ where $U_k^* \sim Ga(1,1)$).

The results could be summarized as follows. On one hand, the new F_1-test outperforms the classical version F_0-test under all simulation settings where the difference in the achieved power of tests can develop up to more than 20%. See Table 1 and Table 2 for detailed numerical results. However, we observe that when the residual errors are misspecified (i.e. non-normally distributed) as Gamma variates, the size of the new F_1-test remains, though not very close to 0.05, less interrupted than the classical one.

Table 1. Proportions of rejecting the null hypothesis under normally distributed residual errors (Nominal level 5%)

m	n_{ij}	*test*	$r_i = 2$			$r_i = 4$		
			σ_2^2			σ_2^2		
			0.0	0.2	0.5	0.0	0.2	0.5
10	4	F_1	5.21	35.8	73.7	5.05	65.2	97.6
		F_0	5.20	22.5	56.7	5.08	56.0	94.9
		eLRT	4.86	51.4	64.8	5.03	83.5	96.5
	7	F_1	4.98	63.0	92.8	4.93	92.4	100
		F_0	5.06	48.0	86.0	4.97	88.6	99.9
		eLRT	5.08	64.4	75.6	5.04	93.0	98.5
	10	F_1	4.89	76.5	97.7	5.02	99.0	100
		F_0	4.91	65.9	95.9	5.02	98.3	100
		eLRT	5.09	66.8	78.2	5.01	94.0	98.9
30	4	F_1	4.98	69.1	97.9	5.00	96.0	100
		F_0	5.06	44.1	92.5	5.10	92.2	100
		eLRT	5.08	69.5	81.2	5.15	96.3	98.5
	7	F_1	4.96	93.6	100	5.10	99.9	100
		F_0	4.93	82.9	99.9	5.23	99.8	100
		eLRT	4.99	80.3	83.5	4.87	100	100
	10	F_1	5.21	99.1	100	5.03	100	100
		F_0	4.99	96.2	100	5.08	100	100
		eLRT	4.91	85.0	89.5	5.04	100	100

On the other hand, the two aforementioned exact tests show a competing performance when compared to the eLRT. Consider the case where $r_i = 2$ with normally distributed residual errors. The power curves indicate that the eLRT is always dominant when σ_2^2 departs slightly from nullity (e.g. when $\sigma_2^2 = 0.1$), regardless of the values of m and n_{ij}. As both m and n_{ij} increase, the situation flips, and the exact F-tests show a considerably higher power compared to the eLRT. For example, when $n_{ij} = 10$, the F_1-test is highly dominant for values of σ_2^2 starting from 0.2. In general, a demerit that is exhibited by the eLRT is that its increasing power rates of convergence towards unity are not as fast as the F-tests for slighter increase in the sub-cluster size.

Table 2. Proportions of rejecting the null hypothesis under non-normally distributed residual errors (Nominal level 5%)*

m	n_{ij}	*test*	$r_i = 2$			$r_i = 4$		
			σ_b^2			σ_b^2		
			0.0	0.2	0.5	0.0	0.2	0.5
10	4	F_1	5.90	25.9	57.5	5.83	45.7	89.4
		F_0	7.70	21.4	41.5	6.70	39.0	82.4
		eLRT	4.50	44.5	61.5	4.56	73.5	91.5
	7	F_1	6.10	45.7	80.5	6.07	78.1	99.3
		F_0	8.32	34.4	68.2	7.40	69.0	98.3
		eLRT	6.50	51.5	65.5	6.01	83.0	96.5
	10	F_1	6.31	61.4	91.5	5.86	93.7	100
		F_0	10.2	46.0	83.9	8.50	88.9	99.8
		eLRT	5.02	54.0	70.3	4.07	87.5	97.0
30	4	F_1	6.62	48.1	88.3	5.89	83.8	99.9
		F_0	11.7	35.0	70.8	7.40	71.1	99.3
		eLRT	4.03	67.0	79.6	4.00	89.2	99.0
	7	F_1	5.98	80.1	99.6	6.38	99.3	100
		F_0	11.3	59.9	94.9	8.10	97.4	100
		eLRT	6.72	70.7	80.3	6.05	97.1	100
	10	F_1	6.26	93.9	100	6.00	100	100
		F_0	11.4	78.8	99.5	8.05	99.6	100
		eLRT	5.51	80.0	87.8	5.82	100	100

* Residuals errors follow Gamma (1,1).

The above conclusions drawn about the power of the competing tests do not change that much when the residual errors are Gamma-distributed. We only observe that when both m and n_{ij} increase the power gap slightly decreases with the eLRT. Obviously, the larger variability in the outcomes variable due to the misspecified normality is one of the obstacles restraining the power increase of the F-test. A final point to highlight is related to the possibility of using the bootstrap method to assess the performance of both test statistics F_0 and F_1 when the normality assumption could be hardly presumed in a given data application. This further point is out of the scope of this article and can also be left as a future point of investigation.

5. Conclusion

In this article, we presented a new version of the exact F-test for zero variance of the sub-clustering random effect in the 2FNE-model. Although our simulation experiments emphasized the use of balanced models, the applicability of the results when the sub-clusters sizes are unbalanced is also permissible. The proposed test yet entertains the attractive properties of its exactness to attain the nominated Type I error rate for any sample size, and unlike modern resampling-based techniques (e.g. the eLRT) are faster to implement due to its known finite sample behavior. The proposed derivation of the test statistic is also valid regardless of the correlation between the main and sub-clustering effects. It remains to assess the performance of the new test statistic when the normality assumption for the residual errors does not hold. We emphasize that the bootstrap or generally resampling methods can be useful tools to refer the reader to alternative tests that may be as powerful or even higher than the eLRT when there is enough data within clusters as required by F-tests.

References

Abo-El-Hadid, S.M., El-Horbaty, Y.S. and Hanafy, E.M. (2021). Testing the need for random effects in mixed one-way analysis of variance models – A comparative study. *Far East Journal of Mathematical Sciences*, 129 (2), 169–192.

Cai, S., Rao, J.N.K., Dumitrescu, L. and Chatrchi, G. (2020). Effective transformation-based variable selection under two-fold subarea models in small area estimation. *Statistics in Transitions*, 21, 68–83.

Crainiceanu, C., and Ruppert D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of Royal Statistical Society, Series B*, 66, 165-185.

Drikvandi, R., Verbeke, G., Khodadadi, A. and Nia, V.P. (2013). Testing multiple variance components in linear mixed-effects models. *Biostatistics*, 14, 144-159.

Fitzmaurice, G.M., Lipsitz, S.R. and Ibrahim, J.G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics*, 63,942-946.

El-Horbaty, Y.S. (2015). Model Checking Techniques for Small Area Estimation. Unpublished Ph.D. Dissertation, University of Southampton.

El-Horbaty, Y.S. (2018). On testing random effects in the nested-error regression model. *Communications in Statistics - Simulation and Computation*, 47(6), 1670-1676.

El-Horbaty, Y.S., and Hanafy, E.M. (2020). On testing variance components using permutation tests. *Journal of Applied Probability and Statistics*, 2020, 15(1), 113-126.

El-Horbaty, Y.S. (2022a). Testing the absence of random effects in the nested-error regression model using orthogonal transformations. *Communications in Statistic-Simulation and Computation*, 51(5), 2736-2746.

El-Horbaty, Y.S. (2022b). A note on covariance decomposition in linear models with nested-error structure: new and alternative derivations of the F-test. *Journal of Statistical Theory and Practice*, doi: 10.1007/s42519-022-00291-7.

El-Horbaty, Y.S. (2023a). A Monte Carlo Permutation Procedure for Testing Variance Components in Generalized Linear Regression Models. *Computational Statistics*, 2023b; doi.org/10.1007/s00180-023-01403-y

El-Horbaty, Y.S. (2023b). A note on using Kruskal-Wallis test in linear mixed-effects models with heavily skewed error components. The 34th annual Conference on Statistics and Modeling in Human and Social Sciences. Faculty of Economics and Political Sciences, Cairo University, Egypt. 1-12.

El-Horbaty, Y.S., and Hanafy, E.M. (2024a). A Monte Carlo Permutation Procedure for Testing Variance Components Using Robust Estimation Methods. *Statistical Papers*, 65, 335-356.

El-Horbaty, Y.S. (2024b) A new matrix-based formulation for computing the variance components F-test in linear models with crossed random effects. *Quality and Quantity,* 58, 3001-3020. doi.org/10.1007/s11135-023-01784-8.

Goldstein, H. (2011). Multilevel statistical models. John Wiley & Sons.

Hui, F.K.C., Muller, S. and Welsh, A.H. (2019). Testing random effects in linear mixed models: another look at the F-test (with discussion). *Australian and New Zealand Journal of Statistics*, 61, 61-84.

Khuri, A.I. (1987). An exact test for the nesting effect's variance component in an unbalanced random two-fold nested model. *Statistics and Probability Letters*, 5, 305-311.

Khuri, A.I., and Littell, R.C. (1987). Exact tests for the main effects variance components in an unbalanced random two-way model. *Biometrics*, 43, 545-560.Ofversten, J. (1993). Exact Tests for Variance Components in Unbalanced Mixed Linear Models. *Biometrics*, 49, 45–57.

Ofversten, J. (1993). Exact Tests for Variance Components in Unbalanced Mixed Linear Models. *Biometrics*, 49:45-57.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28, 40–68.

Rao, J.N.K., and Molina, I. (2015). Small Area Estimation. New Jersery: Wiley.

Samuh, M.H., Grilli, L., Rampichini, C., Salmaso, L. and Lunardon, N. (2012). The use of permutation tests for variance components in linear mixed models. *Communications in Statistics – Theory and Methods*, 41, 3020-3029.

Scheipl, F., Greven, S. and Kuchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics and Data Analysis*, 52, 3283–3299.

Seely, J., and El-Bassiouni, Y. (1983). Applying Wald's variance component test. *Annals of Statistics*, 11, 197-201.

Self. S.G., and Liang. K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605–610.

Snijders, T.A.B., and Bosker, R.J. (2012). Multilevel Analysis: An Introduction to basic and advanced multilevel modeling. 2nd Edn. Sage, London.

Stram, D.O., and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50, 1171–1177.

Wald, A. (1941). On the analysis of variance in case of multiple classifications with unequal class frequencies. *Annals of Mathematical Statistics*, 12, 346-350.

Wald, A. (1947). A note on regression analysis. *The Annals of Mathematical Statistics*, 18, 586-589.

Zhang, Y., Staicu, A.M. and Maity, A. (2006). Testing for additivity in non-parametric regression. *Canadian Journal of Statistics*, 44, 445–462.

Appendix A

Proof of Lemma 1

Let $rank(\boldsymbol{Z}_2) = q$. Since $rank([\mathbf{1}_N, \boldsymbol{Z}_2, \boldsymbol{Z}_1]) - rank([\mathbf{1}_N, \boldsymbol{Z}_2]) = 0$, the number of zero rows of $\boldsymbol{QZ}_1$ is $N - q$. Premultiplying model (2) by matrix $\boldsymbol{Q}$ yields the following structure

$$\boldsymbol{QY} = \begin{bmatrix} \tilde{y}_1 \\ \boldsymbol{Y}_2 \\ \boldsymbol{Y}_3 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 & \widetilde{\boldsymbol{d}}_1^T & \widetilde{\boldsymbol{d}}_2^T \\ \mathbf{0} & \boldsymbol{D}_1 & \boldsymbol{D}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a \\ \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \end{bmatrix}$$

Assuming that $\boldsymbol{Y}$ is normally distributed, we have $\boldsymbol{Y}_2 \sim N[\mathbf{0}, \boldsymbol{D}_1\boldsymbol{D}_1^T\sigma_1^2 + \boldsymbol{D}_2\boldsymbol{D}_2^T\sigma_2^2 + \boldsymbol{I}_{q-1}\sigma^2]$. Since $q - 1 > m$ by construction, an idempotent matrix $\boldsymbol{H}$ can be constructed such that $\boldsymbol{HD}_1 = \mathbf{0}$ and hence $V(\boldsymbol{HY}_2) = \boldsymbol{H}(\boldsymbol{D}_2\boldsymbol{D}_2^T\sigma_2^2 + \boldsymbol{I}_{q-1}\sigma^2)\boldsymbol{H}$. Further, $Cov(\boldsymbol{Y}_2, \boldsymbol{Y}_3^T) = \mathbf{0}$. Then, the quadratic forms $\boldsymbol{Y}_2^T\boldsymbol{HY}_2$ and $\boldsymbol{Y}_3^T\boldsymbol{Y}_3$ are independent with respective null distributions $\chi^2_{q_H}$ and χ^2_{N-q}, which completes the proof.

Power of Test

Following Ofversten (1993), the power of the proposed test using F_1 is given as follows. Let ψ denote the power of this test and let c be the critical value corresponding to a test of size α. That is, $\alpha = 1 - P(F_1 < c)$. Then, under the alternative hypothesis

$$\begin{aligned}\psi &= Pr(F_1 \geq c|\sigma_u^2 > 0)\\ &= Pr\left(\frac{\boldsymbol{Y}_2^T\boldsymbol{H}\,\boldsymbol{Y}_2/q_H}{\boldsymbol{Y}_3^T\boldsymbol{Y}_3/(N-q)} \geq c\right)\\ &= Pr\left(\frac{\boldsymbol{Y}_2^T\boldsymbol{H}\boldsymbol{H}^-\boldsymbol{H}\,\boldsymbol{Y}_2/q_H}{\boldsymbol{Y}_3^T\boldsymbol{Y}_3/(N-q)} \geq c\right)\\ &\geq Pr\left(\frac{\boldsymbol{Y}_2^T\boldsymbol{H}\left[\boldsymbol{H}\left(\boldsymbol{D}_2\boldsymbol{D}_2^T\sigma_2^2+\boldsymbol{I}\sigma^2\right)\boldsymbol{H}\right]^-\boldsymbol{H}\,\boldsymbol{Y}_2/q_H}{\boldsymbol{Y}_3^T\sigma^{-2}\boldsymbol{Y}_3/(N-q)} \geq \frac{\sigma^2 c}{\delta(\sigma_2^2+\sigma^2)}\right)\end{aligned} \tag{A1}$$

where δ denotes the smallest positive eigenvalue of $\boldsymbol{H}_* = [\boldsymbol{H}(\boldsymbol{D}_2\boldsymbol{D}_2^T\sigma_2^2 + \boldsymbol{I}\sigma^2)\boldsymbol{H}]$. Under the alternative hypothesis in [3], the last random variable in (A1) is a ratio of two mutually independent χ^2 random variables each divided by its respective degrees of freedom. Then,

$$\psi \geq 1 - F_{q_H,(N-q)}\left\{\frac{\sigma^2 c}{\delta(\sigma_2^2+\sigma^2)}\right\}. \tag{A2}$$

By (A2), for any true values of the variance components, we always have $\psi \geq 1 - F_{q_H,(N-q+1)}(c) = \alpha$. By (A2), with fixed σ^2 and σ_2^2, $\psi \to 1$ if $\delta \to \infty$. Hence the test is consistent as $N \to \infty$. It also follows from (A2) that with fixed values for δ and σ^2, then $\psi \to 1$ as $\sigma_2^2 \to \infty$.

Figures

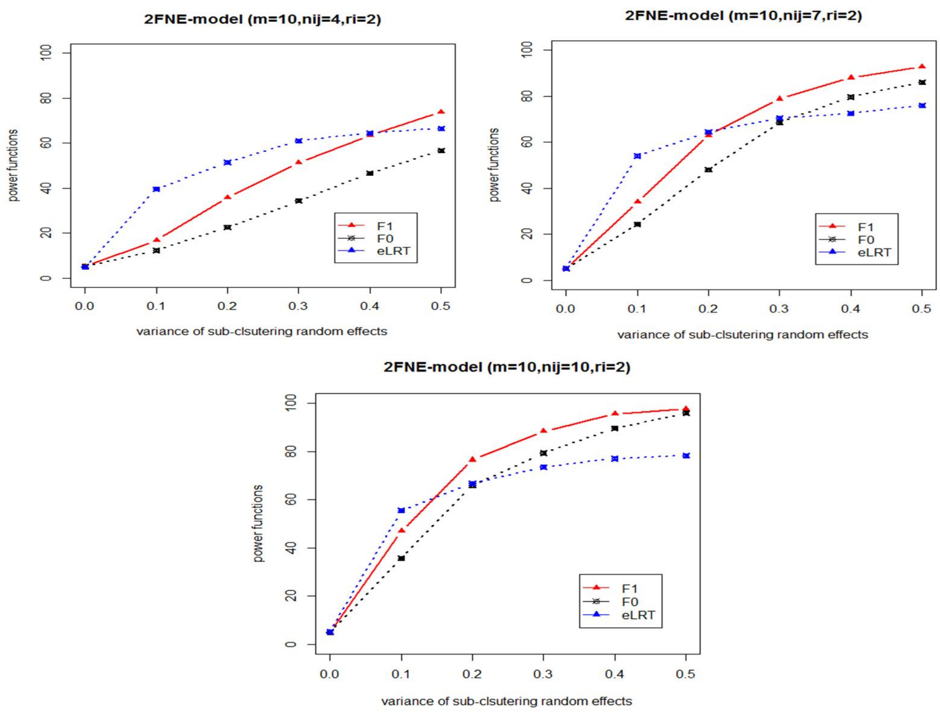

Figure 1. Empirical power curves for the F-tests and the eLRT ($m = 10,\ n_{ij} = 4,7,10,\ r_i = 2$) under normally distributed residual errors

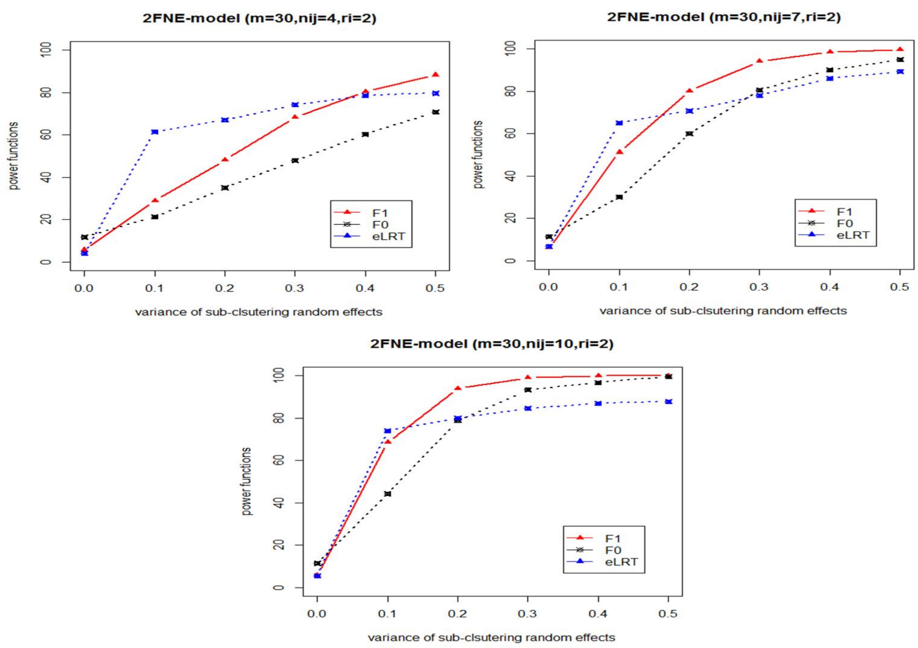

Figure 2. Empirical power curves for the F-tests and the eLRT ($m = 30,\ n_{ij} = 4,7,10,\ r_i = 2$) under normally distributed residual errors

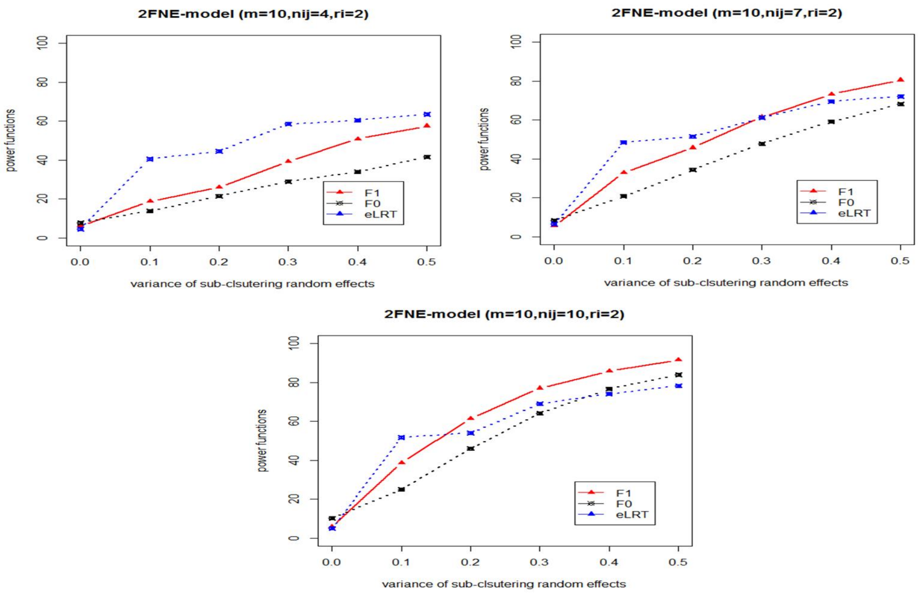

Figure 3. Empirical power curves for the F-tests and the eLRT ($m = 10,\ n_{ij} = 4,7,10, r_i = 2$) under $Gamma(1,1)$ distributed residual errors

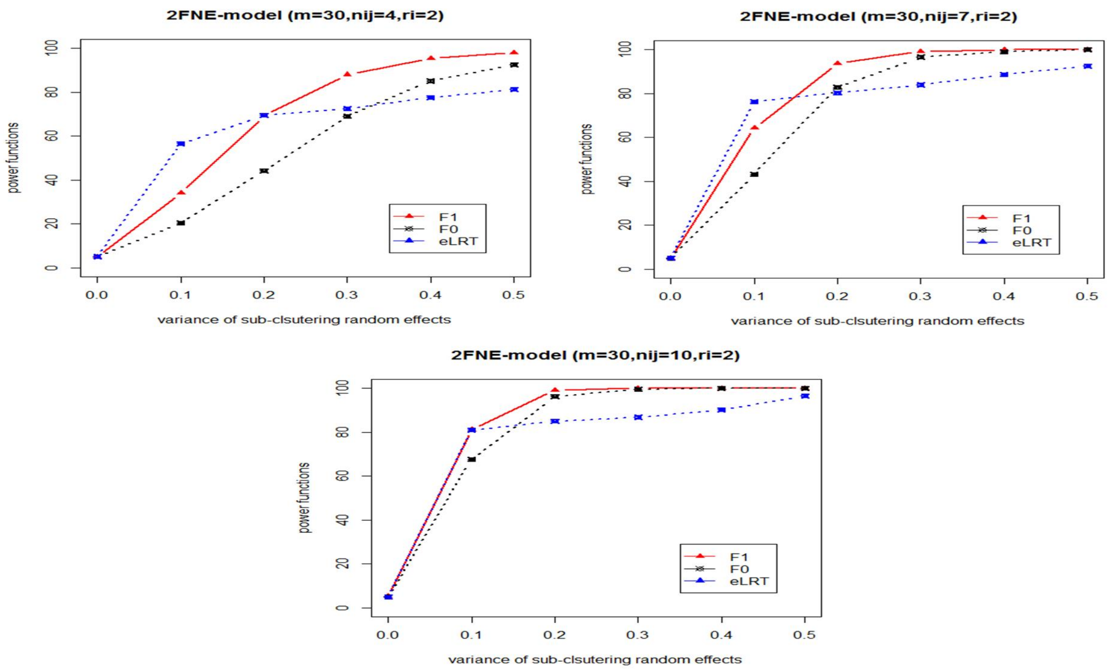

Figure 4. Empirical power curves for the F-tests and the eLRT ($m = 30,\ n_{ij} = 4,7,10, r_i = 2$) under $Gamma(1,1)$ distributed residual errors