Electronic Journal of Linear Algebra

Volume 33 *Volume 33: Special Issue for the International Conference on Matrix Analysis and its Applications, MAT TRIAD 2017* Article 6

2018

Testing Hypotheses Of Covariance Structure In Multivariate Data

Miguel Fonseca
NOVA University of Lisbon, fmig@fct.unl.pt

Arkadiusz Koziol
University of Zielona Gora, a.koziol@wmie.uz.zgora.pl

Roman Zmyslony
University of Zielona Gora, r.zmyslony@wmie.uz.zgora.pl

Recommended Citation

Fonseca, Miguel; Koziol, Arkadiusz; and Zmyslony, Roman. (2018), "Testing Hypotheses Of Covariance Structure In Multivariate Data", *Electronic Journal of Linear Algebra*, Volume 33, pp. 53-62.
DOI: https://doi.org/10.13001/1081-3810.3743

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

TESTING HYPOTHESES OF COVARIANCE STRUCTURE IN MULTIVARIATE DATA*

MIGUEL FONSECA†, ARKADIUSZ KOZIOŁ‡, AND ROMAN ZMYŚLONY‡

Abstract. In this paper there is given a new approach for testing hypotheses on the structure of covariance matrices in double multivariate data. It is proved that ratio of positive and negative parts of best unbiased estimators (BUE) provide an F-test for independence of blocks variables in double multivariate models.

Key words. Quadratic subspace, Testing hypotheses, Structure of covariance matrices, Positive and negative part of estimator, Block compound symmetric covariance structure, Double multivariate data.

AMS subject classifications. 62J10, 62F03 62F05, 62H15.

1. Introduction. Blocked compound symmetric (BCS) covariance structure for doubly multivariate observations (m dimensional observation vector repeatedly measured over u locations or time points), which is a multivariate generalization of compound symmetry covariance structure for multivariate observations, was introduced by [8, 9] while classifying genetically different groups, and then [7] studied BCS covariance structure while developing general linear model with exchangeable and jointly normally distributed error vectors.
We consider the following multivariate data:

$$\underset{num\times 1}{\boldsymbol{y}} = \mathrm{vec}\Big(\underset{um\times n}{\boldsymbol{Y}}\Big) \sim N\big((\mathbf{1}_n \otimes \boldsymbol{I}_{um})\boldsymbol{\mu}, \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_{um}\big), \tag{1.1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ are unknown, $\boldsymbol{\mu}$ is $um \times 1$ vector and

$$\boldsymbol{\Gamma} = \boldsymbol{I}_u \otimes \boldsymbol{\Gamma}_0 + (\boldsymbol{J}_u - \boldsymbol{I}_u) \otimes \boldsymbol{\Gamma}_1.$$

Here $\boldsymbol{I}_{\boldsymbol{u}}$ is an identity matrix of size u and $\boldsymbol{J}_u = \mathbf{1}_u \mathbf{1}_u'$ is the matrix with all elements equal to one, of size u. $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ are $m \times m$ unknown parameters. Equivalently the structure of covariance can be written as the sum of strong orthogonal matrices i.e. product of matrices is equal zero:

$$\boldsymbol{\Gamma} = \left(\boldsymbol{I}_u - \frac{1}{u}\boldsymbol{J}_u\right) \otimes (\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \frac{1}{u}\boldsymbol{J}_u \otimes (\boldsymbol{\Gamma}_0 + (u-1)\boldsymbol{\Gamma}_1)\,. \tag{1.2}$$

The paper is divided into five sections: the present introduction, sections dedicated to best unbiased and maximum likelihood estimation for the considered model, and to the proposed hypothesis test, followed by the simulation study and by final remarks.

*Received by the editors on February 24, 2018. Accepted for publication on July 30, 2018. Handling Editor: Natalia Bebiano. Corresponding Author: Miguel Fonseca.
†Center of Mathematics and Applications, Faculty of Sciences and Technology, NOVA University of Lisbon, Monte da Caparica, 2829-516 Caparica, Portugal (fmig@fct.unl.pt).
‡Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Szafrana 4a, 65-516 Zielona Góra, Poland (a.koziol@wmie.uz.zgora.pl, r.zmyslony@wmie.uz.zgora.pl).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

2. Best unbiased estimators and Maximum Likelihood Estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$. In [2] the best unbiased estimators (BUE) for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ are given. It follows that BUE for $\boldsymbol{\Delta}_0 = \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1$ and $\boldsymbol{\Delta}_1 = \boldsymbol{\Gamma}_0 + (u-1)\boldsymbol{\Gamma}_1$ are given by:

$$\widetilde{\boldsymbol{\Delta}}_0 = \widetilde{\boldsymbol{\Gamma}}_0 - \widetilde{\boldsymbol{\Gamma}}_1,$$
$$\widetilde{\boldsymbol{\Delta}}_1 = \widetilde{\boldsymbol{\Gamma}}_0 + (u-1)\widetilde{\boldsymbol{\Gamma}}_1.$$

On the other hand the maximum likelihood (ML) estimators for $\boldsymbol{\Delta}_0$ and $\boldsymbol{\Delta}_1$ are (see [3]) given by

$$\widehat{\boldsymbol{\Delta}}_0 = \frac{(n-1)(u-1)}{n(u-1)}\widetilde{\boldsymbol{\Delta}}_0,$$
$$\widehat{\boldsymbol{\Delta}}_1 = \frac{n-1}{n}\widetilde{\boldsymbol{\Delta}}_1.$$

Moreover, from [4] $\widetilde{\boldsymbol{\Delta}}_0$ and $\widetilde{\boldsymbol{\Delta}}_1$ are independent and

$$n(u-1)\widehat{\boldsymbol{\Delta}}_0 = (n-1)(u-1)\widetilde{\boldsymbol{\Delta}}_0 \sim \mathcal{W}_m\big(\boldsymbol{\Delta}_0, (n-1)(u-1)\big),$$
$$n\widehat{\boldsymbol{\Delta}}_1 = (n-1)\widetilde{\boldsymbol{\Delta}}_1 \sim \mathcal{W}_m\big(\boldsymbol{\Delta}_1, n-1\big),$$

where $\mathcal{W}_m\big(\boldsymbol{\Sigma}, n\big)$ stands for Wishart distribution with $m \times m$ scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom parameter n.

3. Testing hypotheses about structure of covariance. The maximum likelihood for the distribution of (1.1) is given by

$$\ell_1 = \big|\widehat{\boldsymbol{\Delta}}_0\big|^{-\frac{n(u-1)}{2}}\big|\widehat{\boldsymbol{\Delta}}_1\big|^{-\frac{n}{2}} e^{-\frac{num}{2}}.$$

Considering

$$H_0 : \boldsymbol{\Gamma}_1 = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\Gamma}_1 \neq \mathbf{0} \Leftrightarrow H_0 : \boldsymbol{\Delta}_1 = \boldsymbol{\Delta}_0 \text{ vs } H_1 : \boldsymbol{\Delta}_1 \neq \boldsymbol{\Delta}_0 \ :$$

From definitions of $\boldsymbol{\Delta}_0$ and $\boldsymbol{\Delta}_1$, it is clear that under H_0 we must necessarily have $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}_1$. Moreover, the maximum likelihood under H_0 is given by

$$\ell_0 = \Big|(nu)^{-1}\big(n(u-1)\widehat{\boldsymbol{\Delta}}_0 + n\widehat{\boldsymbol{\Delta}}_1\big)\Big|^{-\frac{nu}{2}} e^{-\frac{num}{2}}.$$

Finally, the likelihood ratio test is given by

$$\Lambda = \frac{\big|\widehat{\boldsymbol{\Delta}}_0\big|^{\frac{n(u-1)}{2}}\big|\widehat{\boldsymbol{\Delta}}_1\big|^{\frac{n}{2}}}{\Big|(nu)^{-1}\big(n(u-1)\widehat{\boldsymbol{\Delta}}_0 + n\widehat{\boldsymbol{\Delta}}_1\big)\Big|^{\frac{nu}{2}}}. \tag{3.3}$$

Now we prove the following:

Lemma 3.1. *If $\boldsymbol{W}_1 \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n_1)$ and $\boldsymbol{W}_2 \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n_2)$ (independent), then, for every fixed vector $\boldsymbol{x} \neq 0 \in \mathbb{R}^m$:*

$$T = \frac{n_2\boldsymbol{x}'\boldsymbol{W}_1\boldsymbol{x}}{n_1\boldsymbol{x}'\boldsymbol{W}_2\boldsymbol{x}} \sim F_{n_1,n_2}.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

Proof. According to the theorem in [1], if $\boldsymbol{W} \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n)$ then for every $\boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^m$:

$$\frac{\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}} \sim \chi^2_n.$$

Now if we calculate ratio of $\frac{\boldsymbol{x}'\boldsymbol{W}_1\boldsymbol{x}}{n_1}$ and $\frac{\boldsymbol{x}'\boldsymbol{W}_2\boldsymbol{x}}{n_2}$ we get:

$$\frac{\frac{\boldsymbol{x}'\boldsymbol{W}_1\boldsymbol{x}}{n_1}}{\frac{\boldsymbol{x}'\boldsymbol{W}_2\boldsymbol{x}}{n_2}} = \frac{\frac{\boldsymbol{x}'\boldsymbol{W}_1\boldsymbol{x}}{n_1\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}}{\frac{\boldsymbol{x}'\boldsymbol{W}_2\boldsymbol{x}}{n_2\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}} \sim \frac{\frac{\chi^2_{n_1}}{n_1}}{\frac{\chi^2_{n_2}}{n_2}} \sim F_{n_1,n_2}. \qquad \square$$

Under the framework of [6] and [10], the positive part of $\widetilde{\boldsymbol{\Gamma}}_1$ is given by $\widetilde{\boldsymbol{\Gamma}}_{1+} = \frac{\widetilde{\boldsymbol{\Delta}}_1}{u}$ and negative part is given by $\widetilde{\boldsymbol{\Gamma}}_{1-} = \frac{\widetilde{\boldsymbol{\Delta}}_0}{u}$.

To build a test for the nullity of $\widetilde{\boldsymbol{\Gamma}}_1$ we can use Lemma 3.1. To do this, first we find the test statistic.

Theorem 3.2. *Noting that the estimator of $\boldsymbol{\Gamma}_1$ is given by*

$$\widetilde{\boldsymbol{\Gamma}}_1 = \widetilde{\boldsymbol{\Gamma}}_{1+} - \widetilde{\boldsymbol{\Gamma}}_{1-} = \frac{\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0}{u}.$$

The test statistic

$$T = \frac{\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1+}\boldsymbol{v}}{\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1-}\boldsymbol{v}}, \tag{3.4}$$

with $\boldsymbol{v} \neq \boldsymbol{0}$, is distributed as an F random variable with $(n-1)$ and $(n-1)(u-1)$ degrees of freedom under $H_0 : \boldsymbol{\Gamma}_{1+} = \boldsymbol{\Gamma}_{1-}$.

Proof. Unbiased estimator of $\boldsymbol{\Gamma}_1$ can be expressed as

$$\widetilde{\boldsymbol{\Gamma}}_1 = \frac{(n-1)(u-1)\widetilde{\boldsymbol{\Delta}}_1 - (n-1)(u-1)\widetilde{\boldsymbol{\Delta}}_0}{(n-1)u(u-1)} = \frac{\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0}{u}.$$

Under null hypothesis $\boldsymbol{\Gamma}_1 = 0$:

$$(n-1)u\widetilde{\boldsymbol{\Gamma}}_{1+} = (n-1)\widetilde{\boldsymbol{\Delta}}_1 \sim \mathcal{W}_m(\boldsymbol{\Gamma}_0, n-1),$$
$$(n-1)u(u-1)\widetilde{\boldsymbol{\Gamma}}_{1-} = (n-1)(u-1)\widetilde{\boldsymbol{\Delta}}_0 \sim \mathcal{W}_m(\boldsymbol{\Gamma}_0, (n-1)(u-1)).$$

Now from Lemma 3.1 it follows that:

$$T = \frac{\frac{(n-1)u\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1+}\boldsymbol{v}}{(n-1)}}{\frac{(n-1)u(u-1)\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1-}\boldsymbol{v}}{(u-1)(n-1)}} = \frac{\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1+}\boldsymbol{v}}{\boldsymbol{v}'\widetilde{\boldsymbol{\Gamma}}_{1-}\boldsymbol{v}}. \qquad \square$$

Remark 3.3. *Under the null hypothesis and using $\boldsymbol{v} = \boldsymbol{1}$, the expectation of the numerator and the denominator of the statistic in (3.4) are equal, while under the alternative hypothesis and assuming that all elements of $\boldsymbol{\Gamma}_1$ are non-negative, the expectation of the numerator is greater than the expectation of the denominator. If the elements in $\boldsymbol{\Gamma}_1$ are non-positive then we reject the null hypothesis when the value of the test statistic is small enough.*

The next section will show simulation results using $\boldsymbol{v} = \boldsymbol{1}$, which is to take the sum of the elements of $\boldsymbol{\Gamma}_1$ as null.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

4. Simulation Study. To test the hypothesis described earlier, a simulation study using the R statistics software was performed and we compare the power function of F test ,one and two-sided, with the likelihood ratio test (LRT). The chosen parameters were $m = 3$, $u = 2$ and

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} 0.01221 & 0.02172 & 0.00901 \\ 0.02172 & 0.07492 & 0.01682 \\ 0.00901 & 0.01682 & 0.01108 \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 = \begin{bmatrix} 0.01038 & 0.01931 & 0.00824 \\ 0.01931 & 0.06678 & 0.01529 \\ 0.00824 & 0.01529 & 0.00807 \end{bmatrix},$$

where λ is a multiplier such that $\boldsymbol{\Gamma} = \boldsymbol{I}_u \otimes \boldsymbol{\Gamma}_0 + (\boldsymbol{J}_u - \boldsymbol{I}_u) \otimes \lambda \boldsymbol{\Gamma}_1$ is positive definite. In this case, $-0.280274 < \lambda < 1.12096$. For this instance, we take $\boldsymbol{v} = \mathbf{1}$, which means that the sum of the elements of $\boldsymbol{\Gamma}_1$ is equal to 0 under the alternative hypothesis. Generating 10000 observation vectors using sample sizes of $n = 5, 10, 15, 20$ and taking the significance level as 5%, the power of the LRT and one and two sided versions of the F test were compared. The results are shown in figures 1, 2, 3 and 4.

Consider another case, a very special one, presented in [5], where $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$ are scalars, with $m = 1$. Let $\boldsymbol{\Gamma}_0 = 2$ and $\boldsymbol{\Gamma}_1 = 1$. Additionally, it is assumed that $u = 2$, and parameter n will be one of the values from the set $\{3, 5, 10, 25\}$.

Matrix $\boldsymbol{\Gamma}$ has the following form:

$$\boldsymbol{\Gamma} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

From conditions of positive definiteness of matrix $\boldsymbol{\Gamma}$ it is easy to show that values of multiplier λ should be from interval $[-2, 2]$. The results are shown in figures 5, 6, which are included after the bibliography.

5. Final Remarks. The test presented in this paper provides a valid alternative to the likelihood ratio test for hypothesis tests on covariance components in multivariate models with BCS covariance structure, given that some previous knowledge on the covariance components is provided in the form of a linear restriction.

Given this condition, the presented test is more powerful than the LRT, as it is shown by the simulation study. It also offers computational advantages over the computationally expensive LRT distribution.

Acknowledgment. Miguel Fonseca acknowledges that his work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

REFERENCES

[1] Tõnu Kollo and Dietrich von Rosen. Advanced Multivariate Statistics with Matrices. *Springer*, 2010.
[2] Anuradha Roy, Roman Zmyślony, Miguel Fonseca and Ricardo Leiva. Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure. *Journal of Multivariate Analysis*, 144:81–90, 2016.
[3] Anuradha Roy and Miguel Fonseca. Linear Models with Doubly Exchangeable Distributed Errors. *Communications in Statistics - Theory and Methods*, 41:2545–2569, 2012.
[4] Anuradha Roy, Ricardo Leiva, Ivan Žežula and Daniel Klein. Testing the equality of mean vectors for paired doubly multivariate observations in blocked compound symmetric covariance matrix setup. *Journal of Multivariate Analysis*, 137:50–60, 2015.
[5] Edward Gasiorek, Andrzej Michalski and Roman Zmyślony. Tests of independence of normal random variables with known and unknown variance ratio . *Discussiones Mathematicae Probability and Statistics*, 20:233–247, 2000.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

[6] Andrzej Michalski and Roman Zmyślony. Testing Hypotheses for Variance Components in Mixed Linear Models. *Statistics: A Journal of Theoretical and Applied Statistics*, 27:297–310, 1996.

[7] Steven F. Arnold. Linear-Models with Exchangeably Distributed Errors. *Journal of the American Statistical Association*, 74:194–199, 1979.

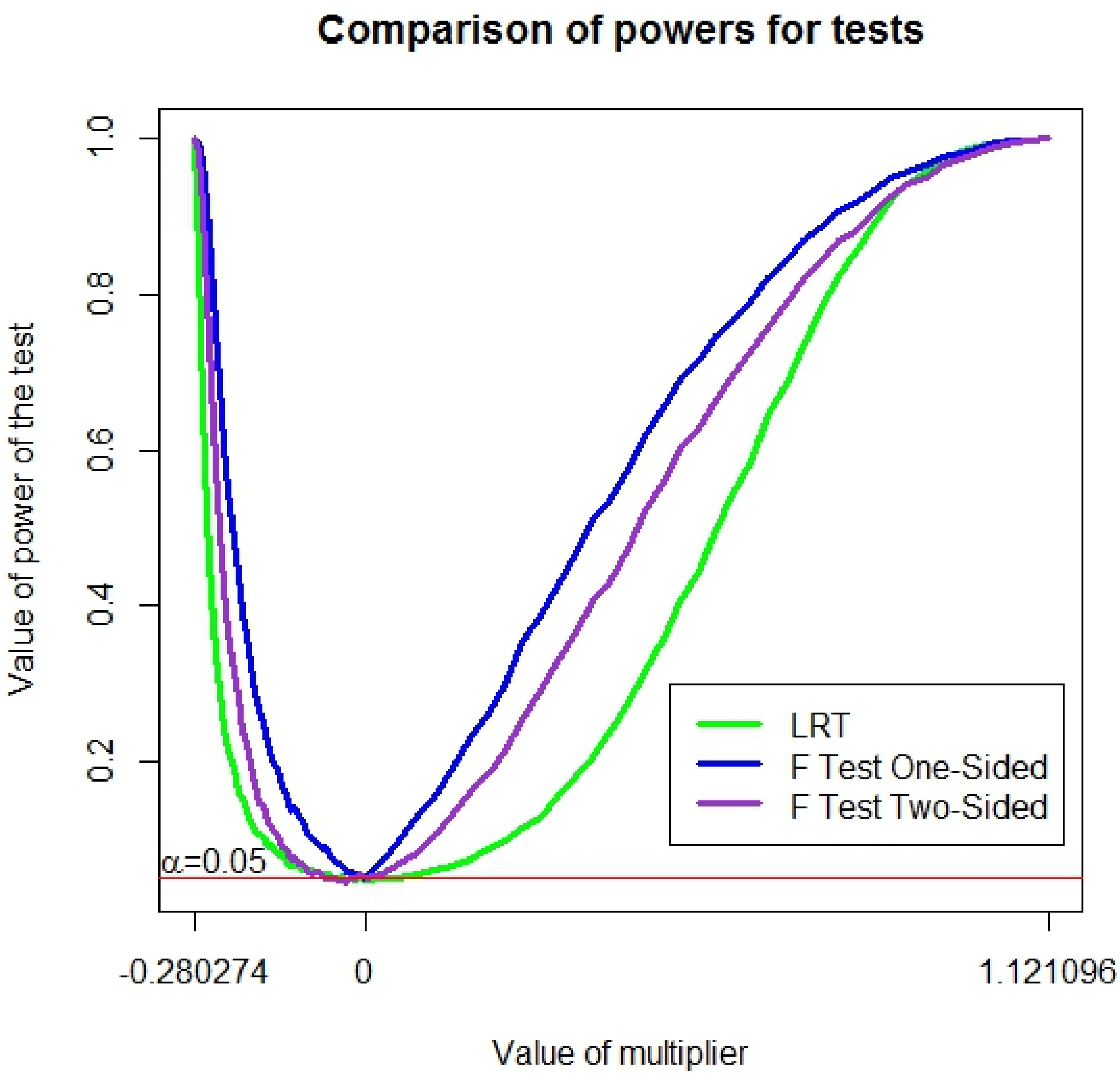
[8] Calyampudi Radhakrishna Rao. Familial correlations or the multivariate generalizations of the intraclass correlation. *Current Science*, 14:66–67, 1945

[9] Calyampudi Radhakrishna Rao. Discriminant Functions for Genetic Differentiation and Selection (Part IV of Statistical Inference Applied to Classificatory Problems). *Sankhyā: the Indian Journal of Statistics*, 12:229–246, 1953.

[10] Dário Ferreira, Sandra S. Ferreira and Célia Nunes. Linear-Models with Exchangeably Distributed Errors. *10th International Conference on Numerical Analysis and Applied Mathematics. AIP Conf. Proc.*, 1479:1682–1685, 2012.

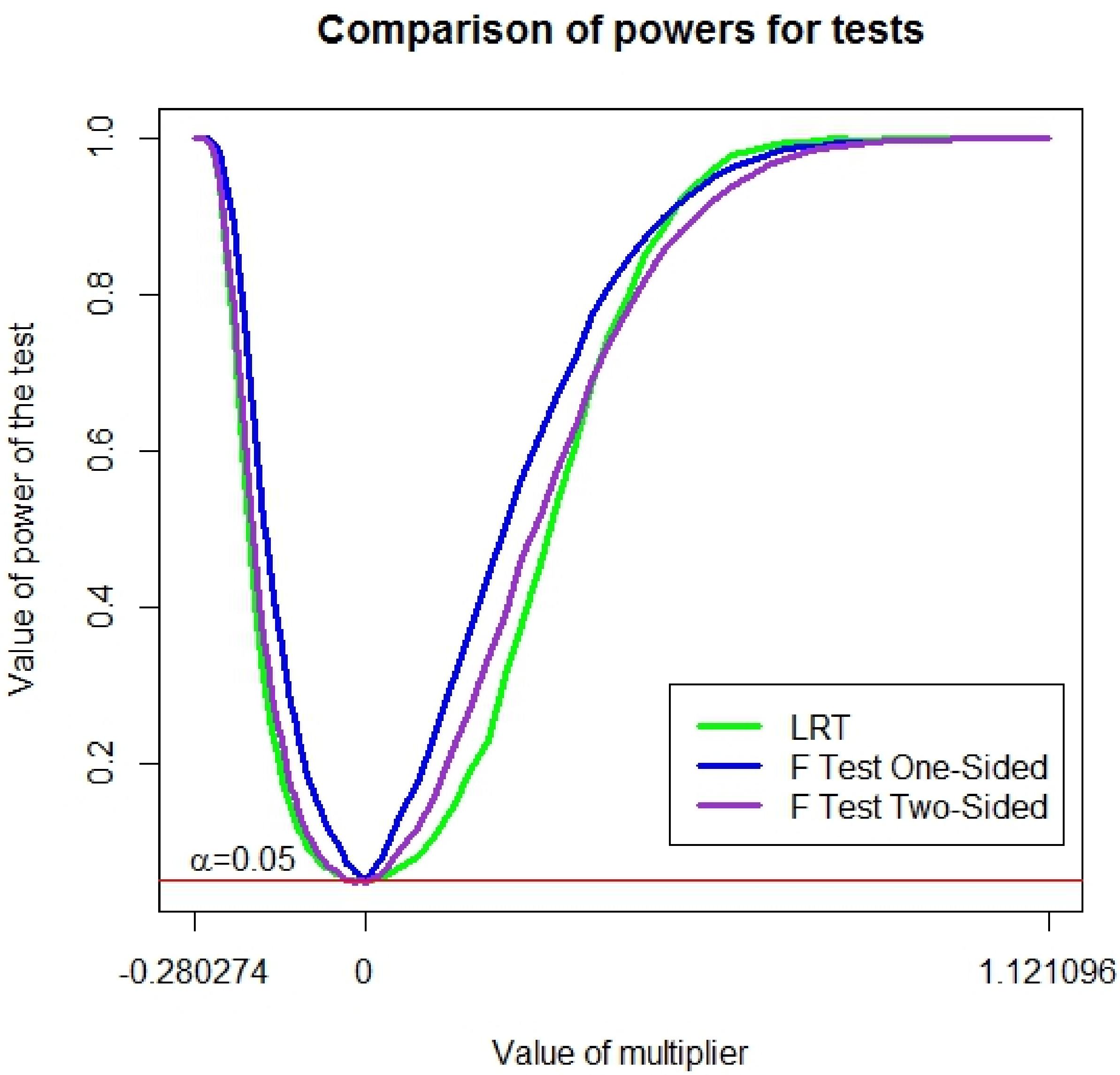
Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

FIGURE 1. $n = 5$



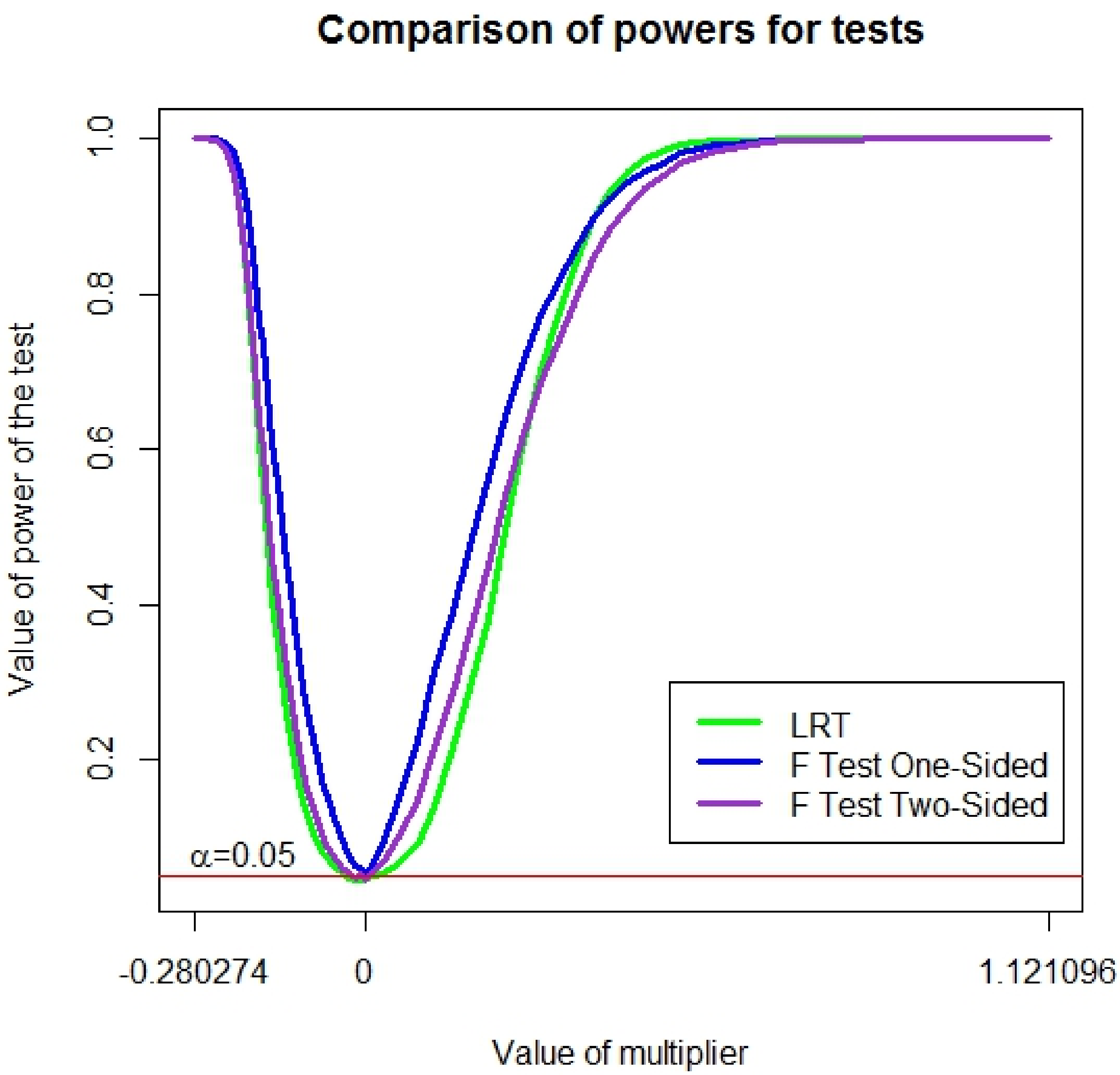
Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

Figure 2. $n = 10$



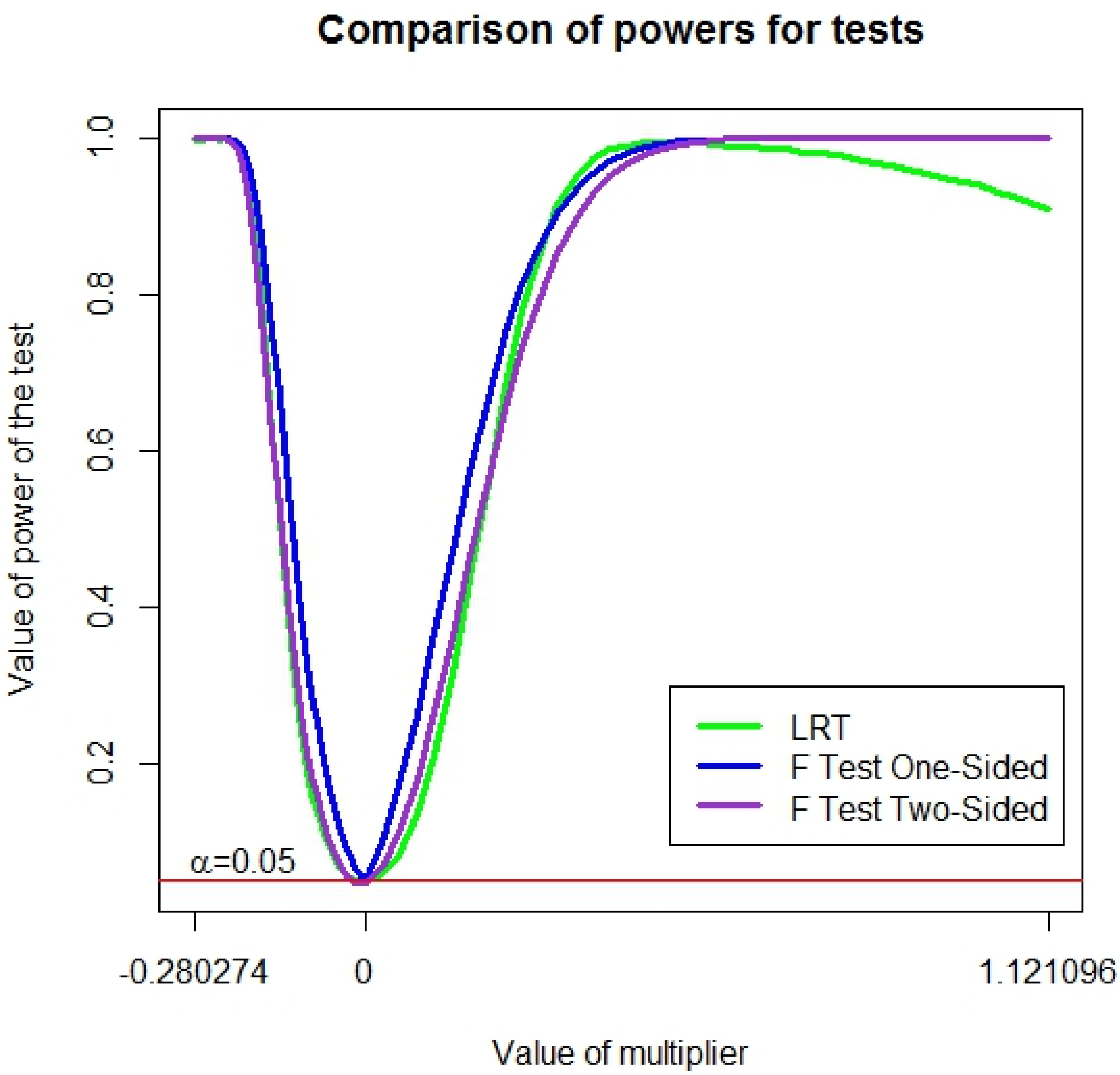
Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

FIGURE 3. $n = 15$



Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

Figure 4. $n = 20$



Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 33, pp. 53-62, August 2018.
http://repository.uwyo.edu/ela

Figure 5. $n = 3$ *and* $n = 5$

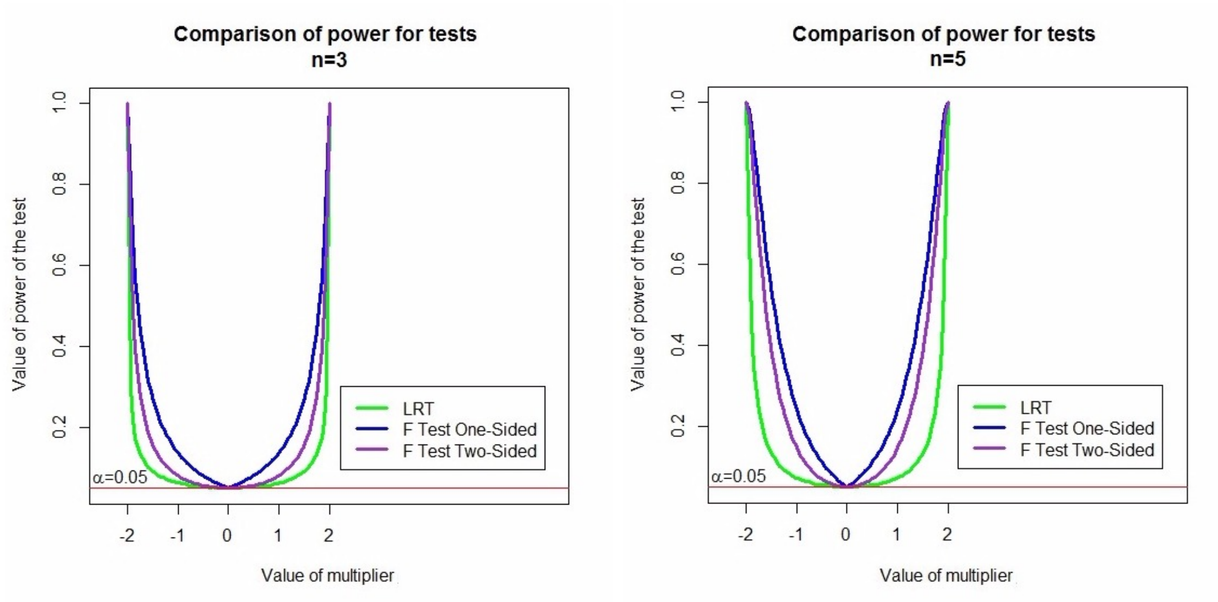


Figure 6. $n = 10$ *and* $n = 25$