

Bootstrapping F test for Testing Random Effects in Linear Mixed Models

P.Y. O'Shaughnessy, F. Hui, S. Muller, A.H. Welsh



UNIVERSITY OF WOLLONGONG AUSTRALIA

INTRODUCTION

Testing the significance of the random effects in the mixed models remains a crucial step in data analysis practise and a topic of research interest. Hui et. al. (2018) revisited the F-test, which was originally proposed by Wald (1947), and generalized the application to test subsets of random effects in a mixed model framework. They allows correlation between random effects, and showed that F test is an exact test when the first two moments of the random effects are specified.

We extend Hui et al. (2018)'s work and examine F test under non-normality of the residual errors. The model considered is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

which satisfies the following:

- The distributional assumptions for random effect \mathbf{b} are $E(\mathbf{b})=\mathbf{0}$ and $\text{Cov}(\mathbf{b})=\boldsymbol{\Sigma}$.
- The distributional assumptions for residual $\boldsymbol{\epsilon}$ are $E(\boldsymbol{\epsilon})=\mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon})=\boldsymbol{\Omega}$.

01 F test and its Bootstrap Counterparts

F Test for Testing a Subset of Random Effects
Suppose we wish to test

$$H_0 : \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \text{ vs } H_a : \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{11}$ is a $r_0 \times r_0$ non-zero and positive-definite matrix. $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{22}$ are non-zero matrices. The F test statistic is

$$F = \frac{(\mathbf{u}_0^T \mathbf{u}_0 - \mathbf{u}_1^T \mathbf{u}_1) / \{rk(\mathbf{X}, \mathbf{Z}) - rk(\mathbf{X}, \mathbf{Z}_0)\}}{(\mathbf{u}_1^T \mathbf{u}_1) / \{N - rk(\mathbf{X}, \mathbf{Z})\}},$$

where $\mathbf{u}_0 = \mathbf{L}_0^T \mathbf{y}$ and $\mathbf{u}_1 = \mathbf{L}_1^T \mathbf{y}$.

Bootstrap Hypothesis Test for F test

Residual Bootstrap

We constructed the null resampled dataset $(\mathbf{y}^*, \mathbf{X}, \mathbf{Z})$ and obtain the residual bootstrap test statistic

$$F^* = \frac{(\mathbf{u}_0^{*T} \mathbf{u}_0^* - \mathbf{u}_1^{*T} \mathbf{u}_1^*) / \{rk(\mathbf{X}, \mathbf{Z}) - rk(\mathbf{X}, \mathbf{Z}_0)\}}{(\mathbf{u}_1^{*T} \mathbf{u}_1^*) / \{N - rk(\mathbf{X}, \mathbf{Z})\}},$$

where we specify $\mathbf{u}_0^* = \mathbf{L}_0^T \mathbf{y}^*$ and $\mathbf{u}_1^* = \mathbf{L}_1^T \mathbf{y}^*$. The bootstrap response \mathbf{y}^* is constructed under the null

using $\hat{\boldsymbol{\epsilon}}^*$ elements resampled with replacement from $\hat{\boldsymbol{\epsilon}}$.

Fast Double Bootstrap

To obtain further asymptotic refinement, we use the fast double bootstrap method discussed by Davidson and MacKinnon (2007). This procedure corrects p-values by producing one 2nd-level bootstrap sample for each 1st-level sample to calculate a critical value at the nominal level equal to the first-level bootstrap p-value. A brief description is given below:

- Generate B bootstrap samples under the null hypothesis, and obtain the 1st-level bootstrap statistic F_b^* and p-value \hat{p}^* .
- Generate a single 2nd-level bootstrap sample for every bootstrap sample and obtain the $1 - \hat{p}^*$ quantile of F_b^{**} , defined as \hat{Q}^{**} .
- Calculate the *fast double bootstrap* p-value as

$$\hat{p}_F^{**} = \frac{1}{B} \sum_{k=1}^B I\{F_k^* > \hat{Q}_B^{**}\}.$$

02 Simulation Setting

The non-normality features we consider include spread, skewness and asymmetry. In particular, the following three specific cases of non-normal errors are considered:

- Student's t distribution with 3 df;
- Zero-mean chi-squared distribution with 3 df;
- Contamination model with mixture normal and a 20% contamination probability.

The random effects for each cluster was generated from a multivariate normal distribution with mean $\mathbf{0}$ and a defined covariance \mathbf{D} . The results show in **Figure 1** is generated using the following setting:

Simulation Design	Sample Size	H_0
Independent Cluster with 2 fixed covariates and 2 uncorrelated random covariates	$n=\{10,15\}$; $m=\{3,5\}$	$R1=R2=0^*$

* $R1=R2=0$ means the null hypothesis tests random effects 1 and 2 are both equal to zero.

04 CONCLUSION

Bootstrapping F test using the residual bootstrap or the fast double bootstrap both performs well for normal and non-normal residual errors in the following aspects:

✓ **Type I error rate:**
Both residual bootstrap and fast double bootstrap F tests achieve the correct Type I error rate of 5%.

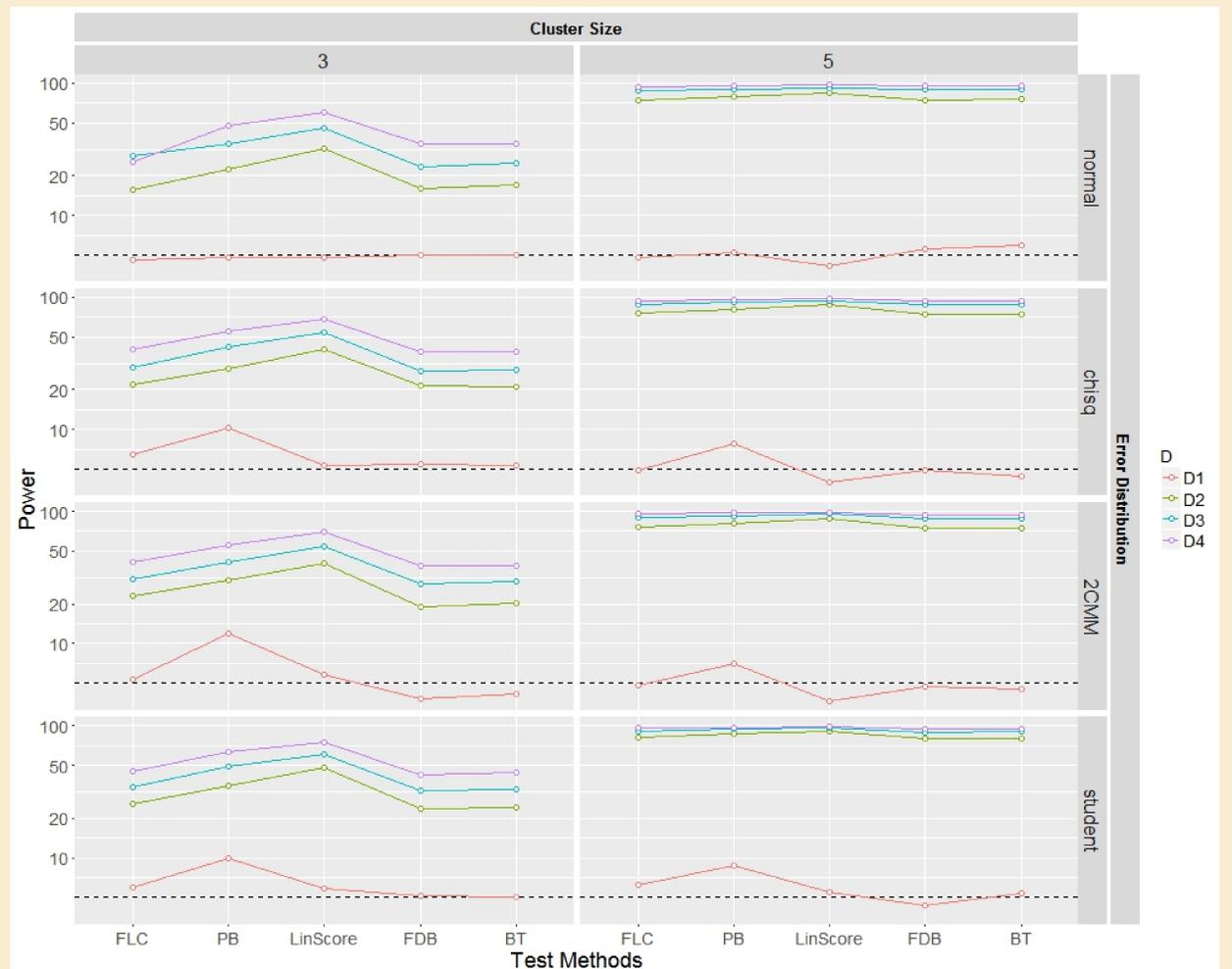
✓ **Power:**
Both residual bootstrap and fast double bootstrap produce comparable results in term of power.

✗ **Computational time:**
Bootstrap methods requires B times more computing time than the original F test, where B is the number of bootstrap samples.

SUMMARY

03 RESULTS

Figure 1: Type I error and power for four underlying distributions with different cluster sizes when the number of cluster $n=10$. The four distributions considered in the simulations are the *standard normal distribution* (normal), *Student's t distribution* with 3 df (student), the *zero-mean chi-squared distribution* with 3 df (chisq) and the *two-component normal mixture distribution* (2CMM). The methods compared included several well-known methods, i.e., the F test (FLC), the parametric bootstrap likelihood ratio test (PB), the linear score (LinScore), the fast double bootstrap FLC test (FDB) and the residual bootstrap FLC test (BT). Performance was assessed in terms of percentage of datasets where the method rejected the null hypothesis. Each colored line represents one of four covariance matrices. The y-axis is plotted on a log-10 scale and the dashed line is the reference line for the nominated 5% significance level. The bootstrap sample size B is 500.



The covariance matrix \mathbf{D} is defined as $\mathbf{D1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{D2} = \begin{pmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{pmatrix}$, $\mathbf{D3} = \begin{pmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{pmatrix}$ and $\mathbf{D4} = \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.10 \end{pmatrix}$. Here $\mathbf{D1}$ represents the case in which null hypothesis is correct and result shows the type I error at the nominal 5% level. The results for $\mathbf{D2}$, $\mathbf{D3}$ and $\mathbf{D4}$ show the power.

Bibliographie / Sources

- Davidson, R. and J. G. MacKinnon (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics and Data Analysis* 51, 3259.
- Hui, F. K., S. Mueller, and A. Welsh (2018). Testing random effects in linear mixed models: Another look at the F-test. *Aust. N. Z. J. Stats.* 1-18.
- Wald, A. (1947). A note on regression analysis. *Ann. Math. Statist.* 18 (4), 586-589.

Presenter: Pauline O'Shaughnessy
Email: poshaugh@uow.edu.au

DATE :12/2018