

## A Comparison of Confidence Interval Methods of Fixed Effect in Nested Error Regression Model

Hatice Tül Kübra AKDUR\*<sup>1</sup>, Deniz ÖZONUR<sup>1</sup>, Hülya BAYRAK<sup>1</sup>

<sup>1</sup> Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06504, Ankara

(Alınış / Received: 15.02.2016, Kabul / Accepted: 01.07.2016, Online Yayınlanma / Published Online: 26.07.2016)

### Keywords

Mixed-effect models,  
Confidence interval,  
Parametric bootstrap,  
Profile likelihood.

**Abstract:** Linear mixed-effects models are very popular and powerful tools in many scientific fields such as zoology, biology, and education. Estimators of fixed effects do not only depend on the variances of error terms but they also depend on random terms in mixed-effect models. When the distributions of random effects are unknown or enough sample size cannot be obtained, standard methods may fail. This study aims to determine a promising confidence interval method among existing methods in terms of coverage probability of true value of parameter. Standard and parametric bootstrap-based confidence interval methods for nested error regression model were compared in the simulation study under small samples. It is observed that parametric bootstrap-based method provides better coverage rates for small intra-correlation and profile likelihood method usually provides better results for moderate and strong correlation.

## İç İçe Hata Regresyon Modelinde Sabit Etki Parametresi için Güven Aralığı Metodlarının Karşılaştırılması

### Anahtar Kelimeler

Karışık etki modelleri,  
Güven aralığı,  
Parametrik bootstrap,  
Profil olabilirlik.

**Özet:** Doğrusal karışık etki modelleri, zooloji, biyoloji ve eğitim gibi bilimin birçok alanında popüler ve güçlü bir araçtır. Bu modellerde sabit etki parametrelerinin tahminleri hata ve rasgele etki terimlerinin varyanslarına bağlıdır. Rasgele etkilerin dağılımları bilinmediğinde veya yeterli sayıda örnek bulunamadığında standart metodlar doğru sonuç vermeyebilir. Karışık etki modellerinin özel bir hali olan iç içe hata regresyon modelinin sabit etki parametresi için varolan güven aralığı metodları arasında parametrenin gerçek değerini kapsama olasılığı bakımından en iyi olan güven aralığı metodu aranmıştır. Standart ve parametrik bootstrap-tabanlı güven aralığı metodları iç içe hata regresyon modeli için simülasyon çalışmasında küçük örnek çaplarında karşılaştırılmıştır. Zayıf korelasyonda parametrik-bootstrap metodları daha iyi sonuçlar vermiştir. Profil olabilirlik metodu orta ve güçlü korelasyonlarda daha iyi sonuçlar sağlamıştır.

### 1. Introduction

Linear mixed-effects model is called with many names in statistics literature. As Demidenko [1] said in his book, "big ideas have many names and applications". These names are such as mixed-effects model, linear mixed model, mixed linear model, hierarchical model, linear mixed-effects model, multilevel model.

Linear mixed-effects model (LMM) is derived from the fact that these models are linear in terms of parameters in the models and they are mixed because these models include independent terms that can be random and fixed. Mixed model as a terminological

term was firstly introduced by Eisenhart [2], random and fixed effect terms are formally distinguished in his study. LMMs do not require the assumption of independent and identically distributed sample as required in the typical statistics models such as linear regression. More complex dataset structures such as nested, hierarchical, multilevel can be handled with LMM. Between and within units or clusters variations are included in LMM. In LMM, observations within each unit or cluster are dependent whereas observations between units or clusters are independent. This flexibility attribute gives LMMs appropriateness for using repeated measures, longitudinal, panel and spatial datasets. With the existence of random effects in the model, theoretical

approach to estimation and inference of fixed effects become more complicated than standard linear models.

In order to make inference for fixed effect parameters of LMM, several confidence interval methods are suggested by researchers in statistics literature. Before constructing confidence interval, estimator of fixed effect and variance estimation of random effects must be found. For variance estimation, best known methods are ANOVA, maximum likelihood (ML), and restricted (or residual) maximum likelihood (REML) methods. For LMM, REML estimation method was first introduced by Patterson and Thompson [3]. Theoretical explanation of these methods are broadly shown in Searle et al [4].

For a good performance of confidence interval, a researcher needs to have a good estimator of fixed effect that has a small mean square error (MSE) . After a careful determination of these estimators, pivot and its distribution must be found to obtain percentiles to construct a confidence interval. Kackar and Harville [5] proposed MSE estimators of fixed and random effects in LMM. They indicated that if true values of variances of fixed and random effects are replaced by estimated values, the MSEs of estimators of fixed and random effects increase in size. Kenward and Roger [6] also proposed MSE estimator by combining Kackar and Harville approximations and their method to avoid bias in small samples for restricted maximum likelihood estimator of fixed effect in LMMs.

Harville and Fenech [7] investigated confidence interval methods for a variance ratio in unbalanced mixed linear models. Savin [8] constructed confidence interval for common mean in one-way classification model. Hall and Maiti [9] studied parametric bootstrap methods to construct confidence interval in mixed-effects model. Staggs [10] proposed parametric bootstrap approach to construct confidence intervals for fixed effect parameters of several mixed-effects models. Burch [11] constructed confidence intervals for variance components in unbalanced one-way random effects model using non-normal distributions. Liu [12] compared some confidence interval methods for function of parameters in repeated measures degradation model. Even though, there are several comparison studies for confidence interval methods, our study attempts to cover all existing methods for fixed effect of LMM.

In this paper, main focus is to be nested error regression model which is a special form of LMMs, given by

$$Y_{ij} = X_{ij}\beta + u_i + e_{ij} \tag{1}$$

A canonical form of nested error regression model is shown as follows

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m \tag{2}$$

$Y_{ij}$  is a response for  $j$  of unit  $i$  ,  $n$  denotes the total number of experimental units or clusters, and  $m$  is the number of replications or observations within each unit or cluster.  $\beta_0$  is common mean,  $\beta_1$  denotes the fixed effect parameter related with fixed regressor  $x_{ij}$  .  $u_i$  is normally distributed random effect corresponding to unit  $i$  .  $\sigma_u^2, \sigma_\varepsilon^2, \rho$  terms can be parameterized for the variance components of the model.  $\sigma_u^2$  is variance of the random effect  $u_i$  ,  $\sigma_\varepsilon^2$  is variance of the error term  $\varepsilon_{ij}$  , and  $\rho$  is intra-class correlation coefficient.  $\rho$  is defined as  $\rho = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_\varepsilon^2)}$  . General form of mixed-effects model for a particular experimental unit  $i$  is shown in Eq.(3).

$$Y_i = X_i\beta + Z_i u_i + \varepsilon_i, \quad u_i \sim N(0, D_{sub}), \quad \varepsilon_i \sim N(0, R_i) \tag{3}$$

$R_i$  with the dimension  $m \times m$  and  $D_{sub}$  with the dimension  $q \times q$  are covariance matrices of respectively error terms and random effects.  $p$  and  $q$  are respectively number of fixed effect and random effect terms in LMMs.  $X_i$  is the design matrix for fixed effects and has the dimension  $m \times p$  for experimental unit  $i$  . It specifies values of fixed effects corresponding to each parameter for each observation. When fixed effect is categorical effect, the values of zero and one are used to denote the absence and presence of effect categories. For covariate effects, the variable values themselves are used in  $X_i$  . Similarly,  $Z_i$  is the design matrix for random effects with the dimension of  $m \times q$  for experimental unit  $i$  .  $u_i$  is  $q \times 1$  vector of random effects for experimental unit  $i$  with zero mean and covariance matrix  $D_{sub}$  .  $\varepsilon_i$  is  $m \times 1$  vector of error terms for experimental unit  $i$  with zero mean and covariance matrix  $R_i$  .

In addition, we assume that  $cov(\varepsilon_i, u_i) = 0$  . The assumptions of LMM above imply that, marginally  $Y_i \sim N(X_i\beta, V_i)$  , where  $V_i = Z_i D_{sub} Z_i' + R_i$  .

In small area models, nested error regression model is used to involve unit-level predictors. The main purpose of this paper is to compare performances of several confidence interval methods for fixed effect parameter of nested error regression model under the situations of small samples.

## 2. Confidence Interval Methods of Fixed Effect in Nested Error Regression Model

In many applications, researchers usually take fixed effect into focus in mixed-effects model. Many researchers suggest constructing confidence interval (CI) for making inference of fixed effect parameters in LMMs. In order to perform statistical inference in this paper, we used three common maximum likelihood-based asymptotic statistics; they are Wald-type statistics and likelihood ratio statistics. Neyman and Pearson introduced likelihood ratio statistics [13]. Wald statistic was firstly proposed by Wald [14]. All these statistics are known as the first-order asymptotic statistics. They are derived from functions of maximum likelihood estimators, Fisher information matrix (FIM), or a consistent estimator of FIM. These statistics and their application to confidence intervals are described in the next chapters. Most common method among them is the Wald-type methods of confidence interval.

### 2.1. Likelihood ratio test-based confidence interval method

Likelihood Ratio Test-based confidence interval is also known as profile likelihood confidence interval (PLCI) in statistics literature. Since profile likelihood confidence interval method does not require asymptotic normality of the estimator, they usually perform better than Wald-type confidence intervals for small to moderate sample sizes. Even though they do not require normality assumption of the estimators, they still depend on an asymptotic approximation of likelihood ratio test statistic that is believed to have chi-square distribution. Likelihood function of linear mixed model in Eq. (3) is written as below

$$L = C - \frac{1}{2} \sum_{i=1}^n \log |R_i| - \frac{1}{2} \sum_{i=1}^n (Y_i - X_i \beta)' R_i^{-1} (Y_i - X_i \beta)$$

where  $C$  is a constant that does not depend on model parameters. This likelihood function expression of linear mixed model is given in detail by Pinheiro and Bates [15]. The model of interest is nested error regression model in this study. Parameter vector of this model is  $\theta^\dagger = (\beta_0, \beta_1, u_i, \sigma_\varepsilon^2, \sigma_u^2)$ .  $\delta = (\beta_0, u_i, \sigma_\varepsilon^2, \sigma_u^2)$  is an additional parameter vector for the model.  $L(\beta_1, \delta)$  is the likelihood function and then profile likelihood function for  $\beta_1$  is  $L_1(\beta_1) = \max_{\delta} L(\beta_1, \delta)$ .  $L_1(\beta_1)$  is the maximum likelihood function over the remaining parameters for each fixed value of  $\beta_1 = \beta_{1(0)}$ . In order to obtain PLCI for parameter of interest, the likelihood ratio test of a two-sided hypothesis should be inverted. For a two-sided test of null hypothesis  $H_0 : \beta_1 = \beta_{1(0)}$ , the likelihood ratio test statistic is the

difference between log-likelihood of full model and reduced model:

$$2[\log L(\hat{\beta}_1, \hat{\delta}) - \log L(\hat{\beta}_{1(0)}, \hat{\delta}_0)] ,$$

where  $\hat{\beta}_1$  and  $\hat{\delta}$  are the MLEs for the full model and  $\hat{\delta}_0$  is the MLE of  $\delta$  for the reduced model with  $\beta_1 = \beta_{1(0)}$ . Hence, a  $100(1-\alpha)\%$  CI for  $\beta_1$  includes the values of  $\beta_{1(0)}$  where the test is non-significant at the  $\alpha$  level as shown below in the inequalities:

$$2[\log L(\hat{\beta}_1, \hat{\delta}) - \log L(\hat{\beta}_{1(0)}, \hat{\delta}_0)] < \chi_{1-\alpha,1}^2$$

or

$$\log L(\hat{\beta}_{1(0)}, \hat{\delta}_0) > \log L(\hat{\beta}_1, \hat{\delta}) - \frac{\chi_{1-\alpha,1}^2}{2} ,$$

where  $\chi_{1-\alpha,1}^2$  is the  $(1-\alpha)$  quantile of  $\chi^2$  distribution with one degree of freedom.

### 2.2. Wald-type confidence interval methods

In appropriate settings where the assumptions of mixed model are valid, Wald-type confidence interval is preferred for fixed effect parameter. This test is called Wald Test based on the large sample theory and maximum likelihood theory [14]. MLEs have useful properties such as consistency and asymptotic normality. These properties are valid if certain conditions defined by Lehmann are hold [16]. If these conditions are hold, Wald-type statistics are reliable to construct confidence interval. For sufficiently large  $n$  under these conditions,

$$\hat{\theta}^\dagger \overset{\cdot}{\square} N(\theta^\dagger, (I(\theta^\dagger))^{-1}) \tag{4}$$

equation (4) indicates that  $\hat{\theta}^\dagger$  is consistent estimator of  $\theta^\dagger$  and  $\hat{\theta}^\dagger$  is “approximately distributed as normal” for sufficiently large  $n$ . Another crucial result is that  $I(\theta^\dagger) = -E \left[ \frac{\partial \ln(L(\theta^\dagger))}{\partial \theta^\dagger \partial \theta^{\dagger'}} \right]$  which is FIM.

By using this result, covariance matrix of MLEs is calculated. Usually, observed information matrix is used because obtaining expected information matrix may not be possible. Observed information matrix is shown as below,

$$I^0(\theta^\dagger) = - \left[ \frac{\partial \ln(L(\theta^\dagger))}{\partial \theta^\dagger \partial \theta^{\dagger'}} \right] \tag{5}$$

Wald statistic is from Eq.(4) for sufficiently large  $n$ ,

$$W = (\theta^\dagger - \hat{\theta}^\dagger)' \left[ I(\hat{\theta}^\dagger) \right] (\theta^\dagger - \hat{\theta}^\dagger) \square \chi_p^2 \tag{6}$$

with  $p$  degrees of freedom,  $\chi_p^2$  is chi-squared distribution and  $I(\hat{\theta}^\dagger)$  is  $I(\theta^\dagger)$  evaluated at the MLE  $\hat{\theta}^\dagger$ .  $W$  is used to construct approximate  $100(1-\alpha)\%$  confidence regions by replacing values of  $\theta^\dagger$  that satisfy  $W = (\theta^\dagger - \hat{\theta}^\dagger)' [I(\hat{\theta}^\dagger)] (\theta^\dagger - \hat{\theta}^\dagger) \leq \chi_p^2$ , where  $\chi_{1-\alpha, p}^2$  is the  $(1-\alpha)$  percentile of  $\chi_p^2$ . In order to construct Wald confidence interval, approximate percentile of the pivot:  $\frac{\hat{\theta}^\dagger - \theta^\dagger}{SE(\hat{\theta}^\dagger)}$  needs to be obtained.

Based on asymptotic normality of the maximum likelihood estimator, the distribution of the pivot has asymptotically standard normal distribution in large sample sizes and its percentiles can be obtained from standard normal distribution percentiles.

In linear mixed model context, FIM is written as negative expectation of the Hessian matrix. It is used for obtaining standard error of the MLE  $\theta^\dagger$  of  $\hat{\theta}^\dagger$ . Hessian matrix is given below:

$$H = \begin{pmatrix} H_{\beta\beta} & H_{\beta V} \\ H_{V\beta} & H_{VV} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta} & \frac{\partial^2 l}{\partial \beta \partial V} \\ \frac{\partial^2 l}{\partial V \partial \beta} & \frac{\partial^2 l}{\partial V \partial V} \end{pmatrix}$$

where  $l$  is the log likelihood,  $\beta = (\beta_0, \beta_1)$  is the vector of fixed effect parameters and  $V = (\sigma_u^2, \sigma_e^2, \rho)$  is the vector of variance component parameters for the model in Eq.(2).

Since we only take  $\beta_1$  fixed effect parameter of nested error regression model into consideration, Wald confidence region is a CI for  $\beta_1$ , an approximate  $(1-\alpha)$  Wald confidence interval written as  $\hat{\beta}_1 \pm z_{\frac{\alpha}{2}} SE(\hat{\beta}_1)$ , where  $z_{\frac{\alpha}{2}}$  is the  $(1-\frac{\alpha}{2})$  percentile of standard normal distribution.

### 2.2.1. t-Naive confidence interval method

t-Naive confidence interval is a Wald-type method used for small sample sizes under the situation that asymptotic distribution of maximum likelihood estimator is not converged. In practice, the variance of maximum likelihood estimator of fixed effects from FIM is usually obtained. Hence, when the variance of the estimator is unknown, it is ideal to use Student-t distribution percentiles for pivot percentiles explained above. Given the degrees of freedom for the pivot statistic, an approximate  $(1-\alpha)$  t-Naive confidence interval can be constructed using the formula  $\hat{\beta}_1 \pm t_{1-\frac{\alpha}{2}} SE(\hat{\beta}_1)$  where  $t_{1-\frac{\alpha}{2}}$  is a  $t$  distribution

cutoff [17]. Demidenko [1] indicated that one can compute confidence interval under this approach by taking variance estimator of fixed effect as the naive approximation and taking  $n-p$  degrees of freedom (where  $p$  is number of fixed effect terms in the model) for the approximating  $t$  distribution.

### 2.3. Parametric bootstrap method

Bootstrap is a statistical concept firstly introduced by Efron [18]. Diccio and Romano approached the bootstrap for constructing confidence interval [19, 20]. In order to obtain confidence interval and standard error of estimates for multilevel linear model, Goldstein [21] used bootstrap technique. Zeger and Liang [22] provided an example of parametric bootstrap using generalized estimating equations. Very broad bootstrap literature can be found by Chernick [23]. Efron's idea about bootstrap is that he considers the original bootstrap method as a maximum likelihood approach in nonparametric framework. If  $F$  was assumed to have a parametric form such as Gaussian distribution, estimator of  $F$  would be Gaussian distribution with the MLE of  $\mu$  and  $\sigma^2$  used for the respective unknown parameter of  $F$ . According to Fisher's MLE theory, resampling with replacement from parametric estimates of  $F$  provides bootstrap estimates that are also ML estimates.

In this paper, one of the most important parametric bootstrap confidence interval methods called as Bootstrap-t introduced by Efron in his monograph was used [24].

Among available bootstrap confidence interval methods, bootstrap-t usually provides the best balance between performance and cost when standard error of the estimate is convenient and accurate [10]. Therefore, this method was taken into consideration to obtain parametric bootstrap confidence interval for fixed effect parameter of nested error regression model. Two parametric bootstrap confidence intervals were constructed: one of them is based on Kenward & Roger MSE approximation [6] and the other is based on conventional MSE obtained by FIM to see the effect of MSEs on the coverage rates of the same method.

### 2.3.1. Bootstrap -t confidence interval method

Bootstrap-t interval [25] is derived from bootstrap estimate of the sampling distribution of the pivot  $\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}$ .  $q^* = (\hat{\beta}_1^* - \hat{\beta}_1) / SE(\hat{\beta}_1^*)$  denotes the bootstrap pivot approximation and  $q_\alpha^*$  denotes the  $\alpha$  cutoff for the bootstrap distribution of  $q^*$ . The probability approximation is shown below:

$$1-\alpha \approx P\left[q_{\alpha/2}^* < \left(\hat{\beta}_1^* - \hat{\beta}_1\right) / SE\left(\hat{\beta}_1^*\right) < q_{1-\alpha/2}^*\right] \\ \approx P\left[q_{\alpha/2}^* < \left(\hat{\beta}_1 - \beta_1\right) / SE\left(\hat{\beta}_1\right) < q_{1-\alpha/2}^*\right].$$

An approximate bootstrap-t confidence interval for  $\beta_1$  is given by the interval below:

$$\left[\hat{\beta}_1 - q_{1-\alpha/2}^* SE\left(\hat{\beta}_1^*\right), \hat{\beta}_1 - q_{\alpha/2}^* SE\left(\hat{\beta}_1^*\right)\right].$$

Because double bootstrap is quite time consuming to obtain standard error of  $\hat{\beta}_1$  as a second layer of bootstrap, two different analytic approximations of standard error were used. One of them is conventional standard error obtained by FIM explained in Section 2.1 and this bootstrap interval is called as BS-t in the simulation study. The other is adjusted estimator of the variance covariance matrix of the fixed effect estimators. This is known as Kenward- Roger approximation (BS-t-KR) and is less biased than the conventional estimator for small or moderate number of experimental units [6]. The confidence interval based on Kenward- Roger approximation of standard error [6] with bootstrap-t procedure is shown below:

### 3. Simulation Study

Confidence interval methods are compared in terms of coverage probability of true parameter value. The datasets generated from one-way random effect ANOVA model,  $Y_{ij} = \beta_0 + u_i + \varepsilon_{ij}$ , are fitted to nested error regression model,  $Y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij}$ .

Scores for a fixed predictor  $x_{ij}$  are simulated from normal distribution,  $N(0,100)$ . It is important to indicate that the datasets were generated independently of the  $x_{ij}$ , and the value of  $\beta_1$  in the generating model (one-way random effect ANOVA model) was taken as zero. For each dataset, three analytic confidence intervals for  $\beta_1$  at %95 confidence level were constructed based on Section 2.1 and 2.2. Wald confidence interval, maximum likelihood estimator of fixed effect and its standard error based on FIM were obtained by performing lme4 package [27] in R. All simulation procedures were implemented in R 3.1.1. Kenward- Roger MSE approximation was obtained by using pbkrtest package [26] in R for constructing parametric bootstrap CI (BS-t-KR). There can be several factors that are expected to affect coverage properties of interval estimation methods for the fixed effect parameter  $\beta_1$  in nested error regression model. The design of simulation study to compare the five types of interval methods given in Section 2.1, 2.2, 2.3 were determined. Simulation outline is given in the following:

The number of experimental unit or cluster was denoted by  $n$  and number of observation within each cluster or unit was denoted by  $m$ . In order to investigate the effects of these two factors, all the combinations of levels of  $n$  ( $n = 2, 4, 6, 8, 10, 12, 16$ ) and  $m$  ( $m = 3, 9, 15$ ) were taken into consideration.

**Table 1.** Parameter values selection summary for simulation design.

	$n = 2, 4, 6, 8, 10, 12, 16, 20$ $m = 3, 9, 15$		
	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
$\beta_0 \square U(-10,10)$	$u \square N(0,1)$	$u \square N(0,3)$	$u \square N(0,9)$
$x_{ij} \square N(0,100)$			
$\varepsilon \square N(0,3)$			

The values in the parameter vector  $(\sigma_u^2, \sigma_\varepsilon^2, \rho)$  were chosen differently to examine their effects on the coverage probability. The summaries of these values are given in Table 1.

### 3.1. Simulation algorithm

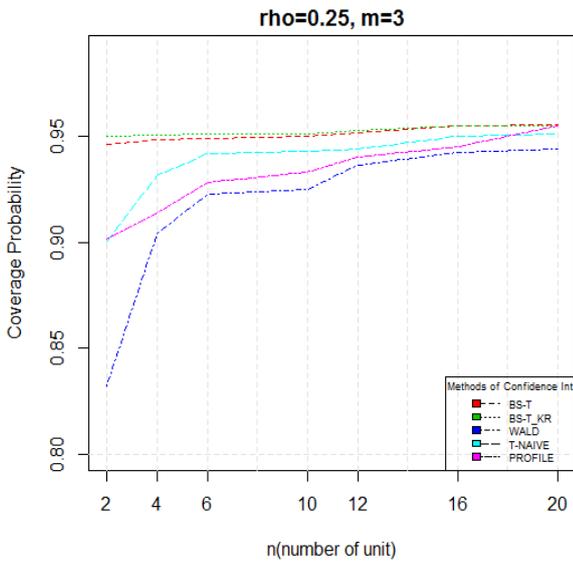
The two-sided %95 confidence interval methods described in Section 2.1, 2.2, 2.3 were examined. For each method, combination of  $(n, m, \rho)$ , the coverage probabilities were obtained by the following steps below:

1. Simulate a dataset for each combination of factors from the simulation design. Construct Wald-type, profile likelihood and parametric bootstrap confidence intervals given in Section 2.1, 2.2, 2.3.
2. Using the one-way random effect model, parametric bootstrap samples were generated. Using the generated samples, bootstrap-t, bootstrap-t based on Kenward-Roger MSE approximation confidence intervals given in Section 2.4 were constructed. The number of bootstrap samples generated in the simulation study was chosen as  $B = 2000$  as Efron and Tibshirani [25] suggested.
3. In step 2 and step 3, check the confidence intervals obtained if the true value of  $\beta_1$  was included, assign an indicator a value of 1; if not included, and then assign an indicator a value of 0.
4. Repeat Step 1 to 4 for a total of  $N = 2000$  times and compute the proportion of confidence intervals which contain the true value of  $\beta_1 = 0$ .

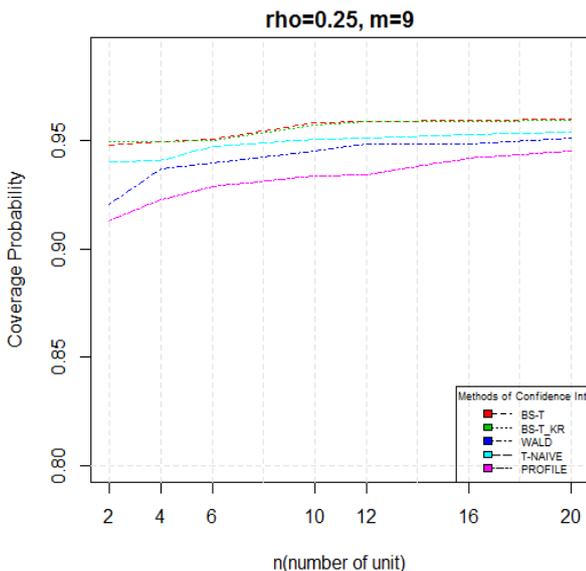
### 3.2. Simulation results

Coverage rates of confidence interval methods are summarized in Figure 1 to 3 for each combination of  $(n, m, \rho)$ .

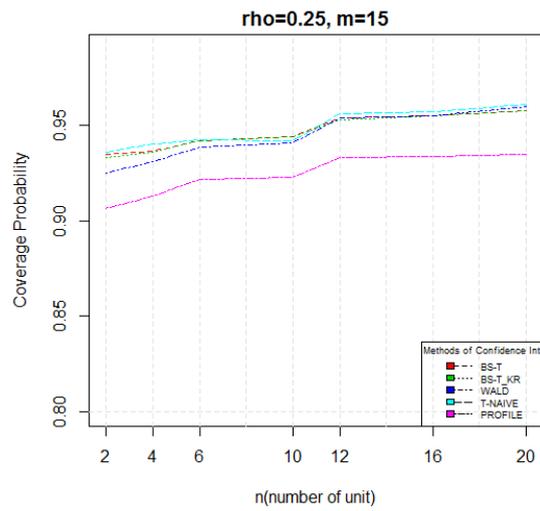
For the combination of  $\rho = 0.25$  and  $m = 3$  for all  $n$  values, it is observed that parametric bootstrap-based confidence intervals methods provide the highest coverage rate among five interval methods and close to nominal level of alpha. BS-t-KR provides the highest coverage rate for small and moderate  $n$  values as seen in Figures 1(a), 1(b), 1(c). For combination of  $\rho = 0.25$  and  $m = 9$  the same result is valid for all  $n$  values stated above. Especially for small  $n$  values, BS-t-KR has the highest coverage rate. For combination of  $\rho = 0.25$  and  $m = 15$  it can be said that t-naïve method provides usually better coverage rates among the other methods as seen in Figure 1(c).



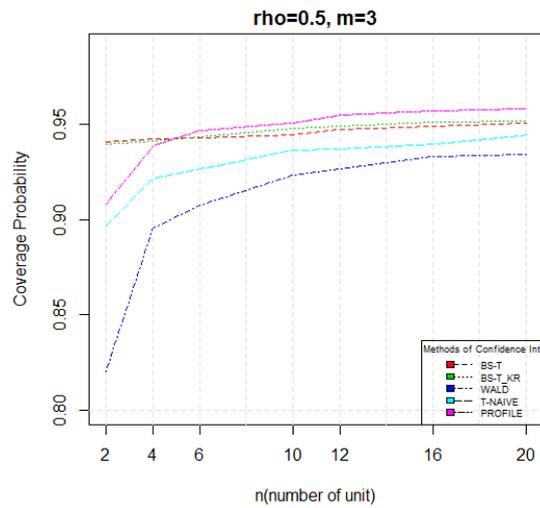
**Figure 1(a).** Coverage rate of confidence interval methods for the combination of  $m = 3$  &  $\rho = 0.25$  for different  $n$  values.



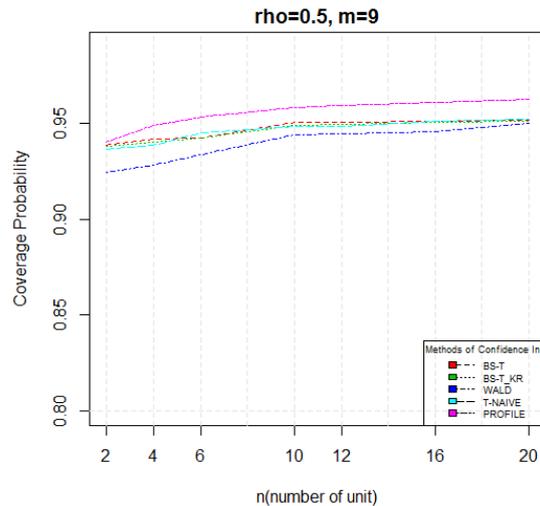
**Figure 1(b).** Coverage rate of confidence interval methods for the combination of  $m = 9$  &  $\rho = 0.25$  for different  $n$  values.



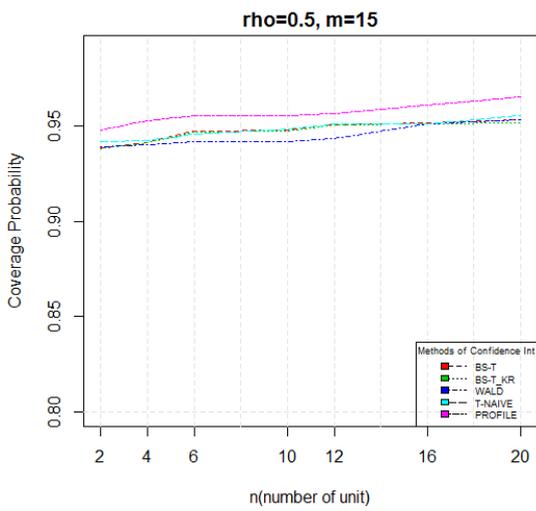
**Figure 1(c).** Coverage rate of confidence interval methods for the combination of  $m = 15$  &  $\rho = 0.25$  for different  $n$  values.



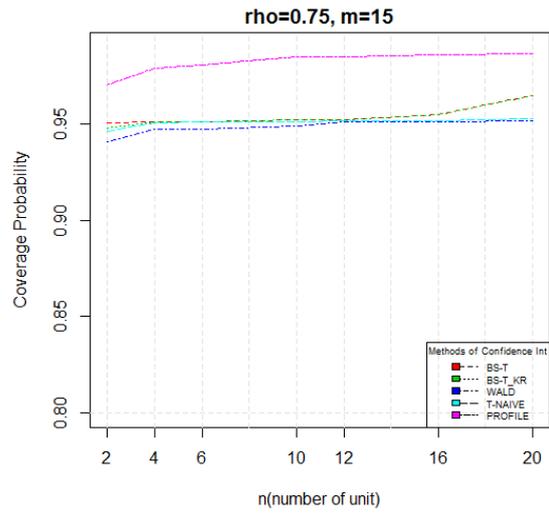
**Figure 2(a).** Coverage rate of confidence interval methods for the combination of  $m = 3$  &  $\rho = 0.5$  for different  $n$  values.



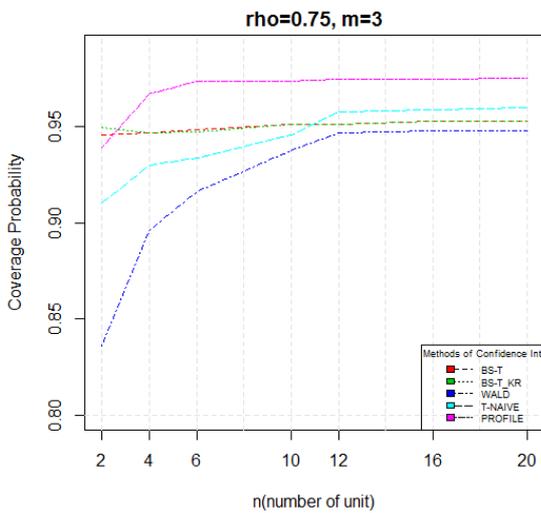
**Figure 2(b).** Coverage rate of confidence interval methods for the combination of  $m = 9$  &  $\rho = 0.5$  for different  $n$  values.



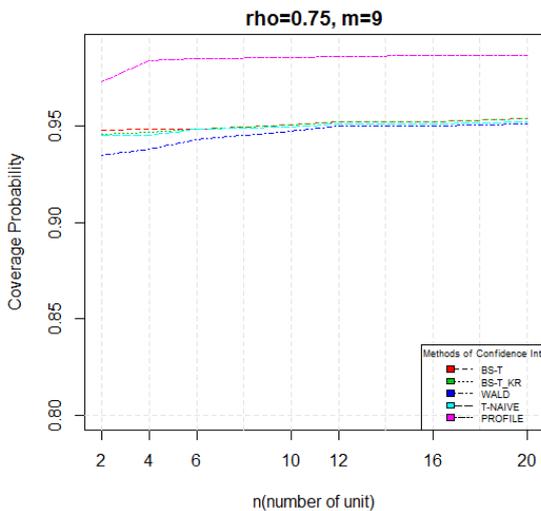
**Figure 2(c).** Coverage rate of confidence interval methods for the combination of  $m=15$  &  $\rho=0.5$  for different  $n$  values.



**Figure 3(c).** Coverage rate of confidence interval methods for the combination of  $m=15$  &  $\rho=0.75$  for different  $n$  values.



**Figure 3(a).** Coverage rate of confidence interval methods for the combination of  $m=3$  &  $\rho=0.75$  for different  $n$  values.



**Figure 3(b).** Coverage rate of confidence interval methods for the combination of  $m=9$  &  $\rho=0.75$  for different  $n$  values.

Parametric bootstrap-based CI methods provide the highest coverage rates for small sample size combinations  $n=2,4$  and  $m=3$  with  $\rho=0.5$ . Under the combination  $n=6,8,10,12,16,20$  and  $m=3$  with  $\rho=0.5$ , it can be reported that parametric bootstrap-based CI and PLCI methods provide similar results to each other, but PLCI method has slightly better coverage rates.

For the combination  $m=9,15$  of  $\rho=0.5$ , PLCI method yields slightly the highest coverage rates among all other methods

For the combination of  $\rho=0.75$  and  $m=3,9,15$  except for  $n=2$  and  $m=3$  combination, profile likelihood method provides the highest coverage rates among five methods. Bootstrap-based methods provide similar results to each other and nominal level of alpha. In addition, they provide better results than Wald-type intervals for small sample sizes as shown in Figure 3.

### 3. Discussion and Conclusion

Even with small sample sizes of experimental unit or cluster, the parametric bootstrap interval method demonstrates a crucial alternative to standard methods of inference for fixed effects in the mixed model especially for small intra-class correlation coefficient. The bootstrap-t method based on Kenward-Roger MSE approximation represents the best coverage rates for small experimental unit sizes despite of its dependency of standard error estimates.

When the intra-class correlation coefficient is increased to  $\rho=0.5,0.75$ , profile likelihood method generally provides the highest coverage rates for all combinations of  $n$  and  $m$ . For this reason, it is the

most conservative method among the others in terms of type 1 error. The highest coverage rates of profile likelihood method may be the result of wide length of confidence intervals. Since average expected length of confidence interval methods were not reported, it is not ideal to state that profile likelihood method among them as the best method without the knowledge of average expected lengths for moderate and strong intra-correlation coefficients.

As a suggestion and future direction of this study, investigating and comparing nonparametric and parametric bootstrap interval approaches in terms of coverage probability rate and average expected length for different linear mixed models especially when normality assumption of random effects of the model cannot be verified can be the main focus. It is known that parametric bootstrap is limited to the normality assumption of random effects in the linear mixed model. For this situation, finding a promising confidence interval method as an alternative to standard methods of inference will be the main attention for the authors of the paper.

Furthermore, a two-sided, 5% level test of hypothesis  $H_0: \beta_1 = 0$  can be carried out by constructing a  $1 - \alpha$  confidence interval for  $\beta_1$  and  $H_0$  is rejected if the interval does not include zero. Since the dataset of the simulation design is generated with  $\beta_1 = 0$ , false rejection rate of null hypothesis can be obtained by the proportion of confidence intervals that do not include zero, or by one minus the observed coverage probability rate of the confidence intervals, equivalently.

Among five interval methods, Wald-type methods are the fastest. The profile likelihood is a more time consuming method than Wald-type methods. However, the most computation time is required by bootstrap-based methods which are the only disadvantage of this method. However, they are the most accurate methods for especially small to moderate sample sizes of small intra-class correlation coefficient. Even though Wald-type methods are the fastest, for the small sample sizes they may not provide accurate results of coverage rates. Also, it can be stated that as the number of observation within unit and the number of experimental unit are increased, Wald-type methods improve their coverage rates for all simulation settings.

All combination factors have a quiet effect on the coverage probability of the methods and especially sample sizes have a great impact on coverage probability rate of the interval methods. Our study focused on confidence interval methods of fixed effect in nested error regression model under small samples. Similar results in a linear mixed model are expected; however, this is too early to generalize the comment obtained through this study for all situations of linear mixed model. In order to make

more generalized suggestions, the present study requires more careful attention and simulation as a future research. The goal of the next study is to present a method that performs well in a broad variety of settings and is easier to implement. The authors can supply a document containing R code to show and obtain these confidence interval methods reported in this article.

## References

- [1] Demidenko, E. 2004. Mixed Models: Theory and Applications. Wiley, 726p.
- [2] Eisenhart, C. 1947. The Assumptions Underlying the Analysis of Variance. *Biometrics*, 3(1), 1-21.
- [3] Patterson, H.D., Thompson, R. 1971. Recovery of Inter-block Information When Block Sizes are Unequal. *Biometrika*, 58(3), 545-554.
- [4] Searle, S.R., Casella, G., McCulloch, C.E., 2009. Variance Components (Vol. 391). John Wiley & Sons, 509p.
- [5] Kacker, R. N., Harville, D. A. 1984. Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association*, 79(388), 853-862.
- [6] Kenward, M. G. Roger, J. H. 1997. Small Sample Inference for Fixed Effects From Restricted Maximum Likelihood. *Biometrics*, 1997, 983-997.
- [7] Harville, D. A., Fenech, A. P. 1985. Confidence Intervals for a Variance Ratio, or for Heritability, in an Unbalanced Mixed Linear Model. *Biometrics*, 1985, 137-152.
- [8] Savin, A., 2005. Confidence Intervals for Common Mean in One-way Classification Model with Fixed Effects. *Measurement*, 2005, 15-19.
- [9] Hall, P., Maiti, T. 2006. On Parametric Bootstrap Methods for Small Area Prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2), 221-238.
- [10] Staggs, V. 2009. Parametric Bootstrap Interval Approach to Inference for Fixed Effects in the Mixed Linear Model, PhD Thesis, University of Kansas, 73 p, Lawrence.
- [11] Burch, B. D. 2011. Confidence Intervals for Variance Components in Unbalanced One-way Random Effects Model Using Non-normal Distributions. *Journal of Statistical Planning and Inference*, 141(12), 3793-3807.
- [12] Liu, J. 2013. Statistical Inference for Functions of the Parameters of a Linear Mixed Model, PhD Thesis, Iowa State University, 75p, Ames.
- [13] Neyman, J., Pearson, E. S. 1928. On the Use and Interpretation of Certain Test Criteria for

Purposes of Statistical Inference: Part I. *Biometrika*, 1928, 175-240.

- [14] Wald, A. 1943. Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54(3), 426-482.
- [15] Bates, D. M., Pinheiro, J. C. 1994. *Model Building for Nonlinear Mixed Effects Models*, University of Wisconsin, Department of Biostatistics.
- [16] Lehmann, E. L. 1999. *Elements of Large-Sample Theory*. Springer Science & Business Media, 645p.
- [17] Hulting, F. L., Harville, D. A. 1991. Some Bayesian and Non-Bayesian Procedures for the Analysis of Comparative Experiments and for Small-area Estimation: Computational Aspects, Frequentist Properties, and Relationships. *Journal of the American Statistical Association*, 86(415), 557-568.
- [18] Efron, B. 1992. *Bootstrap Methods: Another Look at the Jackknife*. Breakthroughs in Statistics. 1992, 569-593, New York.
- [19] DiCiccio, T. J., Romano, J. P. 1988. A Review of Bootstrap Confidence Intervals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 338-354.
- [20] DiCiccio, T. J., Romano, J. P. 1990. Nonparametric Confidence Limits by Resampling Methods and Least Favourable Families. *International Statistical Review/Revue Internationale de Statistique*, 1990, 59-76.
- [21] Goldstein, H. 2011. *Multilevel Statistical Models (Vol. 922)*. John Wiley & Sons, 358p.
- [22] Zeger, S. L., Liang, K. Y. 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 1986, 121-130.
- [23] Chernick, M. R. 2011. *Bootstrap Methods: A Guide for Practitioners and Researchers (Vol. 619)*. John Wiley & Sons, 369p.
- [24] Efron, B., Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans (Vol. 38)*. Philadelphia: Society for Industrial and Applied Mathematics, 92p.
- [25] Efron, B., Tibshirani, R. J. 1994. *An Introduction to the Bootstrap*. CRC press, 436p.
- [26] Halekoh, U., Højsgaard, S. 2014. A kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models—the R Package `pbkrtest`. *Journal of Statistical Software*, 59(9), 1-30.
- [27] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R Package Version, 1(7).