Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2020, 2020:65
https://doi.org/10.28919/cmbn/4955
ISSN: 2052-2541

FIRTH BIAS CORRECTION FOR ESTIMATING VARIANCE COMPONENTS OF LOGISTICS LINEAR MIXED MODEL USING PENALIZED QUASI LIKELIHOOD METHOD

RESTU ARISANTI[1,*], I. MADE SUMERTAJAYA[2], KHAIRIL ANWAR NOTODIPUTRO[2], INDAHWATI[2]

[1]Department of Statistics, Padjadjaran University, Bandung, Indonesia

[2]Department of Statistics, IPB University, Bogor, Indonesia

Abstract: Firth bias correction originally was applied to correct bias of the variance components estimator that obtained by the maximum likelihood method. Extensive research has shown that Firth bias correction is powerful to reduce bias for normal distributed response model. Questions have been raised about the use of Firth bias correction in binomial distributed response model which has under dispersion problem. The motivation of this study is giving contribution to exploring the Firth bias correction for binomial distributed response model. The binomial distributed response model which is estimated by the maximum likelihood method obtain an under-dispersion estimator. Therefore, the Penalized Quasi-Likelihood (PQL) is used as alternative numerical method to estimate the model. This paper aims to investigate whether the Firth method can reduce bias of the variance components using the PQL technique in longitudinal data.

Keywords: logistics linear mixed model; penalized quasi likelihood; Firth-adjusted penalized quasi likelihood; unadjusted penalized quasi likelihood; longitudinal data.

2010 AMS Subject Classification: 62J05.

*Corresponding author:
E-mail address: r.arisanti@unpad.ac.id
Received August 18, 2020

1. INTRODUCTION

In Generalized Linear Mixed Models (GLMMs), it is critical to estimate the random effects separated of fixed effects. Random effect parameters are known as the variance components [1]. Both effects are generally estimated by Maximum Likelihood Estimation (MLE). A number of studies have begun to examine the impact of MLE method in estimating the fixed effect and random effect in GLMMs. MLE is known as an asymptotic method that produces an unbiased parameter estimator when the sample size is large or going to infinity. However, there is evidence that for a small sample number, the estimation of MLE would be downward biased. The downward biased estimator produces an overdispersion estimator described by [2]. Bias estimator defined as the difference between an estimator's expected value and the true value of the parameter being estimated [3].

Discussed by [4], the bias of variance components seen clearly in case of binary data. A special case of GLMMs that has dichotomous response variable and involves fixed effects and random effects is known as Logistics Linear Mixed Model (LLMM). The LLMM usually used to model the heterogeneity among subjects and correlations in repeated observations [5].

According to [6], the likelihood function of GLMMs is analytically difficult to obtain the closed-form solutions. It happens because the random effects of likelihood function is not integrable. Moreover, the LLMM which is estimated by maximum likelihood method obtain an overdispersion estimator (Lin). This happens because the variance's value of dichotomous response variables is not the same as the variance assumptions. To overcome this problem, an alternative numerical method to estimate the model using Penalized Quasi Likelihood (PQL) suggested by [6]. In addition, [4] claims that the PQL in estimating the variance component can overcome heterogeneity of random effects from various sources of variance. Other than that, Capanu et.al. [7] points out that PQL able to accommodate complex correlation structures for dichotomous response variable model. In all the studies reviewed here, this paper attempts to show that the PQL technique can be used to estimate the variance components for the LLMM.

The estimator of variance components from PQL are downward biased has been proved by

[8]. Therefore, the bias of variance components needs to be reduced. Modifying the score function from the likelihood function is proposed to reduce bias by [9]. In the previous studies, [1] proved that the Firth method can reduce the bias of variance components of GLMMs by maximum likelihood estimation (MLE) method. Several studies have used Firth Method to reduce the bias, but none has explained how to reduce the bias in LLMM especially using the PQL techniques.

This paper aims to investigate whether the Firth method can reduce bias for the LLMM using the PQL technique in longitudinal data. The study is conducted in the form of three issues. This paper begins by the analytical studies that carried out to develop the iterative procedures to estimate the corrected bias. The second is simulation studies to evaluate the performance of Firth method. The third issue is the application of the longitudinal study using SUSENAS data that collected annually.

This paper is organized as follows. Section 2 presents the Logistics Linear Mixed Model. In section 3, we discuss the Analytical Studies of Firth Bias Correction Method for Logistics Linear Mixed Model Via Penalized Quasi Likelihood. Section 4 presents the Bias Simulation Studies. In the section 5, we describe the Illustration on longitudinal Study. Finally, the Conclusion is presented in section 6.

2. LOGISTIC LINEAR MIXED MODEL (LLMM)

Mixed model involves fixed effect and random effect in the model. if we assume the response variable has the normal distribution, it is known as Linear Mixed Model (LLM). The common form of linear mixed model as follow:

(1)
$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$$

Where $\boldsymbol{y}$ is N x 1 column vector, the response variable; $\boldsymbol{X}$ is a N x p matrix of the p predictor variables; $\boldsymbol{\beta}$ is a p x 1 column vector of the fixed-effects regression coefficients; $\boldsymbol{Z}$ is the N x q design matrix for the q random effects; $\boldsymbol{b}$ is a q x 1 vector of the random effects; and $\boldsymbol{\varepsilon}$ is a N x 1 column vector of the residuals, that part of $\boldsymbol{y}$ that is not explained by the model. $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma})$, $\boldsymbol{b} \sim N(0, \mathbf{D})$. $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$ are independent. The variance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{D}$ are parameterized by an unknown variance component parameter $\boldsymbol{\theta}$. $\mathbf{D} = \mathrm{diag}(\sigma_{b_j}^2)$

However, when the condition of response variable is non normal, then the model known as Generalized Linear Mixed Models (GLMMs). The means of the response variable $\boldsymbol{y}$ conditional on random effects $\boldsymbol{b}$ as follow

(2) $$E(\boldsymbol{y}|\boldsymbol{b}) = \boldsymbol{X\beta} + \boldsymbol{Zb}$$

Where $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V} = \boldsymbol{\Sigma} + \mathbf{Z\Omega Z'})$. The matrix form of GLMMs can be written as:

(3) $$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zb}$$

This model has a distribution of $\boldsymbol{y}$ population that depend on $\boldsymbol{\eta}$. The special case of GLMMs where the response variable has a binomial distribution is called Logistics Linear Mixed Model (LLMM). LLMM can be written as follow:

(4) $$log\left(\frac{\boldsymbol{p}}{1-\boldsymbol{p}}\right) = \boldsymbol{X\beta} + \boldsymbol{Zb}$$

Log likelihood function of binomial distribution can be written as

(5) $$logL(\boldsymbol{y};\boldsymbol{p}) = log\binom{n}{\boldsymbol{y}} + \boldsymbol{y}\log\left(\frac{\boldsymbol{p}}{1-\boldsymbol{p}}\right) + n\log(1-\boldsymbol{p})$$

2.1 BEST LINEAR UNBIASED PREDICTOR OF LLMM VIA PENALIZED QUASI LIKELIHOOD

As mentioned before, this study uses longitudinal data design. Let $\boldsymbol{y} = \{\boldsymbol{y}_{kti}\}$ defined as the vector if sample values of the variable Y. The subscripts $k = 1, 2, \cdots, K; t = 1,2, \cdots, T; i = 1, 2, \cdots, n_{kt}$ denotes area, time and sample unit respectively. The Best Linear Unbiased Predictor (BLUP) procedure consists of maximizing the sum of two component loglikelihoods. Let l_1 is the loglikelihood function of the binomial vector $\boldsymbol{y}$ conditional on fixed $\boldsymbol{\beta}$ and l_2 is the log of probability density function of $\boldsymbol{b}$. $l = l_1 + l_2$ represents the loglikelihood based on the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{b}$. From equation (**3**) and (**4**), we find

(6) $$\boldsymbol{p}_{kti} = \frac{e^{\boldsymbol{\eta}_{kti}}}{1+e^{\boldsymbol{\eta}_{kti}}}$$

then

(7) $$l_1 = const. + \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{I}[\boldsymbol{y}_{kti}\,\boldsymbol{\eta}_{kti} - n_{kti}\log(1+e^{\boldsymbol{\eta}_{kti}})]$$

$$l_2 = -(0.5)[const. + \ln|\boldsymbol{D}| + \boldsymbol{b}'\boldsymbol{D}^{-1}\boldsymbol{b}]$$

whereas $\boldsymbol{D}$ is the variance covariance matrix of $\boldsymbol{b}$.

To estimate $\boldsymbol{\beta}$ and $\boldsymbol{b}$, different method has been proposed by Saei and McGilchrist [10]. This method is one of the practical ways of estimating $\boldsymbol{\beta}$ and $\boldsymbol{b}$ because it involves the log-likelihood function directly. In this paper, this method is extended to be applied in LLMM. The method requires the first and second derivatives of l_1 with respect to $\boldsymbol{\eta}, \boldsymbol{\beta},$ and $\boldsymbol{b}$ and l_2 with respect to $\boldsymbol{\beta}$ and $\boldsymbol{b}$ as follows:

$$\frac{\partial l_1}{\partial \boldsymbol{\eta}} = \boldsymbol{y}_{kti} - \frac{n_{kti} e^{\boldsymbol{\eta}_{kti}}}{e^{\boldsymbol{\eta}_{kti}} + 1} = \boldsymbol{y}_{kti} - n_{kti}\boldsymbol{p}_{kti}; \ \frac{\partial^2 l_1}{\partial \boldsymbol{\eta}_{kti} \partial \boldsymbol{\eta}'_{kti}} = \text{diag}[-n_{kti}\boldsymbol{p}_{kti}(1 - \boldsymbol{p}_{kti})]$$

$$\frac{\partial l_1}{\partial \boldsymbol{\beta}} = \frac{\partial l_1}{\partial \boldsymbol{\eta}} \cdot \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}} = (\boldsymbol{y}_{kti} - n_{kti}\boldsymbol{p}_{kti})\boldsymbol{X}_{kti}; \frac{\partial l_1}{\partial \boldsymbol{b}} = \frac{\partial l_1}{\partial \boldsymbol{\eta}} \cdot \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{b}} = (\boldsymbol{y}_{kti} - n_{kti}\boldsymbol{p}_{kti})\boldsymbol{Z}_{kti}$$

since l_2 does not consist of parameter $\boldsymbol{\beta}$, then derivatives with respect to $\boldsymbol{\beta}$ are equal to zero

$$\frac{\partial l_2}{\partial \boldsymbol{b}} = \boldsymbol{D}^{-1}\boldsymbol{b}; \ \frac{\partial^2 l_2}{\partial \boldsymbol{b} \partial \boldsymbol{b}'} = \boldsymbol{D}^{-1}\boldsymbol{b}\boldsymbol{b}' = \boldsymbol{D}^{-1}$$

Assumed **H** is a Hessian matrix which is known as the derivative matrix of the log likelihood function. Let $\boldsymbol{V}$ is a minus of Hessian matrix, then **V** can be written as follows:

$$\boldsymbol{V} = \begin{pmatrix} -\left(\frac{\partial^2 l_1}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} + \frac{\partial^2 l_2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right) & -\left(\frac{\partial^2 l_1}{\partial \boldsymbol{\beta} \partial \boldsymbol{b}'} + \frac{\partial^2 l_2}{\partial \boldsymbol{\beta} \partial \boldsymbol{b}'}\right) \\ -\left(\frac{\partial^2 l_1}{\partial \boldsymbol{b} \partial \boldsymbol{\beta}'} + \frac{\partial^2 l_2}{\partial \boldsymbol{b} \partial \boldsymbol{\beta}'}\right) & -\frac{\partial^2 l_1}{\partial \boldsymbol{b} \partial \boldsymbol{b}'} + \frac{\partial^2 l_2}{\partial \boldsymbol{b} \partial \boldsymbol{b}'} \end{pmatrix}$$

then

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{bmatrix} \left(\frac{-\partial^2 l_1}{\boldsymbol{\eta}_{kti} \boldsymbol{\eta}'_{kti}}\right) [\boldsymbol{X} \quad \boldsymbol{Z}] + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{bmatrix} \quad \text{or}$$

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{bmatrix} [\text{diag}[n_{kti}\boldsymbol{p}_{kti}(1 - \boldsymbol{p}_{kti})]] [\boldsymbol{X} \quad \boldsymbol{Z}] + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{bmatrix}$$

If $\boldsymbol{V}$ is evaluated at $\boldsymbol{\beta}_0$ and $\boldsymbol{b}_0$, then the procedure for estimating $\boldsymbol{\beta}$ and $\boldsymbol{b}$ is

(8) $$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{b}_0 \end{bmatrix} + \boldsymbol{V}^{-1} \begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{bmatrix} \left(\frac{\partial l_1}{\eta_{kti}}\right) - \boldsymbol{V}^{-1} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{D}^{-1}\boldsymbol{b}_0 \end{pmatrix}$$

Let $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}$ and $\boldsymbol{V}^{-1} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix}$ as the partitioning of the matrix $\boldsymbol{V}$.

Define $\boldsymbol{P} = -(\partial^2 l / \boldsymbol{\eta}_{kti} \boldsymbol{\eta}'_{kti})$ and $\boldsymbol{Q}^* = \boldsymbol{Q}^*_{jj} = (\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1}$. Then the variance components of $\boldsymbol{b}$ can be defined as

(9) $$\widehat{\boldsymbol{\sigma}}^2_{b_{(j)ML}} = r_j^{-1}\left(tr\left(\boldsymbol{Q}^*_{jj}\boldsymbol{D}_j^{-1}\right) + \sigma^{-2}\boldsymbol{b}'_j\boldsymbol{D}_j^{-1}\boldsymbol{b}_j\right)$$

where the r_j is the rank of the matrix $\boldsymbol{Z}_j$.

The variance components of $\boldsymbol{b}$ also can be obtained using the restricted estimation maximum likelihood (REML) method [11]. For the REML estimator's method, the $\boldsymbol{Q}^*$ using the $\boldsymbol{Q}_{22}$ submatrix of the $\boldsymbol{Q}^*$ matrix. This matrix can be expressly determined by utilizing the formula as below:

(10) $$\boldsymbol{Q}_{22} = \boldsymbol{Q}^* + \boldsymbol{Q}^*\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{X}\boldsymbol{Q}_{11}\boldsymbol{X}'\boldsymbol{B}'\boldsymbol{Z}\boldsymbol{Q}^*$$

while $\boldsymbol{Q}_{11} = (\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{B}\boldsymbol{Z}\boldsymbol{Q}^*\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{X})^{-1}$. Therefore, the variance components of $\boldsymbol{b}$ that using REML method can be written as

(11) $$\hat{\boldsymbol{\sigma}}^2_{b_{(j)REML}} = r_j^{-1}\left(tr\left(\boldsymbol{Q}_{22}\boldsymbol{D}_j^{-1}\right) + \boldsymbol{\sigma}^{-2}\boldsymbol{b}'_j\boldsymbol{D}_j^{-1}\boldsymbol{b}_j\right)$$

where the r_j is the rank of the matrix $\boldsymbol{Z}_j$.

3. ANALYTICAL STUDIES OF FIRTH BIAS CORRECTION METHOD FOR LLMM VIA PENALIZED QUASI LIKELIHOOD

The Basic idea of the Firth method to reduce bias is substituting the smaller bias to the score function (Firth 1993). Here is the illustration

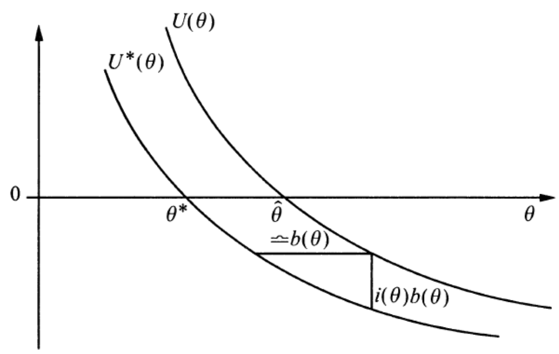

FIGURE 1. **Modification of Score Function**

Based on Figure-1 above, if $\hat{\boldsymbol{\theta}}$ is the subject of positive bias *b(θ)*, then the score function is shifted at each value of ***θ*** equal to *i(θ)**b**(θ)*, where *-i (θ) = U '(θ)* is local gradient. The modified score function is

(12) $$U^*(\boldsymbol{\theta}) = U(\boldsymbol{\theta}) - i(\boldsymbol{\theta})b(\boldsymbol{\theta})$$

So that the modified estimator is obtained $\hat{\boldsymbol{\theta}}$ with the solution $U^*(\boldsymbol{\theta})$= 0, where *i* (**θ**) is the Fisher information matrix. Firth's Method is using Jeffrey prior penalty to reduce bias. Common form of modified score function for LLMM can be written as:

$$U^*(\boldsymbol{p}) = U(\boldsymbol{p}) - A_r(\boldsymbol{p})$$
$$A_r(\boldsymbol{p}) = \frac{1}{2} tr \left\{ i^{-1} \left(\frac{\partial i}{\partial p_r} \right) \right\}$$
$$= \frac{\partial i}{\partial p_r} \left\{ \frac{1}{2} \; log \; [i(\boldsymbol{p})] \right\}$$

because $U_r^* = U_r + A_r = 0$, then the location stationer point is

(13) $$l^*(\boldsymbol{p}) = l(\boldsymbol{p}) + \frac{1}{2} \; log \; [i(\boldsymbol{p})]$$

or similar with penalized likelihood functions

$$L^*(\boldsymbol{p}) = L(\boldsymbol{p}) \; [i(\boldsymbol{p})]^{1/2}$$

Penalty function $|i(\boldsymbol{p})|^{1/2}$ is known as Jeffrey Invariant prior.

To modified the loglikelihood function, we begin from binomial distribution as the basic distribution of the LLMM. The likelihood function of binomial distribution as follow

$$L(\boldsymbol{p}|n, \boldsymbol{y}) = \binom{n}{\boldsymbol{y}} \boldsymbol{p}^{\boldsymbol{y}} (1 - \boldsymbol{p})^{n - \boldsymbol{y}}$$

the loglikelihood function is

$$l(\boldsymbol{p}) = \boldsymbol{y} \; log \left(\frac{\boldsymbol{p}}{1 - \boldsymbol{p}} \right) + n \log(1 - \boldsymbol{p}) + log \binom{n}{\boldsymbol{y}}$$

Then we obtain Fisher information matrix

(14) $$i(\boldsymbol{p}) = -E_p \left(\frac{\partial^2 l}{\partial \boldsymbol{p}^2} \right) = \frac{1}{\boldsymbol{p}(1 - \boldsymbol{p})}$$

Based on (**13**) and (**14**) then the penalized likelihood function which is known as Firth penalty:

(15) $$\frac{1}{2} log[i(\boldsymbol{p})] = \frac{1}{2} log \; \left(\frac{1}{\boldsymbol{p}(1 - \boldsymbol{p})} \right)$$
$$= \frac{1}{2} log \left(\frac{(1 + e^{\boldsymbol{\eta}})^2}{e^{\boldsymbol{\eta}}} \right)$$

Let $\boldsymbol{y} = \{\boldsymbol{y}_{kti}\}$ defined as the vector if sample values of the variable $\mathbf{Y}$. The subscripts $k = 1, 2, \cdots, K; t = 1,2, \cdots, T; i = 1, 2, \cdots, n_{kt}$ denotes area, time and sample unit respectively. Let $l^* = l_1 \; + l_2 +$ penalty represents the modification of loglikelihood based on the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{b}$. Let l_1^* is the loglikelihood function of the binomial vector $\boldsymbol{y}$ conditional on fixed $\boldsymbol{\beta}$ and penalized likelihood function of Firth method, $l_2$is the log of probability density function $\boldsymbol{b}$. Then the modification of loglikelihood function for binomial distribution is

$$l_1^* = const. + \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{I}[(\boldsymbol{y}_{kti} - 0.5)\,\boldsymbol{\eta}_{kti} - (1 - n_{kti})\log(1 + e^{\boldsymbol{\eta}_{kti}})]$$

$$l_2 = -(0.5)[const. + \ln|\boldsymbol{\Omega}| + \boldsymbol{b}'\mathbf{D}^{-1}\boldsymbol{b}]$$

The matrix $\boldsymbol{V}$ is the matrix of second-order derivatives

$$\boldsymbol{V} = \begin{pmatrix} -\left(\frac{\partial^2 l_1^*}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} + \frac{\partial^2 l_2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) & -\left(\frac{\partial^2 l_1^*}{\partial\boldsymbol{\beta}\partial\boldsymbol{b}'} + \frac{\partial^2 l_2}{\partial\boldsymbol{\beta}\partial\boldsymbol{b}'}\right) \\ -\left(\frac{\partial^2 l_1^*}{\partial\boldsymbol{b}\partial\boldsymbol{\beta}'} + \frac{\partial^2 l_2}{\partial\boldsymbol{b}\partial\boldsymbol{\beta}'}\right) & -\left(\frac{\partial^2 l_1^*}{\partial\boldsymbol{b}\partial\boldsymbol{b}'} + \frac{\partial^2 l_2}{\partial\boldsymbol{b}\partial\boldsymbol{b}'}\right) \end{pmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}\left(\frac{-\partial^2 l_1^*}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right)[\boldsymbol{X} \quad \boldsymbol{Z}] + \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}^{-1}\end{bmatrix} \quad \text{or}$$

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}[\text{diag}[(n_{kti} - 1)\boldsymbol{p}_{kti}/(1 + e^{\boldsymbol{\eta}_{kti}})]][\boldsymbol{X} \quad \boldsymbol{Z}] + \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}^{-1}\end{bmatrix}$$

If $\boldsymbol{V}$ is evaluated at $\boldsymbol{\beta}_0$ and $\boldsymbol{b}_0$, then the procedure for estimating $\boldsymbol{\beta}$ and $\boldsymbol{b}$ is

(16)
$$\begin{bmatrix}\widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\beta}_0 \\ \boldsymbol{b}_0\end{bmatrix} + \boldsymbol{V}^{-1}\begin{bmatrix}\boldsymbol{X}' \\ \boldsymbol{Z}'\end{bmatrix}\left(\frac{\partial l_1^*}{\partial\boldsymbol{\eta}}|_{\boldsymbol{\beta}_0,\boldsymbol{b}_0}\right) - \boldsymbol{V}^{-1}\begin{pmatrix}\mathbf{0} \\ \boldsymbol{D}^{-1}\boldsymbol{b}\end{pmatrix}$$

Let $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22}\end{bmatrix}$ and $\boldsymbol{V}^{-1} = \begin{bmatrix}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22}\end{bmatrix}$ as the partitioning of the matrix $\boldsymbol{V}$. Define $\boldsymbol{P} = -(\partial^2 l/\boldsymbol{\eta}_{kti}\boldsymbol{\eta}_{kti}')$ and $\boldsymbol{Q}^* = \boldsymbol{Q}_{jj}^* = (\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1}$. Then the variance components of $\boldsymbol{b}$ can be defined as

(17)
$$\widehat{\boldsymbol{\sigma}}_{b_{(j)ML}}^2 = r_j^{-1}(tr(\boldsymbol{Q}_{jj}^*\boldsymbol{\Omega}_j^{-1}) + \sigma^{-2}\boldsymbol{b}'_j\boldsymbol{D}_j^{-1}\boldsymbol{b}_j)$$

where the r_j is the rank of the matrix $\boldsymbol{Z}_j$.

The variance components of $\boldsymbol{b}$ also can be obtained using the restricted estimation maximum likelihood (REML) method [11]. For the REML estimator's method, the $\boldsymbol{Q}^*$ using the $\boldsymbol{Q}_{22}$ submatrix of the $\boldsymbol{Q}^*$ matrix. This matrix can be expressly determined by utilizing the formula as below:

(18)
$$\boldsymbol{Q}_{22} = \boldsymbol{Q}^* + \boldsymbol{Q}^*\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{X}\boldsymbol{Q}_{11}\boldsymbol{X}'\boldsymbol{B}'\boldsymbol{Z}\boldsymbol{Q}^*$$

while $\boldsymbol{Q}_{11} = (\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{B}\boldsymbol{Z}\boldsymbol{Q}^*\boldsymbol{Z}'\boldsymbol{B}\boldsymbol{X})^{-1}$. Therefore, the variance components of $\boldsymbol{b}$ that using REML method can be written as

(19)
$$\widehat{\boldsymbol{\sigma}}_{b_{(j)REML}}^2 = r_j^{-1}(tr(\boldsymbol{Q}_{22}\boldsymbol{D}_j^{-1}) + \sigma^{-2}\boldsymbol{b}'_j\boldsymbol{D}_j^{-1}\boldsymbol{b}_j)$$

where the r_j is the rank of the matrix $\boldsymbol{Z}_j$.

4. BIAS SIMULATION STUDIES

The simulations conducted to determine the behavior of the Firth-adjusted PQL (Firth method which is applied to the PQL). The purpose of the simulation is to assess and compare the performance of the Firth method in reducing bias of variance components. Specifically, the data generation model can be written as follows:

$$log\left(\frac{\boldsymbol{p}_{kti}}{1-\boldsymbol{p}_{kti}}\right) = \boldsymbol{x}'_{kti}\boldsymbol{\beta} + \boldsymbol{b}_{0k} + \boldsymbol{b}_{1t}$$

$$\begin{bmatrix} b_{0k} \\ b_{1t} \end{bmatrix} \sim MVN\left(\boldsymbol{0}, \begin{bmatrix} \sigma^2_{b_{0k}} & 0 \\ 0 & \sigma^2_{b_{1t}} \end{bmatrix}\right)$$

where $\log\left(\frac{\boldsymbol{p}_{kti}}{1-\boldsymbol{p}_{kti}}\right)$ is the canonical parameter for the binomial distribution for linear models. $\boldsymbol{p}_{ikt}$ is the probability that $y_{kti} = 1$ for the *k*th area, at the *t*th time and *i*th sample unit.

The estimation of the binomial parameter is interesting to be a research question. For example, such as estimating the proportion of an individual population with a particular scope, that the researcher randomizes the location where the individual lives. The comparison between unadjusted PQL method and the Firth-adjusted PQL method can be shown from the two tables and figures below.

TABLE 1. Mean of random effects estimates of Firth-adjusted PQL and unadjusted PQL Methods

True Variances $(V(\hat{b}_{0k}), V(\hat{b}_{1t}))$	Unadjusted PQL		Firth-Adjusted PQL	
	$\hat{\sigma}^2_{b_{0k}}$ Mean est. (rmse)	$\hat{\sigma}^2_{b_{1t}}$ Mean est. (rmse)	$\hat{\sigma}^2_{b_{0k}}$ Mean est. (rmse)	$\hat{\sigma}^2_{b_{1t}}$ Mean est. (rmse)
(1,2)	0.42476 (0.57526)	1.51092 (1.48962)	0.43014 (0.56985)	1.54077 (1.45954)
(1,4)	0.58698 (0.413)	2.7203 (1.28)	0.5888 (0.4111)	3.7432 (1.257)
(4,4)	2.0777 (1.922)	3.1684 (1.833)	3.13257 (1.885)	3.2465 (1.773)

Table 1 compares the results which is obtained from the analysis of variance components from unadjusted PQL and Firth-adjusted PQL. One interesting finding is for the homogenous variances, the variance components in the Firth-adjusted PQL obtain the variances that are close to the true variances. Similarly, for the heterogenous variances, it can be seen from the data in the results of this study in Table 1 indicate that the variance components in Firth-adjusted PQL obtain better variances than unadjusted PQL.

TABLE 2. Biases of variance components estimates of Firth-adjusted PQL and unadjusted PQL Methods

True Variances	Unadjusted PQL		Firth-Adjusted PQL	
$(V(\hat{b}_{0k}), V(\hat{b}_{1t}))$	$\hat{\sigma}^2_{b_{0k}}$ Mean bias est.	$\hat{\sigma}^2_{b_{1t}}$ Mean bias est.	$\hat{\sigma}^2_{b_{0k}}$ Mean bias est.	$\hat{\sigma}^2_{b_{1t}}$ Mean bias est.
(1,2)	0.57523	1.48907	0.56985	1.45923
(1,4)	0.41301	3.27969	0.41117	3.25679
(4,4)	2.9222	2.83153	2.86743	2.75348

Table 2 shows the biases of variance component estimates of the unadjusted PQL and the Firth-adjusted PQL. These results reflect those the result from Table 1 which also found that there is a significant difference between the two conditions. The bias of variance components is shown that Firth-adjusted PQL obtain lower biases of variance components than unadjusted PQL method. The comparison of the result from the two method is clearly seen in the figures below.

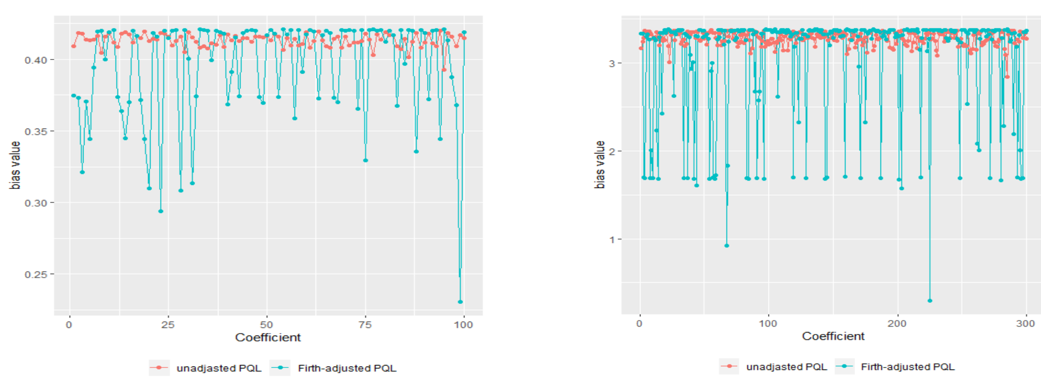

FIGURE 2. Comparison of Variance Estimation's bias for $b_{0k} \sim N(0,1)$ and $b_{1t} \sim N(0,4)$

($\hat{\sigma}^2_{b_{0k}}$ on the left side and $\hat{\sigma}^2_{b_{1t}}$on the right side)

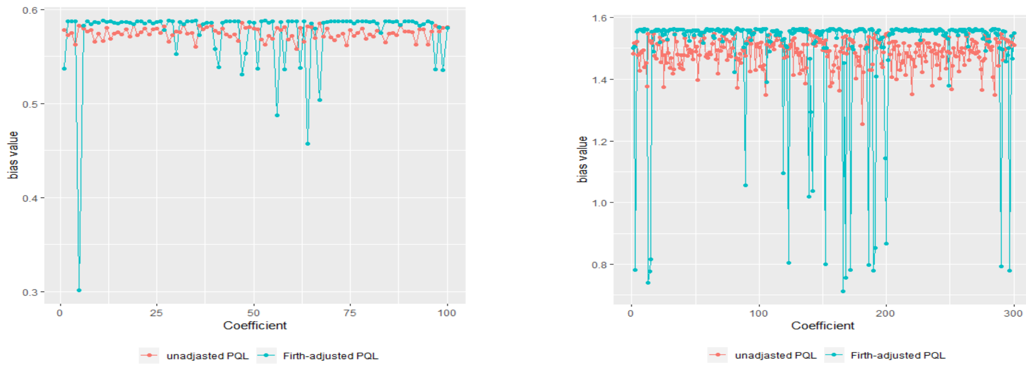

FIGURE 3. Comparison of Variance Estimation's bias for $b_{0k}{\sim}N(0{,}1)$ and $b_{1t}{\sim}N(0{,}2)$

($\hat{\sigma}^2_{b_{0k}}$ on the left side and $\hat{\sigma}^2_{b_{1t}}$on the right side)

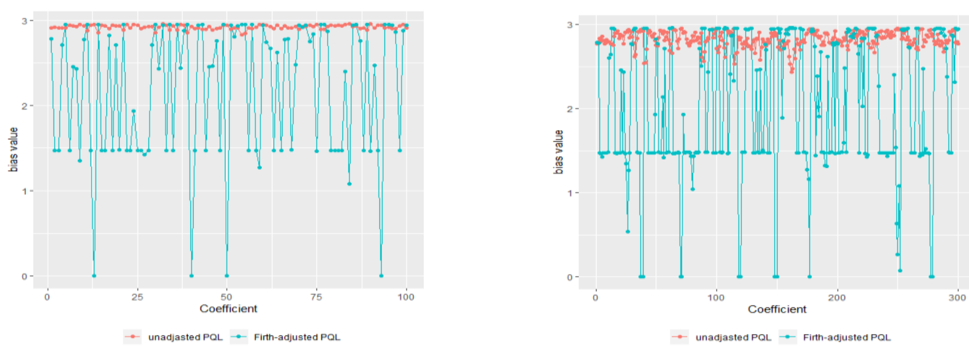

FIGURE 4. Comparison of Variance Estimation's bias for $b_1{\sim}N(0{,}4)$ and $b_2{\sim}N(0{,}4)$

($\hat{\sigma}^2_{b_{0k}}$ on the left side and $\hat{\sigma}^2_{b_{1t}}$on the right side)

In summary, comparing the two methods, it can be seen that the firth-adjusted PQL having less biases of variance component estimates leads to have better random effect's variability estimates.

5. ILLUSTRATION ON LONGITUDINAL STUDY

To make this more concrete, let's consider the illustration from the poverty dataset. The Poverty is still one of the complicated problems in every country, especially for developing countries like Indonesia. To measure poverty, the Statistics of Indonesia (BPS) uses the concept of ability to meet the basic needs (basic needs approach). With this approach, poverty is seen as an inability on the economic side to meet the basic needs of food and non-food measured from the expenditure side. So, the poor population means the population that has an average monthly per capita expenditure under the poverty line.

The analysis of the illustration is focused on estimating using LLMM via PQL. We assumed that y_{ikt}, whether poor or not for the *k*th household on the *i*th block and *t*th time, was (conditionally) binomial-distributed with mean μ_{ij}^{b}. In equation form, the model can be written as follow:

$$log\left(\frac{\boldsymbol{p}_{kti}}{1-\boldsymbol{p}_{kti}}\right) = \boldsymbol{x}'_{kti}\boldsymbol{\beta} + \boldsymbol{b}_{0k} + \boldsymbol{b}_{1t}$$

$$\begin{bmatrix} b_{0k} \\ b_{1t} \end{bmatrix} \sim MVN\left(\boldsymbol{0}, \begin{bmatrix} \sigma_{b_{0k}}^2 & 0 \\ 0 & \sigma_{b_{1t}}^2 \end{bmatrix}\right)$$

$\boldsymbol{p}_{kti}$ is the probability of the *i*th household in poverty. β_1 and β_2 are the coefficient of the fixed effects. To illustrate how the Firth method, reduce the bias of variance components in the LLMM, we will use an illustration of longitudinal data (SUSENAS) from the Statistics of Indonesia (BPS). Data is taken from 2011-2013.

The responses in longitudinal data are usually correlated. The head of household as respondents was taken from 38 cities and districts of East Java province. There are 2910 census blocks with 10 households for each census block. The response variable is determination whether the household is categorized as poor or not. Further, suppose we had two fixed effects predictors, the household predictors are floor area of the house (X_1) and number of the household members (X_2).

The random effects for each household's block (b_{0k}) and for each household's time in block (b_{1t}) are including into the model. Assumed that block household's random effects are i.i.d draws from $N\sim(\boldsymbol{0}, \widehat{\boldsymbol{\sigma}}_{b_{0k}}^2)$, where $\widehat{\boldsymbol{\sigma}}_{b_{0k}}^2$ is an unknown parameter to be estimated. Similarly, for the household's time in block random effects are i.i.d. draws from $N\sim(\boldsymbol{0}, \widehat{\boldsymbol{\sigma}}_{b_{1t}}^2)$, where $\widehat{\boldsymbol{\sigma}}_{b_{1t}}^2$ is an unknown parameter to be estimated. Finally, the block and time random effects are assumed independent of one another. The reason in choosing the random effects in this study is because we expect that the variation within block may be correlated. There are many reasons why this could be. For example, the determination of the block is taken from the same region that have the same poverty line, such that within a block, the households are more homogeneous than they are between blocks.

TABLE 3. Summary result of LLMM

Parameter	Unadjusted PQL		Firth adjusted PQL	
	Estimate	Odds Ratio Estimate	Estimate	Odds Ratio Estimate
Intercept	0.0214 (< 0.001)	1.021	0.1684 (< 0.001)	1.183
Floor Area of the house	-0.8223 (< 0.001)	0.439	-0.3994 (< 0.001)	0.671
Number of the Household Members	1.1397 (< 0.001)	3.125	0.5663 (< 0.001)	1.761
$\hat{\sigma}^2_{b_{0k}}$	0.283		0.4201	
$\hat{\sigma}^2_{1t}$	0.335		0.4759	

The table above illustrates the result of unadjusted PQL and Firth adjusted PQL method. Firth adjusted PQL explain that for floor area of the house, a one-unit increase in floor area of the house is associated with 0.671 unit decrease in the expected log odds of poverty. Similarly, the household who are has bigger number of household members are expected to have 1.761 higher log odds of being in poverty than household who has smaller number of the household members.

Turning to the odds ratio here is the conditional odds ratio for the household with the floor area of the house and number of the household members constant as well as for the household with either the same block, or blocks with identical random effects. When there is large variability between blocks, the relative impact of the fixed effects may be small.

To proof whether there are the differences between the two variances from the two kinds of random effects, there are two hypotheses:

Hypothesis 1:

H_0: $\hat{\sigma}^2_{b_{0k\,(PQL)}} = \hat{\sigma}^2_{b_{0k\,(PQLF)}}$ (two variances are equal)

H_1: $\hat{\sigma}^2_{b_{0k\,(PQL)}} \neq \hat{\sigma}^2_{b_{0k\,(PQLF)}}$

Hypothesis 2:

H_0: $\hat{\sigma}^2_{b_{1t\,(PQL)}} = \hat{\sigma}^2_{b_{1t\,(PQLF)}}$ (two variances are equal)

H_1: $\hat{\sigma}^2_{b_{1t\,(PQL)}} \neq \hat{\sigma}^2_{b_{1t\,(PQLF)}}$

the *F-ratio*:

$$F_1\text{-tests}=\frac{\hat{\sigma}^2_{b_{0k\,(PQLF)}}}{\hat{\sigma}^2_{b_{0k\,(PQL)}}} = \frac{0.4201}{0.2883} = 1.457$$

$$F_2\text{-tests}=\frac{\hat{\sigma}^2_{b_{0k\,(PQLF)}}}{\hat{\sigma}^2_{b_{0k\,(PQL)}}} = \frac{0.4759}{0.335} = 1.421$$

F-table from the two kinds of variances are $F_{0.05(87299,87229)} = 1$. If the two population have equal variances, then the F-test is close to one, but if F-test is more than one, then the evidence is against the null hypothesis. Therefore, we can conclude that the variances between unadjusted PQL and Firth adjusted PQL are different. Comparing the two results, it can be seen that the firth-adjusted PQL obtain greater of variance components estimate leads to better estimates of the variability of the random effects estimates.

6. CONCLUSION

The aim of the present study is to examine whether the Firth method can reduce bias for the LLMM using the PQL technique in longitudinal data. On the LLMM with multiple random effects, the simulation of this study shows that the Firth-adjusted PQL improves the bias of the variance components estimate. In general, the result of the simulation from this study indicate that the variance components of Firth adjusted PQL are leads to the true value.

The limitation of this study is the assumption of independent between $\sigma^2_{b_{0k}}$and $\sigma^2_{b_{1t}}$, but in practice the assumption may not be realistic. For future research more complicated models need to be investigated especially in the case where the random effects are correlated. These random effects also require estimation. Based on the result of this study, the Firth-adjusted PQL is preferable to the unadjusted PQL for the model studied. Future work will determine whether the Firth-adjusted PQL is a suitable choice for other models.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] E.A. Claassen, A Reduced Bias Method of Estimating Variance Components on Generalized Linear Mixed Models. PhD Dissertation. Lincoln, NE: University of Nebraska, 2014.

[2] W.J. Browne, D. Draper, A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal. 1 (2006), 473–514.

[3] I. Kosmidis, brglm: Bias reduction in binary-response Generalized Linear Models. R package version 0.5-9. 2013.

[4] W. Jang, J. Lim, PQL estimation biases in generalized linear mixed models, Institute of Statistics and Decision Sciences, Duke University, Durham, NC, USA, 2006, 5–21.

[5] M.-L. Feddag, I. Grama, M. Mesbah, Generalized Estimating Equations (GEE) for Mixed Logistic Models, Commun. Stat. Theory Meth. 32 (2003), 851–874.

[6] N.E. Breslow, D.G. Clayton, Approximate Inference to Generalized Linear Mixed Model. J. Amer. Stat. Assoc. 88 (1993), 9-25.

[7] M. Capanu, M. Gönen, C.B. Begg, An assessment of estimation methods for generalized linear mixed models with binary outcomes, Stat. Med. 32 (2013), 4550–4566.

[8] N.E. Breslow, X. Lin, Bias Correction in Generalized Linear Mixed Models With a single Components of Dispersion. Biometrika, 82 (1) (1995), 81-91.

[9] D. Firth, Bias Reduction of Maximum Likelihood Estimates. Biometrika, 80 (1) (1993), 27-38.

[10] A. Saei, C.A. McGilchrist, Longitudinal Threshold Models with Random Components. J. R. Stat. Soc. Ser. D. 47 (2) (1998), 365-375.

[11] D.A. Harville, Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. J. Amer. Stat. Assoc. 72 (1977), 320-338.