The Application of Firth Bias Correction in Variance Components Estimation of Clustered Random Intercept Binary Model

R Arisanti¹*, I M Sumertajaya², K A Notodiputro² and Indahwati²

¹ Statistics Department, Padjadjaran University, Bandung, Indonesia
² Statistics Department, IPB University, Bogor, Indonesia

*E-mail: r.arisanti@unpad.ac.id

Abstract. Firth Bias Correction has been discussed in literatures as an alternative method for reducing the bias of variance components in generalized linear mixed model. This paper discusses the application of Firth’s correction in a clustered random intercept binary model in which the penalized quasi likelihood method is utilized. The clustered random intercept binary model is usually applied in clustered data in which the data can be naturally clustered. In this paper the Firth’s correction has been applied to low birth weight data in Indonesia to investigate factors affecting the incidence of the low birth weight as well as its variation within and between clusters.

1. Introduction

In several studies, researchers are given data that having clustered structure. This clustered data violating the assumption of independence which are commonly used in the statistical methods such as analysis of variance models. In the clustered data, there are at least two levels in the data namely level 1 units (e.g. person, household members, time points) nested within level 2 units or cluster (e.g. households, location). Therefore, this study will use the ‘person’ to refer to level 1 unit and ‘cluster’ to refer to level 2 unit, respectively. Assumed the measurement within a cluster to be dependent and be more homogeneous than measurements from different clusters. Henceforth, difference systematically from each other cluster that causes there is heterogeneity between clusters.

To perform an analysis which ignores the heterogeneity between clusters or dependency within cluster and assume that all measurements as independent, Generalized Linear Mixed Models (GLMMs) can be applied [1]. GLMMs is a statistical model that can be applied to analyse the clustered data which contains both ordinary regression parameters to all clusters and cluster-specific parameters that is called as mixed models. The cluster-specific effects are assumed as the random sample from a population distribution. GLMMs also can be applied in binary outcome. Henceforth, the correlation between subjects in the same cluster can be described and analysed. When the outcome of GLMMs is binary, then the regression coefficient is estimated by using conditional on the random effect [2] and has a subject-specific interpretation [3].

Estimating parameters in GLMMs by maximizing exact likelihood contains an intractable integration. Several methods have been proposed to obtain the estimation. A commonly method that can be used is Penalized Quasi Likelihood (PQL), proposed by Breslow and Clayton [5]. PQL method
tends to underestimate variance components. Numerically, biases of variance component estimation of PQL are systematically related to the biases of regression coefficient estimates [4].

The aim of this paper is to give an alternative method to reduce the bias of variance components estimation in clustered random intercept binary model. The motivation for this study are three folds. First, many measurements in the social sciences are categorical (binary, discrete data). For clustered data, GLMMs can be used for analysing the clustered data. The second motivation is PQL method can be used to obtain the estimation of variance components. The most advantage of PQL is very straightforward and computationally simple [6]. The third motivation of this study is bias correction method that is Firth correction can be implemented to the clustered data case.

This paper is organized as follows. Section 2 introduces the inference of PQL estimation for clustered random intercept binary model. Section 3 describes the Firth correction of PQL estimation on variance components. Section 4 presents an analysis of low birth weight in Indonesia. Conclusion are given in Section 5.

2. Overview of PQL Inference for Clustered Random Intercept Binary Model
Suppose the area under study is divided into m clusters labelled \( i = 1, \ldots, m \). Repeated measures \( y_{ij} \) is the outcome observed variable for subject \( j \) from cluster \( i \), and \( j \) denotes the subject within the cluster. Conditional on the random effects, the low birth weight in each area is assumed to be binomial distributed. Laplace approximation is commonly implemented to the PQL. In GLMMs, for response variable \( y_{ij} \) given random effect \( b = (b_1, \ldots, b_j)' \), then

\[
E(y_{ij}|b) = h(x_{ij}'\beta + z_{ij}'\theta)
\]  

(1)

\( \beta \) is a column vector of the fixed-effects regression coefficients parameter, \( x_{ij} \) and \( z_{ij} \) are the design matrixes for the fixed and random effects and \( b \) is a vector of the random effects.

The link function is called \( g(\cdot) \). If the link function that relates the outcome \( y \) to the linear predictor \( \eta \), then the invers link function

\[
h(\cdot) = g^{-1}(\cdot)
\]

Therefore, the conditional expectation of \( y_{ij} \) from the model (conditional because it is the expected value is depending on the level of the predictors) is

\[
g(E(y_{ij})) = \eta_{ij}
\]

Let the linear predictor, \( \eta_{ij} \), be the combination of the fixed and random effects, then

\[
\eta_{ij} = g(\mu_i) = x_{ij}'\beta + z_{ij}'\theta
\]

(2)

PQL usually involves Laplace approximation on the integrated quasi-likelihood. Assume that \( b \) has a multivariate normal distribution with mean 0 and (co)variance matrix \( G \) depending on variance component \( \theta \). The integrated quasi-likelihood function used to estimate \( (\beta, \theta) \) is defined by

\[
e^{q(\beta, \theta)} \propto |G|^{-\frac{1}{2}} \int \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} d_i - \frac{1}{2} b'G^{-1}b \right\} db
\]

(3)

where

\[
d_{ij} = -2 \int_{y_{ij}}^{u_{ij}} \frac{y_{ij} - b}{a_{ij}(\phi)v(b)} db
\]

\( d_{ij} \) is defined as the deviance. The estimation procedure consists of maximizing the sum of two component loglikelihoods. Let \( l_1 \) is the loglikelihood function of the binomial vector \( y \) conditional on fixed \( \beta \) and \( l_2 \) is the log of probability density function of \( b \) [8].

\( l = l_1 + l_2 \) represents the loglikelihood based on the joint distribution of \( y \) and \( b \). From equation (1) and (2), we find

\[
P_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}
\]

(4)

then

\[
l_1 = \text{const.} + \sum_{i=1}^{m} \sum_{j=1}^{l} \left[ y_{ij} \eta_{ij} - n_{ij} \log(1 + e^{\eta_{ij}}) \right]
\]

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

(5)
whereas $D$ is the variance covariance matrix of $b$.

To estimate $\beta$ and $b$, different method has been proposed by Saei and Chambers (2003). This method is one of the practical ways of estimating $\beta$ and $b$ because it involves the log-likelihood function directly. In this paper, this method is extended to be applied in Clustered Random Intercept Binary Model. The method requires the first and second derivatives of $l_1$ with respect to $\eta$, $\beta$, and $b$ and $l_2$ with respect to $\beta$ and $b$ as follows:

$$\frac{\partial l_1}{\partial \eta} = y_{ij} - \frac{ni_{ij} \eta_{ij}}{e^{\eta_{ij}} + 1} = y_{ij} - ni_{ij}p_{ij}; \quad \frac{\partial^2 l_1}{\partial \eta^2_{ij} \partial \eta_{ij}^{*}} = \text{diag}[-ni_{ij}p_{ij}(1-p_{ij})]$$

$$\frac{\partial l_1}{\partial b} = \frac{\partial l_1}{\partial \eta} \frac{\partial \eta}{\partial b} = (y_{ij} - ni_{ij}p_{ij})X_{ij}; \quad \frac{\partial^2 l_1}{\partial \eta \partial b} = \frac{\partial l_1}{\partial \eta} \frac{\partial \eta}{\partial b} = (y_{ij} - ni_{ij}p_{ij})Z_{ij}$$

since $l_2$ does not consist of parameter $\beta$, then derivatives with respect to $\beta$ and $b$ are equal to zero

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

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

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

then

$$V = \begin{bmatrix} X' \left(-\frac{\partial^2 l_1}{\partial \eta_{ij}^2}\right)X & Z \\ Z' \left(\frac{\partial^2 l_2}{\partial \eta_{ij}^2}\right)Z & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix}$$

or

$$V = \begin{bmatrix} X' \left(-\frac{\partial^2 l_1}{\partial \eta_{ij}^2}\right)X & Z \\ Z' \left(\frac{\partial^2 l_2}{\partial \eta_{ij}^2}\right)Z & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix}$$

(6)

If $V$ is evaluated at $\beta_0$ and $b_0$, then the procedure for estimating $\beta$ and $b$ is

$$\begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ b_0 \end{bmatrix} + V^{-1} \begin{bmatrix} X' \left(-\frac{\partial^2 l_1}{\partial \eta_{ij}^2}\right)X & Z \\ Z' \left(\frac{\partial^2 l_2}{\partial \eta_{ij}^2}\right)Z & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial l_1}{\partial \eta} \\ \frac{\partial l_1}{\partial b} \end{bmatrix} - V^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(7)

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

Define $P = -\left(\frac{\partial^2 l_1}{\partial \eta_{ij}^2}\right)$ and $Q^* = Q_{ij}^* = (Z' BZ + D^{-1})^{-1}$. Then the variance components of $b$ can be defined as

$$\sigma_b^2 = \eta_{ij}^{-1} (tr(Q_{ij}^* D_{ij}^{-1}) + \sigma^{-2} b'_j D_{ij}^{-1} b_j)$$

(8)

where the $\eta_{ij}$ is the rank of the matrix $Z_{ij}$.

3. Firth Correction Performance of PQL Estimates

Firth correction was originally developed by David Firth. Firth purposes to reduce the small sample bias of maximum likelihood estimates. Firth correction is known as penalized maximum likelihood. This method using the penalty that involves to the likelihood. The penalty for the exponential family models, the method using Jeffreys’ invariant prior [7].

The Basic idea of the Firth method to reduce bias is substituting the smaller bias to the score function [7]. Here is the illustration
As shown in Figure-1, if \( \hat{\theta} \) is the subject of positive bias \( b(\theta) \), then the score function is shifted at each value of \( \theta \) equal to \( i(\theta)b(\theta) \), where \( -i(\theta) = U'(\theta) \) is local gradient. The modified score function is

\[
U^*(\theta) = U(\theta) - i(\theta)b(\theta)
\]

Then, the modified estimator is obtained \( \hat{\theta} \) with the solution \( U^*(\theta) = 0 \), where \( i(\theta) \) is the Fisher information matrix.

4. Analysis of Low Birth Weight

Low Birth Weight (LBW) is a baby with a birth weight of less than 2500 grams who is weighed 1 (one) hour after the baby is born (Indonesian Doctors Association, 2018). LBW is still a public health problem in many developing countries, because it is one of the factors causing infant mortality.

According to WHO in 2019, the prevalence of LBW is estimated to be around 15-20% of all births in the world, as many as more than 20 million babies were born with low birth weight, namely below 2500 grams (equivalent to 5.5 pounds) in 2015. WHO also states that more than 80% of babies are born in the world, there are 2.5 million of the babies die every year because they born with low weight. This occurs because babies born with low body weight have a high risk of not surviving and are also at risk of developing health problems.
Low birth weight is closely associated with neonatal mortality and morbidity, impaired cognitive development of the baby, and also the risk of chronic disease later in life. Reducing the incidence of low birth weight requires a global strategy that includes improving the nutritional status of pregnancy, care for pregnancy conditions such as high blood pressure during pregnancy, to social support.

The research that related to low birth weight has been carried out in several regions in Indonesia. In Sampang Madura, the incidence of LBW is 9.6-11.6% [12]. Research in Ujung Berung, West Java province obtained LBW of 14.3% [13]. Papua and West Papua Provinces have LBW 8-16% [14]. According to the Riskesdas data in 2013, the prevalence of LBW in Indonesia was 10.2%. The highest LBW rate was in the province of Central Sulawesi at 16.9%, which was then followed by Papua Province (15.6%) and East Nusa Tenggara (15.5%). Indonesia is ranked 6th out of seven countries in Southeast Asia with the highest LBW prevalence, namely 7.1%.

Based on the impact of this LBW, it is necessary to make efforts to reduce the BBLR cases and anticipate that LBW cases does not increase again. One of the ways that can be done to prevent LBW case is early predicting on the weight of the foetus in the womb.

The research that related to LBW in Indonesia uses secondary data that source from IDHS (Indonesian Demographic and Health Survey/SDKI) data in 2017. IDHS is collected data from BPS (Central Bureau of Statistics) in collaboration with BKKBN (National Population and Family Planning Agency) whose surveys are carried out in all provinces, in Indonesia. The sample units are divided into clusters for each province. This study aims to see how the variances of LBW between clusters in Indonesia and also to determine the factors that influence to LBW. The analysis does not only involve fixed factors, but also involve the random effect to know the variation between clusters. There are four fixed effects which are involved in this study namely wealth index combined of the family (x1), mother that have smoking habit (x2), literacy of mother (x3) and total children ever born (x4). The random intercepts are clusters which are contained in the correlated data and assumed have normal distribution.

To examine the relationship between fixed effects and the outcome (response variable) visually, we can see from the conditional density plot. Figure 2 below presents the conditional density plot for each fixed effect on the outcome (response variable). As shown in figure 2, when the wealth index combined of family (x1) increases, then the low baby weight decreases. It is also happened on x3, when the mother literacy ability increases, then the low baby weight decreases. Contrary to the other two fixed effects. When there is increasing of mother smoking habit (x2), then the low baby weight also increasing. The same condition when the total children ever born (x4), then the probability for having low baby weight will also increase.
Figure 2. Conditional density plot of fixed effects

The results of the estimation of Clustered Random Intercept Binary Model can be seen from the table 1 below. The estimate for the wealth index combined (x1) of -0.2474 means that a 1 unit increase in wealth index combined of the family (x1) is associated with a 0.2474 decrease in the log-odds of low baby weight (y) being 1, compared to low baby weight (y) being 0. If we exponentiate this number then we obtain the odds ratio of 0.78082. It means that for a 1 unit increase in wealth index combined of the family, then we expect to see (approximately) a 24.74% decrease in the odds of low baby weight (y) being 1.

On the other hand, the estimation of mother that have smoking habit (x2) of 0.3874 means that a 1 unit increase in mother that have smoking habit (x2) is associated with a 0.3874 increase in the log-odds of low baby weight (y) being 1, compared to low baby weight (y) being 0. If we exponentiate this number then we obtain the subject-specific odds ratio of 1.47314. It means that for a 1 unit increase in mother that have smoking habit (x2) we expect to see (approximately) a 38.74% increase in the odds of low baby weight (y) being 1.

The result obtained from the estimation of literacy of mother (x3) of 0.1536 means that a 1 unit increase in literacy of mother (x3) is associated with a 0.1536 also increase in the log-odds of low baby weight (y) being 1, compared to low baby weight (y) being 0. If we exponentiate this number, then we obtain the subject-specific odds ratio of 1.16602. It means that for a 1 unit increase in literacy of mother (x3) we expect to see (approximately) a 15.36% increase in the odds of low baby weight (y) being 1.

It is apparent from the table 1 that the estimation of total children ever born (x4) of 0.3491 means that a 1 unit increase in total children ever born (x4) is associated with a 0.3491 also increase in the log-odds of low baby weight (y) being 1, compared to low baby weight (y) being 0. If we exponentiate this number, then we obtain the odds ratio subject-specific of 1.41779. This result means that for a 1 unit increase in total children ever born (x4) we expect to see (approximately) a 34.91% increase in the odds of low baby weight (y) being 1.

| Table 1. Estimation of fixed effects and variance component for LBW |
|---------------------------------------------------------------|
| | Unadjusted PQL | Firth adjusted PQL |
|----------------|-----------------|-----------------|
| Estimate | p-value | Estimate | p-value | OR |
A case-study approach is used to determine the factors that affect low baby weight cases. As we have previously seen, there is evidence that mother smoking habit, literacy of mother and total children ever born have associated to the response (low baby weight). It can be seen from the probability of observing data which have significant p-value. The two of method (unadjusted PQL and Firth adjusted PQL) show the same result. This result is significant at the p-value 0.05 level. Further analysis, the variability of low baby weight showed that the Firth adjusted PQL improve the variance of random effects that is \( \sigma^2_b = 0.788 \). In the case of low weight birth, a small p-value for each cluster’s random intercept would indicate that the cluster’s typical low baby weight probability is significantly different form an average cluster’s typical low baby weight probability. Clearly, this information could be interest to know the factors that affect low baby weight in cluster.

### 5. Conclusion

This study applied the performance of adjusted Firth to the PQL estimation in clustered random intercept binary model when the data are clustered. The results show that the expected odds ratio holding all the other predictors fixed. This makes sense as we are often interested in statistically adjusting for other effects, such as wealth index combined, to get the “pure” effect of low baby weight or whatever the primary predictor of interest is. The same is true with intercept binary model, with the addition that holding everything else fixed includes holding the random effect fixed. that is, the odds ratio here is the conditional odds ratio for the low baby weight that have wealth index combined of family constant as well as mother who has total baby ever born with either the same cluster, or cluster with identical random effects. Constant here indeed, in clustered random intercept binary model and because of the nonlinear link function that is used to connect the mean of the outcome with the linear predictor, the fixed effects coefficients have an interpretation conditional on the random effects. Because the components of fixed effects have interpretation that depend upon holding the \( i \)th subject’s random effects), then referred to as subject-specific regression coefficients. As a result, clustered random intercept binary model is most useful when the main objective to make inferences about individuals rather than population averages.

### References

[1] Tuerlinckx F, Rijmen F, Verbeke G, and Boeck PD 2006 Statistical inference in generalized linear mixed model: A review (British Journal of Mathematical and Statistical Psychology) 59:225-255.

[2] Diggle P, Heagerty P, Liang K Y, and Zeger S L 2002 Analysis of Longitudinal Data (Oxford: Oxford University Press).

[3] Neuhaus JM, Hauck WW, and Kalbfleisch JD 1992 The Effect of Mixture Distribution Misspecification When Fitting Mixed Effects logistic Models (Biometrika) 79:755-762.

[4] Jang W and Lim J 2009 A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects (Communications in Statistics-Simulation and Computation) 38: 692-702.

[5] Breslow NE and Clayton DG 1993 Approximate Inference to Generalized Linear Mixed Model (Journal of the American Statistical Association) 88:9-25.

[6] Schall R 1991 Estimation in Generalized Linear Models With Random Effects (Biometrika) 78: 719-27.
[7] Firth D 1993 *Bias Reduction of Maximum Likelihood Estimates* (Biometrika) 80(1) 27-38.

[8] McGilChrist 1994 *Estimation in Generalized Mixed Models* (JSTOR) 56(1) 61-69.

[9] Saei A and Chambers R 2003 *Small Area Estimation Under Linear and Generalized Linear Mixed Models with Time and Area Effects* (University of Southhampton) Provide at https://www.researchgate.net/publication/242359115_Small_Area_Estimation_Under_Linear_and_Generalized_Linear_Mixed_Models_With_Time_and_Area_Effects?enrichId=rqreq6cb615ffe34e34c681141d24309a539d8-XXX&enrichSource=Y292ZXJQYWdlOi0MjMjMjOTExNTRtBUzoxMDQxMTMzNTg1MDgwMzhAMTQwMTgzMzk2NzQzOA%3D%3D&el=1_x_2&_esc=publicationCoverPdf.

[10] Wand M P 2007 *Fisher Information for Generalised Linear Mixed Models* (Journal of Multivariate Analysis) 98: 1412-1416.

[11] BKKBN, BPS, Kemenkes 2018 *Survei Demografi dan kesehatan Indonesia 2017* (Jakarta: Badan Kependudukan dan Keluarga Berencana Nasional).

[12] Kardjati S, Kusin J A and De With C 1985 *Factors Influencing Birth Weight* (Jakarta: Seminar Iptek Gizi dan Kesehatan Ibu Hamil).

[13] Alisjahbana A 1985 *Kematian Perinatal dan Faktor yang Berhubungan dengan Masalah Gizi* (Jakarta: Yayasan Obor).

[14] Biro Pusat Statistik 2011 *Multiple indicator cluster survey (MICS) selected Province in Papua and West Papua (diseminasi seminar)* Provide at https://dokumen.tech/document/multiple-indicator-cluster-survey-kabupaten-terpilih-di-papua-dan-.html. Accessed on September 2, 2020.