AMSI VACATIONRESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Robust Adjustments to Random Effects Models for Dispersed Counts

Wilson Lorensyah

Supervised by Dr Alan Huang

The University of Queensland

Vacation Research Scholarships are funded jointly by the Department of Education, Skills and Employment and the Australian Mathematical Sciences Institute.

Table of Contents

Abstract

The purpose of data analysis using mixed model is to account both random and fixed effects in the observations. Traditional method for using generalized linear model (GLM) is not offering capability to look at clustered data (correlation in the data) and random factor that each cluster might have. However, under GLM, there is a large sample property that could give robust standard error. Then, one could think, “Is this idea could be implemented in the mixed model?”. This project goal is looking at applying the idea in the robust variance linear model to be implemented in the generalized linear mixed model (GLMM). However, based on the simulation, it is not fully transferable to apply the idea under GLM to get robust variance in GLMM. It would require a bit of adjustment in the formula to find the correct variance of fixed effects model to account for the random effects in the model, or approach differently with likelihood-based approach.

1 Introduction

The popular GLM has been widely used to analyze data to account fixed effects which is common across observations. However, analysis would be more sophisticated by incorporating random effect. Also, if one would like to have correlated observations in the data in a cluster, it would be feasible to use GLMM. One application of this would be to look at the Salamanders count throughout different sites. Each site would have common factor that could affect count output on Salamanders, for example, humidity, water temperature, season, etc., but each site could have random factor that could affect the count output of the Salamanders. For example, particular site might have higher population of Salamanders (or seen as ‘inherent ability’ of the sites) at the beginning before the observations. Thus, to account for this random factor, this random effect on each site is treated by taking a random draw from normal distribution with mean 0 and some variance value.

Subsequently, generalized estimating equations (GEE) has been established that robust variance under GLM could be achieved by applying sandwich estimator. Thus, in a large sample, regardless family or distribution is correctly specified, it will specify the correct variance of the fixed effects. Motivated by this idea, one would like to produce the robust variance for model that considers both fixed and random effects. Aim of this project is to apply the idea of sandwich estimator from GEE to GLMM to account for the random effect (particularly count model under Poisson model) and get the robust variance of the fixed effects after incorporating extra information about the random effects. Furthermore, the model can be extended for dispersed counts specified by Conway-Maxwell Poisson (CMP) model. Numerical computation has been conducted using the software R Studio.

2 Statement of Authorship

Alan Huang and Wilson Lorensyah developed an implementation of enhancement of existing project of generalized version of Poisson regression. Alan Huang explained the ideas, possible approach, demonstrate samples R code and proofread the report. Wilson Lorensyah wrote the report, R code, and conducting numerical experiments. The project was conducted during end of 2020/ beginning of 2021.

3 Background - Introduction and Model

Linear Mixed Model (LMM) is a method for analysing non-independent, multilevel / hierarchical, or correlated data using mean model given response variables are distributed as Gaussian. This would be an extension of simple linear model (LM) to allow both *fixed* and *random* effects and response variables to be correlated. General form of the model is given by:

LMM	LM
$\boldsymbol{y}_{N\times 1} = \boldsymbol{X}_{N\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{Z}_{N\times q}\boldsymbol{u}_{q\times 1} + \boldsymbol{\epsilon}_{N\times 1}$	$\boldsymbol{y}_{N\times 1} = \boldsymbol{X}_{N\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{N\times 1}$
Fixed and Random effects	Fixed effects

where

$\boldsymbol{y}$ is the column vector for outcome variables

$\boldsymbol{X}$ is design matrix of p predictor variables

$\boldsymbol{\beta}$ is a column vector of fixed effects regression coefficients

$\boldsymbol{Z}$ is design matrix of q random effects

$\boldsymbol{u}$ is a column vector of random effects (random complement for fix $\boldsymbol{\beta}$)

$\boldsymbol{\epsilon}$ is column vector of residuals

Suppose one is shown a linear model from dataset of health risk against time:

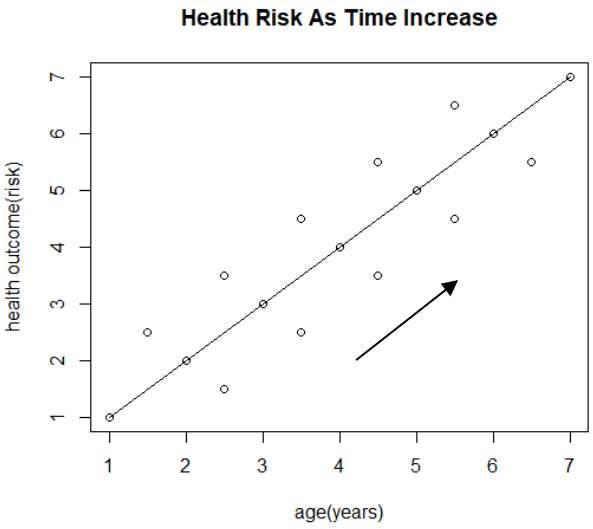

Chart 1: Population Model / Marginal Model

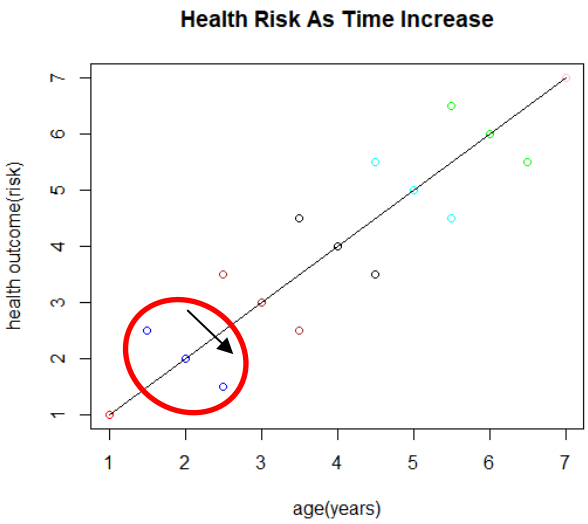

Chart 2: Subject-Specific Model (Within Clusters)

From Chart 1 and Chart 2, the straight line shows that it uses the mean model. From Chart 1, the trend of the data is positive in the overall population. This model will depict how the mean response in population is related to the covariates. Now, suppose an additional information has arrived that the data is consisting points of measurements from the same subjects (denoted within same color in Chart 2). This will change the trend to become negative within each person. This means that each subject (or known as cluster) has their own intercept and slopes (direction shown in the red circle and arrow in Chart 2) which are drawn from population intercepts and slopes. Note that in this case variability within clusters are not independent, while observations across different clusters (groups) are independent.

However, rather than modelling direct mean, often used in this case would be a *link* function. The generic link function relates the mean of outcome $\boldsymbol{y}$ to the linear predictor $\boldsymbol{X\beta + Zu}$. By using this idea, LMM would become generalized LMM (GLMM) which could be produced not only for Gaussian, but also for binary and count response.

Marginal models are attractive because it could specify within cluster association without a probabilistic model to generate the association. This means, estimation methods can be done using *estimating equations* which will become underlying model for the data and is an alternative to likelihood-based methods. When dealing with correlated model – it cannot find likelihood function explicitly unless doing some numerical integration to integrate out the random effects in the model.

The marginal model of clustered data would have a form of observation Y_{ji} which denotes jth cluster and ith observation. The standard error of the marginal models needs 3 things to be correctly assumed (Huang, 2020):

- Condition of mean-model: this can be correctly specified mostly from the plots or residual analysis.

  For a given vector of covariates of explanatory variables $\boldsymbol{X_{ji}}$, the mean-model is given by

  $$\mathbb{E}(Y_{ji}|\boldsymbol{X_{ji}}) = \mu_{ji} = \mu(\boldsymbol{X_{ji}^T\beta})$$

  where the mean function $\mu(\cdot)$ would be a link function on how Y_{ji} depends on the covariates $\boldsymbol{X_{ji}}$ which could be identity link= $\boldsymbol{X_{ji}^T\beta}$, log link= $e^{(\boldsymbol{X_{ji}^T\beta})}$, or logistic link= $1 + e^{(\boldsymbol{X_{ji}^T\beta})}$ and the slope $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots)$
- Conditional variance for each Y_{ji} often incorrectly specified due to various data structure, and it is given by:

  $$\text{Var}(Y_{ji}|\boldsymbol{X_{ji}}) = \phi V(\mu_{ji})$$

  where ϕ is a constant and $V(\cdot)$ is some variance function (note that when $\phi = 1, V(\mu_{ji}) = \mu_{ji}$ it will follow Poisson distribution since it has same mean and variance)
- Association between two observations within same cluster

  $$\text{Association}(Y_{ji}, Y_{jk}) = \rho(\mu_{ji}, \mu_{jk}; \gamma)$$

  where $\rho(\mu_{ji}, \mu_{jk}; \gamma)$ is some function that contains association parameter γ

In this case, association refers to relation of continuous and discrete variables, whereas correlation refers to the relation of 2 continuous variables. In addition, the association/correlation itself needs not to be specified correctly and can be adjusted for any initial correlation structure.

After having these properties set up, estimation of the marginal models would be carried out using generalized estimating equations (GEE) was developed by Liang & Zeger (1986). This is an extension of GLM which only considers the fixed effect. This method is commonly used when underlying joint distribution cannot be explicitly written for the data. Thus, this idea is used as a starting point to be implemented for creating robust standard error for model with fixed and random effects.

There are two main properties for the GEE estimator ($\hat{\beta}_{\text{GEE}}$):

1. $\hat{\beta}_{\text{GEE}}$ is consistent estimator for $\boldsymbol{\beta}$ even if *working variance-covariance* matrix W_j is incorrectly specified.

2. For large sample sizes, $\hat{\beta}_{\text{GEE}}$ distribution is approximately given by mean $\boldsymbol{\beta}$ and covariance matrix:

$$\text{Var}(\hat{\beta}_{\text{GEE}}) = \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \left(\sum_{j=1}^{J} D_j W_j^{-1} \text{Var}(\boldsymbol{Y}_j) W_j^{-1} D_j^T\right) \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1}$$

where for each cluster j,

D_j is the *"derivative matrix"* of the mean-model

$$D_j = \frac{\partial \mu(\boldsymbol{X}_{\boldsymbol{ji}}^{\boldsymbol{T}} \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

W_j is the *working variance-covariance* matrix under the assumed model.

$Var(\boldsymbol{Y}_j)$ is the true variance-covariance matrix for the responses $\boldsymbol{Y}_j$ which is unknown. This can be estimated using regression residuals

$$\text{Var}(\boldsymbol{Y}_j) = (\boldsymbol{Y}_j - \widehat{\boldsymbol{\mu}}_j)(\boldsymbol{Y}_j - \widehat{\boldsymbol{\mu}}_j)^T$$

When this is substituted to the covariance matrix $Var(\hat{\beta}_{GEE})$, the term will be:

$$\left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \left(\sum_{j=1}^{J} D_j W_j^{-1} (\boldsymbol{Y}_j - \widehat{\boldsymbol{\mu}}_j)(\boldsymbol{Y}_j - \widehat{\boldsymbol{\mu}}_j)^T W_j^{-1} D_j^T\right) \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1}$$

Which is so-called ***"sandwich estimator"*** of variance (it is thought of the first and third term as the "bread", and the second term as the "meat")

Note that if the working variance-covariance matrix happens to be correctly specified (i.e. when $Var(\boldsymbol{Y}_j) = W_j$). Then the covariance matrix $Var(\hat{\beta}_{GEE})$ is given by:

$$\begin{aligned}
\text{Var}(\hat{\beta}_{GEE}) &= \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \left(\sum_{j=1}^{J} D_j W_j^{-1} W_j W_j^{-1} D_j^T\right) \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \\
&= \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right) \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \\
&= \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1} \dots (*)
\end{aligned}$$

The sandwich estimator of variance is always true covariance of $\hat{\beta}$ and consistent whether *working variance-covariance* model is correctly specified. GEE framework will offer asymptotically correct inferences for β. This property is robust for large sample property, so care needs to be taken in small/moderate sample size for choosing a good starting model for the data to get accurate variance. Hence, this idea of sandwich estimator is deemed as ***"robust adjustments"*** to variance.

4 Discussion / Findings

In this application, the relevant model to use would be clustered count data for its marginal model. Then, suitable model to use would be (discrete) Poisson model which has some properties:

$$Y \sim \text{Poi}(\lambda),\ \mathbb{E}(Y) = \text{Var}(Y) = \lambda,\ \text{where } \lambda \in (0, \infty)$$

More generally, it can also use the Conway-Maxwell-Poisson (CMP) model which is the generalized version of Poisson distribution which allow **dispersion parameter** (ν) has some properties:

$$Y \sim \text{CMP}(\lambda, \nu),\ \mathbb{E}(Y) = \sum_{k=0}^{\infty} \frac{k\lambda^k}{(k!)^\nu Z(\lambda, \nu)}, \text{Var}(Y) = [\sum_{k=0}^{\infty} \frac{k^2\lambda^k}{(k!)^\nu Z(\lambda, \nu)}] - [\mathbb{E}(Y)]^2$$

$$\text{PMF} = \Pr(Y = k) = \frac{\lambda^k}{(k!)^\nu Z(\lambda, \nu)}$$

$$Z(\lambda, \nu) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu},\ \text{where } \lambda, \nu > 0 \text{ or } \lambda \in (0,1), \nu = 0$$

There is no closed form for the variance for CMP, and when $\nu = 1$, the CMP distribution becomes Poisson distribution. Over (Under) dispersion would create greater (smaller) variability in a dataset given a statistical model. Note that when $\nu = 0$ and $\lambda < 1$ it will follow geometric distribution starting at $k = 0$. Also, when $\nu \to \infty$, the CMP model converges in distribution to Bernoulli distribution with mean $\lambda(1 + \lambda)^{-1}$.

To simulate this, glmmTMB package from R is simulated since it accommodates both Poisson and CMP distribution comparing to the other available options. Simulation will be based on Poisson model and will be generalized to CMP afterwards.

The datasets used were from glmmTMB package which is called “Salamanders” data which consists of 644 observations with the following 9 variables and response variable “Count”:

- Site: name of a location where repeated samples was taken
- Mined: factor indicating whether the site was affected by mountain top removal coal mining
- Cover: amount of cover objects in the stream (scaled)
- Sample: repeated sample
- DOP: Days since precipitation (scaled)
- Wtemp: water temperature (scaled)
- DOY: day of year (scaled)
- Spp: abbreviated species name, possibly also life stage
- Count: number of salamanders observed

In this simulation the call was to see the random effects, the mixed model was run depending on 7 different species, sampled 4 times for each species (total of 28 observations in each site) with the random effect given by 23 sites (clusters). Under Poisson distribution, the hierarchical model for jth cluster and subject ith can be specified as:

$$Y_{ji}|\boldsymbol{X_{ji}}, \alpha_j \sim \text{Poi}(\lambda_{ij} = \exp(\boldsymbol{X_{ij}^T\beta} + \alpha_j))$$

$$\lambda_{ij}|\alpha_i \sim \exp(\boldsymbol{X_{ij}^T\beta} + \alpha_j)$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

where

the number of sites $j = 1, \dots, 23$, the number of species $i = 1, \dots, 7$. (could have second digits $k = 1, \dots, 4$ after i to account for 4 times repetition). For each $\mathbb{I}(\cdot)$ will return 1 if $(\cdot)$ is satisfied and 0 otherwise:

$$\boldsymbol{X_{ji}^T\beta} = \beta_0 + \beta_1\mathbb{I}(i = 2) + \beta_2\mathbb{I}(i = 3) + \cdots + \beta_6\mathbb{I}(i = 7)$$

$$Y_{jik}|\alpha_j \sim \text{Pois}(\exp(\beta_0 + \beta_1\mathbb{I}(j = 2) + \beta_2\mathbb{I}(j = 3) + \cdots + \beta_6\mathbb{I}(j = 7) + \alpha_j))$$

each α_j represent the random effect / inherent ability of each site as thought of coming from Normal distribution, but common through all measurements in i.

Thus, the conditional expectation and conditional variance under Poisson model are given by:

$$\mathbb{E}(Y_{ji}|\alpha_j) = \text{Var}(Y_{ji}|\alpha_j) = e^{(\boldsymbol{X_{ji}^T\beta} + \alpha_j)}$$

In this structure, the mean model uses the log-link as this is considered as marginal model for clustered count data. The data structure would be shown as:

$$\begin{pmatrix} Y_{111}|\alpha_1 \\ Y_{112}|\alpha_1 \\ Y_{113}|\alpha_1 \\ Y_{114}|\alpha_1 \\ Y_{121}|\alpha_1 \\ Y_{122}|\alpha_1 \\ \vdots \\ Y_{174}|\alpha_1 \end{pmatrix}_{28\times1} \quad \cdots \quad \begin{pmatrix} Y_{2311}|\alpha_{23} \\ Y_{2312}|\alpha_{23} \\ Y_{2313}|\alpha_{23} \\ Y_{2314}|\alpha_{23} \\ Y_{2321}|\alpha_{23} \\ Y_{2322}|\alpha_{23} \\ \vdots \\ Y_{2374}|\alpha_{23} \end{pmatrix}_{28\times1}$$

To find the variance of β, there are two ways to obtain it: vcov output from glmmTMB package (which gives the "bread" part of the sandwich estimator denoted in notation $(*)$) or create manually by finding W_j and D_j matrix and assemble them. For variance matrix $vcov = \text{Var}(\widehat{\boldsymbol{\beta}})$, it is defined as:

$$vcov = \begin{bmatrix} \text{Var}(\widehat{\beta_0}) & \text{Cov}(\widehat{\beta_0}, \widehat{\beta_1}) & \cdots & \text{Cov}(\widehat{\beta_0}, \widehat{\beta_6}) \\ \text{Cov}(\widehat{\beta_1}, \widehat{\beta_0}) & \text{Var}(\widehat{\beta_1}) & \text{Cov}(\widehat{\beta_1}, \widehat{\beta_2}) & \vdots \\ \vdots & \text{Cov}(\widehat{\beta_2}, \widehat{\beta_1}) & \ddots & \text{Cov}(\widehat{\beta_5}, \widehat{\beta_6}) \\ \text{Cov}(\widehat{\beta_6}, \widehat{\beta_0}) & \cdots & \text{Cov}(\widehat{\beta_6}, \widehat{\beta_5}) & \text{Var}(\widehat{\beta_6}) \end{bmatrix}_{7\times7} = \left[\sum_{i=1}^{23} D_{i_{7\times28}} W_{i_{28\times28}}^{-1} D_{i_{28\times7}}^T\right]_{7\times7}^{-1}$$

In order to find D_j^T, it would require the expectation properties. The mean-model for site 1 is given by:

$$\mu_{ji} = \mathbb{E}\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{114} \\ Y_{121} \\ Y_{122} \\ \vdots \\ Y_{174} \end{pmatrix} = \mathbb{E}\begin{pmatrix} \mathbb{E}(Y_{111}|\alpha_1) \\ \mathbb{E}(Y_{112}|\alpha_1) \\ \mathbb{E}(Y_{113}|\alpha_1) \\ \mathbb{E}(Y_{114}|\alpha_1) \\ \mathbb{E}(Y_{121}|\alpha_1) \\ \mathbb{E}(Y_{122}|\alpha_1) \\ \vdots \\ \mathbb{E}(Y_{174}|\alpha_1) \end{pmatrix} = \mathbb{E}\begin{pmatrix} \exp(\beta_0 + \alpha_1) \\ \exp(\beta_0 + \alpha_1) \\ \exp(\beta_0 + \alpha_1) \\ \exp(\beta_0 + \alpha_1) \\ \exp(\beta_0 + \beta_1 + \alpha_1) \\ \exp(\beta_0 + \beta_1 + \alpha_1) \\ \vdots \\ \exp(\beta_0 + \beta_6 + \alpha_1) \end{pmatrix}$$

$$D_j^T = \frac{\partial \mu_{ji}}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial}{\partial \beta_0} & \frac{\partial}{\partial \beta_1} & \dots & \frac{\partial}{\partial \beta_6} \end{pmatrix}$$

$$D_j^T = \begin{pmatrix} e^{\beta_0}\mathbb{E}(e^{\alpha_1}) & 0 & \cdots & \cdots & 0 \\ e^{\beta_0}e^{\beta_1}\mathbb{E}(e^{\alpha_1}) & e^{\beta_0}e^{\beta_1}\mathbb{E}(e^{\alpha_1}) & 0 & \cdots & \vdots \\ e^{\beta_0}e^{\beta_2}\mathbb{E}(e^{\alpha_1}) & 0 & e^{\beta_0}e^{\beta_2}\mathbb{E}(e^{\alpha_1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ e^{\beta_0}e^{\beta_6}\mathbb{E}(e^{\alpha_1}) & \cdots & 0 & 0 & e^{\beta_0}e^{\beta_6}\mathbb{E}(e^{\alpha_1}) \end{pmatrix}_{28\times 7} \quad \dots (**)$$

For simplification, note that each row above is duplicated 4 times to make the matrix size becomes 28×7. Also, the expected value $\mathbb{E}(e^{\alpha_1}) = \exp\left(\frac{\sigma^2}{2}\right)$ due to log-normal distribution. D_j can be found by transposing D_j^T. Since observations between sites are independent, it would be repeatable for all 23 sites and each matrix would have the same structure just changing the first index until 23. Then, it would require the working variance-covariance W_j is given by:

$$W_j = \begin{pmatrix} \mathrm{Var}(\mathrm{Y}_{111}) & \mathrm{Cov}(\mathrm{Y}_{111}, \mathrm{Y}_{112}) & \cdots & \mathrm{Cov}(\mathrm{Y}_{111}, \mathrm{Y}_{174}) \\ \mathrm{Cov}(\mathrm{Y}_{112}, \mathrm{Y}_{111}) & \mathrm{Var}(\mathrm{Y}_{112}) & \mathrm{Cov}(\mathrm{Y}_{112}, \mathrm{Y}_{113}) & \vdots \\ \vdots & \mathrm{Cov}(\mathrm{Y}_{113}, \mathrm{Y}_{112}) & \ddots & \mathrm{Cov}(\mathrm{Y}_{173}, \mathrm{Y}_{174}) \\ \mathrm{Cov}(\mathrm{Y}_{174}, \mathrm{Y}_{111}) & \cdots & \mathrm{Cov}(\mathrm{Y}_{174}, \mathrm{Y}_{173}) & \mathrm{Var}(\mathrm{Y}_{174}) \end{pmatrix}_{28\times 28} \quad \dots (***)$$

In this case, one can find the pattern of the first few trials and can be generalized to other entry. The calculation can be found using the *law of total variance*:

$$\begin{aligned} \mathrm{Var}(\mathrm{Y}_{111}) &= \mathbb{E}\big(\mathrm{Var}(\mathrm{Y}_{111}|\alpha_1)\big) + \mathrm{Var}\big(\mathbb{E}(\mathrm{Y}_{111}|\alpha_1)\big) \\ &= \mathbb{E}\big(e^{(\beta_0+\alpha_1)}\big) + \mathrm{Var}\big(e^{(\beta_0+\alpha_1)}\big) \\ &= e^{\beta_0+(\sigma^2/2)} + e^{2\beta_0}\mathrm{Var}\big(\mathrm{e}^{(\alpha_1)}\big) \\ &= e^{\beta_0+(\sigma^2/2)} + e^{2(\beta_0)}\Big(\mathrm{e}^{\sigma^2}\big(\mathrm{e}^{\sigma^2}-1\big)\Big) \end{aligned}$$

Note that the variance of log-normal distribution $Var\big(e^{(\alpha_1)}\big) = \Big(\mathrm{e}^{\sigma^2}\big(\mathrm{e}^{\sigma^2}-1\big)\Big)$. Simplifying to the other variance entry, then it would just replace β_0 with $\beta_0 + \beta_{i'}$, $i' = 1, \dots, 6$.

For covariance within the same Species, it can be found ($k' = 2,3,4$):

$$\begin{aligned}\mathrm{Cov}(\mathrm{Y}_{111}, \mathrm{Y}_{11k'}) &= \mathbb{E}\big(\mathrm{Cov}(\mathrm{Y}_{111}, \mathrm{Y}_{11k'}|\alpha_1)\big) + \mathrm{Cov}\big(\mathbb{E}(\mathrm{Y}_{111}|\alpha_1), \mathbb{E}(\mathrm{Y}_{11k'}|\alpha_1)\big) \\ &= 0 + \mathrm{Cov}_{\alpha_1}\big(\mathrm{e}^{\beta_0+\alpha_1}, \mathrm{e}^{\beta_0+\alpha_1}\big) = \mathrm{e}^{2\beta_0}\mathrm{Var}\big(\mathrm{e}^{(\alpha_1)}\big) = \mathrm{e}^{2(\beta_0)}\left(\mathrm{e}^{\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right)\right)\end{aligned}$$

For covariance within the different Species, it can be found ($i, i' = 1, \dots, 7,\ i \neq i', k = 1, \dots, 4$):

$$\mathrm{Cov}(\mathrm{Y}_{1ik}, \mathrm{Y}_{1i'k}) = \mathbb{E}\big(\mathrm{Cov}(\mathrm{Y}_{1ik}, \mathrm{Y}_{1i'k}|\alpha_1)\big) + \mathrm{Cov}\big(\mathbb{E}(\mathrm{Y}_{1ik}|\alpha_1), \mathbb{E}(\mathrm{Y}_{1i'k}|\alpha_1)\big)$$

For example, comparing species 4 second observation and species 6 third observation would be:

$$\begin{aligned}\mathrm{Cov}(\mathrm{Y}_{142}, \mathrm{Y}_{163}) &= \mathbb{E}\big(\mathrm{Cov}(\mathrm{Y}_{142}, \mathrm{Y}_{163}|\alpha_1)\big) + \mathrm{Cov}\big(\mathbb{E}(\mathrm{Y}_{142}|\alpha_1), \mathbb{E}(\mathrm{Y}_{163}|\alpha_1)\big) \\ &= 0 + \mathrm{Cov}_{\alpha_1}\big(\mathrm{e}^{\beta_0+\beta_3+\alpha_1}, \mathrm{e}^{\beta_0+\beta_5+\alpha_1}\big) \\ &= e^{\beta_3}e^{\beta_5}e^{2\beta_0}\mathrm{Var}\big(e^{(\alpha_1)}\big) = e^{\beta_3}e^{\beta_5}e^{2\beta_0}\left(\mathrm{e}^{\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right)\right)\end{aligned}$$

Then, once we get the expression for D_j, D_j^T, W_j, one can apply the idea of $(*)$ and sum over all sites $j = 1, \dots, 23$, then inverse, to compare the output with the results given by 'vcov' function in glmmTMB package which is thought to output

$$\mathrm{Var}\big(\hat{\beta}_{\mathrm{glmmTMB}}\big) = \mathrm{Var}\big(\hat{\beta}_{\mathrm{GEE}}\big) = \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1}$$

Then, after comparing the results for both methods (Appendix), it proves that all entries are correct (slightly off due to numerical rounding error) except first row, first column entry. Thus, it could be seen that

$$\mathrm{Var}\big(\hat{\beta}_{\mathrm{glmmTMB}}\big) \neq \mathrm{Var}\big(\hat{\beta}_{\mathrm{GEE}}\big) = \left(\sum_{j=1}^{J} D_j W_j^{-1} D_j^T\right)^{-1}$$

5 Conclusion and Further Research

The idea of sandwich estimator from GEE cannot be directly applied to GLMM (requires some adjustment to account for the random-effects / different methods e.g. likelihood-based to get robust standard errors). There are some areas for further research: once replication from vcov matrix output can be replicated from glmmTMB, robust estimation could be obtained using modified version of the "Sandwich Estimator". The results using Poisson model would then be applied to CMP model to account for dispersion in counts.

6 Acknowledgements

I would like to convey my special thanks to Alan Huang for supervising me throughout the summer research project, being available for regular consultation sessions and giving more exposure on the available topics developed in statistics. I would also like to thank the University of Queensland for facilitating the completion of the project as well as AMSI for the funding and great opportunity to be involved in a Vacation Research Scholarship (VRS).

Appendix

Algorithm to manual setup vcov or $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{GEE}})$:

- Run glmmTMB, count depending on the fixed effects of species and random effect of sites using the Poisson model
- Extract the betas values ($\boldsymbol{\beta}$), Variance of the random effect model (σ^2)
- Set up D_j^T the transpose of derivative matrix with size (28×7) based on $(**)$
- Transpose D_j^T to get D_j with size (7×28)
- Find the W_j matrix with size (28×28) based on $(***)$
- Do matrix multiplication $D_j W_j^{-1} D_j^T$ with size (7×7)
- Sum $D_j W_j^{-1} D_j^T$ across all 23 sites, and inverse the result to get manual vcov

Output of manual setup:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	0.00959251	-0.00926500	-0.00926500	-0.00926500	-0.00926500	-0.00926500	-0.00926500
[2,]	-0.00926500	0.04632422	0.00926500	0.00926500	0.00926500	0.00926500	0.00926500
[3,]	-0.00926500	0.00926500	0.01662246	0.00926500	0.00926500	0.00926500	0.00926500
[4,]	-0.00926500	0.00926500	0.00926500	0.02927735	0.00926500	0.00926500	0.00926500
[5,]	-0.00926500	0.00926500	0.00926500	0.00926500	0.01424315	0.00926500	0.00926500
[6,]	-0.00926500	0.00926500	0.00926500	0.00926500	0.00926500	0.01396271	0.00926500
[7,]	-0.00926500	0.00926500	0.00926500	0.00926500	0.00926500	0.00926500	0.01781728

Output of vcov from glmmTMB:

Conditional model:

	(Intercept)	sppPR	sppDM	sppEC-A	sppEC-L	sppDES-L	sppDF
(Intercept)	0.100414787	-0.009259333	-0.009259333	-0.009259333	-0.009259334	-0.009259334	-0.009259333
sppPR	-0.009259333	0.046295871	0.009259332	0.009259332	0.009259333	0.009259333	0.009259332
sppDM	-0.009259333	0.009259332	0.016612285	0.009259333	0.009259333	0.009259333	0.009259333
sppEC-A	-0.009259333	0.009259332	0.009259333	0.029259437	0.009259333	0.009259333	0.009259333
sppEC-L	-0.009259334	0.009259333	0.009259333	0.009259333	0.014234439	0.009259333	0.009259333
sppDES-L	-0.009259334	0.009259333	0.009259333	0.009259333	0.009259333	0.013954163	0.009259333
sppDF	-0.009259333	0.009259332	0.009259333	0.009259333	0.009259333	0.009259333	0.017806382

Any queries or development ideas are welcome email at w.lorensyah@uqconnect.edu.au

References

Huang A. (2020). “19. Marginal Models for Correlated Data and Estimating Equations”, Problems & Applications in Modern Statistics, Lecture Notes, STAT3500, The University of Queensland, delivered 20 Dec 2020.

Liang K, & Zeger SL (1986). “Longitudinal Data Analysis Using Generalized Linear Models”, Biometrika, Vol. 73, No.1, pp.13-22, https://doi.org/10.2307/2336267. (originally: http://www.biostat.jhsph.edu/~fdominic/teaching/bio655/references/extra/liang.bka.1986.pdf)

Price SJ, Muncy BL, Bonner SJ, Drayer AN, Barton CD (2016) Effects of mountaintop removal mining and valley filling on the occupancy and abundance of stream salamanders. *Journal of Applied Ecology* **53** 459--468. http://dx.doi.org/10.1111/1365-2664.12585

Price SJ, Muncy BL, Bonner SJ, Drayer AN, Barton CD (2015) Data from: Effects of mountaintop removal mining and valley filling on the occupancy and abundance of stream salamanders. *Dryad Digital Repository*. http://dx.doi.org/10.5061/dryad.5m8f6