

Prediction of Number of Claims using Poisson Linear Mixed Model with AR(1) random effect

Fia Fridayanti Adam^{1,3}, Anang Kurnia^{1*}, I Gusti Putu Purnaba², I Wayan Mangku²
{fia@vokasi.ui.ac.id, anangk@apps.ipb.ac.id, purnaba@gmail.com}

¹Department of Statistics, IPB University, Bogor, 16680, Indonesia

²Department of Mathematics, IPB University, Bogor, 16680, Indonesia

³Program Vokasi, Universitas Indonesia, Depok, 16424, Indonesia

Abstract. This study focuses on the number of claims data in an insurance company in Indonesia in 35 locations. The approach taken is a linear Poisson mixed model with two random effects. The response variable is number of claims, the fixed variable is deductibles and random effects are the area and the month of occurrence which is assumed to follow the first-order autoregressive process. Fixed and random component estimation is carried out based on MPQL while estimating component variance is using REML which the initial values are $\beta_0 = 0, \beta_1 = 0, \sigma_v^2 = 0.5$, and $\sigma_u^2 = 1$. Modeling is carried out on training data which is 75% of observations and predictions carried out with testing data which is 25% of the observations. Modeling on training and testing data produces accurate models in almost all regions included in the model. This are indicated by the MAPE values which are less than 20% in all regions.

Keywords: autoregressive process, MAPE, number of claims, Poisson mixed models.

1 Introduction

To pay obligations to its customers, insurance companies must be able to predict how much the obligation. Insurance companies must be able to predict the amount of claims and the number of claims. The amount of claims or claim severity is a measure of how much the claim has occurred. Whereas number of claims or claim frequencies is a measure of how many claims occur. If the size of the claim is approached as a continuous variable, number of claims constitute data on counts. The distribution approach used usually is Poisson distribution.

This study focuses on number of claims data in an insurance company in Indonesia in several regions in Indonesia. Number of claims was observed in these areas of events for one year. The approach taken is the Generalized Linear Mixed Model approach (GLMM). The GLMM is an extension of the linear mixed model with the variable response including the exponential family, in other words GLMM is an extension of the Generalized Linear Model (GLM) but has a fixed and random effects. Poisson regression and binomial-logit models are the generalized linear models used for count data [3]. In this model the assumption of linearity is changed to the existence of a link function that connects the expected value of observation

with the covariate variable. The assumption of normality is also changed to an exponential family distribution. In [2] some of the applications of the generalized linear mixed models in the insurance sector are explained, especially in the model of credibility and ratemaking in general insurance. Previously on [1] used a mixed lognormal model to estimate the reported reserves of claims with periods of events and periods of delays as fixed effects and calendar periods as random effects.

Meanwhile, a longitudinal response data analysis method was explained with correlations between observations on the same subject [10]. To analyze longitudinal count data, Poisson linear mixed models are commonly used [12]. [4] used Poisson mixed model with one of its random effects is the time that follows the first-order autoregressive process. The solution was made by using the moment method. The linear Poisson mixed model with random variables approached by the ARMA process is then used [6].

[7] used linear mixed model with the random effect of area and time following the first-order autoregressive process with application in the unemployment rate which is assumed to be a normal distribution. While the iterative approach is used to obtain estimators of parameters in the generalized mixed linear mixed models [5]. The approach as done by [5] and [7] will be used in this study to model number of claims.

Besides that, deductibles and area codes are shown in [9] as variables that significantly influence the size of claims. Suppose that these variables also influence number of claims, the research question arises how to get predictions of number of claims using the linear Poisson mixed model approach with variable occurrences of events as a random effect and event periods as a random effect of time which are considered to follow the first-order autoregressive process? Next, how to get a prediction of the total number of claims in a particular event? Thus, the purpose of this study is to obtain a prediction of the total number of claims of an insurance company in Indonesia in a particular area of incidence using a linear Poisson mixed model approach. The data used is sourced from motor vehicle insurance risk and loss profile report data for a general insurance company in Indonesia.

2 Material and Methods

2.1 Material

The material used comes from the Financial Services Authority (OJK), which is a one year report on the risk and loss profile of motor vehicle insurance in a general insurance company in Indonesia. Reported claims data are all claims that occur in the reported calendar year with due regard to the date of claim and date claim agreement. The data is then extracted so that it contains the code of the area of occurrence, the month of occurrence, the size of the deductible, and the number of claims. Area code is a standard code that indicates the location of the claim or the address of the nearest police station in the area of the claim. The code is divided based on 35 observation areas in Indonesia.

The event period is the month, from month 1 to month 12, and hereinafter referred to as the month of occurrence. The month is the month of the accident and not the month the claim is approved or paid. Deductible is the amount of rupiah that is borne by the vehicle owner applied to the claim, whereas number of claims are calculated from how many of the same policy number make claims in the month of events and regions of a particular event.

The number of claims becomes the variable of the observed response, the deductible becomes a fixed effect and the area code of the incident and the month of occurrence become random effect with the month of the incident considered to follow the first-order autoregressive process. For the purposes of this study data was taken in part through simple random sampling from actual data of 175,000 items. In each region and for every month the sample is taken as much as 5. The initial value are $\beta_0 = 0, \beta_1 = 0, \sigma_v^2 = 0.5$, and $\sigma_u^2 = 1$ which refer to the research of [5] and assuming the value of $\rho = 0.5$. Of the 35 registered incident areas, there are no occurrences of claims in the codes for events 12 and 28 for all months. While in certain areas of the incident, the occurrence of claims only occurred in a few months. This happens, for example, in the areas of events 9, 10, 11, 15, 17, 18, 23, 29, 30, and 35. Data in the occurrence areas are not included in the modeling.

Further modeling was carried out in part on observations with simple random sampling of 5 in each area of the incident and the month of occurrence. Then the data is divided in two based on the month of occurrence. Observations on month 1 to month 9 become training data and observations on the 10th month until the 12th month becomes testing data. Modeling is done on training data and applied to testing data.

2.2 Methods

The number of claims are count data which are assumed to have Poisson distribution, so the proposed model is

$$E(y_{itj} | v_i, u_{it}) = \mu_{itj}$$

where $\eta_{itj} = x_{itj} \beta + v_i + u_{it}$, $u_{it} = \rho u_{i,t-1}, |\rho| < 1$,
and $\log \mu_{itj} = \eta_{itj}$ (1)

with $\mathbf{y} = \{y_{itj}\}$ is a vector of observations from random variable Y with $i = 1, 2, \dots, m, t = 1, 2, \dots, T$ and $j = 1, 2, \dots, n_{it}$, respectively is the region of occurrence, period of events, and unit of observation. Therefore y_{itj} is assumed to be Poisson (μ_{itj}), or $E(y_{itj} | v_i, u_{it}) = \mu_{itj}$,
then the probability density function of y_{itj} with the terms v_i and u_{it} is

$$f(y_{itj} | v_i, u_{it}) = \frac{1}{y_{itj}!} \exp[y_{itj} \log(\mu_{itj}) - \mu_{itj}] \quad (2)$$

with $v_i \sim iid N(0, \sigma_v^2)$ is the random effect of the event region i , and the component $u_{it} \sim iid N(0, \sigma_u^2)$ is the random influence of region-time assumed to follow the first-order autoregressive process. The random effects v_i and u_{it} are assumed to be mutually independent.

Suppose there is one covariate variable, then for each region i at time t equation (1) can be denoted in the form of the following matrix:

$$\begin{bmatrix} \log \mu_{it1} \\ \log \mu_{it2} \\ \vdots \\ \log \mu_{itn_{it}} \end{bmatrix} = \begin{bmatrix} 1 & x_{it1} \\ 1 & x_{it2} \\ \vdots & \vdots \\ 1 & x_{itn_{it}} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + v_i \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + u_{it} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \eta_{it1} \\ \eta_{it2} \\ \vdots \\ \eta_{itn_{it}} \end{bmatrix} = \begin{bmatrix} 1 & x_{it1} \\ 1 & x_{it2} \\ \vdots & \vdots \\ 1 & x_{itn_{it}} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + v_i \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + u_{it} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

For each region i , Equation (1) can be denoted in the form of the following matrix

$$\begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \vdots \\ \eta_{iT} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{i1} \\ \mathbf{X}_{i2} \\ \vdots \\ \mathbf{X}_{iT} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + v_i \begin{bmatrix} \mathbf{1}_{n_{i1}} \\ \mathbf{1}_{n_{i2}} \\ \vdots \\ \mathbf{1}_{n_{iT}} \end{bmatrix} + \begin{bmatrix} \mathbf{1}_{n_{i1}} & \mathbf{0}_{n_{i1}} & \cdots & \mathbf{0}_{n_{i1}} \\ \mathbf{0}_{n_{i2}} & \mathbf{1}_{n_{i2}} & \cdots & \mathbf{0}_{n_{i2}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_{iT}} & \mathbf{0}_{n_{iT}} & \cdots & \mathbf{1}_{n_{iT}} \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}$$

Or

$$\boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta} + v_i \mathbf{1}_{n_i} + (\text{blockdiagonal}_{1 \leq t \leq T} \mathbf{1}_{n_{it}}) \mathbf{u}_i, \quad (3)$$

where $\boldsymbol{\eta}_i = (\boldsymbol{\eta}'_{i1}, \boldsymbol{\eta}'_{i2}, \dots, \boldsymbol{\eta}'_{iT})$, $\boldsymbol{\eta}_{it} = (\eta_{it1}, \dots, \eta_{itn_{it}})'$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{iT})'$, $\mathbf{X}_{it} = (1, x_{itj})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{iT})'$, and $n_i = \sum_{t=1}^T n_{it}$. Component $\mathbf{1}_{n_i}$ dan $\mathbf{1}_{n_{it}}$ respectively are vector of size n_i and n_{it} which all element are 1.

For the case of the number of balanced observations in each event area at time t or $n_{i1} = n_{i2} = \dots = n_{iT}$, model (1) becomes

$$\boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta} + v_i \mathbf{1}_{n_i} + (\mathbf{I}_T \otimes \mathbf{1}_{n_{it}}) \mathbf{u}_i \quad (4)$$

with \mathbf{I}_T is an identity matrix measuring $T \times T$ and \otimes is a kronecker product.

If defined $\mathbf{Z}_{1i} = \mathbf{1}_{n_i}$ and $\mathbf{Z}_{2i} = \mathbf{I}_T \otimes \mathbf{1}_{n_{it}}$ then the model (4) above becomes

$$\boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{1i} v_i + \mathbf{Z}_{2i} \mathbf{u}_i. \quad (5)$$

Model for all regions to be

$$\begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{11} & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_{1m} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{21} & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_{2m} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_m \end{bmatrix}$$

or

$$\boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{v} + \mathbf{Z}_2 \mathbf{u} \quad (6)$$

where $\mathbf{Z}_1 = \mathbf{I}_m \otimes \mathbf{Z}_{1i}$, $\mathbf{Z}_2 = \mathbf{I}_m \otimes \mathbf{Z}_{2i}$, $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \boldsymbol{\eta}'_2, \dots, \boldsymbol{\eta}'_m)'$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)'$, $\mathbf{v} = (v_1, v_2, \dots, v_m)'$, $\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_m)'$, and \mathbf{I}_m is identity matrix size $m \times m$.

On model (1) assumed $v_i \sim iid N(0, \sigma_v^2)$, so the expected value and covariance matrix of vector $\mathbf{v} = (v_1, \dots, v_m)'$ are $E(\mathbf{v}) = \mathbf{0}$ dan $Cov(\mathbf{v}) = \mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m$.

On model (5) \mathbf{u}_i is assumed independent so the expected value and covariance matrix of vector $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_m)'$ are $E(\mathbf{u}) = \mathbf{0}$ dan $Cov(\mathbf{u}) = \mathbf{G}_2 = \mathbf{I}_m \otimes \sigma_u^2 \boldsymbol{\Gamma}$. With $\boldsymbol{\Gamma}$ is symmetrical matrix size $T \times T$ which element (t, t') is $\frac{\rho^{|t-t'|}}{(1-\rho^2)}$, $t = 1, \dots, T$ and $t' = 1, \dots, T$.

The form of matrix Γ is
$$\Gamma = \frac{1}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix}.$$

Suppose $\mathbf{y} = \{y_{itj}\}$ is a vector of observations from random variable \mathbf{Y} with $i = 1, 2, \dots, m, t = 1, 2, \dots, T$, and $j = 1, 2, \dots, n_{it}$. Let $f_1(\mathbf{y}|\mathbf{v}, \mathbf{u})$, the probability density function of \mathbf{y} with the conditions $\{\mathbf{v}, \mathbf{u}\}$, $f_2(\mathbf{v})$ the probability density function \mathbf{v} , and $f_3(\mathbf{u})$, the probability density function \mathbf{u} . $\mathbf{y}_i, i = 1, \dots, m$, have Poisson distribution while $f_2(\mathbf{v})$ and $f_3(\mathbf{u})$ each have normal distribution with zero mean and variance $\mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m$ and $\mathbf{G}_2 = \mathbf{I}_m \otimes \sigma_u^2 \mathbf{\Gamma}$. The log-likelihood function of $f_1(\mathbf{y}|\mathbf{v}, \mathbf{u})$, $f_2(\mathbf{v})$ dan $f_3(\mathbf{u})$, is

$$l_1 = \ln f_1(\mathbf{y}|\mathbf{v}, \mathbf{u}) = \text{Const} + \sum_{i=1}^m \{\mathbf{y}_i \boldsymbol{\eta}_i - \exp(\boldsymbol{\eta}_i)\} \quad (7)$$

$$l_2 = \ln f_2(\mathbf{v}) = \text{Const} - \frac{1}{2} \{\ln |\mathbf{G}_1| + \mathbf{v}^T \mathbf{G}_1^{-1} \mathbf{v}\}$$

$$l_3 = \ln f_3(\mathbf{u}) = \text{Const} - \frac{1}{2} \{\ln |\mathbf{G}_2| + \mathbf{u}^T \mathbf{G}_2^{-1} \mathbf{u}\}.$$

For \mathbf{G}_1 and \mathbf{G}_2 which are fixed, $\boldsymbol{\beta}, \mathbf{v}$, dan \mathbf{u} which maximize the value of $l = l_1 + l_2 + l_3$ are referred to as the maximum penalized quasi likelihood (MPQL) estimators. Entering the MPQL estimator value into Equation (4) produces an estimate of $\boldsymbol{\eta}$.

If σ_v^2, σ_u^2 , and ρ are known, for instance $\sigma_v^2 = 0.5, \sigma_u^2 = 1, \rho = 0.5$ and the iterative procedure to obtain MPQL estimators from $\boldsymbol{\beta}, \mathbf{v}$, dan \mathbf{u} is as follows:

1. Iteration initialization. Starting from $k = 0$ and the initial values $\boldsymbol{\beta} = \mathbf{0}$.

2. Get
$$\Gamma = \frac{1}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix}, \mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m, \text{ and } \mathbf{G}_2 = \mathbf{I}_m \otimes \sigma_u^2 \mathbf{\Gamma}.$$

3. Get the initial value $\mathbf{v} \sim N(\mathbf{0}, \mathbf{G}_1)$, and $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}_2)$.

4. Update the values $\boldsymbol{\beta}, \mathbf{v}$, and \mathbf{u} with

$$\begin{bmatrix} \boldsymbol{\beta}_{k+1} \\ \mathbf{v}_{k+1} \\ \mathbf{u}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_k \\ \mathbf{v}_k \\ \mathbf{u}_k \end{bmatrix} + \mathbf{V}_k^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \end{bmatrix} \frac{dl_1}{d\boldsymbol{\eta}_k} - \mathbf{V}_k^{-1} \begin{bmatrix} 0 \\ \mathbf{G}_1^{-1} \mathbf{v}_k \\ \mathbf{G}_2^{-1} \mathbf{u}_k \end{bmatrix}$$

where

$$\mathbf{V}_k = - \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \end{bmatrix} \left(\frac{d^2 l_1}{d\boldsymbol{\eta}_k d\boldsymbol{\eta}_k} \right) \begin{bmatrix} \mathbf{X}' & \mathbf{Z}'_1 & \mathbf{Z}'_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_2^{-1} \end{bmatrix}$$

and $\frac{dl_1}{d\boldsymbol{\eta}_k}, \frac{d^2 l_1}{d\boldsymbol{\eta}_k d\boldsymbol{\eta}_k}$ are the first and second derivative l_1 of $\boldsymbol{\eta}$ evaluated on k

$$\frac{dl_1}{d\boldsymbol{\eta}_k} = \sum_{i=1}^m \mathbf{y}_i - \exp(\boldsymbol{\eta}_i) |_{k} = \sum_{i=1}^m (\mathbf{y}_i - \exp(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{v}_i + \mathbf{Z}_2 \mathbf{u}_i)) |_{k}$$

$$\frac{d^2 l_1}{d\boldsymbol{\eta}_k d\boldsymbol{\eta}_k} = \sum_{i=1}^m -\exp(\boldsymbol{\eta}_i) |_{k} = \sum_{i=1}^m (-\exp(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{v}_i + \mathbf{Z}_2 \mathbf{u}_i)) |_{k}.$$

5. Do step 2 until you get the values $\boldsymbol{\beta}, \mathbf{v}$, dan \mathbf{u} that are convergent.
6. The values of $\boldsymbol{\beta}, \mathbf{v}$, dan \mathbf{u} that are convergent are the values estimating $\boldsymbol{\beta}, \mathbf{v}$, dan \mathbf{u} with known σ_v^2 and σ_u^2 .

Suppose that $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{u}}$ are obtained from the above iterations, to get the estimator of the component variance σ_u^2 and σ_e^2 , the REML method will be used. As was done in [5] who followed [11] defining the adjusted variable

$$\mathbf{y}^* = g(\mathbf{y}) \cong g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}) = \boldsymbol{\eta} + \mathbf{e} \left(\frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}} \right) = \boldsymbol{\eta} + \mathbf{e}^*$$

where $\mathbf{e}^* = \mathbf{e} \left(\frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}} \right)$ so

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{v} + \mathbf{Z}_2\mathbf{u} + \mathbf{e}^*. \quad (8)$$

Variable $\mathbf{e}^* = \{e_i^*\}$ is adjusted error term with variance $Var(\mathbf{e}^*) = \tau \boldsymbol{\Sigma}_e^* = \boldsymbol{\Sigma}_e^*$ since $\tau = 1$ is dispersion parameter in Poisson model and $\boldsymbol{\Sigma}_e^{*-1} = diag \left[V(\mu_i)^{-1} \left\{ \frac{dg^{-1}(\eta_i)^2}{d\eta_i} \right\} \right]$.

Model (8) above is linear mixed models with

$$Var(\mathbf{y}^*) = Cov(\mathbf{Z}_1\mathbf{v}) + Cov(\mathbf{Z}_2\mathbf{u}) + Cov(\mathbf{e}^*) = \mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{G}_2\mathbf{Z}_2' + \boldsymbol{\Sigma}_e^*.$$

Supposed e_i^* is assumed independent so the expectation value and structure of the covariance matrix of $\mathbf{e}^* = (e_1^*, e_2^*, \dots, e_m^*)$ are

$$E(\mathbf{e}^*) = \mathbf{0} \text{ and } Cov(\mathbf{e}^*) = \boldsymbol{\Sigma}_e^* = \mathbf{I}_m \otimes \sigma_e^{2*} \mathbf{I}_{n_i} = \sigma_e^{2*} \mathbf{I}_n.$$

With assumption \mathbf{v} , \mathbf{u} , and \mathbf{e}^* are mutually independent, covariance matrix of \mathbf{y}^* :

$$\begin{aligned} \mathbf{V}^* &= Var(\mathbf{y}^*) = Cov(\mathbf{Z}_1\mathbf{v}) + Cov(\mathbf{Z}_2\mathbf{u}) + Cov(\mathbf{e}^*) \\ &= \mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{G}_2\mathbf{Z}_2' + \boldsymbol{\Sigma}_e^* \\ &= (\mathbf{I}_m \otimes \mathbf{Z}_{1i}) \sigma_v^2 \mathbf{I}_m (\mathbf{I}_m \otimes \mathbf{Z}_{1i})' + (\mathbf{I}_m \otimes \mathbf{Z}_{2i}) \sigma_u^2 \boldsymbol{\Gamma} (\mathbf{I}_m \otimes \mathbf{Z}_{2i})' + \boldsymbol{\Sigma}_e^* \\ &= \sigma_v^2 (\mathbf{I}_m \otimes \mathbf{Z}_{1i} \mathbf{Z}_{1i}') + \sigma_u^2 \boldsymbol{\Gamma} (\mathbf{I}_m \otimes \mathbf{Z}_{2i}) (\mathbf{I}_m \otimes \boldsymbol{\Gamma} \mathbf{Z}_{2i}') + \sigma_e^{2*} \mathbf{I}_n \\ &= \sigma_v^2 (\mathbf{I}_m \otimes \mathbf{Z}_{1i} \mathbf{Z}_{1i}') + \sigma_u^2 (\mathbf{I}_m \otimes \mathbf{Z}_{2i} \boldsymbol{\Gamma} \mathbf{Z}_{2i}') + (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_{ei}^*) \\ &= \mathbf{I}_m \otimes \left(\sigma_v^2 \mathbf{Z}_{1i} \mathbf{Z}_{1i}' + \sigma_u^2 \mathbf{Z}_{2i} \boldsymbol{\Gamma} \mathbf{Z}_{2i}' + \boldsymbol{\Sigma}_{ei}^* \right) \\ &= \mathbf{I}_m \otimes (\sigma_v^2 \mathbf{J}_{ni} + \sigma_u^2 (\boldsymbol{\Gamma} \otimes \mathbf{J}_{nit}) + \sigma_e^{2*} \mathbf{I}_{ni}). \end{aligned}$$

The steps to get variance component estimators are as follows:

1. Set the initial values for each variance component
For instance $\sigma_{v(0)}^2 = 0.5$, $\sigma_{u(0)}^2 = 1$, and $\sigma_{e(0)}^2 = 0.5$, then arrange the initial value as a vector measuring 3×1 , $\boldsymbol{\Omega}_0 = (\sigma_{v(0)}^2, \sigma_{u(0)}^2, \sigma_{e(0)}^2)'$
Make a matrix $\boldsymbol{\Gamma}$ which is symmetrical matrix $T \times T$ with $(t, t)'$'s element is $\frac{\rho^{|t-t'|}}{(1-\rho^2)}$, $t = 1, \dots, T$ and $t' = 1, \dots, T$.
2. Calculate
 $\mathbf{V}_0 = \mathbf{I}_m \otimes (\sigma_{v(0)}^2 \mathbf{J}_{ni} + \sigma_{u(0)}^2 (\boldsymbol{\Gamma} \otimes \mathbf{J}_{nit}) + \sigma_{e(0)}^{2*} \mathbf{I}_{ni})$.
Then calculate \mathbf{V}_0^{-1} .
3. Calculate $\mathbf{Q}_0^{-1} = [\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X}]^{-1}$.
4. Calculate $\mathbf{P}_0 = \mathbf{V}_0^{-1} - \mathbf{V}_0^{-1}\mathbf{X}\mathbf{Q}_0^{-1}[\mathbf{X}'\mathbf{V}_0^{-1}]$.
5. Calculate $\mathbf{P}_0\mathbf{y}^*$ with \mathbf{y}^* is in (8)
6. Calculate
 $s_{1(0)} = -\frac{1}{2} trace(\mathbf{P}_0(\mathbf{I}_m \otimes \mathbf{J}_{ni})) + \frac{1}{2} \mathbf{y}^{*'} \mathbf{P}_0(\mathbf{I}_m \otimes \mathbf{J}_{ni}) \mathbf{P}_0 \mathbf{y}^*$

$$s_{2(0)} = -\frac{1}{2} \text{trace} \left(\mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})) \right) + \frac{1}{2} \mathbf{y}^* \mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})) \mathbf{P}_0 \mathbf{y}^*$$

$$s_{3(0)} = -\frac{1}{2} \text{trace} (\mathbf{P}_0 \mathbf{I}_n) + \frac{1}{2} \mathbf{y}^* \mathbf{P}_0 \mathbf{I}_n \mathbf{P}_0 \mathbf{y}^*.$$

Then arrange as vector 3×1 , $\mathbf{s}_0 = (s_{1(0)}, s_{2(0)}, s_{3(0)})'$.

7. Make Fisher Information matrix \mathbf{F}_0 of 3×3 with each element are

$$\mathbf{F}_{11} = \frac{1}{2} \text{trace} (\mathbf{P}_0 (\mathbf{I}_m \otimes \mathbf{J}_{ni}) \mathbf{P}_0 (\mathbf{I}_m \otimes \mathbf{J}_{ni}))$$

$$\mathbf{F}_{12} = \mathbf{F}_{21} = \frac{1}{2} \text{trace} (\mathbf{P}_0 (\mathbf{I}_m \otimes \mathbf{J}_{ni}) \mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})))$$

$$\mathbf{F}_{13} = \mathbf{F}_{31} = \frac{1}{2} \text{trace} (\mathbf{P}_0 (\mathbf{I}_m \otimes \mathbf{J}_{ni}) \mathbf{P}_0 \mathbf{I}_n)$$

$$\mathbf{F}_{22} = \frac{1}{2} \text{trace} (\mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})) \mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})))$$

$$\mathbf{F}_{23} = \mathbf{F}_{32} = \frac{1}{2} \text{trace} (\mathbf{P}_0 (\mathbf{I}_m \otimes (\Gamma \otimes \mathbf{J}_{nit})) \mathbf{P}_0 \mathbf{I}_n)$$

$$\mathbf{F}_{33} = \frac{1}{2} \text{trace} (\mathbf{P}_0 \mathbf{I}_n \mathbf{P}_0 \mathbf{I}_n).$$

$$\text{And make } \mathbf{F}_0 = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{F}_{33} \end{bmatrix}.$$

8. Calculate \mathbf{F}_0^{-1} .

9. Calculate

$$\Omega_1 = \Omega_0 + \mathbf{F}_0^{-1} \mathbf{s}_0.$$

10. Do it all until Ω_1 convergence.

11. The value of Ω_1 that is convergent are the value estimating Ω and the estimates σ_v^2 and σ_u^2 are the component 1 and 2 of vector Ω .

Use $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ from the REML procedure above as initial value of MPQL iteration to get the new $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{v}}$, and $\hat{\mathbf{u}}$. And now the steps as follows :

1. Update value $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{v} + \mathbf{Z}_2\mathbf{u}$ using $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{v}}$, and $\hat{\mathbf{u}}$ from the latest MPQL iteration
$$\eta_{itj} = \beta_0 + x_{itj} \beta_1 + v_i + u_{it}$$
2. $E(y_{itj} | v_i, u_{it}) = \mu_{itj} = \exp(\eta_{itj})$.
3. $E(\mathbf{y}_{it} | v_i, u_{it}) = \sum_{j=1}^{n_{it}} E(y_{itj} | v_i, u_{it})$.
4. $E(\mathbf{y}_i | v_i, u_{it}) = \sum_{t=1}^T E(\mathbf{y}_{it} | v_i, u_{it})$

3 Result and Discussion

As explained earlier, modeling is done on training data. The variable that is of concern y_{itj} is number of claims to j from the region of events to i and the month of events to t with $i = 1, 2, \dots, 35$, $t = 1, 2, \dots, 10$ and $j = 1, 2, \dots, 5$. y_{itj} is assumed to follow Poisson distribution with $E(y_{itj} | v_i, u_{it}) = \mu_{itj}$ with the proposed model is

$$\log \mu_{itj} = \eta_{itj} = x_{itj} \boldsymbol{\beta} + v_i + u_{it}$$

$$u_{it} = \rho u_{i,t-1}, |\rho| < 1,$$

with x_{itj} is the size of the deductible on the claim to j from the region of occurrence to i and the month of occurrence to t , and ρ is the autoregressive coefficient specified, $v_i \sim iid N(0, \sigma_v^2)$ is the random effect of the event region i and the component $\rho = 0.5$. $u_{it} \sim iid N(0, \sigma_u^2)$ is the random effect of region-time which is assumed to follow the first-order autoregressive process. The random effects v_i and u_{it} are assumed to be mutually independent. The initial value are $\beta_0 = 0, \beta_1 = 0, \sigma_v^2 = 0.5$, and $\sigma_u^2 = 1$ which refers to the research of [5].

In Table 1, there is an estimator of the fixed effect coefficient of the model using the MPQL algorithm. In Table 2, the estimation of variance from random influences is used using the REML method.

Table 1 Fixed Effect Coefficient Estimator

	Estimate
β_0	1.33e-01
β_1	-1.24e-07

Table 2 Estimates of variance from random effects

	Estimate
Variance v	0.002
Variance u	0.001

The parameter estimation results in Table 1 and Table 2 are then used to find the estimation of number of claims to j in the month of occurrence t and the region of occurrence i, \hat{y}_{itj} . Then the estimator \hat{y}_{it} is obtained by summing \hat{y}_{itj} for all j .

To measure the accuracy of the estimates used MAPE. According to [8] the calculation of mean absolute percentage error (MAPE) is done with the following formula:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t} \times 100$$

with n number of observation, \hat{y}_t prediction value at time t , dan y_t observation at time t .

On Table 3 the calculation of MAPE's values is given in each region with training data. According to Palmer *et al.* (2013), MAPE's values that are less than 10% show highly accurate predictive values, MAPE values ranging from 10 to 20% show good predictive values, MAPE's values between 20 and 50% show fairly good predictions, and MAPE's values that are more than 50% shows inaccurate predictions. In Table 3, it can be seen that the MAPE's values in all regions of the event in training data is less than 20%. This shows that in all occurrences the model predictive value is good. Even in the 18 regions of the model showed very accurate predictive values. Only 5 regions have MAPE's values between 10 and 20%.

Table 3 MAPE's value on each region using training data

Area code	MAPE (%)	Area code	MAPE (%)
01	10.3	20	11.75
02	7.56	21	16.31
03	7.25	22	8.44
04	8.59	24	9.71
05	8.66	25	7.56
06	8.76	26	7.53
07	8.34	27	10.48
08	11.89	31	8.65
13	10.01	32	7.78
14	9.17	33	8.59
16	9.79	34	7.76
19	8.13		

Table 4 explain MAPE's value on each region using testing data. As similar to training data, the MAPE's value in the testing data also appears to be less than 20%. Even the MAPE value of testing data is generally smaller than the MAPE's value in the training data. Table 3 and Table 4 show us that the model showed very accurate to predict the number of claim in some region and time of occurrence.

Table 4 MAPE's value on each region using testing data

Area Code	MAPE (%)	Area Code	MAPE (%)
01	10.68	20	8.72
02	6.39	21	13.18
03	6.92	22	9.52
04	7.37	24	6.73
05	8.58	25	6.10
06	9.78	26	8.72
07	9.70	27	8.04
08	5.79	31	8.03
13	6.02	32	7.6
14	9.44	33	9.39
16	4.55	34	6.65
19	7.48		

4 Conclusion

Generalized linear mixed models can be used to predict the number of claims that occur in a particular area of events and time events. Assuming the number of claim is following a Poisson distribution and the timing of events following the first-order autoregressive process, it can be generated number of claim models with estimators of fixed effect parameters $\beta_0 = 1.33e - 01$ and $\beta_1 = -1.24e - 07$ and, estimator component variance $v = 0.02$ and component variance $u = 0.001$. Modeling on training data produces accurate models in almost all regions included in the model. Similarly, the testing data produced accurate models in all regions. This are indicated by the MAPE values which are less than 20% in all regions on both training and testing data.

References

- [1] Antonio K, Beirlant J, Hoedemaker T, Verlaak R : Lognormal mixed models for reported claim reserved. *North American Actuarial Journal*. 5(1) (2006)
- [2] Antonio K, Beirlant J : Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics* 40 58–76(2007)
- [3] Boubeta M, Lombardia MJ, Morales D : Empirical best prediction under area-level Poisson mixed models. *Test* 25:548–569. DOI 10.1007/s11749-015-0469-8 (2016)
- [4] Boubeta M, Lombardia MJ, Morales D : Poisson mixed models for studying the poverty in small areas. *Computational Statistics and Data Analysis* 107 : 32–47 (2017)
- [5] Chandra H., et al.. Small area estimation under a spatially non-linear model. *Computational Statistics and Data Analysis*(2018). <https://doi.org/10.1016/j.csda.2018.04.002>
- [6] Choi J, Lee K : Poisson linear mixed models with ARMA random effects covariance matrix. *Journal of the Korean Data & Information Science Society*. (2017). <http://dx.doi.org/10.7465/jkdi.2017.28.4.927>
- [7] Muchlisoh S : Pengembangan model pendugaan area kecil dengan pengaruh acak waktu mengikuti proses autoregresif ordo pertama [disertasi]. Bogor: Institut Pertanian Bogor.(2017)
- [8] Palmer A, Moreno JJM, Abad AS, Blesco BC : Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, Vol. 25, No. 4, 500-506. (2013)
doi: 10.7334/psicothema2013.23
- [9] Pratama, H : Pemodelan data asuransi mobil melalui regresi gamma [skripsi]. Bandung: Universitas Islam Bandung (2011)
- [10] Saei A, McGilchrist CA : Longitudinal threshold models with random components. *The Statistician* 47, Part 2, pp. 365-375.(1998)
- [11] Schall, R : Estimation in generalized linear models with random effects. *Biometrika* 78, 719–727 (1991)
- [12] Sutradhar B, Qu Z : On approximate likelihood inference in a Poisson mixed model. *The Canadian Journal of Statistics*. 26(1): 169-186. (1998)