Modeling The Amount of Insurance Claim using Gamma Linear Mixed Model with AR (1) random effect

F A Adam¹, A Kurnia*, I G P Purnaba², I W Mangku², A M Soleh¹

¹Department of Statistics, IPB University, Bogor, 16680, Indonesia
²Department of Mathematics, IPB University, Bogor, 16680, Indonesia
³Program Vokasi, Universitas Indonesia, Depok, 16424, Indonesia

*E-mail: anangk@apps.ipb.ac.id

Abstract. The amount of insurance claims is continuous data and is positive so it is usually assumed to have gamma distribution. Usually, Generalized Linear Model (GLM)’s approach is used since the gamma distribution is a member of the exponential family. In case there is a random effect in modeling, then GLM can be extended to Generalized Linear Mixed Model (GLMM). This study models the amount of insurance claims with the most GLMM’s approach using two random effects, namely the region and time of the occurrence which is assumed to follow a first-order autoregressive process. The h-likelihood method is used to estimate the regression parameters and the variance parameters. A simulation study is carried out with an evaluation using the average relative bias and the average MSE. An application study which is conducted to model the amount of insurance claims in a certain region and time based on the 2014 profile of risk and loss of motor vehicle insurance in Indonesia is also carried out.

1. Introduction
To pay obligations to its customers, the insurance company must be able to estimate how much of this obligation. For this reason, the insurance company must be able to estimate number of claims and the claim amount. Number of claim or claim frequency is a measure of how often claims occur. The claim amount or claim severity is a measure of how big the claim is. Number of claims are approached as count data, while the claim amount is considered as continuous data. Because the value is always positive, the claim amount is assumed to have gamma distribution.

Generalized linear model (GLM) is can be used to model a gamma distribution. For example, GLM with gamma distribution is used to model insurance claims as Pratama [8]. If there are fixed and random effects, GLM become generalized linear mixed model (GLMM) which is more flexible in data analysis. Several studies that model gamma distribution with GLMM include Zhang et al. [11], Bossio and Cuervo [2], and Ribeiro et al. [9].

Estimating parameters in GLMM with response variables that have gamma distribution is not analytically easy to do. For this reason, the approach taken is numerical based. Zhang et al. [11] used the Gauss-Newton algorithm to estimate regression parameters and the MCMC method to obtain variance parameters. Another alternative method is the hierarchical likelihood (h-likelihood) developed by Lee and Nelder [6]. The h-likelihood method is not like the ordinary likelihood method. Fixed and random effects combine to form an enlarged likelihood function. And also, the random effects do not have to come from normal distribution as usually random effects in GLMM. This is like in the research
of Lee and Lee [5] which evaluated a model with random effect had gamma distribution. The use of integrals to obtain marginal likelihoods is avoided in this method. Another advantages in this method is the ability to obtain the regression and variance parameter estimators.

GLMM with one random effect is used by Ribeiro et al. [9] to model gamma response variables. Supposed one other random effect is added, can GLMM with response variables that have gamma distribution still be used? So our model is a gamma mixed model with two random effects which are the area and time with first order autoregressive process assumption. This model is similar as Muchlisoh [7] did, unless the distribution used by Muchlisoh [7] was normal distribution. In Adam et al. [1], a similar model was also developed but the response variable used is compound Poisson distribution. Furthermore, to estimate the regression parameters and variance parameters, can the h-likelihood method be used to obtain all these parameters?

The aim of this research is to develop a gamma mixed model with two random effects, which are the area and time with a first-order autoregressive process assumption. To estimate regression and variance parameter, the h-likelihood method is then used. A simulation study to evaluate the goodness of the model is carried out by measuring its relative bias and mean square error (MSE). Application on real data modeling is carried out using data from one of the general insurance companies registered with the FSA (Financial Service Authority) Indonesia on year 2014. The response variable used is the claim amount, deductible is a covariate, area code and month of occurrence with a first-order autoregressive process assumption are the random effects.

In section 2, model and method are presented in detail. Simulation study to measure the model performance is in section 3. Application study in section 4 is done to model the amount of claims of the general insurance companies in some area and time. Section 5 is the conclusion.

2. Model and Methods

2.1. Basic Model

By adding the dispersion parameters, an exponential family distribution is defined as the exponential dispersion model (EDM). In Jorgensen [4], the density function of the random variable $Y$ with the distribution including the EDM family is

$$f(y; \theta, d) = l(y; \phi) \exp \left( \frac{1}{d} \left( y\theta - \kappa(\theta) \right) \right),$$  \hspace{1cm} (1)

where $\kappa$ and $l$ are known function, $d > 0$ is dispersion parameter, $\kappa(\theta)$ is cumulant function, and $\theta$ is the natural parameter. A characteristic of EDM is the mean-variance relation if the dispersion parameter is considered constant. In other words, if $Y$ has EDM distribution with mean $\mu$, variance function $Var(\cdot)$, and dispersion parameter $d$ then $Var(Y) = d Var(\mu)$. If $Var(\mu) = \mu^p$, where $p$ is index parameter, then $Var(Y) = d \mu^p$.

If a random variable $Y$ has gamma distribution with mean $(\mu > 0)$, and shape parameter $\alpha$ $(\alpha > 0)$ then the density of $Y$ is

$$f(y; \mu, \alpha) = \frac{1}{\Gamma(\alpha)} \left( \frac{\alpha y}{\mu} \right)^\alpha e^{-\frac{y}{\mu}} \left( \frac{1}{\alpha} \right) I_{(0, \infty)}(y)$$  \hspace{1cm} (2)

where $\Gamma(\cdot)$ is gamma function. In this situation, $E(Y) = \mu$ dan $r(y) = \frac{\mu^2}{\alpha}$. Supposed $Y$ has the density function as in equation (1) above, then $Y \sim gamma(\mu, \alpha)$. As $Var(y) = \frac{\mu^2}{\alpha}$ then gamma distribution is in EDM family with $p = 2$ and $d = \frac{1}{\alpha}$.

Supposed $Y_1, ..., Y_i$ are independent random variables such that $Y_j \sim gamma(\mu_j, \alpha)$ and if $E(Y_j) = \mu_j$ is linked to covariate variable through

$$\eta_j = g(\mu_j) = x\beta + Z\beta_j$$  \hspace{1cm} (3)
where \( \beta \) is regression parameter vector, \( x_j \) covariate vector of \( i \)th observations, \( Z_j \) is \( j \)th incidence matrix, \( b_j \) is a random effect, and \( g(\cdot) \) is a link function.

For all \( j = 1, ..., i \) then Bossio and Cuervo [2] explained the above equation as

\[
\eta = X\beta + Zb
\]

(4)

where \( g(\mu) = \eta, \beta \) is a fixed effect vector, \( X \) is a covariate matrix, \( Z \) is incidence matrix, and \( b \) is a random effect where \( b \sim N(0, \Sigma) \)

Supposed the variance \( \Sigma \) can be explained as \( \Sigma = \phi \Delta \Delta' \) as in Zhang [10] where \( \Delta \) is the relative covariance factor. The equation (4) is now defined as

\[
\eta = X\beta + Z\Delta b^* = X\beta + Z^*b^*
\]

(5)

where \( b^* \sim N(0, \phi I) \).

2.2 Proposed Model.

The proposed model is a development of the Bossio and Cuervo [2] model by adding a time random effect which is assumed to follow a first-order autoregressive process. Let \( y_{jtk} \) be the observation \( k \) in the area \( j \) and time \( t \), where \( j = 1, ..., i; t = 1, ..., T; k = 1, ..., s_jt \), and \( y_{jtk} \) have gamma distribution assumption and is related to an explanatory variable \( x_{jtk} \) through the model

\[
E(y_{jtk}; u_t) = \mu_{jtk}, \\
\log \mu_{jtk} = \beta_0 + x_{jtk}'\beta_1 + b_j + c_t, \\
c_t = \rho u_{t-1} + \xi_t, |\rho| < 1, \quad (6)
\]

where \( \beta_0 \) and \( \beta_1 \) are fixed effect vectors, \( b_j \sim iid \ N(0, \sigma_b^2) \) is the random effect of the \( j \)-th area, the \( c_t \) is the random effect of the time assumed to follow the first-order autoregressive process where \( \xi_t \) is an error of the \( c_t \) assumed \( \xi_t \sim iid \ N(0, \sigma^2) \), and \( \rho \) is the autoregressive coefficient. The independence assumption is exist between \( b_j \) and \( c_t \).

If there is one explanatory variable, then for each area \( j \) at time \( t \), equation (6) is then stated as

\[
\begin{bmatrix}
\log \mu_{j1t} \\
\log \mu_{j2t} \\
\vdots \\
\log \mu_{jstt}
\end{bmatrix}
= 
\begin{bmatrix}
1 & x_{j1t} \\
1 & x_{j2t} \\
\vdots \\
1 & x_{jstt}
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_{stt}
\end{bmatrix}
+ \begin{bmatrix}
b_j \\
b_j \\
\vdots \\
b_j
\end{bmatrix}
+ \begin{bmatrix}
c_t \\
c_t \\
\vdots \\
c_t
\end{bmatrix}.
\]

(7)

If \( \log \mu_{jtk} = \eta_{jtk} \) then the equation (7) is transformed into

\[
\begin{bmatrix}
\eta_{j1t} \\
\eta_{j2t} \\
\vdots \\
\eta_{jstt}
\end{bmatrix}
= 
\begin{bmatrix}
1 & x_{j1t} \\
1 & x_{j2t} \\
\vdots \\
1 & x_{jstt}
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_{stt}
\end{bmatrix}
+ \begin{bmatrix}
1 \\
1 \\
\vdots \\
1
\end{bmatrix}
+ \begin{bmatrix}
c_t \\
c_t \\
\vdots \\
c_t
\end{bmatrix}.
\]

(8)

So for area \( j \), equation (6) is

\[
\begin{bmatrix}
\eta_{j1} \\
\eta_{j2} \\
\vdots \\
\eta_{jt}
\end{bmatrix}
= 
\begin{bmatrix}
x_{j1} \\
x_{j2} \\
\vdots \\
x_{jt}
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_j \\
\beta_{jt}
\end{bmatrix}
+ \begin{bmatrix}
1_{s_j} \\
1_{s_j} \\
\vdots \\
1_{s_j}
\end{bmatrix}
+ \begin{bmatrix}
1_{s_t} \\
1_{s_t} \\
\vdots \\
1_{s_t}
\end{bmatrix}
.
\]

(9)
If it is assumed that number of observation is balanced and $Z_{ij} = 1_{s_j}$ and also $Z_{2t} = 1_{s_t}$ the equation (9) becomes

$$\eta_j = X_j \beta + Z_{1j} b_j + Z_{2t} c_t.$$  \hspace{1cm} (10)

Model for all area is

$$\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\vdots \\
\eta_l
\end{bmatrix} = \begin{bmatrix} X_1 \\
X_2 \\
\vdots \\
X_l
\end{bmatrix} \begin{bmatrix} \beta_0 \\
\beta_1 \\
\vdots \\
\beta_l
\end{bmatrix} + \begin{bmatrix} Z_{11} & 0 & \cdots & 0 \\
0 & Z_{12} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & Z_{1l}
\end{bmatrix} \begin{bmatrix} b_1 \\
b_2 \\
\vdots \\
b_l
\end{bmatrix} + \begin{bmatrix} Z_{21} & 0 & \cdots & 0 \\
0 & Z_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & Z_{2T}
\end{bmatrix} \begin{bmatrix} c_1 \\
c_2 \\
\vdots \\
c_T
\end{bmatrix}$$

or

$$\eta = X \beta + Z_1 b + Z_2 c$$  \hspace{1cm} (11)

where $Z_1 = I_t \otimes Z_{1j}$, $Z_2 = I_T \otimes Z_{2t}$, $\eta = (\eta_1, \eta_2, \ldots, \eta_l)'$, $X = (X_1, X_2, \ldots, X_l)'$, $b = (b_1, b_2, \ldots, b_l)'$, $c = (c_1, c_2, \ldots, c_T)'$, $I_t$ is identity matrix sized $l \times l$, and $I_T$ is identity matrix sized $T \times T$.

From equation (6) is known that $b_j \sim iid N(0, \sigma_b^2)$. The expectation value and covariance matrix of vector $b = (b_1, b_2, \ldots, b_l)'$ are

$$E(b) = 0 \text{ and } Cov(b) = C_1 = \sigma_b^2 I_t.$$  \hspace{1cm} (12)

On equation (6) is also known that $c_t$ is mutually independent and have the first-order autoregressive process assumption so the expectation value and covariance matrix of vector $c = (c_1', c_2', \ldots, c_T')'$ are

$$E(c) = 0 \text{ dan } Cov(c) = C_2 = \frac{1}{(1-\rho^2)^*} \sigma_c^2 \Gamma,$$  \hspace{1cm} (13)

where $\Gamma$ is symmetrical matrix sized $T \times T$ with element $(t, t')$ is $\rho^{t-t'|}$, $t = 1, \ldots, T$ and $t' = 1, \ldots, T$. Matrix $\Gamma$ is

$$\Gamma = \begin{bmatrix}
1 & \rho & \cdots & \rho^{T-1} \\
\rho & 1 & \cdots & \rho^{T-2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho^{T-2} & \cdots & \rho & 1
\end{bmatrix}.$$  \hspace{1cm} (14)

Equation (12) and equation (13) are changed to the form of relative covariance factor $\Delta_1$ and $\Delta_2$ so that

$$C_1 = \sigma_b^2 \Delta_1 \Delta_1' \text{ and } C_2 = \sigma_c^2 \Delta_2 \Delta_2'.$$

where $\Delta_1 \Delta_1' = I_t$, $\Delta_1$ is a Cholesky decomposition matrix of $I_t$, and $\Delta_2 \Delta_2' = \frac{r}{(1-\rho^2)^*}$ where $\Delta_2$ is a Cholesky decomposition matrix of first-order autoregressive correlation matrix $\frac{r}{(1-\rho^2)^*}$.

As a consequence, the proposed model on equation (6) is transformed into

$$\eta = X \beta + Z_1 \Delta_1 b^* + Z_2 \Delta_2 c^* = X \beta + Z_1 b^* + Z_2 c^*$$  \hspace{1cm} (15)
where \( \mathbf{b}^* \sim N(0, \sigma_b^2 \mathbf{I}) \) and \( \mathbf{c}^* \sim N \left( 0, \frac{1}{(1-\rho^2)} \sigma_c^2 \mathbf{I} \right) \).

If link function is a logarithmic then

\[
\eta = \log \mu = X \beta + Z_1^* \mathbf{b}^* + Z_2^* \mathbf{c}^* \quad \text{or} \quad \mu = \exp \eta = \exp (X \beta + Z_1^* \mathbf{b}^* + Z_2^* \mathbf{c}^*)
\]

where \( \mathbf{b}^* \sim N(0, \sigma_b^2 \mathbf{I}) \) and \( \mathbf{c}^* \sim N \left( 0, \frac{1}{(1-\rho^2)} \sigma_c^2 \mathbf{I} \right) \).

3. Parameter Estimation

The h-likelihood method is used to get the all parameter estimates which are \( \hat{\beta}, \hat{\mathbf{b}}^*, \) and \( \hat{\mathbf{c}}^* \). According to Lee and Nelder [5] h-likelihood is defined as

\[
h = h_1 + h_2 + h_3
\]

where \( h_1 = \log f(\mathbf{y}; \mathbf{b}^*, \mathbf{c}^*) \) is the log-density function for \( \mathbf{y} \) given \( \mathbf{b}^* \) and \( \mathbf{c}^* \), \( h_2 \) is the log-density function for \( \mathbf{b}^* \) with \( \mathbf{b}^* \sim N(0, \sigma_b^2 \mathbf{I}) \), and \( h_3 \) is the log-density function for \( \mathbf{c}^* \) with \( \mathbf{c}^* \sim N \left( 0, \frac{1}{(1-\rho^2)} \sigma_c^2 \mathbf{I} \right) \).

In equation (6), it is assumed \( \mathbf{y} \) has gamma distribution so \( \mathbf{y} \) become a member MDE family with \( p = 2 \) and dispersion parameter \( d = \frac{1}{\alpha} \) with \( \alpha \) is gamma shape parameter.

The parameter estimation values is obtain by maximizing the h-likelihood function \( h \). This is done with the solution of the equation \( \frac{dh}{d\beta} = 0, \frac{dh}{d\mathbf{b}} = 0, \) and \( \frac{dh}{dc} = 0 \) and the approach used is numerical based.

Parameter estimation can be done by solving the Newton-Raphson iteration below

\[
\begin{bmatrix}
\beta_{r+1} \\
\mathbf{b}_{r+1}^* \\
\mathbf{c}_{r+1}^*
\end{bmatrix} =
\begin{bmatrix}
\beta_r \\
\mathbf{b}_r^* \\
\mathbf{c}_r^*
\end{bmatrix} +
H_r^{-1}
\begin{bmatrix}
\frac{dh}{d\beta} \\
\frac{dh}{d\mathbf{b}} \\
\frac{dh}{dc}
\end{bmatrix}
\]

where Hessian matrix \( H \) is

\[
H =
\begin{bmatrix}
-E \left( \frac{d^2h}{d\beta^2} \right) & -E \left( \frac{d^2h}{d\beta d\mathbf{b}} \right) & -E \left( \frac{d^2h}{d\beta dc} \right) \\
-E \left( \frac{d^2h}{d\mathbf{b} d\beta} \right) & -E \left( \frac{d^2h}{d\mathbf{b}^* d\beta} \right) & -E \left( \frac{d^2h}{d\mathbf{b}^* dc} \right) \\
-E \left( \frac{d^2h}{dc d\beta} \right) & -E \left( \frac{d^2h}{dc d\mathbf{b}} \right) & -E \left( \frac{d^2h}{dc^2} \right)
\end{bmatrix}
\]

In every iteration, \( (r + 1)^{th} \) solutions must meet the equation (18) and convergence is achieved.

After the estimates of \( \beta, \mathbf{b}^*, \) and \( \mathbf{c}^* \) are obtained, the next step is to estimate the variance parameters. Lee and Nelder [6] defined adjusted hierarchical likelihood \( h_A \) to find the estimation of variance parameters as

\[
h_A = h - \frac{1}{2} \ln \left\{ \det \left( \frac{H}{2\pi} \right) \right\}
\]

where \( h \) is h-likelihood function from equation (15) and \( H \) is a Hessian matrix from equation (17).

The adjusted profile hierarchical likelihood is \( h_p = h_{A|_{\beta=\hat{\beta}, \mathbf{b}^*=\hat{\mathbf{b}}^*, \mathbf{c}^*=\hat{\mathbf{c}}^*}} \) where \( \hat{\beta}, \hat{\mathbf{b}}^*, \) and \( \hat{\mathbf{c}}^* \) are taken from equation (16).

Variance parameters \( \sigma_b^2 \) and \( \sigma_c^2 \) is solved iteratively by using the equation below
\[
\begin{pmatrix}
\sigma^2_{\tau r+1} \\
\sigma^2_{\xi r+1}
\end{pmatrix} =
\begin{pmatrix}
\sigma^2_{\beta r} \\
\sigma^2_{\zeta r}
\end{pmatrix} + J^{-1}
\begin{bmatrix}
\frac{\partial h_A}{\partial \sigma^2_{\beta r}} \\
\frac{\partial h_A}{\partial \sigma^2_{\zeta r}}
\end{bmatrix}
\begin{pmatrix}
\sigma^2_{\beta r} = \sigma^2_{\beta r} \\
\sigma^2_{\zeta r} = \sigma^2_{\zeta r}
\end{pmatrix}
\] (21)

until convergence is met. The \( J \) matrix itself is a Hessian matrix of the second derivatives of \( h_A \).

4. Simulation Study

4.1. Simulation Design

In this design below, a simulation study is conducted to measure the performance of the proposed model.

1. The data is generated under these following conditions

   - 1.1 The total area is 35 (\( i = 35 \)) and observation time is 12 (\( T = 12 \)). There are 10 observations in each area and time so there are \( S = 4200 \) observations in total.
   - 1.2 Three types of autoregressive coefficient which are \( \rho = 0.2, \rho = 0.5, \) and \( \rho = 0.8 \).
   - 1.3 Three types of variance from the area random effect which are \( \sigma^2_{\beta} = 0.1, \sigma^2_{\beta} = 0.5, \) and \( \sigma^2_{\beta} = 1.0 \).

   - 1.4 Set \( \sigma^2_{\zeta} = 0.5 \), dispersion parameter \( d = 1 \), and index parameter \( p = 1.5 \).

   - 1.5 An explanatory variable \( x \) is generated as \( x \sim \text{Normal}(0, 1) \).

   - 1.6 Assign the initial value \( \beta_0 = (0, 0) \), area \( b_0 \), with \( b_0 \sim \text{N}(0, \sigma^2_\beta I_1) \), and time \( c_0 \), where \( c_0 \sim N \left( 0, \frac{\sigma^2_\zeta}{(1-\rho^2)} I_T \right) \). The values \( \nu_\beta, \sigma^2_{\beta}, \) and \( \sigma^2_{\zeta} \) are from steps 1.2, 1.3, and 1.4 above.

   - 1.7 Matrices \( Z_1 = Z_1 \Delta_1 \) and \( Z_2 = Z_2 \Delta_2 \) are then obtained with \( Z_1 = I_t \otimes Z_{1j}, Z_2 = I_T \otimes Z_{2t} \).

   - 1.8 Response variable \( y \) has been generated by gamma distribution assumption with parameter

\[ \mu(y; b_0, c_0) = \mu = \exp(X \beta_0 + Z_1 \beta + Z_2 c) \] and \( \text{shape parameter } \alpha = 1 \).

2. The parameters of the fixed effect \( \beta \), the area random effect \( \beta^* \), and the time random effect \( \zeta^* \) are estimated by h-likelihood method until convergent.

3. The variance component \( \sigma^2_{\hat{\beta}} \) and \( \sigma^2_{\hat{\zeta}} \) are estimated by adjusted profile likelihood \( h_A \) method until convergent.

4. Rerun stage 2 with the initial values \( \hat{\beta}, \hat{\beta}^*, \hat{\zeta}^* \), \( \sigma^2_{\hat{\beta}} \), and \( \sigma^2_{\hat{\zeta}} \) taken from step 2 and 3.

5. Determine the value \( \hat{\eta} = \hat{X} \hat{\beta} + Z_1 \hat{\beta}^* + Z_2 \hat{\zeta}^* \) then \( \hat{\mu} = \exp \hat{\eta} = E(\hat{y}; \hat{\beta}, \hat{\zeta}) \).

6. Do stages 1 through 5 as much as \( U = 100 \) times.

7. Calculate

- Mean Relative Bias \( = \frac{1}{U} \sum_{i=1}^{U} \frac{(\hat{\mu}_i - \mu_i)}{\mu_i} \).
- \( MSE = \frac{1}{U} \sum_{i=1}^{U} (\hat{\mu}_i - \mu_i)^2 \)

4.2 Simulation Result

The non-reversible Hessian matrix is always generated by the model using direct generation data. It is important to rescale the explanatory and the response variables. Both variables are divided by 10000 in order to make an invertible Hessian matrix. Now, the next process based on the algorithm above can be done. Model 1 as in equation (6) is then compared to Model 2 below.
\[ E(y_{jtk}; b_j, c_t) = \mu_{jtk}, \]
\[ \log \mu_{jtk} = \beta_0 + \gamma_{jtk} \beta_1 + b_j + c_t, \]  

(22)

a model that has no autoregressive assumptions on the time random effect. Model 2 has the same initial values as Model 1.

In this simulation study, the h-likelihood method is done to get all the parameters estimate. The R programming software is used to build a computational program. As in Muchlisoh [7], the autoregressive coefficient values are assumed known. In addition, the shape parameter is also known. Table 1 and Figure 1 has described the simulation results of 100 replications.

| Table 1. Simulation result from 100 replications |
|---|
| Model 1 with time AR (1) |
| \( \rho \) & 0.2 & 0.5 & 0.8 |
| Variance | 0.1 | 0.5 | 1 | 0.1 | 0.5 | 1 | 0.1 | 0.5 | 1 |
| Relative Bias | 0.2577 | 0.0492 | 0.0400 | 0.0890 | 0.0371 | 0.0385 | 0.0701 | 0.0251 | 0.0290 |
| MSE | 4.40E-07 | 1.60E-08 | 6.12E-09 | 4.73E-09 | 1.41E-07 | 4.83E-08 | 0.0008 | 0.0006 | 0.1069 |
| Model 2 without time AR (1) |
| Variance | 0.1 | 0.5 | 1 |
| Relative Bias | 0.0823 | 0.0530 | 0.0433 |
| MSE | 3.06E-08 | 6.10E-09 | 4.03E-08 |

From Table 1 Model 1, for the same autoregressive coefficient, the greater the area variances the smaller the relative bias is. Then, the greater the autoregressive coefficient the smaller the relative bias’s result. As a consequence, the greater the autoregressive coefficient, the more unbiased the estimator can be. In addition, the greater the area variances the more unbiased the estimators produced. Hereinafter, the autoregressive coefficient and area variance have an effect to the biasness of Model 1. Comparing to Model 2 at small variance, Model 1 with a small autoregressive coefficient has more bias than Model 2. This is known by the smaller relative bias value of Model 2. But other than that, Model 1 always has estimators with smaller relative bias values. Figure 1 provides an explanation of this. Generally, it can be seen from Table 1, the MSE values of the two models closed to zero. This means that both models can explain the data well.
5. Application Study

The motor vehicle insurance of a general insurance company in Indonesia in 2014 data sourced from the Financial Services Authority (FSA) is used. The amount of the claim is a response variable and the deductible is the explanatory variable. Selection of response variables and explanatory variables refers to Pratama [8]. The actual area code is 35 areas, but only 17 areas are included in the model because in the other 18 areas there are no claim events. The area code is considered as a random effect and the month of occurrence is a random effect which is considered to follow a first-order autoregressive process. For the purposes of this research, the data was taken in part through simple random sampling from the actual data of 175,000 items. In each area and month of the occurrence 10 policy numbers were taken as samples.

To show that response variable has gamma distribution, fitdistrplus package in R is used. Figure 2 describes data exploration in general. It can be seen that response variable has gamma distribution. Table 2 further explains that the amount of claim has gamma distribution with shape parameters $\alpha = 0.8945$ and scale parameter 1.8142. As is known there is a relationship between the dispersion parameter $d$ in MDE and the gamma distribution form parameter $\alpha$ as $d = \frac{1}{\alpha}$, so that in modelling the amount of claim, the dispersion parameter used is $d = \frac{1}{0.8945} = 1.117943$. 

![Bias Relative](image)

**Figure 1.** Average relative bias from 100 replications
The correlation coefficient between the amount of claim at time $t$ and $t-1$ is equal to $\rho = 0.01$ and the initial value $\sigma_b^2 = 0.12$ is taken from the difference in the amount of claims at time $t$ subtracted by multiplication of the correlation coefficient $\rho$ and the amount of claim at time $t - 1$. The initial value of the area random effect $\sigma_b^2 = 0.5$, and the regression parameters $\beta_0 = 0$, and $\beta_1 = 0$ refer to Chandra et al. [3]. Modeling is carried out based on the model on equation (6).

Modeling based on original data always produces a singular Hessian matrix, so that trial and error scaling of amount of and deductibles is carried out. Dividing the amount of claim and deductibles by $10^{000000}$ ultimately results in a non-singular Hessian matrix so that modeling can be carried out. The estimation results are described in Table 3.

| Table 3. Estimate parameter |
|-----------------------------|
| **Fixed Effect**           |
| Intercept($\beta_0$)       | -1.40272 | 0.22598 |
| **Deductible($\beta_1$)**  | 10.60682 | 0.575839 |
| **Random Effect Variance** |
| Area                       | 0.5727454 | 0.156171 |
| Time                       | 0.195147 | 0.050576 |
| **Residual**               | 1.069567 |          |
From Table 3 above, parameter $\beta_0$ is -1.40272 is $\beta_1$ is 10.60682. It means if the deductible has changed one point, then the expected value of the amount of claim will change 40409.49 times. In the estimation of random effect, it can be shown that the variance of area random effects and time random effects are smaller than the residual variance. This explains that there is a variability of the amount claims within the areas, but there is no variability of the amount claims between areas. Likewise, there is a variability of the amount of claims within the time but no variability over the time.

In Figure 3, it can be seen the comparison of the observed and the estimated of the amount of claim on each area and time. In general, for each time, modeling in each area produces a smaller estimated value than the observed value. This can be seen in almost all areas, except in the area 16 where the estimated value is greater than the observed value.

![Figure 3](image)

**Figure 3.** The Comparison between observed and estimate the amount of claim

![Figure 4](image)

**Figure 4.** Plot and histogram of residual

However, the diagnostic model using the histogram and the residual plot shows that the model still does not match in describing data. As shown in Figure 4, the remaining plots still show a pattern and the remaining histograms do not show the remainder of the normal distribution. This may be due to the small correlation which is only $\rho = 0.01$. In other words, autoregressive assumptions are not fulfilled in the data. So, it is necessary to do a further study using data with a larger autoregressive coefficient.
6. Conclusion
The h-likelihood method can be done to obtain regression and variance estimator in GLMM with gamma response variables using two random effects which are area and time with a first-order autoregressive assumption. The simulation study explains that for a model with a time random effect with a first-order autoregressive assumption, the greater autoregressive coefficient, the smaller the relative bias is and the greater the variance of areas the more unbiased the estimator is. So, the autoregressive coefficient and the variance of area have an effect to the biasness. The MSE value which is approaching to zero in almost all simulation designs explains that the model explains the data well. An application study using motor vehicle insurance of a general insurance company data in Indonesia in 2014 shows that the amount of claim is influenced by the deductible. It can be shown that if the deductible has changed one point, the expected value of the amount of claim will change by 40409.49 times. In the variance analysis, in the estimation of random effect, it can be seen that the variance of area random effects and time random effects are smaller than the residual variance. It can be seen that there is a variability of the amount claims within the areas, but there is no variability of the amount claims between areas. Likewise, there is a variability of the amount of claims within the time but no variability over the time. The diagnostic model using the histogram data plot and the residual plot shows that the model still does not match the data. This can be seen from the residual plot which still shows patterns and the residual histograms do not show normal distribution. The reason may be due to the small correlation coefficient which is only \( \rho = 0.01 \). For this reason, it is necessary to further study using data with a larger autoregressive coefficient.

References
[1] Adam F F, Kurnia A, Purnaba I G P, Mangku I W, Soleh A M 2020 Tweedie compound Poisson model with first order autoregressive time random effect Commun. Math. Biol. Neurosci
[2] Bossio M L C, Cuervo E C 2019 A Bayesian approach to mixed gamma regression models Revista Colombiana de Estadística Volume 42 Issue 1 pp. 81 to 99 http://dx.doi.org/10.15446/rec.v42n1.69334
[3] Chandra H, Salvati N, Chambers R. 2018 Small area estimation under a spatially non-linear model Spatial Statistics Computational Statistics and Data Analysis https://doi.org/10.1016/j.csda.2018.04.002
[4] Jørgensen B 1987 Exponential dispersion models (with discussion) J. R. Stat. Soc. B 49 127–162
[5] Lee C, Lee Y 1998 Sire evaluation of count traits with a Poisson-gamma hierarchical generalized model. AJAS Vol 11 No 6
[6] Lee Y, Nelder J A 1996 Hierarchical generalized model J.R. Statist. Soc. B 58 619-678
[7] Muchlisoh S 2017 Pengembangan model pendugaan area kecil dengan pengaruh acak waktu mengikuti proses autoregresif ordo pertama [disertasi]. Bogor: Institut Pertanian Bogor
[8] Pratama H 2011 Pemodelan data asuransi mobil melalui regresi gamma [skripsi] Bandung: Universitas Islam Bandung
[9] Ribeiro M H D, Santiago A N, Milani H, Oliveira R M W, and Prevideli I 2019 Longitudinal modeling using log-gamma mixed model: case of memory deterioration after chronic cerebral hypoperfusion associated with diabetes in rats Acta Scientiarum. Technology 41 Doi: 10.4025/actascitechvol.v41i1.35789
[10] Zhang Y 2013 Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models Stat. Comput. 23:747-757 DOI: 10.1077/s11222-012-9343-7.
[11] Zhang P, Luo D, Li P, Sharpsten L, and Medeiros F A 2015 Log-gamma linear-mixed effects models for multiple outcomes with application to a longitudinal glaucoma study Biom J. 57(5): 766–776. DOI:10.1002/bimj.201300001