Nested Generalized Linear Mixed Model with Ordinal Response: Simulation and Application on Poverty Data in Java Island

Yekti Widyaningsih*, Asep Saefuddin†, Khairil A. Notodiputro† and Aji H. Wigena†

*Department of Mathematics, Faculty of Mathematics and Natural Sciences,
University of Indonesia, Indonesia.
yekti@sci.ui.ac.id
†Department of Statistics, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, Indonesia.
asaefuddin@gmail.com, khairiln@bima.ipb.ac.id, ajihamim@yahoo.com

Abstract. The objective of this research is to build a nested generalized linear mixed model using an ordinal response variable with some covariates. There are three main jobs in this paper, i.e. parameters estimation procedure, simulation, and implementation of the model for the real data. At the part of parameters estimation procedure, concepts of threshold, nested random effect, and computational algorithm are described. The simulations data are built for 3 conditions to know the effect of different parameter values of random effect distributions. The last job is the implementation of the model for the data about poverty in 9 districts of Java Island. The districts are Kuningan, Karawang, and Majalengka chose randomly in West Java; Temanggung, Boyolali, and Cilacap from Central Java; and Blitar, Ngawi, and Jember from East Java. The covariates in this model are province, number of bad nutrition cases, number of farmer families, and number of health personnel. In this modeling, all covariates are grouped as ordinal scale. Unit observation in this research is sub-district (kecamatan) nested in district, and districts (kabupaten) are nested in province. For the result of simulation, ARB (Absolute Relative Bias) and RRMSE (Relative Root of mean square errors) scale is used. They show that prov parameters have the highest bias, but more stable RRMSE in all conditions. The simulation design needs to be improved by adding other condition, such as higher correlation between covariates. Furthermore, as the result of the model implementation for the data, only number of farmer family and number of medical personnel have significant contributions to the level of poverty in Central Java and East Java province, and only district 2 (Karawang) of province 1 (West Java) has different random effect from the others. The source of the data is PODES (Potensi Desa) 2008 from BPS (Badan Pusat Statistik).

Keywords: generalized linear mixed model, ordinal response, nested
PACS: 12.40.Ee

INTRODUCTION

Variable associated with levels often expressed as an ordinal scale category variable. Besides appearing as a result of direct measurement of an object under study, ordinal scale variables can also arise as a result of numerical clustering measurements. One advantage of using data with ordinal-scale variables is more easily to be understood in the interpretation. Objects that be observed in the study did not always stand alone, it is always a member of a larger group. Usually, groups (clusters) in the data emerge based on some reasons, such as the existence of homogeneity or heterogeneity of individuals, the influence of geographical location (spatial clustering), or even the existence of different treatments[1]. Nested in this section is identical with the understanding of multilevel or group or cluster of the data.

Determination of statistical modeling is influenced by many aspects. One of these aspects is homogeneity or heterogeneity of the data, as well as the location of the individual being observed. The closer the location of two individuals being observed, the greater their correlation [2]. These conditions are often become the basis for the nested models formation. Statistical modeling that will be discussed in this paper is a modeling for ordinal response in nested conditions. A link function is needed to connect the ordinal response variable and explanatory variables. The model will be implemented for poverty data in Java Island. The relationship between poverty level, as response variable, and some explanatory variables will be analyzed. The explanatory variables that related to the poverty are the number of bad nutrition cases, the number of poverty, the number of farmer families, and the number of health personnel [3]. All variables used in the model are grouped into 3 levels, the ordinal scale variables. The nested condition is clear, that the data comprises of sub-districts, districts, and provinces, where sub-districts nested in a district, and some districts nested in a province.

The 5th International Conference on Research and Education in Mathematics
AIP Conf. Proc. 1450, 147-154 (2012); doi: 10.1063/1.4724132

As a model with non-continue and non-Gaussian response variable, an appropriate link function is needed to connect the ordinal response variable and explanatory variables [4]. The explanatory variables also being assessed through the model, to know how they contribute to determine the poverty level of areas. In this case, the districts have chosen randomly, so the model should be a mixed linear model. The basic assumption in the development of the model is: the observations in the same group are more homogeneous compared to the observations from different groups. The existence of these groups lead to the selection of nested models and the existence of random effect lead to mixed linear model. In other words, the basic development of this model is the nonlinearity (ordinal response), the nested condition, and randomness of an effect in the model.

METHOD

Generalized Linear Mixed Model with Ordinal Response

Generalized Linear Mixed Models with Ordinal Response is a development of Generalized Linear model with continuous response. This study developed a generalized linear mixed model with ordinal response by adding the nested consideration of the data. The steps undertaken in this research are: first, discuss the threshold model and parameter estimation; second, carry on assessing (the values of model parameter estimation using simulated data with some given conditions.

Threshold Model

Threshold is a latent variable that made the difference between the linear models with ordinal response and the linear models with non-ordinal responses. Threshold model is explained as follows. In logistic and probit regression models, there are assumptions about an unobserved latent variable (y) associated with the actual responses through the concept of threshold [5]. For the dichotomy model, it is assumed there is a threshold value and for the ordinal model with K categories (polytomi), it is assumed there are $K-1$ threshold values, namely $\gamma_1, \gamma_2, \ldots, \gamma_{K-1}$, with $\gamma_0 = -\infty$ and $\gamma_K = \infty$. Response occurs in category k ($Y = k$), if the latent response y is greater than the threshold γ_{k-1}, but not greater than the threshold value γ_k. Let a model has ordinal response with K categories, assume Y_i is unobserved, and the i-th observation is in a category, say category $Z_i, i = 1, \ldots, N$. The relationship between Y_i and Z_i is taken to be

$$\gamma_{k-1} < Y_i \leq \gamma_k \Longleftrightarrow Z_i = k \tag{1}$$

where $k \in 1, \ldots, K, \gamma_0 = -\infty, \gamma_K = +\infty$ and $\gamma_1, \ldots, \gamma_{K-1}$ are unknown boundary points that define a partitioning of the real line into K intervals. Thus, when the realized value of Y_i belongs to the kth interval, we observe that $Z_i = k$. Under that assumptions, the probability-mass function of $Z_1, \ldots, Z_N$ is

$$\begin{aligned} P(z_1, \ldots, z_N) &= Pr\{Z_i = z_i\,(i = 1, \ldots, N)\} \\ &= Pr\{\gamma_{z_i - 1} < Y_i \leq \gamma_{z_i}\,(i = 1, \ldots, N)\} \end{aligned} \tag{2}$$

This model is called the threshold model[6].

Linear Model

In general, a linear model can be described below. Let y be a response variable of a linear mixed model

$$y = X\beta + W\alpha + e \tag{3}$$

where **X** and **W** are given matrices of dimensions $N \times q$ and $N \times r$, respectively, β is an unknown parameter vector $qx1$, α is a random effect vector $r \times 1$, and **e** is a random residuals vector $Nx1$ that are distributed independently of α. Moreover, $\alpha \sim \text{MVN}(0, \sigma^2\mathbf{D})$and e $\sim \text{MVN}(0, \sigma^2\mathbf{I})$,$\sigma$ is an unknown positive parameter, and the elements of **D** are functions of an unknown parameter $\xi = (\xi_1, \ldots, \xi_c)'$. Assume that ξ is restricted to a given subset, Ξ, of Euclidean c-space, **D** is positive definite for all $\xi \in \Xi$. Furthermore, can be obtained that y $\sim \text{MVN}(X\beta, \sigma^2 V)$, with $V = I + WDW'$. The linear mixed model for nested data is

$$y_{ijm} = x'_{ijm}\beta_{j(i)} + w'_{ijm}\alpha_{j(i)} + \varepsilon_{ijm} \tag{4}$$

where $i = 1, 2, \ldots, I; j = 1, 2, \ldots, J; m = 1, 2, \ldots, n_j$ and y_{ijm} is a value of response variable for mth unit, at jth category level 1 in ith category level 2. $\mathbf{x}_{ijm}$ is a covariate vector $qx1$ and $\mathbf{w}_{ijm}$ is the $rx1$ design vector for r random effects, both vectors being for the mth unit, at jth category level 1 in ith category level 2. $\beta_{j(i)}$ is the a vector of unknown fixed parameters for jth category level 1 in ith category level 2, $\alpha_{j(i)}$ is the vector of unknown random effects for jth category level 1 in ith category level 2, and ε_{ijm} are model residuals.

This paper discusses a modeling with ordinal response variables which is assumed to be multinomial distributed, nested level 1 is the sub-districts within a county (district), nested level 2 is the districts within a province. Characteristics of sub-districts are used to represent the profile of a district. If associated with longitudinal data, the value of covariates of sub-district is the result of repeated measurements for a given district, or as a sample of observations from a cluster. Modeling theory is explained in the following sub-section.

Nested Random-Effects Ordinal Regression Model

If the data has ordinal response with more than two categories, then the response variable will be assumed to be multinomial distributed. Furthermore, if the data is a nested data, then the model in equation (4) should be a Nested Generalized Linear Mixed Models.

Generalized Linear Model (GLM) is a class of regression models with fixed effects for several types of response variables, i.e. continuous, dichotomous, and discrete [7]. Some of the common GLM are linear regression, logistic regression and Poisson regression. There are 3 specifications in GLMs: the exponential distribution families, the linear predictor, and link function. Linear predictor denoted by $\eta_i = \mathbf{x}_i'\beta$ For nested model, which observations m nested in i and i nested in j, the model becomes

$$\eta_{ijm} = \mathbf{x}'_{ijm}\beta_{j(i)} \tag{5}$$

Link function g(·) is used to convert the expected value of the response variable Y_{ijm} to be a linear estimator η_{ijm}

$$g(\mu_{ijm}) = \eta_{ijm} \tag{6}$$

For mixed nested model,

$$\eta_{ijm} = \mathbf{x}'_{ijm}\beta_{j(i)} + \mathbf{w}'_{ijm}\alpha_{j(i)} + \varepsilon_{ijm} \tag{7}$$

Nested Generalized Linear Mixed Model with ordinal response can be expressed in a cumulative logit model as the following

$$log\left[\frac{P(Y_{ijm} \le k)}{1 - P(Y_{ijm} \le k)}\right] = \gamma_k + \mathbf{x}'_{ijm}\beta_{j(i)} + \mathbf{w}'_{ijm}\alpha_{j(i)} + \varepsilon_{ijm} \tag{8}$$

γ_k is a threshold value of laten variable that related to the true response value Y_{ijm}. For a single random effect, the expected values can be written as follows:

$$E\left[log\frac{P(Y_{ijm} \le k)}{1 - P(Y_{ijm} \le k)}\right] = \gamma_k + \mathbf{x}'_{ijm}\beta_{j(i)} + \alpha_{j(i)} \tag{9}$$

or

$$P(Y_{ijm} \le k) = \frac{1}{1 + exp\left[-(\gamma_k + \mathbf{x}'_{ijm}\beta_{j(i)} + \alpha_{j(i)} + \varepsilon_{ijm})\right]} \tag{10}$$

Equation (8) is called the ordered logit model that depends upon the idea of the cumulative logit. This in turn relies on the idea of the cumulative probability. Cumulative probability $P(Y_{ijm} \le k)$ means the probability that the *m*-th individual of the *i*-th unit level 1 and the *j*-th unit level 2 is in or lower than category k,

$$P(Y_{ijm} \le k) = \sum_{h=1}^{k} P(Y_{ijm} = h).$$

The next is applying the model to the poverty data in the 3 provinces, with 3 districts in every province, and m_j sub-districts in every district j. Ordinal response variable is the level of poverty of the sub-district, while the explanatory variables are number of bad nutrition cases, number of farmer families, number of health personnel. The goal of modeling is want to know the linkages between poverty level of sub-districts and multiple covariates. All the values of covariates are grouped into three levels (1, 2, 3): the province (X_1), bad nutrition cases (X_2), farmer families (X_3), health personnel(X_4), and districts (u). Districts are assumed to be chosen randomly and should be a random effects in the model. These effects are taken to be normally distributed random effects with means of μ and covariances $\sigma^2 a_{ff'}$(*f*',*f*=1,2,3). Covariates in the model consist of three categories, therefore dummy variables is needed, and the model for nested design should be

$$\begin{aligned} y_{ijm} = {} & \beta_1 prov1 + \beta_2 prov2 + \beta_3 badnut11 + \beta_4 badnut12 + \beta_5 badnut21 + \beta_6 badnut22 + \\ & \beta_7 badnut31 + \beta_8 badnut32 + \beta_9 farm11 + \beta_{10} farm12 + \\ & \beta_{11} farm21 + \beta_{12} farm22 + \beta_{13} farm31 + \beta_{14} farm32 + \\ & \beta_{15} med11 + \beta_{16} med12 + \beta_{17} med21 + \beta_{18} med22 \\ & + \beta_{19} med31 + \beta_{20} med32 + u_{j(i)} + e_{ijm} \end{aligned} \tag{11}$$

or in the conditional probability,

$$P(Y_{ijm} \le k) = \frac{1}{1 + exp\left[-(\gamma_k + \omega_{ijm}\right]} \tag{12}$$

where ω_{ijm} is similar to right side of the equation (11), and m, j and i are indexes for sub-districts, districts, and provinces, respectively.

Threshold model for this data can be written as a special case of the general threshold model with $r = 9, q = 20$; that is $\mathbf{u} = (u_{11}, u_{12}, u_{13}, u_{21}, u_{22}, u_{23}, u_{31}, u_{32}, u_{33})$, $\beta = (\beta_1, \beta_2, \ldots, \beta_{20})$, and $\mathbf{e} = (e_{111}, \ldots, e_{11M}, \ldots, e_{331}, \ldots, e_{33M})$, M is the number of sub-districts in every district. $\mathbf{D} = \xi\mathbf{A}$ (where $\xi_i = \sigma^2_{u_i}/\sigma^2$, and $\mathbf{A}$ is the matrix whose *ff'*th element is $a_{ff'}$).

Prediction Procedure

This section builds a prediction procedure for a standardized threshold model. It is assumed that ξ is a known quantity. The procedure can be applied also if it is unknown by first estimating ξ and by then acting as it is a true value. It will be focused on the estimation of a linear combination $l'\gamma + x'\beta + w'u$, where $\gamma = (\gamma_2, \ldots, \gamma_{M-1})$, i.e. on a linear combination of the unknown boundary points and the fixed and random effects. Denoted that the probability density function (pdf) and the cumulative distribution function (cdf) of the N(0,1) distribution by $\phi(\cdot)$

and $\Phi(\cdot)$, respectively. Further, for an arbitrary kx1 vector μ and a positive definite matrix Σ, $\phi_k(\cdot\,;\mu,\Sigma)$ represent the pdf of the MVN(μ, Σ) distribution.

Approach. If y were observable, Handerson's BLUP procedure could be used to estimate $\mathbf{x}'\beta+\mathbf{w}'\alpha$. The BLUP would be $\mathbf{x}'\tilde{\beta}+\mathbf{w}'\tilde{\alpha}$, where $\tilde{\beta}$ is any solution to $\mathbf{X'V^{-1}X}\tilde{\beta}$ = $\mathbf{X'V^{-1}y}$ and $\tilde{\alpha}=\mathbf{DW'V^{-1}}(\mathbf{y}-\mathbf{X}\tilde{\beta})$ or, equivalently, $\tilde{\beta}$ and $\tilde{\alpha}$ represent any solution to the system of linear equations,

$$\begin{bmatrix} X'X & X'W \\ W'X & D^{-1}+W'W \end{bmatrix}\begin{bmatrix} \tilde{\beta} \\ \tilde{\alpha} \end{bmatrix}=\begin{bmatrix} X'y \\ W'y \end{bmatrix} \tag{13}$$

known as the mixed-model equations[6].The BLUP of β and α could be found by maximizing

$$\begin{aligned}&\phi_{N+q}\left(\begin{bmatrix} \mathbf{y} \\ \alpha \end{bmatrix};\begin{bmatrix} X'\beta \\ \mathbf{0} \end{bmatrix},\begin{bmatrix} V & WD \\ DW' & D \end{bmatrix}\right)\\ &=\phi_N\left(y;X\beta+W'\alpha,I\right)\phi_q\left(\alpha;0,D\right)\\ &=\phi_q\left(\alpha;0,D\right)\prod_{i=1}^{N}\phi\left(Y_i-x_i'\beta-w_i'\alpha\right)\end{aligned} \tag{14}$$

where $\mathbf{x_i'}$ and $\mathbf{w_i'}$ represent the tth rows of $\mathbf{X}$ and $\mathbf{W}$. Note that (15) represents the joint pdf of $\mathbf{y}$ and $\mathbf{a}$, as a random variable of α. Equation (13) are obtained upon equating to 0 the partial derivatives (with regard to β and α) of the logarithm of (14).

An approach that analog to the maximization of (14) is applied to the standardized threshold model as follows. Define

$$\begin{aligned}&\psi(z_{111},\ldots,z_{IJM};\gamma,\beta,\alpha)\\ &=\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{m=1}^{M}\int_{\gamma_{z_i-1}}^{\gamma_{z_i}}\phi\left[y_{ijm}-x_{ijm}'\beta_{j(i)}-w_{ijm}'\alpha_{j(i)}\right]dy_{ijm}\\ &=\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{m=1}^{M}\left(\Phi(\gamma_{z_i}-\omega_{ijm})-\Phi(\gamma_{z_i-1}-\omega_{ijm})\right)\end{aligned} \tag{15}$$

where $\omega_{ijm}=x_{ijm}'\beta_{j(i)}+w_{ijm}'\alpha_{j(i)}$. The proposed procedure is to estimate $\mathbf{l}'\gamma+\mathbf{x}'\beta+\mathbf{w}'\alpha$ by $\mathbf{l}'\hat{\gamma}+\mathbf{x}'\hat{\beta}+\mathbf{w}'\hat{\alpha}$, where $\hat{\gamma}=(\hat{\gamma}_2,\ldots,\hat{\gamma}_{M-1})'$, $\hat{\beta}$, and $\hat{\alpha}$ are any values of γ, β, and α that maximize

$$\psi(Z_{111},Z_{112},\ldots,Z_{IJM};\gamma,\beta,\alpha)\phi_q(\alpha;\mathbf{0},\mathbf{D}) \tag{16}$$

Quantity (15) represents the conditional probability (given $\mathbf{a}=\alpha$) dan $Z_{111}=z_{111},\ldots,Z_{IJM}=z_{IJM}$, and thus that the function (16) represents the joint-probability mass density function of $Z_{111},\ldots,Z_{IJM},\mathbf{a}$, so that

$$P(\mathbf{z_i})=\prod_{j=1}^{J}\prod_{m=1}^{M}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi(z_{ijm};\gamma,\beta,\alpha)\phi_q(\alpha;\mathbf{0},\mathbf{D})d\alpha \tag{17}$$

In the special case where model (3) reduces to a fixed-effects model, i.e. where $\mathbf{y}=\mathbf{X}\beta+\mathbf{e}$, the proposed estimation procedure is identical to maximum likelihood (ML) estimation. The estimator $l'\hat{\gamma}$+x'$\hat{\beta}$+w'$\hat{\alpha}$ has a Bayesian interpretation. If the joint prior distribution of γ and β is taken to be proportional to a constant, then the point $\hat{\gamma}$, $\hat{\beta}$, $\hat{\alpha}$ represents a mode of the joint posterior distribution of γ, β, α.

Computational Algorithm. This section explains the computation of $\hat{\gamma}$, $\hat{\beta}$, and $\hat{\alpha}$, i.e. the numerical problem of maximizing the function (16). Maximizing this function is equivalent to maximizing

$$f(\gamma,\beta,\alpha;z)=f_1(\gamma,\beta,\alpha;z)+f_2(\alpha) \tag{18}$$

where $f_1(\gamma,\beta,\alpha;z)=\ln\psi(Z_{111},\ldots,Z_{IJM})$ and $f_2(\alpha)=\ln[\phi(\alpha;\mathbf{0},\mathbf{D})]$. Actually f_1 is the log-likelihood function for a standardized threshold model in which the underlying linear model is the fixed-effects linear model obtained by replacing $\mathbf{a}$ by α in equation (3). The results on fixed-effects threshold model is given by Mee on his Ph.D dissertation, 1981 [6]. Let $\tau'=(\gamma',\beta',\alpha')$, and define

$$r(\gamma,\beta,\alpha;z)=\partial f/\partial r=r_1(\gamma,\beta,\alpha;z)+r_2(\alpha)$$

where

$$r_1(\gamma,\beta,\alpha;z)=\partial f_1/\partial\tau=\begin{bmatrix} \partial f_1/\partial\gamma \\ \partial f_1/\partial\beta \\ \partial f_1/\partial\alpha \end{bmatrix}=\begin{bmatrix} v(\gamma,\beta,\alpha;z) \\ X'\varepsilon(\gamma,\beta,\alpha;z) \\ W'\varepsilon(\gamma,\beta,\alpha;z) \end{bmatrix}$$

and

$$r_2(\gamma,\beta,\alpha;z)=\partial f_2/\partial\tau=\begin{bmatrix} \partial f_2/\partial\gamma \\ \partial f_2/\partial\beta \\ \partial f_2/\partial\alpha \end{bmatrix}=\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{D}^{-1}\beta \end{bmatrix}$$

Here, $v(\gamma,\beta,\alpha;z)=\partial f_1/\partial\gamma$, is the $(M-2)x1$ vector whose *(k-1)*th element is

$$\begin{aligned}&\sum_{m\in\Omega_k}\left(\frac{\phi_{mk}}{\triangle_{mk}}\right)-\sum_{m\in\Omega_{k+1}}\left(\frac{\phi_{mk}}{\triangle_{m,k+1}}\right)\\ &=\sum_{m\in\Omega_k}\left(\frac{\phi(\gamma_k-\omega_m)}{\Phi(\gamma_k-\omega_m)-\Phi(\gamma_{k-1}-\omega_m)}\right)\\ &-\sum_{m\in\Omega_{k+1}}\left(\frac{\phi(\gamma_k-\omega_m)}{\Phi(\gamma_{k+1}-\omega_m)-\Phi(\gamma_k-\omega_m)}\right)\end{aligned}$$

and $\varepsilon(\gamma,\beta,\alpha;z)$ is the $N\times 1$ vector whose ith element is

$$\varepsilon_m=(\gamma,\beta,\alpha;z)=\delta_{m,Z_m}/\triangle_{m,Z_m}$$

where $\omega_m=x_m'\beta-w_m'\alpha$, $\Omega_k=\{m;Z_m=k\}$,
$\phi_{mk}=\phi(\gamma_k-x_m'\beta-w_m'\alpha)$, $\delta_{mk}=\phi_{m,k-1}-\phi_{mk}$ and

$$\triangle_{mk}=\Phi(\gamma_k-x_m'\beta-w_m'\alpha)-\Phi(\gamma_{k-1}-x_m'\beta-w_m'\alpha)$$

The quantities $\hat{\gamma},\hat{\beta},\hat{\alpha}$ necessarily satisfy the condition

$$\mathbf{r}(\hat{\gamma},\hat{\beta},\hat{\alpha}) = \mathbf{0} \tag{19}$$

It had been showed by Pratt (1981) that f_1 is a concave function of γ,β,α, and since f_2 is a concave function, the sum f is a concave function, implying that Condition (19) is sufficient and necessary, i.e. that any solution to the system (19) of nonlinear equation maximizes f.
Various iterative algorithms are available for solving system of nonlinear equations. To solve the system (19), let $k+1$th iterate $\hat{\gamma}^{k+1}$, $\hat{\beta}^{k+1}$, $\hat{\alpha}^{k+1}$ satisfies

$$\bar{C}(\hat{\gamma}^k,\hat{\beta}^k,\hat{\alpha}^k)\begin{bmatrix}\hat{\gamma}^{k+1}-\hat{\gamma}^k\\ \hat{\beta}^{k+1}-\hat{\beta}^k\\ \hat{\alpha}^{k+1}-\hat{\alpha}^k\end{bmatrix} = r(\hat{\gamma}^k,\hat{\beta}^k,\hat{\alpha}^k;z) \tag{20}$$

Here

$$\bar{C}(\gamma,\beta,\alpha) = E\left[C(\gamma,\beta,\alpha;z)|\mathbf{a}=\alpha\right]$$

where $C(\gamma,\beta,\alpha;z) = -\partial^2 f/\partial\tau\partial\tau'$. Since

$$C(\gamma,\beta,\alpha;z) = C_1(\gamma,\beta,\alpha;z) + C_2$$

with $C_1(\gamma,\beta,\alpha;z) = -\partial^2 f_1/\partial\tau\partial\tau'$ dan $C_2 = -\partial^2 f_2/\partial\tau\partial\tau'$, we have $\bar{C}(\gamma,\beta,\alpha) = \bar{C}_1(\gamma,\beta,\alpha;z) + C_2$, where $\bar{C}_1(\gamma,\beta,\alpha) = E\left[C_1(\gamma,\beta,\alpha;z)|\mathbf{a}=\alpha\right]$. From Bock, 1975 and Mee's dissertation, we have

$$\bar{C}_1(\gamma,\beta,\alpha)=\begin{bmatrix}Q & L'X & L'W\\ X'L & X'RX & X'RW\\ W'L & W'RX & W'RW\end{bmatrix},C_2=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & D^{-1}\end{bmatrix} \tag{21}$$

$\mathbf{L}$ is an $(I\times J\times M)\times(K-1)$ matrix, $\mathbf{Q}$ is an $(K-2)\times(K-2)$ tridiagonal matrix and $\mathbf{R}$ is an $(I\times J\times M)\times(I\times J\times M)$ diagonal matrix. The nonzero elements of $\mathbf{L}$, $\mathbf{Q}$ and $\mathbf{R}$ are functionally dependent on γ, β, and α.

Mean Square Error. Consider the error incurred in estimating $l'\gamma+x'\beta+w'\alpha$ by $l'\hat{\gamma}+$x'$\hat{\beta}+$w'$\hat{\alpha}$. Suppose that

$$E(l'\hat{\gamma}+x'\hat{\beta}+w'\hat{\alpha}-(l'\gamma+x'\beta+w'\alpha)) \simeq 0 \tag{22}$$

and

$$var[l'\hat{\gamma}+x'\hat{\beta}+w'\hat{\alpha}-(l'\gamma+x'\beta+w'\alpha)]$$
$$\simeq E[(l',x',w')\bar{C}^{-}(\gamma,\beta,a)(l',x',w')'] \tag{23}$$

Motivated by the approximate results (22) and (23), mean square error of $l'\hat{\gamma}+x'\hat{\beta}+w'\hat{\alpha}$ is estimated by

$$(l',x',w')\bar{C}^{-}(\hat{\gamma},\hat{\beta},0)(l',x',w')' \tag{24}$$

SIMULATION

The data simulation consists of an ordinal response variable (Z), two fixed effects covariates (X_1,X_2) and one random effect (W). All variables are in ordinal scale, as level of 1,2, and 3 as representative numeric for good, moderate, and bad. The structure of the data is (Z,X_1,X_2,W). Some aspects that should be considered in the model building in this research are nested, homogeneity, and number of observations (sub-districts) in a district. The objective of this simulation is to assess the effects of homogeneity and size of a district (n = the number of sub-districts in a district) to the parameters estimating values. To create the aspect of homogeneity, the data simulation are designed for some conditions (scenarios), as the following: (1) $w \sim Normal(1,\sigma^2)$, with σ^2 = 1, 0.81, and 0.64, for province 1, 2, and 3, respectively. (2) $w \sim Normal(1,0.64)$, for all provinces. (3) $w \sim Normal(\mu,0.64)$, with μ = 1, 2, and 3, for provinces 1, 2, and 3, respectively. All these three conditions are implemented for n = 6 (N = 54), 13 (N = 117), 20 (N = 180), N is the total number of observations. Here, we have 3 provinces with 3 districts in every province.

ARB and RRMSE

Two important scales for estimators are ARB and RRMSE. The ARB is defined as the absolute value of the relative bias of the estimate over the realized finite population value, to obtain the accurate scale of an estimator. The relative root-mean-square error (RRMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. The formulas of ARB and RRMSE are as follows [8]

$$ARB = \left|\frac{1}{M}\sum_{s=1}^{M}\frac{\hat{\beta}_p^{(s)}-\beta_p}{\beta_p}\right| \tag{25}$$

$$RRMSE = \frac{\sqrt{\frac{1}{M}\sum_{s=1}^{M}\left(\hat{\beta}_p^{(s)}-\beta_p\right)^2}}{\beta_p} \tag{26}$$

These two scales were applied to compare the performance of parameters estimator for three different conditions simulation data. The two explanatory variables (prov and x) in the simulations are independent each other (not correlated). They have three level categories, so the model needs two dummy variables for each category variable. The data is in nested condition (districts are in province), so we have 10 parameters that should be estimated, they are intercept1, intercept2, prov1, prov2, b11, b12, b21, b22, b31, and b32.

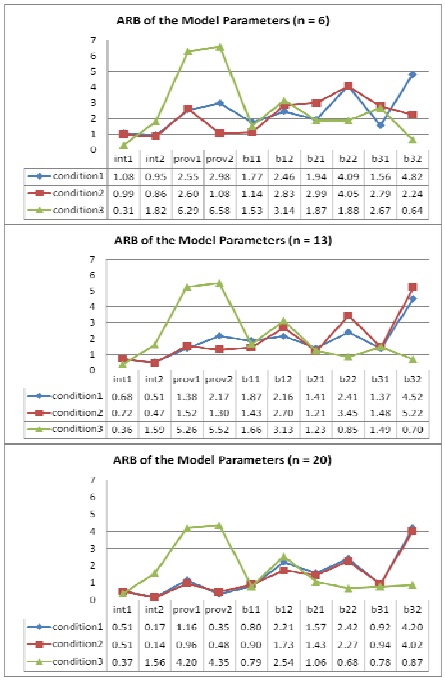

FIGURE 1. ARB of the Model Parameters

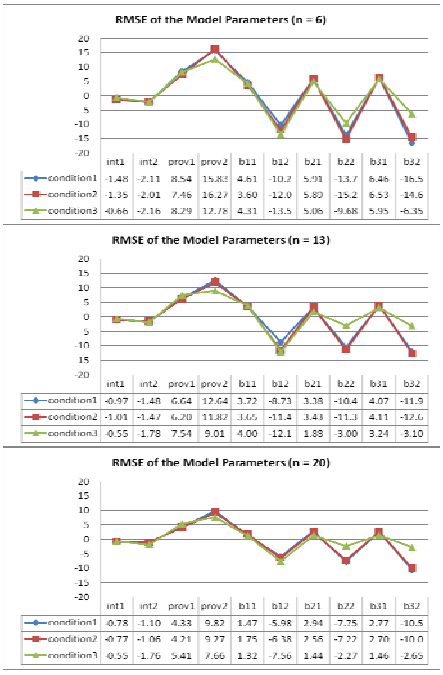

FIGURE 2. RRMSE of the Model Parameters

Simulation Results

The results are showed by Figure 1 and 2, each for the ARB and the RRMSE with n = 6, 13, and 20.
According to the ARB graphs on the Figure 1, biases (absolute values of relative bias from the realized finite population values) of prov1 and prov2 parameters for condition 3 are much higher than the bias of other parameters, but biases are relatively lower for b21, b22, b31, and b32 for this condition. It occurs at the simulation data with n = 6, 13, and 20. For n = 20, biases are almost same for conditions 1 and 2, while biases for condition 3 are still looks different from conditions 1 and 2.

Figure 2 shows the RRMSE (relative root means square error), the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. It shows that the parameter estimations for all conditions are nearly equal. For condition 3, prov2 is smaller but b22 and b32 are higher than those of conditions 1 and 2. As well as ARB, the larger the n, the narrower the interval. For n = 6, 13, and 20, the intervals are (-15, 15), (-13, 13), and (-10, 10), respectively.

APPLICATION

The procedures of generalized linear mixed model with ordinal response described above could be used to compare the districts and subdistricts with respect to the poverty level. The model is implemented to the poverty data in 9 districts in Java Island which took in 2008. The data consists of an ordinal response variable, four fixed effects and a random effect. The ordinal response variable is the level of poverty, while the covariates are province (prov), number of bad nutrition cases (badnut), number of farmer families (farm), and number of health personnel (med) which each of them has three levels values, and district is as a random effect. The values of covariates are grouped into the ordinal scale. The names of districts in the data are Kuningan, Karawang, and Majalengka that had been chosen randomly from West Java; Temanggung, Boyolali, and Cilacap from Central Java; and Blitar, Ngawi, and Jember from East Java. Unit observation in this research is sub-district (kecamatan) nested in district, and districts (kabupaten) are nested in province. Actually, the districts had been chosen randomly from a collection of ordering dually (ranking method) results, furthermore, they are nested in a province and will be considered in the model. The Equation (11) is the model for the data.

TABLE 1. The Categories and the Overall Observed Frequencies

Variable	Category or Number of cases	Level	number of districts
Y	Not Poor	1	70
	Moderate	2	59
	Poor	3	48
Prov	West Java	1	56
(X_1)	Central Java	2	54
	East Java	3	67
badnut	[0, 9]	1	76
(X_2)	[10, 24]	2	50
	[25, 666]	3	51
farm	[0, 2600]	1	36
(X_3)	[2601, 5000]	2	46
	[5001, 16404]	3	95
med	[0, 29]	1	55
(X_4)	[30, 49]	2	78
	[50, 1336]	3	44

RESULTS AND DISCUSSION

As the computation results, Table 1 shows the categories and overall observed frequencies and Table 2 shows the variances of fixed effects in every district and every province. According to the variances of poverty level (Y), in province 1 (West Java), var(Y) = 0.68 is greater than variances of districts 1, 2, and 3 (0.13, 0.38, and 0.25, respectively) in the same province. It means that the poverty levels of sub-districts in the same district are more homogeneity than those in different districts. The same descriptions are for province 2 (Central Java)(var(Y) = 0.61) and province 3 (East Java) (var(Y) = 0.58), which these variances are greater than the variances of their districts. This homogeneity condition comply to the assumption for nested model, even though this condition is not complied by the explanatory variables.

Table 3 shows the correlations of variables in the model. Bad nutrition cases has no correlation to the response variable (Y) (p-value = 0.8775), while the remaining variables have statistically significant correlation to the response. Table 4 shows the solution for fixed effects. According to the computation result, intercept1 is almost significant (p-value = 0.0508) and just farmer family (farm21, p-value = 0.0217) and medical (health) personnel are significant (med21, p-value = 0.0264). The table shows that farmer families from province 2 with category 1 is significant different from category 3 (as the base level). It means in province 2, number of farmer families with category 1 and 3 cause the different poverty level of sub-districts. Furthermore, in province 2, number of medical personnel with category 1 and 3 cause the different poverty level of sub-districts, as well as in province 3, for medical personnel with category 2 and 3. An interpretation of the modeling result is as follows. The ratio of the odds A over B (let A is a sub-district in province 2 with category 1 for farmer families and B is a sub-district in province 2 with category 3 for farmer families, and have the same categories for other covariates) can be obtained by the calculation,

$\theta = \omega_A/\omega_B =$

$[P(y_A \le k)/(1-P(y_A \le k))]/[P(y_B \le k)/(1-P(y_B \le k))]$

$= e^{2.0883} = 8.071.$

It means the odds of A response are about eight times of the odds of B response or the risks of A become better sub-district are eight times higher than the risks of B.

The solution for random effect is shown by Table 5 which district 2 (Karawang) in Province 1 (West Java) is significantly different from other districts (p-value = 0.0103), while the other districts are not statistically different each other.

TABLE 2. The Variances of the Ordinal Variables

prov	dist	Y	badnut	farm	med
1	1	0.13	0.68	0.66	0.62
	2	0.38	0.51	0.49	0.40
	3	0.25	0.58	0.49	0.36
2	1	0.23	0.69	0.27	0.58
	2	0.51	0.45	0.54	0.62
	3	0.53	0.70	0.45	0.34
3	1	0.43	0.33	0.71	0.69
	2	0.47	0.70	0.45	0.34
	3	0.58	0.45	0.22	0.54
1		0.68	0.51	0.52	0.46
2		0.61	0.61	0.30	0.61
3		0.58	0.68	0.40	0.52
overall		0.66	0.69	0.42	0.59

TABLE 3. The correlations of the variables and the p-values

	Y	prov	badnut	farm	med
Y	1	0.2654	-0.0117	0.3807	0.1791
		(.0004)	(.8775)	(<.0001)	(.0171)
prov		1	-0.3574	0.2348	-0.1129
			(<.0001)	(.0017)	(.1348)
badnut			1	-0.0063	0.1091
				(.9339)	(.1483)
farm				1	0.0655
					(.3865)
med					1

TABLE 4. Solution for Fixed Effects

	Estimate	StdErrEst	t Value	Pr > |t|*
intercept1	-3.0756	1.2631	-2.43	0.0508
intercept2	-0.7189	1.2328	-0.58	0.5810
prov1	2.2371	1.7749	1.26	0.2543
prov2	0.5727	0.7993	0.32	0.7610
badnut11	0.7138	1.0054	0.71	0.4788
badnut12	0.3129	0.7498	0.42	0.6771
badnut21	-0.4245	0.7941	-0.53	0.5938
badnut22	0.7651	0.8509	0.90	0.3700
badnut31	0.8637	0.7214	1.20	0.2331
badnut32	-0.4148	0.8328	-0.50	0.6191
farm11	1.0739	0.9763	1.10	0.2731
farm12	0.2750	0.8587	0.32	0.7492
farm21	2.0883	0.9004	2.32	0.0217
farm22	0.6587	0.7308	0.90	0.3688
farm31	0.7781	0.8759	0.89	0.3758
farm32	0.6949	0.7013	0.99	0.3233
med11	0.5739	1.0053	0.57	0.5689
med12	0.7225	0.8353	0.86	0.3885
med21	2.2131	0.9871	2.24	0.0264
med22	1.2190	0.6730	1.81	0.0721
med31	0.9243	0.6964	1.33	0.1864
med32	1.7934	0.7147	2.51	0.0132

* SAS Output

TABLE 5. Solution for Random Effects

	Estimate	StdErrEst	t Value	Pr > |t|*
district11	1.1031	1.0825	1.02	0.3098
district12	-2.7568	1.0615	-2.60	0.0103
district13	1.6537	1.0760	1.54	0.1264
district21	0.8236	1.0767	0.76	0.4455
district22	0.01994	1.0372	0.02	0.9847
district23	-0.8435	1.0502	-0.80	0.4231
district31	0.9874	1.0254	0.96	0.3371
district32	-0.2911	1.0348	-0.28	0.7789
district33	-0.6963	1.0162	-0.69	0.4943

* SAS output

CONCLUSION

In condition 3, where the means of random effect are different for the 3 districts, the ARBs are different from ARBs other conditions of the simulations data. Generally, the larger the n, the smaller the bias. The larger the n, the better the parameters values or it is closer to the realized finite population values.

The Figure 2 shows that the amplitudes of RRMSE are narrower for the larger n. It means the estimations are better for larger n. Condition 3 gives slightly different values estimator from conditions 1 and 2.

The Simulations data need improvement to obtain the better results in parameter estimations, especially for the computation of prov parameters, which the ARBs show that prov parameters are the highest values for conditions 3 of simulations data, where the mean of 3 provinces are different.

Furthermore, the condition for correlated data between explanatory variables need to be carried on to know this effect for the parameter estimating. In addition, the random effects of the simulation data need to be assessed. As a conclusion of the application of the model for the data, implementation of the model for the poverty data in Java Island needs more explanatory variables as well as interaction between them to have broader view of the applications.

ACKNOWLEDGMENTS

This paper is based upon work partially supported by the Direktorat Jenderal Pendidikan Tinggi (DIKTI) through Hibah Pasca IPB Geoinformatik. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of agencies.

REFERENCES

1. R.P. Haining, (1990), *Spatial data analysis in the social and environmental sciences*, Cambridge University Press, Reading, NY.
2. N.A.C. Cressie, (1993), *Statistics for spatial data*, John Wiley and Sons Inc, Reading, NY.
3. G. Tampubolon, (2009), *Esai tentang kemiskinan dan kesehatan masyarakat di Indonesia selama dwiwindu*, preprint, available at `http://kemisan.files.wordpress.com/2009/02/esai-kemiskinan-kesehatan.pdf`.
4. P. McCullagh, (1980), Regression Models for Ordinal Data, *Journal of the Royal Statistical Society. Series B, Methodological* **42** : 109 – 142
5. D. Hedeker, and R.D. Gibbons, (1994), A random-effects Ordinal Model for Multilevel Analysis, *Biometrics*, Vol.50, 4, pp. 933–944.
6. D. A. Harville, and R. W. Mee, (1984), A Mixed-Model procedure for analyzing ordered categorical data, *Biometrics*, Vol.40, pp. 393–408.
7. P. McCullagh and J.A. Nelder FRS, (1989), *Generalized Linear Models*, Chapman and Hall, Reading, London, 1989.
8. V. Nekrasaite, (2008), *Small area estimation in practice*, preprint, available at `http://www.ms.ut.ee/samp2008/Presentations/VNekrasaiteLiege.pdf`.