

Nested Generalized Linear Model with Ordinal Response for Correlated Data

Yekti Widyaningsih¹, Asep Saefuddin², Khairil A. Notodiputro², and Aji H. Wigena²

Abstract—In this paper, we discuss the generalized linear models with ordinal response for correlated data in nested area. Some basic concepts are described, that is nested spatial, threshold model, and cumulative link function. Due to correlated data used for this modeling, Generalized Estimating Equation (GEE) is used as model parameters estimation method. Nested is shown by the model building and its application on nested spatially data. In this method, some Working Correlation Matrices (WCM) are able to be specified depend on the nature and type of the data. In this study, 3 types of WCM and 2 types of parameters estimation covariance are used to see the results of parameters estimation from these combinations. As a conclusion, independent WCM is appropriate to the data.

Keywords—nested generalized linear model, ordinal response, working correlation matrix, correlated

Abstrak—Makalah ini membahas generalized linear models dengan respon ordinal untuk data berkorelasi pada area tersarang. Beberapa konsep dasar dibahas, yaitu sedikit pendahuluan mengenai spatial tersarang, model threshold, dan fungsi penghubung kumulatif. Karena ada indikasi data berkorelasi, Generalized Estimating Equation (GEE) digunakan untuk pendugaan parameter model. Pembentukan model disesuaikan dengan kondisi tersarang dan diaplikasikan pada data spasial tersarang. Pada metode pendugaan parameter GEE, beberapa Working Correlation Matrices (WCM) dapat ditentukan tergantung dari kondisi data. Tiga struktur WCM dan 2 jenis pendugaan digunakan untuk melihat pengaruhnya pada hasil pendugaan parameter. Hasil perhitungan memberikan kesimpulan bahwa WCM independent paling sesuai untuk data yang digunakan.

Kata Kunci—nested generalized linear model, respon ordinal, working correlation matrix, berkorelasi

I. INTRODUCTION

As the starting consideration of the nested Generalized Linear Mixed Models (nested GLMM) for ordinal response, this paper works through about nested Generalized Linear Models (nested GLMs) for ordinal response, sub topics about parameter estimation method, and implementation of the model to the data.

Related to the evaluation of regions on poverty alleviated program, comparison among regions is needed. In this work, the score in ordinal scale is prefer than numeric to simplify the interpretation [1]. This study uses ordinal response for modelling, and the unit of observation is sub district. Connected to this region and certain multilevel spatial survey, assumed the regions (e.g., districts or ‘kabupaten’) of one area (e.g., province) are similar but not identical for another area. Such an arrangement is called a *nested*, with levels of district nested under the levels of province. For example, consider the government has a goal to reduce poverty and modeling is used to know the factors that contribute to determine poverty level. The question is: do these factors have the same effects on poverty level in all provinces? This question will be answered through the nested modeling.

There are some districts available from each province. The situation is depicted by Figure 1, which in this problem, a district from particular province has different nature from districts of another province [2]. Every

province has a particular nature and policy especially for a specific province such as ‘Daerah Istimewa’. This situation has an effect on the correlation matrix and parameter estimation. Based on this effect, the nature of the spatial component should be considered especially when a modelling is needed to analyze the effects of districts and province

Spatial data can be viewed as realizations of a spatial stochastic process $\{Z(s): s \in D\}$ where s is the location from which the data is observed and D is a random set in d dimensional Euclidean space [3]. Lattice data is defined as follows. Denote that $Z(s_1), \dots, Z(s_n)$ are lattice data observed at n sites. D is a fixed subset of R^d and it is partitioned into a finite number of lattices (or areal units), while site index s varies continuously over D [4].

Generalized estimating equations as parameters estimation method, were developed to extend generalized linear models to accommodate correlated longitudinal and/or clustered data [5]. In statistics, a Generalized Estimating Equation (GEE) is used to fit the parameters of a generalized linear model where unknown correlation between observations in a cluster is present. This method is usually used for the models of the clustering or longitudinal data. GEE was introduced by [6] as a method of regression model parameters estimation when dealing with correlated or clustered data. To define a regression model using the GEE methodology, one needs to define the following principles: the distribution of dependent variable (which must be a member of the exponential family), the monotonic link function, the independent variables, and the correlation or covariance structure of the repeated (clustered) measurements.

Analogy to the concept of four calcium content measurements on a leaf [7], unit of observation in this research is sub-district. Analogy of sub district is the point of calcium contents measurement, analogy of

Yekti Widyaningsih is with Department of Mathematics, FMIPA, Universitas Indonesia, Depok, 16424, Indonesia. E-mail:yekti@sci.ui.ac.id.

Asep Saefuddin, Khairil A. Notodiputro, and Aji H. Wigena are with Department of Statistics, FMIPA, Institut Pertanian Bogor, Bogor, 16680, Indonesia.

district is the leaf, and analogy of province is the tree where the leaf come from.

Sub district is presented as repeated measurement in a district, and district as level 1 is nested in province as level 2. Ordinal measurement is made on each sub-district within a district [1]; or repeated ordinal measurements are made on each district at different sub-districts. For instance, in a poverty study, a district may be represented by several sub districts at a given value of covariates, furthermore, these districts can be classified as “worst”, “moderate”, or “mild” in poverty [8].

Clusters are an example to represent the correlated observations: assumed that there is a correlation between observations in a cluster, while there is no correlation between observations from different cluster. This structure can be used as an example of longitudinal data or panel data as well as the data of family studies, or data with spatial structure [9].

To capture some of the beneficial aspects of quasi-likelihood estimation in the context of models for correlated data, [6] and [7] established GEE method. Beside robust in misspecification of covariance matrix, estimation using GEE is often easier to quantify than the maximum likelihood estimation [10].

In the parlance of the GEE approach, V_i is known as a “working” covariance matrix to distinguish it from the true underlying covariance among the Y_i . That is, the term “working” acknowledges our uncertainty about the assumed model for the variances and within-subject associations; unless they have been correctly modeled, our model for the covariance matrix may not be correct. The GEE approach allows the modeler to specify an incorrect structure [5]. The objective of this paper is to build generalized linear model of ordinal poverty response in nested area using GEE as the method to estimate the model parameters.

A. Study Area

The study area (D) consist of 3 provinces with 3 districts in every province, and n_{si} sub districts in district i , $i = 1, 2, 3$ of province s , $s=1,2,3$. The districts are chosen random independently in every province, with a uniform distribution. In other words, simple random sampling is used to select 3 districts from every province (without Banten, DKI, and DIY). Figure 2 shows study area in this research.

B. Data

Based on the report of BAPPENAS, there is a relation between the level of severity (poverty) of a region with a number of farmer families, the number of education centers (schools), the number of medical centers and/or health personnel, as well as the number of cases of malnutrition or bad nutrition [11]. To find out how these variables affects poverty level, through the modeling.

Nested GLM in this study are applied to the data on poverty with response variable is the poverty (severity) level of sub district, which has levels (worst, moderate, mild) [1]. The explanatory variables in this modeling are the number of schools, farmer families, health personnel, and malnutrition hotspots status. All explanatory variables are divided into 3 levels (low, moderate, and high), except the hotspot has 2 levels (hotspot = 1, non hotspot = 0). The model is shown by Equation 7.

II. METHOD

A. Threshold Model

Threshold is a latent variable at the model that made the difference between linear models with ordinal response and 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 [12]. For dichotomy model, it is assumed there is a threshold value, while for ordinal model with K categories (polytomy), it is assumed there are $K-1$ threshold values, namely $\gamma_1, \gamma_2, \dots, \gamma_{K-1}$. Response occurs in category k ($Z = k$), if the latent response y is greater than γ_{k-1} and smaller than γ_k . Assumed Y_j is unobserved, and the j -th observation is in a category, say category Z_j , $j = 1, \dots, N$. The relationship between Y_j and Z_j is taken to be,

$$\gamma_{k-1} < Y_j \leq \gamma_k \Leftrightarrow Z_j = k \quad (1)$$

where $k \in \{1, \dots, K\}$, $\gamma_0 = -\infty$, $\gamma_K = +\infty$ and $\gamma_1, \gamma_2, \dots, \gamma_{K-1}$ are unknown boundary points that define a partitioning of the real line into K intervals. Thus, when the realized value of Y_j belongs to the k -th interval, we observe that $z_j = k$. Under that assumptions, the probability-mass function of Z_1, \dots, Z_N is,

$$\begin{aligned} P(z_1, \dots, z_N) &= P_Y \{Z_j = z_j; j = 1, \dots, N\} \\ &= P_Y \{ \gamma_{z_j-1} < Y_j \leq \gamma_{z_j}; j = 1, \dots, N \} \end{aligned} \quad (2)$$

This model is called the threshold model [13].

These models can also be interpreted in terms of a *latent variable*. Specifically, suppose that the manifest response Z_j results from grouping an underlying continuous variable Y_i using cut-points $\gamma_1 < \gamma_2 < \dots < \gamma_{K-1}$, so that Z_j takes the value 1 if Y_j is below γ_1 , the value 2 if Y_j is between γ_1 and γ_2 , and so on, taking the value K if Y_j is above γ_{K-1} . Figure 3 illustrates this idea for the case of five response categories.

B. Cumulative Link Models

All of the models to be considered in this research arise from focusing on the cumulative distribution of the response. Let $\pi_{jk} = \Pr\{Z_j = k\}$ denote the probability that the response of an individual with characteristics x_j falls in the k -th category, and let p_{jk} denote the corresponding cumulative probability,

$$p_{jk} = p_r \{z_j \leq k\}$$

that the response falls in the k -th category or below, so,

$$p_{jk} = \pi_{j1} + \pi_{j2} + \dots + \pi_{jk} \quad (3)$$

$$\pi_{jk} = p_{jk} - p_{j,k-1}$$

Let $g(\cdot)$ denote a link function mapping probabilities to the real line. Then the class of models that will be considered, assumes that the transformed cumulative probabilities are a linear function of the predictors, of the form,

$$g(p_{jk}) = \gamma_k - x_j \beta \quad (4)$$

In this formula, γ_k is a constant representing the baseline value of the transformed cumulative probability for category k , x_j is a row vector of covariates of j -th observation and β is a column vector, represents the effect of the covariates on the transformed cumulative probabilities. Since it is written as the constant explicitly,

we assume that the predictors do not include a column of ones.

Suppose further the underlying continuous variable follows a linear model of the form

$$Y_j = x_j\beta + \varepsilon_j$$

where the error term ε_j has c.d.f. $F(\varepsilon_j)$. Then, the probability response of the j -th individual will fall in the k -th category or below, given x_j , satisfies the equation

$$p_{jk} = P_\gamma \{Y_j < \gamma_k\} = P_r \{\varepsilon_j < \gamma_k - x_j\beta\} = F(\gamma_k - x_j\beta) \tag{5}$$

and therefore follows the general form in Equation (4) with link given by the inverse of the c.d.f. of the error term

$$g(p_{jk}) = F^{-1}(p_{jk}) = \gamma_j - x_j\beta \tag{6}$$

It is assumed that the predictors x_j do not include a column of ones, because the constant is absorbed in the cut-points [15].

C. Model Building

This part is about Nested Generalized Linear Models for ordinal response. Model building in this study is based on spatial concept: the closer the observation, the larger the correlation [3]. Based on this concept, the idea was expanded to the nested of location or area. Furthermore, as the data in the observation is not always continue nor has normal distribution, the model should be in the general form

$$E(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

Generalized linear models for nested data is

$$E(Y_{s(i)}) = \mu_{s(i)} = g^{-1}(X_{s(i)}\beta_s)$$

$E(Y_{s(i)})$ is the expected value of $Y_{s(i)}$, response of observation in province s and district i ; $X_{s(i)}\beta_s$ is the linear predictor, a linear combination of unknown parameters β_s ; g is a link function.

Modelling in this research uses ordinal scale as response variable and some categorical (ordinal) variables as covariates. Index j is for sub district (repeated observations in district i), i is for district (nested in province s), and s is an index for province. District as level-1 is nested in province as the level-2. Link function for multinomial ordinal response is the cumulative logit model [16-17].

$$\log \left[\frac{p(y_{s(i(j))} \leq k)}{1 - p(y_{s(i(j))} \leq k)} \right] = \gamma_k - (x_{s(i(j))}\beta_s + \varepsilon_{s(i(j))}) \tag{7}$$

Where γ_k is threshold. β 's are the fixed effect at the transformed cumulative probabilities, $y_{s(i(j))}$ = the response variable of j th sub subject (sub district), in i th subject (district), in s th center (province), that could be continuous, binary, count, or category., $x_{s(i(j))}$ is the value of covariates x of sub district j , in district i and province s . Let S = number of centers, I = number of subject n_{si} = the number of repeated measurements (observations or sub-subjects) of the response on the subject. The response on the i th district of province s th can be grouped into a $n_{si} \times 1$ vector.

$$y_{s(i)} = \begin{pmatrix} y_{s(i(1))} \\ y_{s(i(2))} \\ \dots \\ y_{s(i(n_{si}))} \end{pmatrix},$$

for $s = 1, 2, \dots, S$ and $i = 1, 2, \dots, I, j = 1, \dots, n_{si}$. $y_{s(i)}$ are assumed to be independent of one another.

$$X = \begin{pmatrix} \mathbf{I}_{K-1} \otimes x'_{1(i(j))} & \dots & 0 \\ \dots & \dots & \dots \\ \mathbf{I}_{K-1} \otimes x'_{1(n_{si})} & \dots & 0 \\ \dots & \dots & \dots \\ \mathbf{1} \sum_{s=1}^S \sum_{i=1}^{n_s} n_{si} \otimes \mathbf{I}_{K-1} & \dots & \dots \\ \dots & \dots & \dots \\ 0 & \dots & \mathbf{I}_{K-1} \otimes -x'_{s(i(j))} \\ \dots & \dots & \dots \\ 0 & \dots & \mathbf{I}_{K-1} \otimes x'_{s(n_{si})} \end{pmatrix}$$

where X size is,

$(\sum_{s=1}^S \sum_{i=1}^{n_s} n_{si})(K-1) \times (K-1 + sp)$. $\varepsilon_{s(i(j))}$ is a spatially unstructured random effect assumed identically independently normally distributed of sub district j in district i and province s . This assumption is based on a theorem, that standardized of Pearson residual Moran's I convergence in distribution to $N(0,1)$ [18].

In this research, the model is developed using ordinal response variable with multinomial distribution for spatially nested area. As the data is nested and correlated, GEE is used as parameters estimation method for GLM. The models are implemented for poverty data in Java Island. Study area comprises of 3 provinces, 9 districts.

D. GEE for Ordinal Response Data

Generalized linear models were first introduced by Nelder and Wedderburn [17] and later expanded by McCullagh and Nelder [16]. The following discussion is based on their works and an extension of GEE from Liang & Zeger [6] for ordinal categorical responses data.

Suppose we have a multinomial response, say z . And for this response, there are K ordered categories with corresponding probabilities $\pi_1, \pi_2, \dots, \pi_K$, that is $\Pr(z = k) = \pi_k$. The proportional odds model is based on the cumulative probabilities, $\phi_k = \pi_1 + \pi_2 + \dots + \pi_k$ for $k = 1$ to $K-1$. Logit link function is used to relate ϕ_k to a linear function of p covariates X . Now let's take a look at the repeated situation. Suppose we have a sample of I subjects. Let z_{ij} be the ordinal response (with K levels) for the i th subject ($i = 1$ to I) at point j ($j = 1$ to n_i). Form of a $(K-1) \times I$ vector $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ij(K-1)})$ where $Y_{ijk} = 1$ if $z_{ij}=k$, and 0 otherwise.

Let's denote the expectation of y_{ij} as $\pi_{ij} = E(y_{ij}) = (\pi_{ij1}, \pi_{ij2}, \dots, \pi_{ij(K-1)})'$ with $\pi_{ijk} = \Pr(y_{ijk} = 1)$. And let x_{ij} denote a $1 \times p$ row vector of covariates for subject i at sub subject j .

The objective of this part is to model the π_{ijk} as a function of x_{ij} and the regression parameters $\theta = (\gamma_1, \gamma_2, \dots, \gamma_{K-1}, \beta)'$ where γ_k are intercept or cut-point parameters (threshold) and β is a $p \times 1$ vector of regression parameters. Let $\phi_{ijk} = \pi_{ij1} + \pi_{ij2} + \dots + \pi_{ijk}$ denote the cumulative probabilities. Then the proportional odds model at sub subject j is: $\log it(\phi_{ijk}) = \gamma_k + x_{ij}\beta$. To establish notation, let $y_i = (y_{i1}, \dots, y_{ini})'$, and $\pi_i = (\pi_{i1}, \dots, \pi_{ini})'$. Then θ can be estimated through solving the estimating equation as follows

$$\psi(\theta) = \sum_{i=1}^I \frac{\partial \pi_i}{\partial \theta} V_i^{-1} (y_i - \pi_i) \quad (8)$$

$$= [\theta]_{(K-1+sp) \times 1}$$

For clustered nested model,

$$\psi(\theta) = \sum_{s=1}^S \sum_{i=1}^{n_s} \frac{\partial \pi_{s(i)}}{\partial \theta} V_{s(i)}^{-1} (y_{s(i)} - \pi_{s(i)}) \quad (9)$$

$$= [\theta]_{(K-1+sp) \times 1}$$

where,

$$\frac{\partial \pi_{s(i)}}{\partial \theta} = \begin{pmatrix} \frac{\partial \pi_{s(i(1)),1}}{\partial \gamma_1} & \dots & \frac{\partial \pi_{s(i(1)),K-1}}{\partial \gamma_{K-1}} & \frac{\partial \pi_{s(i(1)),1}}{\partial \beta_{s1}} & \dots & \frac{\partial \pi_{s(i(1)),K-1}}{\partial \beta_{sp}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial \pi_{s(i(1)),K-1}}{\partial \gamma_1} & \dots & \frac{\partial \pi_{s(i(1)),K-1}}{\partial \gamma_{K-1}} & \frac{\partial \pi_{s(i(1)),K-1}}{\partial \beta_{s1}} & \dots & \frac{\partial \pi_{s(i(1)),K-1}}{\partial \beta_{sp}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial \pi_{s(i(n_s)),1}}{\partial \gamma_1} & \dots & \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \gamma_{K-1}} & \frac{\partial \pi_{s(i(n_s)),1}}{\partial \beta_{s1}} & \dots & \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \beta_{sp}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \gamma_1} & \dots & \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \gamma_{K-1}} & \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \beta_{s1}} & \dots & \frac{\partial \pi_{s(i(n_s)),K-1}}{\partial \beta_{sp}} \end{pmatrix}$$

the size of $\frac{\partial \pi_{s(i)}}{\partial \theta}$ is $n_{si} \times (K-1) \times (K-1+p)$ and for all $j = 1, \dots, n_{si}$,

$$\frac{\partial \pi_s(i(j)),k}{\partial \gamma_k} = \frac{\partial \varphi_s(i(j)),k}{\partial \eta_s(i(j)),k}, k = 1, \dots, K-1$$

$$\frac{\partial \pi_s(i(j)),k}{\partial \gamma_{k-1}} = \frac{\partial \varphi_s(i(j)),k-1}{\partial \eta_s(i(j)),k-1}, k = 1, \dots, K-1$$

$$\frac{\partial \pi_s(i(j)),k}{\partial \gamma_m} = 0, \text{ for } k < m \text{ or } k - m > 1, k = 1, \dots, K-1 \text{ and } m = 1, \dots, K-1$$

$$\frac{\partial \pi_s(i(j)),1}{\partial \beta_q} = -\frac{\partial \varphi_s(i(j)),1}{\partial \eta_s(i(j)),1}, x_s(i(j))q, q = 1, \dots, p,$$

$$\frac{\partial \pi_s(i(j)),1}{\partial \beta_q} = -\left(\frac{\partial \varphi_s(i(j)),k}{\partial \eta_s(i(j)),k} - \frac{\partial \varphi_s(i(j)),k-1}{\partial \eta_s(i(j)),k-1} \right),$$

$$x_s(i(j))q, k = 2, \dots, K-1 \text{ and } q = 1, \dots, p$$

$V_{s(i)}^{-1}$ is an inverse (or generalized inverse) of $V_{s(i)}$. Generalized inverse is used if $V_{s(i)}$ is singular or almost singular due to redundant data. Here

$$V_{s(i)} = \phi A_{s(i)}^{-1/2} R_{s(i)}(\alpha) A_{s(i)}^{-1/2} \quad (10)$$

with little mathematical operations, from equation (10), the working correlation matrix, $R_{s(i)}(\alpha)$ is

$$R_{s(i)}(\alpha) = \frac{1}{\phi} A_{s(i)}^{-1/2} V_{s(i)} A_{s(i)}^{-1/2}$$

$$R_{s(i)}(\alpha) = \frac{1}{\phi} \begin{pmatrix} A_{s(i)}^{-1/2} V_{s(i)} A_{s(i)}^{-1/2} & \rho_{s(i)12} & \dots & \rho_{s(i)1n_{si}} \\ \rho_{s(i)12} & A_{s(i)2}^{-1/2} V_{s(i)2} A_{s(i)2}^{-1/2} & \dots & \rho_{s(i)2n_{si}} \\ \dots & \dots & \dots & \dots \\ \rho_{s(i)n_{sv}1} & \rho_{s(i)n_{sv}2} & \dots & A_{s(i)n_{si}}^{-1/2} V_{s(i)n_{si}} A_{s(i)n_{si}}^{-1/2} \end{pmatrix}$$

For simplifying, index $j = 1, \dots, n_{si}$ is not inside parenthesis, but the same meaning with written before is maintained.

Note that there is a subscript $s(i)$ in $R_{s(i)}(\alpha)$ which means each subject has different working correlation

matrix. In fact, only the diagonal blocks are different for different subjects, the off-diagonal blocks will be the same for all subjects. The diagonal blocks of

$$R_{s(i)}(\alpha), A_{s(i)}^{-1/2} V_{s(i)} A_{s(i)}^{-1/2} \text{ with}$$

$$A_{s(i)j}^{1/2} = \text{diag} \left\{ \left\{ \pi_{s(i)j,1} (1 - \pi_{s(i)j,1}) \right\}^{1/2}, \dots, \left\{ \pi_{s(i)j,K-1} (1 - \pi_{s(i)j,K-1}) \right\}^{1/2} \right\}$$

and

$$V_{s(i(j))} = \left[\text{diag} \left[\pi_{s(i(j))} \right] - \pi_{s(i(j))} \pi_{s(i(j))}' \right]$$

are specified entirely by $\pi_{s(i(j))}$. In particular, the diagonal elements of $A_{s(i(j))}^{-1/2} V_{s(i(j))} A_{s(i(j))}^{-1/2}$ are 1 and off-diagonal

$$(k, m) \text{ element are } \frac{-\pi_{s(i(j))k} \pi_{s(i(j))m}}{\left\{ \pi_{s(i(j))k} (1 - \pi_{s(i(j))k}) \pi_{s(i(j))m} (1 - \pi_{s(i(j))m}) \right\}^{1/2}}$$

which are not constant and depend on the categories k and m at measurement j .

The unknown off-diagonal blocks of $R_{s(i)}(\alpha)$ are the $(K-1) \times (K-1)$ matrix $\rho_{uv}(\alpha)$, $u, v = 1, \dots, n_{si}$ which need to be parameterized and estimated [19].

E. Working Correlation Matrices

This part describes about Working Correlation Matrix (WCM) for ordinal multinomial model. $Y_{sij,k}$ and $\pi_{sij,k}$ are described in part D in this section. WCM uses Pearson-like residuals that defined as follows

$$r_{sij,k} = (y_{sij,k} - \pi_{sij,k}) \sqrt{\frac{\omega_{sij}}{\pi_{sij,k} (1 - \pi_{sij,k})}}$$

and the vector $r_{sij} = [r_{sij,1}, \dots, r_{sij,K-1}]$. The following structures are available [19].

1. Exchangeable

The exchangeable correlation structure is defined as:

$$R_{uv} = \begin{cases} 1, & \text{if } u=v \\ \alpha & \text{otherwise} \end{cases}$$

where,

$$\alpha = \frac{\sum_{s=1}^S \sum_{i=1}^I f_i \sum_{j < j'} \frac{1}{2} (r_{sij} r_{sij'} + r_{sij'} r_{sij})}{\left(\frac{1}{2} \sum_{s=1}^S \sum_{i=1}^I f_i n_{si} (n_{si} - 1) \right) - (K-1 + p_x)}$$

and $\rho_{siv} = \alpha$; $u, v = 1, \dots, n_{si}$ and $u \neq v$ [19].

2. Unstructured

The unstructured correlation structure is defined as:

$$R_{uv} = \begin{cases} 1, & \text{if } u=v \\ \alpha_{uv} & \text{otherwise} \end{cases}$$

where,

$$\alpha_{uv} = \frac{\sum_{s=1}^S \sum_{i=1}^I f_i \sum_{j < j'} \frac{1}{2} (r_{si} r_{siu} + r_{siu} r_{si})}{\left(\sum_{s=1}^S \sum_{i=1}^I f_i I_{si} I_{si,uv} \right) - (J-1 + p_x)}$$

And $\rho_{iv} = \alpha_{uv}$, $u, v = 1, \dots, n_i$ and $u \neq v$ [19].

Some provisions in choosing the working correlation matrix are: (1) if I is small and data are balanced and complete, then an unstructured matrix is recommended, (2) if observations are repeated, then use a structure that accounts for correlation as function of time (stationary, or auto-regressive), (3) if observations are clustered (i.e. no logical ordering) then *exchangeable* may be

appropriate, (4) if the number of clusters is small, then independent may be the best [20].

F. Algorithm for GEE Parameter Estimation

The algorithm for estimating model parameters using GEEs is outlined below [19, 6]. The standard iterative procedure to fit GEE, based on Liang and Zeger is :

- 1) Compute initial estimates for θ , $\theta^{(0)}$ using conventional GLM, i.e., assuming independence.
- 2) Compute the working correlation $R(\alpha)$ based on $\theta^{(0)}$, Pearson residuals and a specified working correlation structure. Check if $R(\alpha)$ is positive definite for exchangeable and unstructured structures. If it is not, revise it to be equal to $\frac{1}{1+\zeta}(R(\alpha) + \zeta I)$, where I is an identity matrix and ζ is a ridge value such that the adjusted matrix is positive definite. If a fixed correlation matrix is specified by the researchers and it is not positive definite, then cannot continue. As an alternative, then compute the initial estimate of the covariance matrix of $y_{s(i)}$ ($V_{s(i)}^{(0)}$), the generalized estimating equation $\Psi^{(0)}$, and generalized Hessian matrix $H^{(0)}$ (see formulae below) based on $\theta^{(0)}$ and $V_{s(i)}^{(0)}$.
- 3) Initialize $v = 0$.
- 4) Set $v = v + 1$.
- 5) Compute estimates of v th iteration (from Taylor series)

$$\theta^{(v)} = \theta^{(v-1)} - (H^{(v-1)})^{-1} \Psi^{(v-1)}$$

If v/N_v is a positive integer, update the working correlation, checking for positive definiteness as above.

- 6) Compute an estimate of the covariance matrix of y_j and its generalized inverse

$$V_{s(i)}^{(v)} = \phi A_{s(i)}^{1/2} R_{s(i)}(\alpha) A_{s(i)}^{1/2} \text{ and}$$

$$(V_{s(i)}^{(v)})^{-1} = \phi^{-1} A_{s(i)}^{-1/2} R_{s(i)}^{-1}(\alpha) A_{s(i)}^{-1/2}$$

For the ordinal multinomial model, replace $R(\alpha)$ with $R_{s(i)}(\alpha)$ in the above equations.

- 7) Revise $s^{(v)}$ and $H^{(v)}$ based on $\theta^{(v)}$ and $V_{s(i)}^{(v)}$

$$\Psi^{(v)} = \sum_{s=1}^S \sum_{i=1}^I f_{s(i)} \left(\frac{\partial \pi_{s(i)}}{\partial \theta} \right)' (V_{s(i)}^{(v)})^{-1} [z_{s(i)} - \pi_{s(i)}]$$

$$H^{(v)} = - \sum_{s=1}^S \sum_{i=1}^I f_{s(i)} \left(\frac{\partial \pi_{s(i)}}{\partial \theta} \right)' (V_{s(i)}^{(v)})^{-1} \left[\frac{\partial \pi_{s(i)}}{\partial \theta} \right]$$

- 8) Check the convergence criteria. If they are met or the maximum number of iterations is reached, then stop. The final vector of estimates is denoted by $\hat{\theta}$. Otherwise, go back to step (4) [19].

1. Convergence Criteria

Let ϵ_p and ϵ_H are given as tolerance levels, then the criteria for parameter convergence can be written as follows:

$$\max_j \left(\frac{|\theta_j^{(v)} - \theta_j^{(v-1)}|}{|\theta_j^{(v-1)}| + 10^{-6}} \right) \in \epsilon_p$$

Hessian convergence: $(s^{(v)})'(H^{(v)})^{-1}(s^{(v)}) < \epsilon_H$ with $\epsilon_H = 10^{-4}$ [19] (may be specified by researcher) after the log-likelihood or parameter convergence criterion has been satisfied.

2. Parameter Estimate Covariance Matrix, Correlation Matrix and Standard Errors

Two parameter estimate covariance matrices can be calculated, that is *model-based* and *robust* estimators. In the generalized linear model, the consistency of the model-based parameter estimate covariance depends on the correct specification of the mean and variance of the response (including correct choice of the working correlation matrix). However, the robust parameter estimate covariance is still consistent even when the specification of the working correlation matrix is incorrect as we often expect [20].

Model-based parameter estimate covariance is $\sum_M = -H_1^{-1}$ where H_1^{-1}

is the generalized inverse of :

$$H_1 = - \sum_{s=1}^S \sum_{i=1}^I \left(\frac{\partial \pi_{s(i)}}{\partial \theta} \right)' V_{s(i)} \left(\frac{\partial \pi_{s(i)}}{\partial \theta} \right)$$

The *robust* parameter estimate covariance is:

$$\sum_R = H_1^{-1} H_2 H_1^{-1}$$

where,

$$H_2 = \sum_{s=1}^S \sum_{i=1}^I f_{s(i)} \left(\frac{\partial n_{s(i)}}{\partial \theta} \right)' V_{s(i)}^{-1} \text{cov}(z_{s(i)}) V_{s(i)}^{-1} \left(\frac{\partial n_{s(i)}}{\partial \theta} \right)$$

and $\text{cov}(z_{s(i)})$ can be estimated by,

$$(z_{s(i)} - \pi_{s(i)})(z_{s(i)} - \pi_{s(i)})'$$

Note that model-based parameter estimate covariance will be affected by how the scale parameter is handled, but the robust parameter estimate covariance will not be affected by the estimate of the scale parameter because ϕ is cancelled in different terms [19].

3. Parameter Estimate Standard Error

For the ordinal multinomial model, let $\hat{\gamma}_k$, $k = 1, 2, \dots, K-1$, be threshold parameter estimates and $\hat{\beta}_r$, $r = 1, \dots, p$ denote non-redundant regression parameter estimates. Their standard errors are the square root of the r th diagonal element of Σ_r : $\hat{\sigma}_{\gamma_k} = \sqrt{\sigma_{kk}}$ and $\hat{\sigma}_{\beta_r} = \sqrt{\sigma_{(K-1+r), (K-1+r)}}$, respectively [19].

G. Wald Statistics

For ordinal multinomial model, the more general test matrix $L = (L(\gamma), L(\beta))$, where $L(\gamma) = (I_1, \dots, I_{K-1})$ consists of columns corresponding to threshold parameters and $L(\beta)$ be the part of L corresponding to regression parameters. Consider matrix $L_0 = (I_0, L(\beta))$, where the column vectors corresponding to threshold parameters are replaced by their sum $I_0 = \sum_{k=1}^{K-1} I_k$. Then $L\theta$ is estimable if and only if $L_0 = L_0 H_0$, where $H_0 = (X_1' \Omega X_1) X_1' \Omega X_1$ is a $(1+p) \times (1+p)$ matrix constructed using $X_1 = (1, X)$. Ω is the scale weight matrix with i th diagonal element ω_i and such that $L_0 \theta$ is estimable. The Wald statistic for testing $L\theta = K$, where L is a $r \times (K-1+p)$ full row rank hypothesis matrix and K is a $r \times 1$ resulting vector, is defined by

$$S = (L\hat{\theta} - K)' (L \sum L')^{-1} (L\hat{\theta} - K)$$

where $\hat{\theta} = (\hat{\gamma}, \hat{\beta})$ is the GEE estimate and Σ is the estimated covariance matrix (Σ could be the model based or robust estimator). The asymptotic distribution of S is $\chi_{r_c}^2$, where $r_c = \text{rank}(L \Sigma L')$ [10].

1. Wald Confidence Intervals

For the ordinal multinomial model, the $100(1 - \alpha)\%$ Wald confidence interval for parameter θ_r is given by

$$(\hat{\theta}_r - z_{1-\alpha/2}\sigma_{\theta_r} + z_{1-\alpha/2}\hat{\sigma}_{\theta_r})$$

the estimate of $\exp(\theta_r)$ is $\exp(\hat{\theta}_r)$, the standard error estimate of $\exp(\hat{\theta}_r)$ is $(\exp(\hat{\theta}_r) \cdot \hat{\sigma}_{\theta_r})$ and the corresponding $100(1 - \alpha)\%$ Wald confidence interval for $\exp(\theta_r)$ is

$$\left(\exp(\hat{\theta}_r - z_{1-\alpha/2}\sigma_{\theta_r}), \exp(\hat{\theta}_r + z_{1-\alpha/2}\hat{\sigma}_{\theta_r})\right)$$

III. RESULTS AND DISCUSSION

Appendix 1 shows the parameter estimates and their standard errors and Appendix 2 presents significance of the parameters. The Following is the explanation about the output.

A. Standard Error of Parameters Estimate

Figure 4 shows the standard errors of model based GLM parameters. Standard errors unstructured tend smaller than standard errors independent and exchangeable, while standard errors independent are the highest.

Moreover, according to Figure 5, there is no particular pattern between exchangeable, unstructured, and independent for robust parameter estimates. For some parameters (prov2, farm11, farm21, sch12, sch22), standard errors unstructured are extremely high.

From Figure 6, robust estimation with independent WCM has the lowest standard error, but according to the nature of the data, unstructured WCM is the most appropriate due to high correlation between sub districts in the same district.

It is desirable to compare the fit of different working correlation structures within a GEE for Nested GLM. An informal comparison is to compare the standard error of the robust (SE_R) and model-based (SE_M). There are no guidelines regarding the size of the ratio, but higher ratios reflect poorer model fit. This comparison is qualitative, but it is the best approach available at this time [21]. Table 1 presents the averages of this ratios (SE_R/SE_M). It shows independent WCM has the smallest ratio. It means independent WCM is most appropriate to the data.

B. Significance (p-values) (Appendix 2)

From Appendix 2, it is believed that province 2 (West Java) is different from province 3 (East Java) for all models (p-values ≈ 0.000). Number of farmer families in province 2 (Central Java) is significant as a contribution to determine the poverty level. Furthermore, number of school is also significant as a contribution to determine the poverty level in West Java.

In addition, exchangeable working correlation matrix has low values, i.e. $corr(\gamma_{11}, \gamma_{11}) = 0.161$, $corr(\gamma_{11}, \gamma_{12}) = corr(\gamma_{12}, \gamma_{11}) = 0.061$, and $corr(\gamma_{12}, \gamma_{12}) = 0.063$, while

unstructured working correlation matrix has varies values, with the full range of correlation values or (0, 1). Some correlations are 1 in some locations (not in the main diagonal).

1. Some interpretation of parameters

School1 in *prov1* with model based and unstructured WCM: in West Java, probability a sub district with category of school1 to be the more severe is $\exp(2.05) = 7.77$ times of sub district with category of school3, with other values of variables are fixed.

2. Classification Tables

Appendix 3 shows classifications result of observed and predicted result of Nested GLM. The true classification is always expected to have high value. The good models have high true classification. True classification for exchangeable WCM is $((35+1+27)/128) \times 100\% = 49.2\%$. For unstructured WCM is also $((35+1+27)/128) \times 100\% = 49.2\%$, while for independent WCM is the highest, $((39+4+35)/128) \times 100\% = 60.9\%$ for both estimations, *robust* and *model-based*. Figure 6 also shows that standard errors of *robust* independent tends lower than other standard errors.

IV. CONCLUSION

Even though based on the nature of the data, unstructured WCM is the most appropriate due to high correlation between sub districts in the same district, but according to the result of modeling, independent WCM is the most appropriate to the data. Ratio of SE_R/SE_M is the smallest, 0.56 and percentage of the true classification is the highest, 60.9%

Using robust estimation with unstructured WCM should be avoided, due to unstable standard error, but model based unstructured is fine (stable).

In general, parameters estimate of model based are more stable than those of robust.

Related to the nested area using in the modeling, the significance results (p-values) give the indication that characteristic of Central Java is different from West Java and East Java, which is appropriate to research of "Civilization Java" by Rahardjo [22]. Rahardjo said, characteristic of geographical areas in Central Java is more closed than in East Java. In addition, number of farmer families in Central Java is significant as contribution to poverty level, but not significant in other provinces. In addition, number of schools in West Java is significant to determine the poverty level, but not significant in East Java.

It is recommended to conduct simulation data to look the differences among some working correlation matrices. Furthermore nested area based on geographical conditions, such as northern coast, inland, and southern coast is still a challenge as an open problem to be studied.

ACKNOWLEDGEMENT

We wish to give our gratitude to DP2M DG DIKTI DEPDIKNAS who has supported this research through Hibah Penelitian Pascasarjana Nasional 2009-2011.

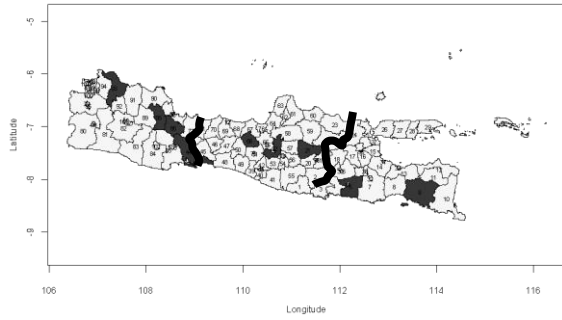


Figure 1. Nested areas (dark color are districts sample)

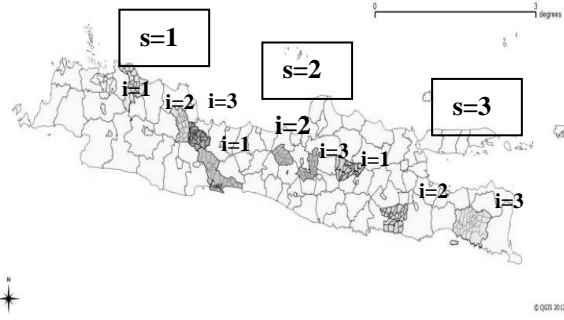


Figure 2. Study area with 3 provinces $\{s = 1, 2, 3\}$, 3 districts $\{i = 1, 2, 3\}$ are randomly chosen from each province, *i.e.* West, Central and East Java. There are n_{si} sub districts in district i of province s

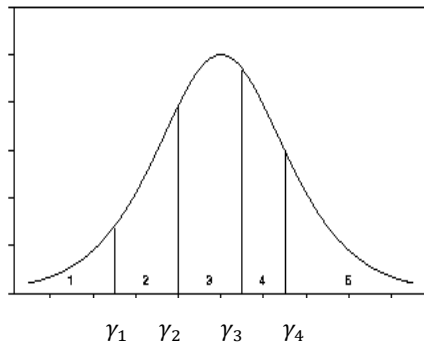


Figure 3. An ordered response and its latent variable [14]

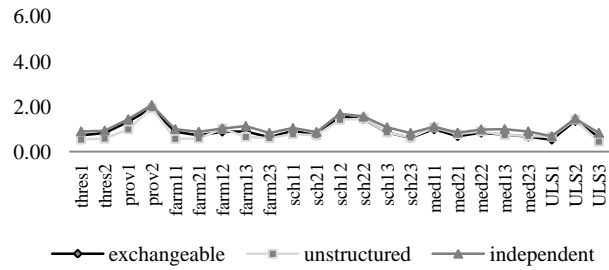


Figure 4. Standard errors of model based GLM parameters

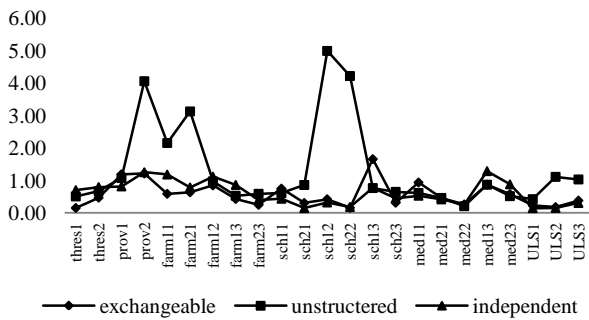


Figure 5. Standard errors of robust GLM parameters

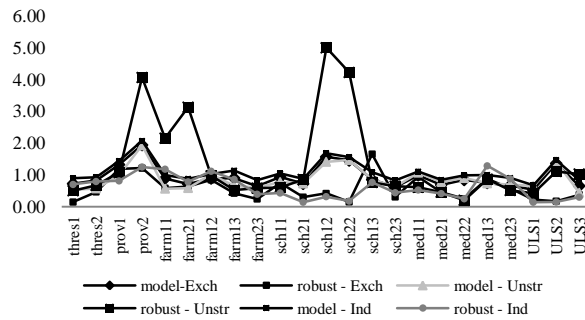


Figure 6. Standard error of all models

TABLE 1
AVERAGES OF SE_R/SE_M OF NESTED GLM PARAMETER ESTIMATES

	Working Correlation Matrix		
	Exchangeable	Unstructured	Independent
Average	0.63	1.5	0.56

APPENDIX 1
PARAMETERS AND STANDARD ERRORS OF NESTED GENERALIZED LINEAR MODEL

Parameter	PARAMETERS AND STANDARD ERRORS OF NESTED GENERALIZED LINEAR MODEL												
	Exchangeable				Unstructured				Independent				
	model		robust		model		robust		Model		robust		
	β	Std. Error	β	Std. Error	β	Std. Error	β	Std. Error	β	Std. Error	β	Std. Error	
Threshold	[ordb=1]	0.30	0.73	0.30	0.15	-0.04	0.54	-0.04	0.51	-0.21	0.90	-0.21	0.70
	[ordb=2]	1.89	0.83	1.89	0.47	1.18	0.58	1.18	0.67	1.80	0.93	1.80	0.80
[prov=1]		0.41	1.32	0.41	1.19	0.31	0.99	0.31	1.07	-0.70	1.44	-0.70	0.82
[prov=2]		-21.85	1.94	-21.85	1.21	-27.13	1.93	-27.13	4.06	-28.65	2.07	-28.65	1.25
[prov=3]		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[farm=1]([prov=1])		1.14	0.89	1.14	0.59	0.39	0.57	0.39	2.15	2.68	1.00	2.68	1.18
[farm=2]([prov=1])		0.55	0.72	0.55	0.64	0.86	0.59	0.86	3.13	2.28	0.88	2.28	0.79
[farm=3]([prov=1])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[farm=1]([prov=2])		24.45	0.89	24.45	0.85	29.78	1.02	29.78	0.97	30.82	1.03	30.82	1.12
[farm=2]([prov=2])		22.367 ^b		22.367 ^b		29.178 ^b		29.178 ^b		28.935 ^b		28.935 ^b	
[farm=3]([prov=2])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[farm=1]([prov=3])		0.63	0.90	0.63	0.43	0.08	0.66	0.08	0.53	0.32	1.14	0.32	0.86
[farm=2]([prov=3])		0.39	0.65	0.39	0.25	0.59	0.60	0.59	0.59	0.14	0.83	0.14	0.40
[farm=3]([prov=3])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[school=1]([prov=1])		1.86	0.93	1.86	0.75	2.05	0.78	2.05	0.61	0.68	1.05	0.68	0.43
[school=2]([prov=1])		0.69	0.72	0.69	0.31	1.05	0.73	1.05	0.86	-0.38	0.87	-0.38	0.15
[school=3]([prov=1])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[school=1]([prov=2])		-0.16	1.57	-0.16	0.42	-0.69	1.41	-0.69	4.99	-0.10	1.69	-0.10	0.32
[school=2]([prov=2])		-1.13	1.46	-1.13	0.16	-1.43	1.47	-1.43	4.22	-1.16	1.56	-1.16	0.18
[school=3]([prov=2])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[school=1]([prov=3])		-0.84	0.86	-0.84	1.66	-0.30	0.86	-0.30	0.77	-0.46	1.08	-0.46	0.77
[school=2]([prov=3])		-0.41	0.62	-0.41	0.31	-0.19	0.63	-0.19	0.65	-0.53	0.83	-0.53	0.44
[school=3]([prov=3])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[medis=1]([prov=1])		-1.67	0.99	-1.67	0.94	-0.89	1.07	-0.89	0.61	0.24	1.11	0.24	0.53
[medis=2]([prov=1])		-0.34	0.69	-0.34	0.46	-0.93	0.78	-0.93	0.46	0.60	0.84	0.60	0.42
[medis=3]([prov=1])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[medis=1]([prov=2])		-		-		-		-		60.619 ^b		60.619 ^b	
[medis=2]([prov=2])		2158.590 ^b		2158.590 ^b		15220.791 ^b		15220.791 ^b		0.22	0.84	0.22	0.25
[medis=3]([prov=2])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
[medis=1]([prov=3])		0.46	0.75	0.46	0.87	-0.04	0.74	-0.04	0.87	-0.48	1.00	-0.48	1.29
[medis=2]([prov=3])		-0.22	0.67	-0.22	0.58	-0.02	0.70	-0.02	0.52	-0.28	0.90	-0.28	0.88
[medis=3]([prov=3])		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a		0 ^a	
ULS([prov=1])		0.52	0.53	0.52	0.23	1.27	0.66	1.27	0.42	0.41	0.68	0.41	0.15
ULS([prov=2])		2.94	1.36	2.94	0.18	1.94	1.42	1.94	1.11	2.98	1.47	2.98	0.15
ULS([prov=3])		-0.65	0.66	-0.65	0.38	-1.05	0.44	-1.05	1.03	-0.55	0.84	-0.55	0.30
(Scale)		1.19		1.19		1.19		1.19		1.19		1.19	

Dependent Variable: ordb

Model: (Threshold), prov, farm(prov), school(prov), medis(prov), ULS(prov)

a. Set to zero because this parameter is redundant.

b. Hessian matrix singularity is caused by this parameter. The parameter estimate at the last iteration is displayed.

APPENDIX 2
SIGNIFICANCE OF NESTED GLM PARAMETER ESTIMATES

Parameter		Exchangeable		Unstructured		Independent	
		model	robust	model	robust	model	Robust
Threshold	[ordb=1]	.678	.049	.940	.937	.817	.767
	[ordb=2]	.022	.000	.041	.079	.053	.024
[prov=1]		.754	.728	.755	.774	.628	.393
[prov=2]		0.000	0.000	0.000	.000	0.000	0.000
[prov=3]							
[farm=1]([prov=1])		.199	.051	.490	.855	.007	.024
[farm=2]([prov=1])		.442	.383	.145	.783	.010	.004
[farm=3]([prov=1])							
[farm=1]([prov=2])		0.000	0.000	0.000	0.000	0.000	0.000
[farm=2]([prov=2])							
[farm=3]([prov=2])							
[farm=1]([prov=3])		.488	.144	.907	.885	.781	.713
[farm=2]([prov=3])		.545	.114	.328	.316	.865	.721
[farm=3]([prov=3])							
[school=1]([prov=1])		.047	.013	.008	.001	.519	.120
[school=2]([prov=1])		.337	.023	.153	.223	.663	.009
[school=3]([prov=1])							
[school=1]([prov=2])		.920	.710	.623	.889	.951	.748
[school=2]([prov=2])		.438	.000	.332	.735	.457	.000
[school=3]([prov=2])							
[school=1]([prov=3])		.327	.613	.731	.699	.671	.549
[school=2]([prov=3])		.512	.193	.765	.772	.527	.237
[school=3]([prov=3])							
[medis=1]([prov=1])		.092	.074	.405	.146	.826	.645
[medis=2]([prov=1])		.620	.454	.234	.043	.479	.151
[medis=3]([prov=1])							
[medis=1]([prov=2])							
[medis=2]([prov=2])		.796	.398	.772	.206	.806	.340
[medis=3]([prov=2])							
[medis=1]([prov=3])		.544	.602	.962	.968	.630	.708
[medis=2]([prov=3])		.742	.705	.979	.972	.756	.750
[medis=3]([prov=3])							
ULS([prov=1])		.332	.022	.054	.003	.550	.005
ULS([prov=2])		.031	0.000	.173	.080	.042	0.000
ULS([prov=3])		.323	.088	.018	.309	.511	.070
(Scale)							

APPENDIX 3
CLASSIFICATION RESULT OF NESTED GLM

Y * \hat{Y} Crosstabulation, Exchangeable, model based

		\hat{Y}			Total
		1	2	3	
1	Count	35	5	10	50
	% within Y	70.0%	10.0%	20.0%	100.0%
Y 2	Count	25	1	11	37
	% within Y	67.6%	2.7%	29.7%	100.0%
3	Count	5	9	27	41
	% within Y	12.2%	22.0%	65.9%	100.0%
Total	Count	65	15	48	128
	% within Y	50.8%	11.7%	37.5%	100.0%

Y * \hat{Y} Crosstabulation, Exchangeable, robust

		\hat{Y}			Total
		1	2	3	
1	Count	35	5	10	50
	% within Y	70.0%	10.0%	20.0%	100.0%
Y 2	Count	25	1	11	37
	% within Y	67.6%	2.7%	29.7%	100.0%
3	Count	5	9	27	41
	% within Y	12.2%	22.0%	65.9%	100.0%
Total	Count	65	15	48	128
	% within Y	50.8%	11.7%	37.5%	100.0%

Y * \hat{Y} Crosstabulation, Unstructured, model based

		\hat{Y}			Total
		1	2	3	
1	Count	35	5	10	50
	% within Y	70.0%	10.0%	20.0%	100.0%
Y 2	Count	25	1	11	37
	% within Y	67.6%	2.7%	29.7%	100.0%
3	Count	5	9	27	41
	% within Y	12.2%	22.0%	65.9%	100.0%
Total	Count	65	15	48	128
	% within Y	50.8%	11.7%	37.5%	100.0%

Y * \hat{Y} Crosstabulation, Unstructured, robust

		\hat{Y}			Total
		1	3		
1	Count	27	23		50
	% within Y	54.0%	46.0%		100.0%
Y 2	Count	14	23		37
	% within Y	37.8%	62.2%		100.0%
3	Count	6	35		41
	% within Y	14.6%	85.4%		100.0%
Total	Count	47	81		128
	% within Y	36.7%	63.3%		100.0%

Y * \hat{Y} Crosstabulation, Independent, model based

		\hat{Y}			Total
		1	2	3	
1	Count	39	7	4	50
	% within Y	78.0%	14.0%	8.0%	100.0%
Y 2	Count	23	4	10	37
	% within Y	62.2%	10.8%	27.0%	100.0%
3	Count	4	2	35	41
	% within Y	9.8%	4.9%	85.4%	100.0%
Total	Count	66	13	49	128
	% within Y	51.6%	10.2%	38.3%	100.0%

Y * \hat{Y} Crosstabulation, Independent, robust

		\hat{Y}			Total
		1	2	3	
1	Count	39	7	4	50
	% within Y	78.0%	14.0%	8.0%	100.0%
Y 2	Count	23	4	10	37
	% within Y	62.2%	10.8%	27.0%	100.0%
3	Count	4	2	35	41
	% within Y	9.8%	4.9%	85.4%	100.0%
Total	Count	66	13	49	128
	% within Y	51.6%	10.2%	38.3%	100.0%

REFERENCES

[1] Y. Widyaningsih, A. Saefuddin, K.A. Notodiputro, "Generalized Linear Mixed Models with Ordinal Response for Clustered Nested Data". Dissertation. *Bogor Agricultural University*, Indonesia, 2012.

[2] S. Rahardjo, *Peradaban Jawa (Dari Mataram Kuno sampai Majapahit Akhir)*, Penerbit: Komunitas Bambu, 2011.

[3] N.A.C. Cressie, *Statistics for Spatial Data*, New York: John Wiley and Sons Inc, 1993.

[4] O. Schabenberger and C.A. Gotway, *Statistical Method for Spatial Data Analysis*, Washington DC: Chapman & Hall, 2005.

[5] J.W. Hardin and J.M. Hilbe, *Generalized Estimating Equations (GEE)*, Washington DC: Chapman & Hall/CRC, 2003.

[6] K.Y. Liang and S.L. Zeger, *Longitudinal data analysis using generalized linear models*, *Biometrika*. 73. 1. pp. 13-22, 1986.

[7] U. Halekoh, "Generalized Estimating Equations (GEE)". [lecture.tex. April 10] *Unit of Statistics and Decision Analysis*,

Faculty of Agricultural Sciences, University of Aarhus. 2008, [Online]. Available : <http://gbi.agrsci.dk/statistics/courses/Rcourses/Rcourse-DJF2008/> [26 Februari 2011, 14:48]

[8] Badan Pusat Statistik, "Profil Kemiskinan di Indonesia". *Berita Resmi Statistik*. Num. 45/Th.XIV, 1 July 2011

[9] S.L. Zeger and K.Y. Liang, "Longitudinal data analysis for discrete and continuous outcomes", *Biometrics*. Vol.42. pp. 121-130, 1986.

[10] C.E. McCulloch and S.R.Searle, *Generalized, Linear, and Mixed Models*, Wiley and Sons, 2001.

[11] A.S. Alisjahbana, "Pertumbuhan ekonomi, penyerapan tenaga kerja, kesenjangan dan penanggulangan kemiskinan". *Dokumen Bappenas*, 2011.

[12] D. Hedeker and R.D. Gibbons, "A random-effects Ordinal Model for Multilevel Analysis", *Biometrics*, Vol.50, 4, pp. 933-944, 1994.

[13] D.A. Harville and R.W. Mee, "A Mixed-Model procedure for analyzing ordered categorical data", *Biometrics*, Vol.40, pp. 393-408, 1984.

- [14] W. Simpson, "Ordinal and Multinomial Models". *Research Computing Services*, 2008. [Online]. Available: <http://intranet.hbs.edu/dept/research/statistics/> [5 August 2010]
- [15] G. Rodriguez, "Models for Ordinal Response Data", lecture notes, [Online]. Available : <http://data.princeton.edu/wws509/> [November 15, 11].
- [16] P. McCullagh and J. Nelder, *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5, 1989.
- [17] J.A. Nelder and R.W.M. Wedderburn, "Generalized Linear Models". *Journal of the Royal Statistical Society Series A*, 135, 370–384, 1972.
- [18] T. Zhang and G. Lin, "Cluster Detection Based on Spatial Associations and Iterated Residuals in Generalized Linear Mixed Models". *The International Biometric Society*, DOI: 10.1111/j.1541-0420.2008.01069.x, 2008.
- [19] IBM SPSS Statistics 20 Algorithms. [Online]. Available : ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/en/client/Manuals/IBM_SPSS_Statistics_Algorithms.pdf [April 3, 2012]
- [20] N.J. Horton and S.R. Lipsitz, "Review of Software to Fit Generalized Estimating Equation Regression Models", *The American Statistical Association*, Vol. 53, May 1999.
- [21] N. Koper and M. Manseau, "Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection", *Journal of Applied Ecology*, Vol. 46, 590–599. doi: 10.1111/j.1365-2664.2009.01642.x. 2009.
- [22] S. Rahardjo, *Peradaban Jawa (Dari Mataram Kuno sampai Majapahit Akhir)*, Penerbit: Komunitas Bambu, 2011.