

Modeling Spatial Ordinal Logistic Regression and The Principal Component to Predict Poverty Status of Districts in Java Island

Muhammad Nur Aidi M.S^{1,*}, Tuti Purwaningsih S.Stat²

¹Department of Statistics, Faculty of Mathematics and Science, Bogor Agricultural University, Dramaga, Bogor, West Java, 16680, Indonesia,

²Master Student at Department of Statistics, Faculty of Mathematics and Science, Bogor Agricultural University, Dramaga, Bogor, West Java, 16680, Indonesia

Abstract This study is intended to determine the factors that affect poverty-stricken districts in Java Island by using spatial ordinal logistic regression (SOLR). This study is urgent for local governments since they have a mission for the welfare of their citizens, especially alleviating poverty that occurred in each region. The evaluation is based on the best model and to evaluate SOLR model with and without handling multicollinearity. The principal component (PC) is for handling multicollinearity. This research used secondary data (Kusumaningrum, 2010)^[6]. Results showed the ordinal logistic regression (OLR) model with all explanatory variables spatial has Correct Classification Rate (CCR) value of 50%. Then the OLR with PC model has a CCR of 24%. So it can be concluded that SOLR model is better than OLR with PC model. Based on spatial model without PC, there are five variables that influence the level of district poverty on Java island, namely: the number of farmers, small industries, families without electricity, hospitals and spatial variable on districts poverty. The variables that have negative correlation are the number of farmers, families without electricity and spatial variables. Increase in the number of farmers and families without electricity will reduce the chances of a district becoming rich. The spatial correlation is negative, which means that if a district is surrounded by districts with high poverty levels, the chances of a district to become more wealthy will decline.

Keywords Spatial Ordinal Logistic Regression, Handling Multicollinearity, The Principal Component Analysis, Poverty, CCR.

1. Introduction

1.1. Background

Java Island is the most populous island in Indonesia. Java Island consists of 6 provinces namely the Special Capital Region of Jakarta, West Java, Banten, Central Java, Yogyakarta and East Java. Each province consists of several districts.

Each province has its vision and mission for the welfare of its citizens, especially in terms of alleviating poverty that occurred in each region. Fund available for this program is limited, so it is necessary to have an election as to which district is entitled as an object of this program. One way is to categorize districts into 6 levels of poverty so that the funds can be assigned on a priority scale. Poverty status of a district cannot be separated from the influence of the status of

poverty in the surrounding district. This indicates the existence of spatial effects. Pratama (2008)^[9] and Thaib (2008)^[11] perform spatial modeling logistics against poverty status of a region. Both studies concluded that the estimation of poverty of an area using spatial logistic regression will yield a better prediction than the classical logistic regression. Logistic regression model with the categorization of spatial poverty into two categories: poor and non-poor has been done by Suprpti (2009)^[10]. Therefore, in this research, we use ordinal logistic regression model spatially for six categories.

1.2. Purpose

The purpose of this research is to know the factors that affect poverty-stricken districts in Java based on Spatial ordinal logistic regression model and to evaluate spatial ordinal logistic regression with and without handling multicollinearity.

2. Review Literatures

* Corresponding author:

nuraidi@yahoo.com (Muhammad Nur Aidi M.S)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2013 Scientific & Academic Publishing. All Rights Reserved

2.1. Poverty

Indonesia is a developing country with the majority of the population experiencing poverty problem. In general, poverty is a situation where there is an inability to meet basic needs such as food, clothing, shelter, education and health.

There are several definitions of poverty made by government and non government agencies. According to the Statistics Center a household is poor if it has a per capita income below the poverty line. The poverty line is the minimum size of one's income that can still be used to meet basic human needs. Meanwhile, the World Bank defines extreme poverty as a condition in which a person lives on less than U.S. \$ 1 per day and poverty if someone has an income of less than U.S. \$ 2 a day. Pre-Prosperous Family is a family that cannot meet the minimum standards of basic human needs such as spiritual needs, food, clothing, housing, education and health. Prosperous families are families that can afford to meet basic human needs, but cannot meet higher human needs (BKKBN 2004)^[3]

2.2. Ordinal Logistic Regression

Ordinal logistic regression was used to model the relationship between variables with ordinal-scale response variables with explanatory variables. If the response variable Y is assumed there is an ordinal scale with K categories and $\mathbf{x} = (x_1, x_2, \dots, x_p)$ is the vector of explanatory variables, then the chances of the variable response of the k -th category of explanatory variables X in particular can be expressed by $P[Y=s|\mathbf{x}] = \pi_s(\mathbf{x})$ and the cumulative odds is in Function 1 (Hosmer & Lemeshow 2000)^[5].

$$P[Y \leq s|\mathbf{x}] = \pi_1(\mathbf{x}) + \dots + \pi_s(\mathbf{x}) \dots\dots\dots (1)$$

Cumulative logit model is defined by Function 2.

$$\begin{aligned} L_k(\mathbf{x}) &= \text{logit}(P[Y \leq s|\mathbf{x}]) \\ &= \log \left(\frac{P[Y \leq s|\mathbf{x}]}{1 - P[Y \leq s|\mathbf{x}]} \right) \\ &= \log \left(\frac{\pi_1(\mathbf{x}) + \dots + \pi_s(\mathbf{x})}{\pi_{s+1}(\mathbf{x}) + \dots + \pi_K(\mathbf{x})} \right) \\ &= \alpha_k - \mathbf{x}'\boldsymbol{\beta} \dots\dots\dots (2) \end{aligned}$$

where $s = 1, \dots, S-1$ and $\alpha_1, \dots, \alpha_{S-1}$ is the *threshold model* and $\boldsymbol{\beta}$ is a vector of regression coefficients.

Parameter estimation method that can be used in ordinal logistic regression is the *Maximum Likelihood* method. This method can be done if the observations are assumed to be independent. Its *likelihood* function for a sample with n independent observations, (y_i, x_i) , $i = 1, 2, \dots, n$, can be expressed as follows in Function 3 (Hosmer & Lemeshow 2000)^[5]

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [\pi_1(x_i)^{z_{0i}} \pi_2(x_i)^{z_{1i}} \dots \pi_s(x_i)^{z_{si}}] \dots\dots\dots (3)$$

$$\text{where : } z_{si} = \begin{cases} 1 & \text{for } y = s \\ 0 & \text{for } y \neq s \end{cases}$$

While the *log likelihood* function defined in Function 4 below.

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^n z_{0i} \ln[\pi_1(x_i)] + \\ &+ z_{1i} \ln \pi_2(x_i) + \dots + z_{si} \ln \pi_s(x_i) \dots\dots\dots (4) \end{aligned}$$

Furthermore, to obtain estimators of parameters of the

ordinal logistic regression function is to maximize the *log likelihood* of parameters.

In ordinal regression analysis there are five options for liaison function (*link function*) as found in Table 1. Their use depends on the distribution of the analyzed data. Logit is used in most data distribution. *Complementary log-log* is used for data that have a tendency of high value. *Negative log-log* is used for data that have a tendency of low-value, while probit is used if the spread is normally latent variables, whereas cauchit (Cauchy inverse) is used if latent variables have extreme values. Ordinal regression analysis previously described is an ordinal regression analysis with logit *link function* often called the ordinal logistic regression (Norusis 2010)^[8].

Table 1. Function Liaison Ordinal Regression

Liaison Function	Function Form
Logit	$\text{Log} \left(\frac{\pi}{1 - \pi} \right)$
<i>Complementary log-log</i>	$\text{Log}(-\text{Log}(1 - \pi))$
<i>Negative log-log</i>	$-\text{Log}(-\text{Log}(\pi))$
Probit	$\Phi^{-1}(\pi)$
Cauchit (Cauchy Inverse)	$\tan(\text{phi}(\pi - 0.5))$

2.2.1. Significance Testing Model

Likelihood ratio test of the model used to estimate the parameters *with the hypothesis*:

$$H_0 : \beta_1 = \dots = \beta_p = 0$$

H1 : there is at least one $\beta_i \neq 0$; $i=1, 2, \dots, p$, where i is the number of explanatory variables.

Likelihood ratio test statistic uses the G , where $G = -2 \ln(L_0 / L_k)$ where L_0 is the Likelihood function without variables and L_k is the Likelihood function with variables (Hosmer & Lemeshow 2000)^[5]. If H_0 is true, Statistics G will follow a Chi-square distribution with degrees of freedom p and H_0 will be rejected if the value of $G > \chi^2_{(p, \alpha)}$ or $p\text{-value} < \alpha$.

Wald test is used to test the significance of each coefficient in the model. The hypothesis is:

$$H_0 : \beta_i = 0$$

H1: $\beta_i \neq 0$; $i=1, 2, \dots, p$, where i is the number of explanatory variables.

Wald test calculates a W statistic, defined as

$$W_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

Reject H_0 if $|W| > Z_{\alpha/2}$ or $p\text{-value} < \alpha$ (Hosmer & Lemeshow 2000)^[5].

2.2.2. Assumptions of Logistic Regression

Logistic regression has the following assumptions (Garson 2010)^[4]. The data has no outliers. In logistic regression, outliers can affect the results significantly. It is recommended that there should be no multicollinearity among the explanatory variables.

2.2.3. Interpretation of Logistic Regression

Interpretation for ordinal logistic regression model can be done using the odds ratio value. Suppose for a nominal-scale variables $X(x_1 \text{ and } x_2)$, the odds ratio in the category $Y \leq s$ is defined in Function 5 (Agresti 1990)^[1] and Dillon and Goldstein (1984)^[3].

$$L_s(x_1) - L_s(x_2) = \log \left(\frac{P[Y \leq s | x_1] / P[Y > s | x_1]}{P[Y \leq s | x_2] / P[Y > s | x_2]} \right) = \beta_i(x_1 - x_2) \quad (5)$$

With $i = 1, 2, \dots, p$ (p = number of explanatory variables) and $s = 1, 2, \dots, S-1$. β_i is coefficient logistic regression₂ with meaning is:

$$\text{Ln}[\psi(x_1, x_2)] = \beta_i(x_1 - x_2)$$

So we get estimators for the odds ratio (ψ) as follows (Agresti 1990)^[1]:

$$\hat{\psi} = \exp[\beta_i(x_1 - x_2)]$$

For categorical explanatory variables, if the odds ratio value is > 1 , then the odds x_1 is greater than the odds x_2 or in other words, $P[Y \leq s | x_1]$ will always be greater than $P[Y > s | x_1]$. So it can be said, when x_1 chance for $Y \leq s$ greater than at x_2 .

For the continuous explanatory variables, x , the odds as x increases one unit is equal to $\exp[\beta(x_1 - x_2)]$ multiply odds when x has not increased. If the value of the odds ratio > 1 , then probability for $Y \leq s$ is greater when x be increased than x has not increased. For large-scale continuous variables, required a change of c units to interpretation, with an odds ratio of $\exp[c\beta]$.

2.3. Spatial Analysis

Spatial analysis is an analysis which includes the influence of spatial or space into the analysis. In spatial analysis the correlation between spaces is called spatial correlation. Thus each observation is not free stochastic (Ward & Gleditsch 2008)^[12].

The types of spatial data are data points, data line, polygon data and latis data. Data points are divided into discrete points and continuous point. Line data are for example, road maps, river or coastlines. Polygon data for example is a map of a garden in terms of irregular shapes. The latis data is for example provinces in which there were districts.

2.3.1. Contiguity Matrix

Contiguity matrix is a matrix that describes the relationship between regions, giving the value 1 if the area i neighbor with area j , while a value of 0 is given if area i is not adjacent to the area- j . Lee and Wong (2001)^[7] refer to this matrix as *binary matrix*, and also called the *connectivity matrix*, which is denoted by C , and c_{ij} being the value of matrix in row i and column j . Later in this study, it will be referred to as matrix W containing w_{ij} for row i and column j .

2.3.2. Spatial Ordinal Logistic Regression

Spatial Ordinal Logistic Regression is an analysis which

incorporates spatial effects into the ordinal logistic regression model. Suprapti (2009)^[10] uses spatial logistic regression model of the form:

$$g(x) = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) + \beta\psi$$

with $\psi = \frac{\sum_{j=1}^k w_{ij} \hat{y}_j}{\sum_{j=1}^k w_{ij}}$

Where is a form of auto-covariance and is a weighted average of the number of events among the k_i neighbors. Weighting of the location of the ij -th is through $w_{ij} = 1 / h_{ij}$ where h_{ij} is the euclid distance between the village i and j . \hat{y}_i is the alleged existence of an event.

Another model which could be established to handle the spatial relationships is :

$$\text{Logit} \left(\frac{P[Y \leq s | x]}{1 - P[Y \leq s | x]} \right) = X\beta + \rho W y + \varepsilon$$

Where k is a category s^{th} from all of the dependent variables and W is a weighting matrix that represents the spatial proximity of the region. $w_{ij} = 0$ if regions i and j are not adjacent regions directly while $w_{ij} = 1$ if regions i and j areas are immediately adjacent. Spatial weighting matrix (W) which has been obtained is multiplied by the vector y and the results will be considered as new variables and will be used in ordinal logistic regression analysis. In general, the parameter estimation process includes the testing of hypotheses, drawing conclusions and interpretations, following the rules in ordinal logistic regression

2.4. Model of Validation

Validation of models is using the *Correct Classification Rate* (CCR). CCR is the percentage of correct observations (suitability with the expected value). CCR can be calculated using the following Function 6:

$$\text{CCR} = \frac{\text{number of correct prediction}}{\text{number of observation}} \times 100\% \quad \dots\dots\dots(6)$$

The higher percentage of CCR shows higher accuracy (Hosmer and Lemeshow 2000)^[5].

2.5. Winsorizing Methods

Winsorizing method is one method to handling outliers, without having to lose the observation data. This method is particularly suitable for modeling data in which observations should not be eliminated. In this method the outliers are replaced with a value closest to it.

3. Materials dan Methods

3.1. Material

Data used in this research is secondary data obtained from hotspot poverty districts data in Java which is obtained from the research by Kusumaningrum (2010)^[6].

Dependent variables used in this study are presented in Table 3.

Explanatory variables used in this study are presented in Table 4.

Table 2. List of Number of Districts

No	Province Name	Number of Districts
1	DKI Jakarta	5
2	West Java	22
3	Banten	6
4	Central Java	35
5	DI Yogyakarta	5
6	East Java	37

Table 3. Ordinal Scale Poverty

Poor	Food	Unemployment	Category of the dependent variable
Yes (3)	Yes (2)	Yes (1)	6
Yes (3)	Yes (2)	No(0)	5
Yes (3)	No (0)	Yes (1)	4
No(0)	Yes (2)	Yes (1)	3
Yes(3)	No(0)	No (0)	3
No (0)	Yes (2)	No (0)	2
No (0)	No (0)	Yes (1)	1
No (0)	No (0)	No (0)	0

Description: The district that has a score dependent variable 0, are not included in the model.

Table 4. List of Variables

No	Variables	Explanation
X1	The economic potential of a village: Agriculture, Industry, Trade, Public Services Facility / etc.	The ratio of potential villages / number of villages per district
X2	Number of farmers	The ratio of farmers / number of villages per district
X3	The number of farm workers	The ratio of farm workers / number of villages per district
X4	The average number of education facilities	The ratio of the number of schools / villages per district
X5	The average distance to downtown village	Km
X6	The number of small scale industries	The ratio of the number of industrial / villages per district
X7	Credit facilities	The ratio of total loans / villages per district
X8	The presence of conflict in society	The ratio of registered conflicts / number of villages per district
X9	The average number of families without electricity	The ratio of the number of families / number of villages per district
X10	Slum area	The ratio of slum areas / number of villages per district
X11	Indonesian labor	The ratio of the number of people / number of villages per district
X12	Hospital	The ratio of hospital / number of villages per district
WY	Spatial variable	Results of spatial weighting multiplication with variable response

3.2. Method

To achieve the objectives of this research, namely:

a) To know the factors that affect poverty-stricken districts in Java based on ordinal logistic regression models of spatial. It can be done by the following procedures:

- Selection of explanatory variables that will be used in the analysis, by looking at the closeness of the relationship of these variables.
- Response variables to be used is a hotspot poverty

districts in the island of Java in the form of ordinal data with 6 categories from research that has been done by Kusumaningrum (2010)^[6].

- Looking at the closeness between districts in Java.
- Creating a spatial weighting matrix W which contains 1 if adjacent, and 0 if not directly adjacent.
- Create a new variable result multiplication of matrix W' with Y, namely X spatial (WY)
- Checking the assumption of logistic regression

- Forming ordinal logistic regression model that has incorporated the spatial variables X using 100% data.
 - Identifying whether the coefficient of X spatial is significant or not. If significant, indicate whether the spatial correlation influences poverty. Then identify the significance of the other X variables.
 - Getting the ordinal logistic regression model with spatial variable X that is significant.
- b) To evaluate the spatial ordinal logistic regression model of ordinal logistic regression model. It can be done by following these procedures:
- Estimating the value of the dependent variable with ordinal logistic regression using 100% spatial data. In this case, do not cross validate because this model does not aim to predict the level of poverty from a new emerging district.
 - Measuring the value of CCR from spatial ordinal logistic regression model then comparing with the value of CCR results from ordinal logistic regression model. A higher CCR value indicates that the model has better accuracy.

4. Results and Discussion

4.1. Exploration Data

The data in this study consisted of one response variable in the form of district poverty level category, and 14 explanatory variables. Poverty Distribution Category of the districts with the respective percentages can be seen in Figure 1. The highest percentage occurs in poverty level category 5, meaning that *poor* districts are experiencing problems and insufficient food and that is as much as 33.3% of all districts in Java. While the number of districts with the lowest category 1 means that only the district experiencing the problem of unemployment.

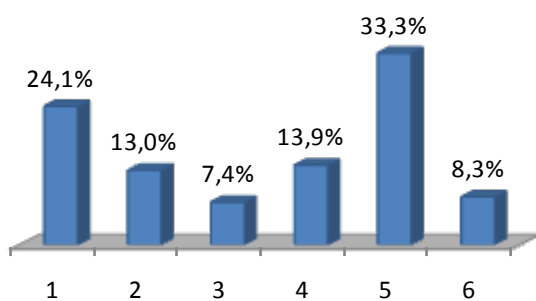


Figure 1. Percentage of district poverty category

4.2. Test of Logistic Regression Assumptions

4.2.1. Detecting Outliers and Handling

Logistic regression analysis assumes that the data has no outliers, then the prior is modeled, while outliers are treated before hand with Winsorized Mean method. Sample distribution of data containing outliers can be seen in Figure 2. Once resolved, the results can be seen in Figure 3.

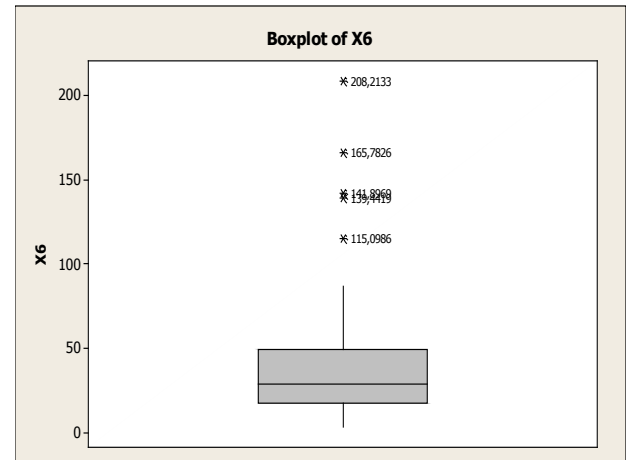


Figure 2. Examples of data contained outliers

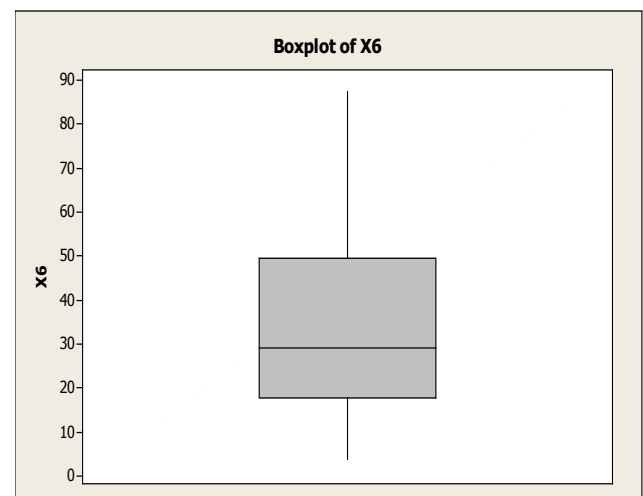


Figure 3. Example of the data after outliers corrected

4.2.2. Detecting Multicollinearity and Handling

Explanatory variables are correlated with each other resulting in so-called multicollinearity. In logistic regression, there must be no multicollinearity because in the presence of multicollinearity the standard error of regression coefficient will be increased so that the possible results of Wald test of each X will not be significant. However, this assumption is actually not very important because multicollinearity does not change its estimate value.

To select the best model, we compute CCR. So we must compare the results of the CCR with the handling of multicollinearity and the results of the CCR without handling multicollinearity.

The value of correlation between explanatory variables can be seen in Table 5. The values show that some variables are correlated highly enough and some quite low. Multicollinearity is handled by using principal components analysis. The 13 explanatory variables were reduced into 3 main components, namely PC1, PC2 and PC3. The process of determining the number of components be used if total of CCR is greater than 70%. Principal component analysis results can be seen in Table 6.

Table 5. The Correlation between Explanatory Variables

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X11	X12	X13	WY
X1	1												
X2	-0,78	1											
X3	-0,63	0,71	1										
X4	0,89	-0,74	-0,59	1									
X5	-0,53	0,68	0,62	-0,47	1								
X6	-0,15	0,31	0,07	-0,17	0,09	1							
X7	0,29	-0,23	-0,23	0,26	-0,41	0,34	1						
X8	0,18	-0,23	-0,15	0,12	-0,29	0,08	0,23	1					
X9	-0,69	0,72	0,73	-0,64	0,65	0,22	-0,3	-0,1	1				
X11	0,7	-0,69	-0,5	0,72	-0,37	-0,25	0,04	0,19	-0,51	1			
X12	-0,36	0,34	0,43	-0,37	0,33	-0,03	-0,15	0,02	0,27	-0,28	1		
X13	0,75	-0,67	-0,56	0,82	-0,42	-0,32	0,08	0,09	-0,59	0,8	-0,32	1	
WY	-0,41	0,24	0,31	-0,43	-0,02	0,26	0,28	-0,07	0,3	-0,48	0,06	-0,61	1

Table 6. Principle Components

Criteria	PC1	PC2	PC3	PC4	PC5
Eigenvalue	6,128	1,908	1,053	0,967	0,749
Proportion	0,471	0,147	0,081	0,074	0,058
Cumulative	0,471	0,618	0,699	0,774	0,831

4.3. Spatial Ordinal logistic Regression Using The Principal Component Analysis

After building a model for ordinal logistic regression with 3 main components, it is known that only the first and second components are significant of influence on the response variable, with a given level of diversity of more than 60%. The result can be seen in Table 7, as follows:

Table 7. Ordinal Logistic Regression Model with Main Components

Predictors	Coef	SE Coef	Wald	P Value	Odds Ratio
Intercept1	-1,45344	0,282551	-5,14	0,000	
Intercept2	-0,483803	0,236442	-2,05	0,041	
Intercept3	-0,0697099	0,22934	-0,3	0,761	
Intercept4	0,66576	0,233446	2,85	0,004	
Intercept5	3,08175	0,386749	7,97	0,000	
PC1	-0,602775	0,107775	-5,59	0,000	0,55
PC2	-0,642587	0,149124	-4,31	0,000	0,53

After learning that the only PC1 and PC2 are significant influence, transformation of the PC towards explanatory variables namely X1, X2 ..., X13 and WY is carried out. The result can be seen in Table 8 below.

Table 8. Results of inverset transformation

Explanatory Variables	coef X	Standardized Variable	coef Z	St dev Z	T -value
X1	0,00	Z1	0,17	0,01	11,61
X2	0,01	Z2	0,10	0,02	6,76
X3	8,57	Z3	0,58	0,02	37,62
X4	0,21	Z4	0,51	0,02	34,63
X5	0,12	Z5	1,94	0,03	75,16
X6	-0,12	Z6	-2,80	0,03	-91,46
X7	-6,66	Z7	-3,56	0,04	-88,05
X8	-65,39	Z8	-1,37	0,02	-84,55
X9	0,02	Z9	0,35	0,01	24,44
X11	4,02	Z11	1,42	0,02	76,04
X12	159,2	Z12	0,76	0,01	63,17
X13	5,15	Z13	1,64	0,02	78,15
WY	-2,26	Zwy	-3,08	0,03	-91,36

Table 8 shows that all the variable X has a T value count greater than 1.96 which means that all explanatory variables influence the level of poverty in the district under 5% significance level (also significant at 15% significance level). The CCR value of this model is then calculated. The result can be seen in Table 9.

Table 9. CCR Spatial Ordinal Logistic Regression Model with Principal Component

Actual	Prediction						Percentage correct
	1	2	3	4	5	6	
1	17	0	0	0	2	7	65,38
2	1	0	0	0	1	12	0,00
3	0	0	0	0	0	8	0,00
4	4	0	0	0	3	8	0,00
5	0	0	0	0	0	36	0,00
6	0	0	0	0	0	9	100,00
Overall Percentage Correct							24,07

Based on table 9, the result *Correct Classification Rate*, is small enough, almost 24%. This shows that this model is less good based on the expected value of 14 explanatory variables. Therefore it is necessary in further study that all explanatory variables are included in order to have a model that is really quite good in classifying district poverty category. Furthermore, ordinal logistic regression modeling was developed using the entire spatial explanatory variables.

4.4. Spatial Ordinal Logistic Regression Models without Handling Multicollinearity

4.4.1. Modelling

In this section, a model of spatial ordinal logistic regression is made by including all explanatory variables, without reducing it into a major component. The result can be seen in Table 10.

Table 10. Spatial Ordinal Logistic Regression Models with all explanatory variables

Predictors	Coef	SE coef	Wald	P value	Ratio Odds
Intercept1	1,717	2,763	0,620	0,534	
Intercept2	2,913	2,766	1,050	0,292	
Intercept3	3,411	2,774	1,230	0,219	
Intercept4	4,309	2,795	1,540	0,123	
Intercept5	7,005	2,863	2,450	0,014	
X2	-0,059	0,023	-2,560	0,011	0,940
X6	0,017	0,010	1,700	0,090	1,020
X9	-0,039	0,024	-1,640	0,102	0,960
X13	1,306	0,897	1,460	0,145	3,690
WY	-0,974	0,244	-3,990	0,000	0,380

4.4.2. Evaluation of Spatial Ordinal Logistic Regression Models

Evaluation models can be made by using the likelihood-ratio test using the statistical G. Based on the analysis, $G = 81.485$ with $p = 0.000$ indicates that H_0 was rejected, meaning that there is one explanatory variable that

significantly influence the level of poverty of a district. Further evaluation is to use *measure of association*, *Correct Classification Rate* (CCR) and the *Goodness of fit test* which will be described further below.

In measuring association, the values of concordant and discordant couples show how good a model is to predict the data. Concordant value of 81% shows that as many as 81% of districts with poverty rates ($< k$) have better chance in predicting the category ($< k$). While the value of discordant 18.7% indicates that as many as 18.7% districts with poverty rates ($> k$) have a better chance predicting the category ($< k$). This means that the size of the association in this model is very good.

Table 11. CCR Spatial Ordinal Logistic Regression Models with All Explanatory Variables

Actual	Prediction						Percentage correct
	1	2	3	4	5	6	
1	20	4	0	0	2	0	76,92
2	6	2	0	0	6	0	14,29
3	2	0	0	0	6	0	0,00
4	4	3	0	0	8	0	0,00
5	3	1	0	0	31	1	86,11
6	0	1	0	0	7	1	11,11
Overall percentage correct							50,00

Table 11 shows that the CCR of spatial ordinal logistic regression model is 50%. This means that 50% of the districts are predicting correctly through the model. The test *Goodness of Fit* using Chi Square yields a value of 464.904 with the value of p of 0.965. Therefore, the spatial logistic regression model has good model fit.

4.4.3. Factors Determining the Impact of Poverty Level

Based on Table 10, it can be seen that the model with no treatment of multicollinearity is better than that with the handling of multicollinearity, because it has a higher CCR value. The conclusion derived is that ordinal logistic regression model without the handling of spatial multicollinearity was used as a reference model to see the variables that influence the level of poverty.

Based on the above explanation, an increase in the number of farmers, the number of families who do not have electricity on a county to reduce the chances of a district to be poor. Meanwhile, by increasing the average number of small industries and the number of hospitals to increase the chances of a district to be poor.

4.4.4. Identifying the Spatial Variable

Looking at the spatial variables (WY), one can say that there is spatial correlation between district poverty levels and this can cause the poverty of a district to influence neighboring districts. Then after further examined, variable spatial relationship with poverty rates assumes a negative

value, This means districts with lower levels of poverty are usually surrounded by districts with high poverty level.

4.5. Comparing Spatial Model with and without Handling Multicollinearity

By comparing the value of CCR, we can say that spatial ordinal logistic regression model is better than the principal component regression because it has higher CCR value that is equal to 50%, while with the principal component model has a CCR value of 24%. Therefore, in this case, handling multicollinearity with principal component is the good choice.

5. Conclusions and Suggestions

5.1. Conclusions

Increasing of the number of farmers, the number of families who do not have electricity on a county to reduce the chances of a district to be poor. Meanwhile, by increasing the average number of small industries and the number of hospitals to increase the chances of a district to be poor.

Based on the analysis that has been done, it can be concluded that there are five variables that affect poverty-stricken districts in Java namely: the number of farmers, the number of families who do not have electricity in a district, the percentage of slum area in a district, the average number small industries, the number of hospitals and the district spatial poverty variable. Therefore, these factors should be prioritized in reducing poverty levels in all districts on the island of Java.

5.2. Suggestion

Based on these results, it is suggested that attention is given to the value of CCR at the time of logistic regression modeling. There is a need for further study regarding the degree of influence of multicollinearity on the results of classification, perhaps by creating a simulation. Last but not least, poverty-stricken districts on Java island are advised to

consider the factors that affect poverty district.

REFERENCES

- [1] Agresti A. *Categorical Data Analysis*. New Jersey: John Wiley and Sons. USA. 1990.
- [2] [BKKBN] National Disaster Coordinating Agency. Family income; Glance.[link] http://www.bkkbn.go.id/article_detail.php[2 October 2011]. 2004.
- [3] Dillon, W R and Goldstein, M. *Multivariate Analysis*. John Wiley & Sons, New York. USA. 1984.
- [4] Garson, GD. *Logistic Regression*. [link] <http://www.chass.ncsu.edu>[1 October 2012]
- [5] Hosmer, DW and Lemeshow, S. *Applied Logistic Regression* Second Edition. New York : John Wiley and Sons. USA. 2000.
- [6] Kusumaningrum D. *Hotspot Analysis on Poverty, Unemployment, and Food Security in Java, Indonesia*. Thesis. Bogor Agricultural University. Indonesia 2010.
- [7] Lee J, and Wong, DWS. *Statistical Analysis ArchView GIS*. New York: John Wiley & Sons, Inc. 2001.
- [8] Norusis, MJ. *SPSS Statistics Guides: Ordinal Regression*. http://www.norusis.com/pdf/ASPC_v13.pdf [25 December 2011]. 2010.
- [9] Pratama, V. *Comparison of Prediction Results of Spatial Regression Model for Different Variogram Models*. Thesis. Bogor Agricultural University. Indonesia. 2008.
- [10] Suprpti. *Weighted distance and optimum Intersection Spatial Logistic Regression for Prediction of Rural Poverty Status in West Java*. Thesis. Bogor Agricultural University. Indonesia. 2009
- [11] Thaib, Z. *Logistic Regression Modeling Spatial Contiguity Matrix Approach*. Thesis. Bogor Agricultural University. Indonesia. 2008
- [12] Ward, MD and Gleditsch, KS. *Spatial Regression Models*. Los Angeles: Sage Publications, Inc. 2008.