Regularized Ordinal Regression with Elastic Net Approach (Case Study: Poverty Modeling in Yogyakarta Province 2018)

Pardomuan Robinson Sihombing[1], Yudhie Andriyana[2], Bertho Tantular[3]

[1]Statistics Indonesia, Jakarta, Indonesia
[1,2,3]Department of Statistics, Padjadjaran University, Bandung, Indonesia

Email: robinson@bps.go.id

ABSTRACT

Generally, modeling poverty aims to obtain the best criteria for assessing poverty status. There are two approaches to model the factors that affect poverty, namely consumption approach and discrete choice model. The advantage of the discrete choice model compared to the consumption approach is that the discrete choice model provides a probabilistic estimate for classifying samples into different poverty categories. The aim of this study is to determine the factors that impact poverty in Yogyakarta through Regularized Ordinal Regression used elastic net approach both for parallel, non-parallel, and semi-parallel models. The data used in this study is Susenas March 2018 for Yogyakarta provinces. The result of this study shows that the best discrete choice model for Yogyakarta's modelling is the parallel model. Households that live in villages, have a large number of household members, are headed by women, have elderly household heads, have low education, and work in the primary sector tend to be more vulnerable to poverty. Therefore, a simultaneous policy with inclusive economic development is needed to reduce cross-border, cross-gender, and cross-sector inequality.

Keywords: elastic net; ordinal regression; parallel; poverty

INTRODUCTION

Poverty is one of the problems in economic development. Every country tries to alleviate poverty with various programs. As an institution that released the official poverty rate in Indonesia, BPS [1] defines poverty as the inability to meet basic needs from an economic perspective, both food and non-food, which is measured in terms of expenditure. Generally, modeling poverty aims to obtain the best criteria for assessing poverty status. Rouband & Razafindrakoto [2] assert that there is a correlation between objective and subjective poverty measures and further argue that various forms of poverty cannot be reduced to one another. A poverty approach is generally a monetary approach, but there is a growing literature that tries to bring up an index of multidimensional aspects of poverty.

The impact factor in poverty approaches with two models. The first uses a regression approach between consumption expenditure per adult equivalent to several potential explanatory variables called the consumption approach. The second model is discrete choice model. The discrete approach is to categorize poverty into three categories based on household consumption expenditure compared to a region's poverty

line. The advantages of the discrete model are the influence of independent variables to vary across poverty categories. One of the most common regression models for ordinal data types is the cumulative logit model [3], also known as the proportional odds model or the ordinal logistic regression model. To improve the prediction accuracy in ordinal regression model with different regression coefficients for each response category, the DOGEV model was introduced to improve prediction accuracy [4].

The DOGEV model has a requirement which is the data that used has extreme values. Moreover, the model has parallel and non-parallel models yet. Wurm, Rathouz, & Hanlon [5] introduced a regression model using different regression coefficients for each response category known as Regulized Ordinal Regression. The data that has ordinal response or dependent should be explained using parallel or non-parallel. When, the number of household observation used maximum likelihood then it is the proper model [5]. After that, both the nonparallel model that includes the parallel model as a particular case and the parallel model will provide an inconsistent estimation coefficient if there are errors in the modeling.

The number of explanatory variables increases, we need a variable selection technique that will reduce some variables. This step is needed because it is impossible to estimate each coefficient with a high degree of accuracy. Then more realistic modeling goal is built a model for out-of-sample prediction and determine the most important explanatory variables. Two variable selection methods that are often used are the lasso and ridge methods. Lasso and ridge regression are techniques that minimize the penalized likelihood objective function. Lasso regression uses the L1 penalty, while ridge regression uses the L2 penalty. Both penalties produce coefficient estimates that are closer to zero than the maximum likelihood estimator; for example, the estimate is "close" to zero yet. The estimation results in an estimation bias towards zero, but a trade-off occurs in terms of reducing the variance, which often reduces the overall mean squared error.

Lasso has properties with some approximate coefficients close to zero. This method provided a natural way to select variables because only the most relevant predictor of the response variable will have a non-zero coefficient. However, it is a group of variables that are highly correlated then, the lasso tends to choose one variable from the correlated group and ignores the others. The elastic net penalty was introduced to overcome those limitations [5]. The elastic net penalty method is the weighted average between lasso and ridge, by dividing the lasso properties and shrinking some coefficients to zero so that it has a unique solution in most cases.

Based on the previous description, the main problem to be examined is how the factors that affect poverty in Yogyakarta through Regularized Ordinal Regression with elastic net approach both for parallel, non-parallel, and semi-parallel models.

METHODS

Data

The data that used in this study is Susenas Consumption Module in March 2018 by BPS. Then used as the response variable and explanatory variables. Base on theoretical studies, residential, community, household and individual characteristics influenced differences in household expenditure. This study uses household data and the variables that related to household characteristics only. The variation in household characteristics will affect the households' expenditure.

Methodology

The aim of poverty modelling is to know factor that influence such as poverty. The focus of study is the family that has characteristics in specific poverty status. The framework of poverty study would be assumed that the real poverty status in household unable observed or unconsidered by well-being ratio, with general mode:

$$logit\ (\boldsymbol{p}_i) = \mathbf{X}_i^T \boldsymbol{\beta} \quad i=1,2,...,n \tag{1}$$

Where:

$\mathbf{X}_i$ = covariate matrix

$\boldsymbol{\beta}$ = vector of regression coefficients

$\boldsymbol{p}_i$ = category probability

The logit model that used is one of the Generalized Linear Model (GLM) models. The GLM model has three components, namely random component, systematic component and link function. The form of the GLM depends on the three components that related to each other.

a. Random Components

Y is a random variable of ordinal response with three categories, which is poor, almost poor, not poor

$$Y \sim \text{Multinomial } (n;\ p_1, p_2, p_3)$$

$$f(y; p_1, p_2, p_3, n) = \frac{n!}{y_1!\, y_2!\, y_3!} p_1^{y_1} p_2^{y_2} p_3^{y_3}$$

b. Systematic component

The systematic component of the model is a set of $\boldsymbol{\beta}$ parameters and a covariate $\mathbf{X}$ that forms a linear combination of $\mathbf{X}_i^T \boldsymbol{\beta}$

The general form of the linear predictor is :

$$\boldsymbol{\eta}_i = \mathbf{X}_i^T \boldsymbol{\beta} \tag{2}$$

According to Wurm, Rathouz, & Hanlon [5], the linear form of the predictor consists of::

i. Parallel Model

$$\mathbf{X}_i = \left(\mathbf{I}_{K \times K} \begin{vmatrix} \boldsymbol{x}_i^T \\ \vdots \\ \boldsymbol{x}_i^T \end{vmatrix} \right)_{K \times (P+K)} , \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b} \end{pmatrix}_{(P+K) \times 1}$$

ii. Nonparallel Model

$$\mathbf{X}_i = \left(\mathbf{I}_{K \times K} \begin{vmatrix} \boldsymbol{x}_i^T & 0 & 0 & 0 \\ 0 & \boldsymbol{x}_i^T & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{x}_i^T \end{vmatrix} \right)_{K \times (PK+K)} , \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{b}_0 \\ \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_k \end{pmatrix}_{(PK+K) \times 1}$$

iii. Semi-parallel Model

$$\mathbf{X}_i = \left(\mathbf{I}_{K \times K} \begin{vmatrix} \boldsymbol{x}_i^T & \boldsymbol{x}_i^T & 0 & \cdots & 0 \\ \boldsymbol{x}_i^T & 0 & \boldsymbol{x}_i^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_i^T & 0 & 0 & \cdots & \boldsymbol{x}_i^T \end{vmatrix} \right)_{K \times (P(K+1)+K)} , \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b} \\ \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_k \end{pmatrix}_{(P(K+1)+K) \times 1}$$

$\boldsymbol{b}_0$ = vector intercept, $\boldsymbol{b}$ = vector slope for parallel model

$\mathbf{B}_i$ = matrix slope for nonparallel Model

$\mathbf{I}_{K \times K}$ = matrix identity
x_i = vektor covariat without intercept

c. Link Function

The link function is a function that connect systematic components and the expected (average) value the random component explained the relationship working $E(y) = np$ with explanatory variable in linier predictor.

We have model p directly or model a monotonous function.

$$E\ (y_i|x_i) = g(p_i)\ = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta} \tag{3}$$

Elastic net penalty

Suppose $\boldsymbol{\beta}$ has the length Q and β_j shows the j element. Wurm, Rathouz, & Hanlon [5] wrote the objective elastic net function as:

$$M(\boldsymbol{\beta}; \alpha, \lambda, c_1, \dots, c_Q) = -\frac{1}{N_*}\ell(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{Q} c_j \left(\alpha|\beta_j| + \frac{1}{2}(1-\alpha)\beta_j^2 \right) \tag{4}$$

where:

$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\beta})$ and $\ell_i(\boldsymbol{\beta})\ =\ L_i(h(\mathbf{X}_i^T \boldsymbol{\beta}))$
in model $\ell(\boldsymbol{\beta})$ is loglikelihood function, $\lambda > 0$ and $0 \leq \alpha \leq 1$.

Wurm, Rathouz, & Hanlon [5] wrote elastic net objective function for each model shape derived from Equation 4 as follows:

Objective function for parallel model is:

$$M(\boldsymbol{b}_0, \boldsymbol{b}; \alpha, \lambda) = -\frac{1}{N_*} \ell(\boldsymbol{b}_0, \boldsymbol{b}) + \lambda \sum_{j=1}^{P} \left(\alpha|b_j| + \frac{1}{2}(1-\alpha)b_j^2 \right)$$

Objective function for nonparallel model is:

$$M(\boldsymbol{b}_0, \mathbf{B}; \alpha, \lambda) = -\frac{1}{N_*} \ell(\boldsymbol{b}_0, \mathbf{B}) + \lambda \sum_{j=1}^{P} \sum_{k=1}^{K} \left(\alpha|B_j| + \frac{1}{2}(1-\alpha)B_{jk}^2 \right)$$

Objective function for semiparallel model is:

$$M(b_0, b, B; \alpha, \lambda, \rho) = -\frac{1}{N_*} \ell(b_0, b, B)$$

$$+ \lambda \left(\rho \sum_{j=1}^{P} \left(\alpha|b_j| + \frac{1}{2}(1-\alpha)b_j^2 \right) + \sum_{j=1}^{P} \sum_{k=1}^{K} \left(\alpha|B_j| + \frac{1}{2}(1-\alpha)B_{jk}^2 \right) \right)$$

when $\lambda \geq 0$ and $\alpha \in [0,1]$ are tuning parameters and $\rho \geq 0$ is tuning parameters which determines the extent to eliminated the parallel term

RESULTS AND DISCUSSION

Firstly, we discuss characteristics of the socioeconomic variables of the household as a general overview of the respondents that used in the study. We use pie charts for the descriptive characteristics of the respondents. It is used to illustrate the frequency of each category in the research variables.

Table 1. Characteristic of Responden

Variable	Category	Poverty Status			Total
		Poor	Almost Poor	Not Poor	
region type	rural	5,99	4,49	54,53	65,01
	urban	5,56	3,57	25,86	34,99
the total number	single	0,39	0,25	7,95	8,59

household member	living with a couple	1,93	1,03	14,05	17,01
	living with a couple with other family members.	9,24	6,78	58,38	74,39
marital status	never marriage	0,25	0,11	4,17	4,53
	marriage	10,16	7,35	65,58	83,10
	divorce	0,07	0,11	2,85	3,03
	divorce by death	1,07	0,50	7,77	9,34
gender	male	10,34	7,49	70,01	87,84
	female	1,21	0,57	10,38	12,16
age	15-64 years old	8,92	6,63	70,86	86,41
	65+ years old	2,64	1,43	9,52	13,59
education	primary and junior	8,84	5,67	39,16	53,67
	senior high school	2,64	2,32	26,64	31,60
	collage	0,07	0,07	14,59	14,73
sector economy	primary	6,13	3,53	19,54	29,21
	secondary	2,92	2,14	17,40	22,47
	tertiary	2,50	2,39	43,44	48,32
	Total	11.05	8,06	80,39	

The first step is to test chi-square independence. Chi-square independence analysis use when it has a relationship between categorical variables. This method has done at first step. Then seeing whether the independent variable/predictor used has a relationship (dependent) with the dependent/response variable. The null hypothesis formulation there is no dependency between poor status and variables the explanation, while the alternative hypothesis there is a dependency between poor status with the explanatory variable. Table 2 is the probability value of the results less than 0.05 then it means all independent variables have a dependent relationship with the dependent variable/response.

Table 2. Independent Test of Category Variables on Poor Status

Category Variables	value chi square	df	p.value
region type	41,086	2	0.000
the total number household member	30,327	4	0.000
marital status	27,660	6	0.000
gender	7,491	2	0.024
age	181,148	4	0.000
education	32,699	2	0.000
sector economy	154,836	4	0.000

In this study, used the ordinalNetCV function on the OrdinalNet software R version 3.61 package. In this study, we compare the results of parallel, non-parallel and semi-parallel models with the AIC, BIC and loglik. In general, parallel and semi-parallel models have similar performance, but non-parallel models are much worse. This model might be due to the unidentified out-of-sample log-likelihood non-parallel model (non-monotonous cumulative probability) in the first few values of λ.

Table 3. Comparison of the AIC, BIC and loglik Values of the Three Ordinal Regression Models

model	AIC	BIC	loglik
parallel	3192.01	3269.21	-319.1258
non-parallel	3403.23	3468.56	-339.732
semi-parallel	3209.88	3358.35	-321.276

Table 3 shows that the values of AIC, BIC, and loglik have the smallest on the parallel model, moreover, the lambda parameters obtained in all three models for each fold, in Table 4. The variability of lambda values is the lowest in the parallel model.

Table 4. Comparison of Lambda Values for the Three Regression Models

Fold	Parallel Model	Nonparallel Model	Semiparallel Model
fold1	0,0015	0,0010	0,0019
fold2	0,0012	0,0229	0,0031
fold3	0,0019	0,0292	0,0009
fold4	0,0025	0,0180	0,0040
fold5	0,0012	0,0372	0,0051

Table 5 shows that five dummy variables have a positive coefficient, six dummy variables that have a negative coefficient and one dummy variable with a zero coefficient. A positive coefficient value means the chance of the understudy category to be poor is higher compared to the reference category, furthermore the negative coefficient means the chance of the category understudy is smaller for the poor status compared to the reference category. The zero coefficient means that the opportunity for the category studied is not significantly different for poor status compared to the reference category.

Table 5. Ordinal Regression

Variabel	Category	logit(P[Y<=1])	logit(P[Y<=2])
	Intercept	-2.389	-1.706
region type (*rural)	region type (urban)	0.042	0.042
the total number	living with a couple	1.249	1.249
household member	living with a couple		
*single	with other family members.	0.539	0.539
marital status	Marriage	0.000	0.000
* never marriage	Divorce	-1.157	-1.157
	divorce by death	-0.367	-0.367
gender (*male)	Gender (female)	0.319	0.319
age (*15-64 years old)	Age (non produktif 65+)	0.390	0.390
education	senior high school	-0.455	-0.455
*primary&junior	Collage	-2.912	-2.912
sector economy	secondary sector	-0.408	-0.408
*primer	tertiary sector	-1.057	-1.057

*baseline category

Discussion of the results

a. Residential Type

Regional type is a category of respondent's residential area; there are two categories: urban and rural areas. The location of the household is one of the factors which is often associated with poverty status. The regional type is due to differences in access to primary facilities such as education and health. The results of this study show that the status of the area of residence significantly affected the poverty status of a household. Rural households have a higher tendency to become poorer than urban households. This result is in line with some previous studies such as [6] and [7] that suggest that rural households are more vulnerable to poverty due to limited access.

b. Household Size

The size of a household indicates the number of people who live in that household. The more people live in a household; then the more resources are needed to keep the household members prosperous. The results of this study show that compared to households consisting of only one person, households of 2 or more people had a higher inclination to live in poverty. The results of this study are in developing countries which

show that as the number of households increases, the average per capita consumption decreases, indicating that households are approaching poor status. The results of the studies that conducted in developing countries such as [6] and [8] show that the larger household's size, the lower the average consumption per capita. It indicates that these households are getting closer to poor status. The problem is even though they live in one household, about 20 percent of the items used together [8]. Therefore, they must allocate limited income to more needs.

c. Marital Status

Marital status is related to responsibility for household expenses. Someone who has a never married status tends to have income that and use it for personal needs. Where, the income generated becomes cumulative from household members, in the results, there is equal opportunities between those who are never married and those who are married. Compared to someone who are never married, household with divorcee-household head have less probability to be poor. According to [9], divorcee usually has economic planning and economic adaptation strategies to align with the amount of income a family needs every day of their life. It proves that from the way a divorcee to save, set aside in part piecemeal revenue that could be used to meet the needs of their child's education and are used for urgent needs.

d. Household Gender

There are characteristic differences between households dominated by men and women. In general, households headed by women are often identified with households with higher chances of poverty. The research results in some regions, both in developed and developing countries, showed that households led by women are more prone to poverty, because female heads of households generally generate lower incomes and generally have more dependencies ([7], [10], [11]). The Yogyakarta data also shows alike. Based on the data collected in this study, female heads of households tend to bear a large number of household members.

e. Head of Household Age

One of the factors that influence a person's level of productivity is age. A person who has at a productive age is likely to have a higher income than someone has at an unproductive age. Therefore, it is a common misconception that households with low-income households are less likely to become poorer. The results of this study support these general assumptions. The result of this research is with the research that has done in [7] and have shown that as a productive age passes, one's income tends to decline, and the risk of becoming poor can higher.

f. Head of Household Education

Education is one of the crucial factors that determine one's well-being. Educational attainment increases potential income of individuals, and as a result, increasing income definitely helped them to out from poverty [12]. In line with previous research, this study showed consistent results. Households headed by a person with a high school education have a higher tendency to be poor compared to those with a lower middle school.

g. Economic Sector of Head of Household

The field of work in which household heads work has an impact on household poverty status. This is due to differences in income levels in each industry sector. The primary

sectors comprising agriculture and mining generally have lower income levels than other sectors. The results of this study show that households with households working in the secondary and tertiary sectors have a lower tendency to be poor compared to those with primary income from the primary sector. Besides that, the results of this study show that households who work in the tertiary sector have a higher chance of being poor than those who work in other sectors. The result of this research is in line with [13] and [14] that state that the shift from the agricultural sector is effective in alleviating poverty.

CONCLUSIONS

Some factors determined poverty such as household size, marital status, the gender of household head, age of head of household, level of education of the head of household, and occupation of the head of household. Based on AIC and BIC criteria, the best model to Yogyakarta poverty data is parallel model. Households that live in villages, have a large number of household members, are headed by women, have elderly household heads, have low education, and work in the primary sector tend to be more vulnerable to poverty. Therefore, a simultaneous policy with inclusive economic development is needed to reduce cross-border, cross-gender, and cross-sector inequality.

REFERENCES

[1]   Badan Pusat Statistik, "Data dan Informasi Kemiskinan Kabupaten/Kota 2018," Jakarta, 2018.

[2]   M. Razafindrakoto and F. Roubaud, "The Multiple Facets of Poverty: the case of urban Africa," in *WIDER Conference on Inequality*, 2003.

[3]   P. Mccullagh, S. Journal, R. Statistical, and S. Series, "Regression Models for Ordinal Data," *J. R. Stat. Soc. Ser. B*, vol. 42, no. 2, pp. 109–142, 1980.

[4]   E. Fissuh and M. Harris, "Modeling Determinants of Poverty in Eritrea: A New Approach," pp. 1–35, 2005.

[5]   M. J. Wurm, P. J. Rathouz, and B. M. Hanlon, "Regularized Ordinal Regression and the ordinalNet R Package," 2017.

[6]   J. C. Anyanwu, "Marital Status, Household Size and Poverty in Nigeria: Evidence from the 2009/2010 Survey Data," *African Dev. Rev.*, vol. 26, no. 1, pp. 118–137, 2014.

[7]   R. Gounder and Z. Xing, "Impact of education and health on poverty reduction: Monetary and non-monetary evidence from Fiji," *Econ. Model.*, vol. 29, no. 3, pp. 787–794, 2012.

[8]   P. Lanjouw and M. Ravallion, "Poverty and Household Size," *Econ. J.*, vol. 105, no. 433, pp. 1415–1434, 1995.

[9]   A. S. Rahayu, "Kehidupan sosial ekonomi single mother dalam ranah domestik dan publik," *J. Anal. Sosiol.*, vol. 6, no. 1, 2017.

[10]   M. Buvinić and G. Rao Gupta, "Female-headed households and female-maintained families: Are they worth targeting to reduce poverty in developing countries?," *Econ. Dev. Cult. Change*, vol. 45, no. 2, pp. 258–280, 1997.

[11]   D. F. Meyer, "Predictors Of Poverty: A Comparative Analysis Of Low Income Communities In The Northern Free State Region, South Africa," *Online) Int. J. Soc. Sci. Humanit. Stud.*, vol. 8, no. 2, pp. 1309–8063, 2016.

[12]   M. Awan *et al.*, "Impact of education on poverty reduction," *Int. J. Acad. Res.*, vol. 3, 2011.

[13] I. D. A. Bagus, E. K. A. Artika, A. A. S. Kencana, I. D. A. Ayu, and K. Marini, "Pergeseran Lapangan Usaha Sektor Pertanian , Pertumbuhan Ekonomi," *J. Unmas Mataram*, pp. 111–117, 2018.

[14] F. Fahar, "Kemiskinan Dan Ketenagakerjaan Di Kepulauan Riau 2014: Permasalahan Dan Implikasi Kebijakan," no. February, 2015.