International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 1, Jan-Feb 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Application of Ordinal Probit Regression Panel Data with Random Effects Approach for Modeling The Environmental Quality Index in Indonesia

Thifalia Nurli Al-Kahfi[1], Suliyanto[2], Nur Chamidah[3], Toha Saifudin[4], M. Fariz Fadillah Mardianto[5]

[1,2,3,4,5]Statistics Study Program, Department of Mathematics, Airlangga University, Surabaya-Indonesia
[1]thifalia.nurli.alkahfi-2019@fst.unair.ac.id
[2]Corresponding author: suliyanto@fst.unair.ac.id
[3]nur-c@fst.unair.ac.id
[4]tohasaifudin@fst.unair.ac.id
[5]m.fariz.fadillah.m@fst.unair.ac.id

--*************************----------------------------------

Abstract:

Environmental problems are inseparable from economic and social development because all three are part of sustainable development that can't be separated. The Environmental Quality Index (EQI) developed by the Kementrian Lingkungan Hidup dan Kehutanan (KLHK) is used to measure and provide information related to environmental quality based on indicators of water quality, air quality, and land cover. This study aims to model EQI in Indonesia using panel data ordinal probit regression with random effects. The assumptions in this regression model are that errors have a standard normal distribution and that the random effects are identically distributed independent of errors. Parameter estimation uses marginal likelihood and to maximize this function uses the Gauss Hermite Quadrature method. The data used is the EQI in Indonesia based on 34 provinces from 2018 to 2021. The modeling results show that the number of poverty, the rate of economic growth, the Human Development Index (HDI) and population density affect EQI with a model classification accuracy of 55.88%.

Keywords **—Environment Quality, Probit Ordinal Regression, Panel Data.**

--*************************----------------------------------

I. INTRODUCTION

Environmental problems are inseparable from economic and social development because all three are part of sustainable development that can't be separated [1]. Based on data in the publication of Badan Pusat Statistik (BPS) urban areas which are centers of the economy and high social mobility have poor environmental quality than rural areas[2]. According to the Kementrian Lingkungan Hidup dan Kehutanan (KLHK), the environment quality in Indonesia is measured using the Environmental Quality Index (EQI) with values ranging from 30 to 100 which are divided into six categories, namely, alert, very unfavorable, unfavorable , quite good, good, and very good, where the higher the IKLH value indicates that the environmental condition of the area is good[3]. EQI is useful as information to support stakeholder decision making in which EQI consists of environmental pollution in the form of water, air, and land cover as indicators. So that the environment becomes a variable that is influenced by social and economic factors [4].

Research from Febriani (2019) economic and social development has an impact on the environment. The damage is caused by human activities which are specifically called externalities. Basically, externalities arise when some of the activities of producers and consumers have indirect effects that can arise positively or negatively [5]. In another studyfrom Shi et. al., (2020) explains that negative externalities can occur due to negative behavior such as disposal of garbage and factory waste which is the result of factory activities with a lot of energy consumption and poor factory management in material processing processes and pollution.[6]. Social factors such as population, affluence, and human activity can affect the structure of environmental change [7].

Rencana Pembangunan Jarak Menengah Nasional (RPJMN) 2020-2024 is a government effort to build a better living environment, but this has several obstacles including weak monitoring and evaluation related to program implementation and inaccurate targets due to weak integration of data and information regarding the factors that drive EQI increase [8]. So, it would be better if analyzed using regression analysis.

On this study using panel data regression analysis is a regression analysis that combines cross section data and time series data so that it has more observations compared to cross section or time series data only[9]. Based on the nature of the response variable, the panel data regression used is ordinal probit regression which is a method for describing the relationship between response variables that are polychotomous and predictor variables that are categorical or continuous. Modeling on panel data probit regression only has the assumption that the error is standard normal distribution and the random effect is independent between individuals with a mean of 0 and a constant variance in the error [10].

II. LITERATURE REVIEW

Panel data is a combination of time series data and cross section data. Panel data can explain two kinds of information. In time series data, observations are made based on suitability of time, whereas in cross section data, values will be taken from one or more variables at one time. According to Gujarati (2009) there are several advantages to be gained by using panel data including the following [9]:

a. Can control individual heterogeneity.
b. More informative, more varied and efficient than combining time periods for each individual observation, so that the relationship between variables is lower and the degrees of freedom are greater.
c. Better at learning dynamic changes.
d. Superior in identifying and measuring effects that cannot be detected when using only time series and cross section data.
e. Panel data makes it easier for researchers to study more complex behavioral models.
f. Panel data can minimize bias generated by aggregation of individuals or companies because there are more data units.

III. METHODOLOGY

A. *Gauss-Hermite Quadrature*

*Gauss-HermiteQuadrature*merupakanmetodependekatanintegralfungsi $f(\bullet)$is a method of approximating the integral function $f(\bullet)$ multiplied by other functions in the form of normal density. In many statistical applications, the Gaussian density is an explicit factor of the integral, so a linear transformation can be made so that this factor takes the form$exp(-x^2)$ [11]. This approach is the sum of the weights that estimate the function at a certain point [12].With Gauss-Hermite Quadrature, the integral form is estimated by the following equation :

$$\int_{-\infty}^{\infty} g(w)\exp(-w^2)dw \approx \sum_{m=1}^{M} w_m^* g(a_m^*) \quad (1)$$

$w_m^*$the quadrature weight given by the*m-point,* whilea_m^* grid from the quadrature, which is the root of the *m-point* Hermite Polynomial. $w_m^*$is the weight corresponding to the mth grid given by the equation

$$w_m^* = \frac{2^{M-1}M!\sqrt{\pi}}{M^2 H_{M-1}(a_m^*)^2} \quad (2)$$

B. *Ordinal Probit Regression Panel Data With Random-Effect Approach*

Ordinal probit regressionis a statistical method for analyzing response variables that have an ordinal scale consisting of 3 or more categories, expressed in the form of the response variable y_{it} with q categories obtained from the continuous latent response variable as follows:

$$y_{it}^{*} = \mathbf{x}_{it}\boldsymbol{\beta} + v_i + \varepsilon_{it} \qquad (3)$$

with

$$y_{it} = ` \begin{cases} 1 \text{ jika} y_{it}^{*} \leq \kappa_1 \\ 2 \text{ jika} \kappa_1 < y_{it}^{*} \leq \kappa_2 \\ \vdots \\ q \text{jika} \kappa_{q-1} < y_{it}^{*} \end{cases} \qquad (4)$$

Errorε_{it}standard normal distribution and independent ofv_i. $v_i$is the random effect of the unit cross section $i$assuming an independent identical distribution$N(0, \sigma_v^2)$While$k$ is the set of cutpoints. The conditional distribution of the response variable$y_{it}$is

$$f(y_{it}, \boldsymbol{\kappa}, \mathbf{x}_{it}\boldsymbol{\beta} + v_i) = \prod_{j=1}^{q} p_{itj}^{I_j(y_{it})}$$

$$= \exp \sum_{j=1}^{q} \{I_j(y_{it}) \log(p_{itj})\} \qquad (5)$$

with

$$I_j(y_{it}) = \begin{cases} 1 & \text{jika } y_{it} = \text{j} \\ 0 & \text{yang lain} \end{cases} \qquad (6)$$

C. *Modeling Inference*

1) ***Partial Test***used to test the effect of the predictor variable on the response variable simultaneously with the test statistic used is the Wald test which follows the Chi-Square distribution by comparing the value of W withχ^2.The test criteria are H$_0$ is rejected if$W > \chi^2{}_{(p;\alpha)}$where p is the number of predictor variables in the model or H$_0$ is rejected if p-value < α. With the Wald Test equation as follows[13]**:**

$$W = (\widehat{\boldsymbol{\beta}} - \mathbf{0})^T I(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \mathbf{0}) = (\widehat{\boldsymbol{\beta}} - \mathbf{0})^T \boldsymbol{V}^{-1}(\widehat{\boldsymbol{\beta}} - \mathbf{0})$$

$$= \widehat{\boldsymbol{\beta}}^T \boldsymbol{V}^{-1} \widehat{\boldsymbol{\beta}} \sim \chi_p^2 \qquad (7)$$

with$\boldsymbol{V}^{-\mathbf{1}} = I(\widehat{\boldsymbol{\beta}}) = -H(\widehat{\boldsymbol{\beta}}) = \frac{\delta^2 l}{\delta \widehat{\boldsymbol{\beta}} \delta \widehat{\boldsymbol{\beta}}^{\boldsymbol{T}}}$

2) ***Individual Test***is a test carried out by testing each βk individually, the test statistic used is the Wald test following a normal distribution so as to obtain a test decision, compared to the valueof $Z_k$with$Z\alpha_{/_2}$. H$_0$ is rejected if the value of$|Z_k| > Z\alpha_{/_2}$or*p-value* < *α*. With the following equation:

$$Z_k = \frac{\widehat{\beta}}{SE(\widehat{\beta}_k)}; k = 1,2, \dots, p \qquad (8)$$

3) ***Model Fit Test***is a statistical test commonly used to compare the fitness of two models. Model suitability testing is used to check whether the model obtained is appropriate or not in accordance with the observed data. The test statistic used is the Likelihood Ratio Test following the Chi-Square distribution with degrees of freedom 1. The test decision is obtained by comparing the Likelihood Ratio Test value and the χ^2 value.H$_0$ is rejected if$\Lambda > \chi^2{}_{(1;\alpha)}$or p-value < α, With the following equation [14]:

$$\Lambda = -2ln\left(\frac{L_{H_0}}{L_{H_1}}\right) \qquad (9)$$

D. *Apparent Error Rate (APPER)*

Apparent Error Rate (APPER) is the value used to see the chance of error in classifying objects. The following is an equation to calculate the value of *APPER*

$$APPER = \frac{\sum_{i \neq j=1}^{q} n_{ij}}{\sum_{i,j=1}^{q} n_{ij}} \times 100\% \qquad (10)$$

IV. RESULT

Based on the calculations from equation (3) ordinal probit regression panel data with random effects on EQI can be written in the form of a linear latent response variable with the response variable y_{it} obtained from the continuous latent response variable with the regression model as follows:

$$\hat{y}_{it}^{*} = -0{,}0006x_{1it} - 0{,}0468x_{2it} - 0{,}1319x_{3it} + 0{,}0002x_{4it}$$

with,

$$\hat{y}_{it} = \begin{cases} 1 \text{ jika} y_{it}^{*} \leq -14{,}6697 \\ 2 \text{ jika} - 14{,}6697 < y_{it}^{*} \leq -13{,}4673 \\ 3 \text{ jika} - 13{,}4673 < y_{it}^{*} \leq -12{,}2513 \\ 4 \text{ jika} - 12{,}2513 < y_{it}^{*} \leq -10{,}0287 \\ 5 \text{ jika} - 10{,}0287 < y_{it}^{*} \leq -7{,}4184 \\ 6 \text{ jika} \qquad - 7{,}4184 < y_{it}^{*} \end{cases}$$

model inference partial test and individual test probit regression ordinal panel data with random effects are explained as follows :

TABLE I
MODELING INFERENCE

Variable	Partial Test	Individual Test	
		Z	P-Value
X$_1$	*P-Value (0,00)<alpha (0,05)*	-3,34	0,001
X$_2$		-1,97	0,049
X$_3$		-2,43	0,015
X$_4$		-2,38	0,018

Based on Table 1 that in thepartial test nilai *P-Value (0,00)<alpha (0,05)* it can be concluded that the predictor variable simultaneously influences the response variable. Meanwhile, in the individual test it can be seen that all variables have absolute values of Z_{hitung}> *dari* Z_{tabel}(1,92)and*P-Value*on the response variable is less than*alpha* (0,05), then it can be concluded that the predictor variableX_1, X_2, X_3,andX_4effect on the response variable.

In the model suitability test based on calculations from equation (9) the value of the likelihood ratio test is equal to 16,69>$\chi^2_{(1;0,05)}$ (3,841), it can be concluded that there is enough inter-provincial variability to meet the model of ordinal probit regression panel data with random effect compared to the standard ordinal probit regression model.

In this study, we compared ordinal probit regression panel data with random effects models (xtoprobit) and standard ordinal probit regression (oprobit). The following is the result of calculation 1 – APPER on the xtoprobit and oprobit models. In the xtoprobit model, a value of 1 – APPER is obtained 0,5588 atau 55,88%. This means that the xtoprobit model has been classified correctly at 55,88% and the remaining 44,12% is classified as less (not the same). Meanwhile, the outsample data obtained a value of 1 - APPER of 0,5073 or 50,73%. This means that the insample data has been classified correctly at 50,73% and the remaining 49,27% is classified as less (not the same). Based on these results it was concluded that the ordinal logistic regression model on panel data with random effects (xtoprobit) is better because it is able to produce a more precise classification model than the standard ordinal logistic regression model.

V. CONCLUSION

Based on the analysis of ordinal probit regression panel data with random effects, all variables of the number of poverty, the rate of economic growth, the human development index, and population density affect EQIboth partially and individually with a classification accuracy of 55,88%.

REFERENCES

[1] B. Giddings, B. Hopwood and G. O'brien, "Environment, Economy and Society: Fitting Then Together into Sustainable Development," *Sustainable Development,* pp. 10(4): 187-196, 2002.

[2] BPS, Statistik Lingkungan Hidup, Jakarta: Badan Pusat Statistik, 2021.

[3] KLHK, IKLH 2017, Jakarta: Kementrian Lingkungan Hidup dan Kehutanan, 2017.

[4] A. Masyruroh and Binyati, "Kajian Indeks Kualitas Lingkungan Hidup Kota Serang," *Jurnal Lingkungan dan Sumberdaya Alam,* vol. 4, no. 2, pp. 151-162, 2021.

[5] S. Febriana, H. C. Diartho and N. Istiyani, "Hubungan Pembangunan Ekonomi Terhadap Kualitas Lingkungan Hidup di Provinsi Jawa Timur," *Jurnal Dinamika Ekonomi Pembangunan,* vol. 4, no. 2, pp. 58-70, 2019.

[6] T. Shi, S. Yang, W. Zhang and Q. Zhou, "Coupling Coordination Degree Measurement and Spatiotemporal Heterogenity Between Economic Development and Ecological Environment--- Emperical Evidence From Tropical and Subtropical Regions of China," *Journal of Cleaner Production,* p. 44, 2020.

[7] A. S. Goudie, Human Impact on The Natural Environment 8th Edition, New Jersey: John Wiley & Sons Ltd, 2018.

[8] KLHK, Rencana Strategis Tahun 2020-2024, Kementrian Lingkungan Hidup dan Kehutanan, 2020.

[9] Gujarati and N. Damodar, Basic Econometrics, Fifth Edition, Singapore: McGraw-Hill Inc. Ewlinger, 2009.

[10] L. Matyas and P. Sevestre, The Econometrics of Panel Data, Berlin: Springer, 2018.

[11] Q. Liu and D. A. Pierce, "A note on Gauss-Hermite quadrature," *Biometrika,* pp. 624-629, 1994.

[12] A. Agresti, Categorical Data Analysis second Edition, New York: John Wiley & Sons, Inc., 2002.

[13] F. E. J. Harrell, Regression Modeling Strategies With Application to Linier Models, Logistic Regression, and Survival Analysis, USA: Springer-Verlag, 2001.

[14] Stata Corp, "Longitudinal-Data/Panel-Data Reference Manual," STATA Press, Texas, 2020.