Environmental quality index modeling in Indonesia using ordinal probit regression approach for panel data with random effect

Suliyanto
Department of Mathematics, Universitas Airlangga, Surabaya, Indonesia

Corresponding Author: suliyanto@fst.unair.ac.id

Abstract. Environmental issues are one of the important global issues and are of concern because they are considered to require a solution. The Ministry of Environment and Forestry has developed an Environment Quality Index (EQI) that provides rapid information of an environmental condition with river water quality, air quality, and forest as indicator. This study aims to model EQI in Indonesia using ordinal probit regression for panel data with random effects. Parameter estimation of ordinal probit regression model for panel data with random effect using the marginal likelihood function, and maximize that function using M-point Gauss-Hermite Quadrature method. The data that be used in this study is EQI from 33 Provinces in Indonesia during 2011 to 2016 which are categorized into three category i.e less good, good and very good. The result of modelling is gotten variable that influence to EQI is Human Development Index (HDI). EQI opportunities in DKI Jakarta province in the year 2016 in the category of less good is high enough, that is equal to 99.89%. This may be due to high population density and less balanced with awareness of the population to preserve the environment. The accuracy of classification of modeling result is 70.202%.

1. Introduction
Environmental issues are one of the important global issues and are of concern in some public discussions because they are deemed to require completion for the survival of a better life. According to the United Nations Environment Program (UNEP) in the 2012 Environmental Program Yearbook, 24% of the world's land area has experienced a decline in quality and productivity due to unsustainable land management patterns. Environmental problems faced by the world are also experienced by developing countries such as Indonesia. In the June 2017 Environment and Forestry Week, Minister of Environment and Forestry, Siti Nurbaya said that Indonesia was facing environmental problems and had reached the top [1]. The Ministry of Environment and Forestry developed a provincial-based environmental index since 2009 which provides rapid information on an environmental condition in a certain period. This Environment Quality Index (EQI) calculation uses river water quality, air quality, and forest cover as indicators. The quality of water, air and land is considered to have suffered a lot of damage over time. Nature, which initially can maintain balance, now the balance is lost and there is damage everywhere because of human hands [2].

Human quality is seen from the health and education aspects listed in the Human Development Index (HDI). Thus the HDI is thought to influence the management of environmental conditions. Environmental problems in Indonesia cannot be separated from existing national development.
Changes in Indonesia's environmental quality are assessed as the impact of economic growth and an increase in population [3]. According to [4], economic development that focuses on growth that often contradicts the principle of environmental conservation, so that economic and environmental development seems contradictory. Natural potential is widely used for national economic development. However, higher economic growth is not accompanied by better quality of the environment.

The research that has been done is analyzing the effect of GRDP as a measure of economic growth on the quality of the environment using panel regression. According to [5] there is a significant relationship between economic growth and environmental quality in Indonesia. There is information that economic growth and the HDI have a negative relationship to environmental quality in Indonesia. Then [6] mentioning that there is a negative relationship between the GRDP of the agricultural sector, the management industry sector and the transportation sector with the EQI of the province on Sumatra Island. Based on the negative relationship between the two variables with EQI, it can be interpreted that any increase in economic growth and HDI can reduce the EQI value of a province. The research that has been conducted assumes that the response variables are continuous, whereas according to the Ministry of Environment and Forestry's publication that EQI is categorized into six categories with very good, good, good enough, bad, very bad, and alert. Based on the categorization, researchers categorized EQI into three categories according to the target of the National Medium Term Development Plan (NMTDP) from 2015 to 2019 by classifying the notions of not good, very poor, and alert. The predicate is quite good and well grouped into good.

According to [7], panel data is good to identify and measure effects that cannot be detected in cross-section data or time series in the form of random effects, besides panel data is considered far more informative. The estimation procedure in the ordinal category generally assumes that to estimate the coefficients of the predictor variable does not change between the category response variables (parallel-lines assumption) [8]. Based on the background and considering the response variables are categorical, the researcher is interested in analyzing the factors that are thought to influence the EQI using the ordinal probit regression approach to panel data with random effect.

2. Methodology

2.1. Gauss-Hermite Quadrature Method

Gauss-Hermite Quadrature is an integral approach method to the function $f(x)$ multiplied by another function that has a normal density form. This approach is the sum of weights that estimate functions at a certain point [9]. According to [10], the Gauss-Hermite Quadrature method is to calculate the integral form using a numerical approach with the following equation:

$$\int_{-\infty}^{\infty} g(w_i)\exp\left(-w_i^2\right)dw_i \approx \sum_{m=1}^{M} w_m^* g(\alpha_m^*)$$  \hspace{1cm} (1)

with

$M$ is a lot of quadrature points

$\alpha_m^*$ is a grid of quadratures, which is the root of the Hermite polynomial $m$th.

$H_M(\alpha^*)$ is Hermite polynomial which is defined as

$$H_M(\alpha^*) = (1)^m e^{-x^2/2} \frac{d^m}{dx^m} e^{-x^2/2} \bigg|_{x = \alpha^*} \text{ or}$$

$$H_M(\alpha^*) = (1)^m e^{-x^2} \frac{d^m}{dx^m} e^{-x^2} \bigg|_{x = \alpha^*}$$  \hspace{1cm} (2)

$w_m^*$ is the quadrature weight that corresponds to the $m$th grid and is stated as follows

$$w_m^* = \frac{2^m m! \sqrt{\pi}}{M^2 H_{M-1}(\alpha_m^*)^2}$$  \hspace{1cm} (3)
2.2. Ordinal Probit Ordinal Model

Ordinal probit regression is a statistical method for analyzing response variables that have an ordinal scale consisting of three or more categories. Given paired data \((Y, X)\), the response variable \(Y\) has an ordinal scale with \(K\) categories and \(X = (X_1, X_2, ..., X_p)\) is a vector of covariates. Suppose \(\pi_k(X) = P(Y = k \mid X)\) is the probability of \(Y = k\) if \(X\) is known, for \(k = 1, 2, ..., K\) and \(y_k = P(Y \leq k \mid X)\) is the cumulative probability of the response variable in the category to \(k\) for \(k = 1, 2, ..., K - 1\). The relationship between probability vectors \(\pi = (\pi_1(X), ..., \pi_K(X))\) with ordinal \(Y\)-response variables can be expressed in the following models

\[
g(y_k) = \theta_k + \sum_{j=1}^{p} \beta_j X_j
\]

for \(k = 1, 2, ..., K - 1\), \(g(y_k) = \Phi^{-1}(y_k)\) is the probit link function, \(\Phi\) is the standard normal distribution function. From (4) is obtained

\[
y_k = \Phi \left( \theta_k + \sum_{j=1}^{p} \beta_j X_j \right), \quad k = 1, 2, ..., K - 1
\]

If \(\pi_k(X) = P(Y = k \mid X)\) is the probability of the response variable in category \(k\) with the condition that the predictor variable \(X\) is known, then from (5) is obtained the following equation

\[
y_k = \sum_{s=1}^{k} \pi_s (X) = \Phi \left( \theta_k + \sum_{j=1}^{p} \beta_j X_j \right) \text{ and } y_K (X) = 1
\]

2.3. Ordinal Probit Regression Model for Panel Data with Random Effect

Ordinal probit regression models for panel data with random effects can be expressed in the form of linear latent response variables, where ordinal response variables \(y_{it}\) with \(K\) categories are obtained from continuous latent response variables as follows

\[
y_{it}^* = X_{it}\beta + v_i + \epsilon_{it}; \quad i = 1, 2, ..., n \text{ and } t = 1, 2, ..., n_i
\]

with

\[
y_{it} = \begin{cases} 
1 & \text{if } y_{it}^* \leq \kappa_1 \\
2 & \text{if } \kappa_1 < y_{it}^* \leq \kappa_2 \\
\vdots & \text{if } \kappa_{k-1} < y_{it}^* \\
K & \text{if } \kappa_{K-1} < y_{it}^*
\end{cases}
\]

\(X_{it} = (X_{1it}, X_{2it}, ..., X_{pit})\) is the vector of predictor variables in the unit of cross section \(i\) and time \(t\), \(\beta = (\beta_1, \beta_2, ..., \beta_p)\) is a vector of parameters, \(v_i\) is the random effect of the unit of the cross section \(i\) assuming an identical independent distribution \(N(0, \sigma_v^2)\), and \(\epsilon_{it}\) is the error of the unit cross section \(i\) and time \(t\) which is assumed to have a standard normal distribution and independent of \(v_i\).

Suppose that \(\kappa\) is the cutpoint set of \(\kappa_1, \kappa_2, ..., \kappa_{K-1}\) and \(\Phi\) is the standard normal distribution function. From equations (7) and (8) it is found that the probability of observing in category \(k\) for the response variable \(y_{it}\) is

\[
p_{itk} = \Pr (y_{it} = k \mid \kappa, X_{it}, v_i)
\]

\[
= \Pr (\kappa_{k-1} < X_{it}\beta + v_i + \epsilon_{it} \leq \kappa_k) \\
= \Pr (\kappa_{k-1} - X_{it}\beta - v_i < \epsilon_{it} \leq \kappa_k - X_{it}\beta - v_i) \\
= \Phi (\kappa_k - X_{it}\beta - v_i) - \Phi (\kappa_{k-1} - X_{it}\beta - v_i)
\]

with \(\kappa_0\) is considered \(-\infty\) and \(\kappa_K\) is considered \(+\infty\). \(X_{it}\) has no constants, due to the effect being absorbed into the cutpoint. The conditional distribution of the response variable \(y_{it}\) is
\begin{align*}
f(y_{it}, \kappa, X_{it}\beta + v_i) &= \prod_{k=1}^{K} p_{itk}^{l_k(y_{it})} \\
&= \exp \sum_{k=1}^{K} \{ l_k(y_{it}) \log (p_{itk}) \} \quad (9)
\end{align*}

with \( l_k(y_{it}) = \begin{cases} 1 & \text{if } y_{it} = k \\ 0 & \text{otherwise} \end{cases} \). For panel \( i \), with \( i = 1, 2, \ldots, M \) is obtained conditional distribution from \( y_{it}, y_{i2}, \ldots, y_{in_t} \) is

\begin{align*}
\prod_{t=1}^{n_t} f(y_{it}, \kappa, X_{it}\beta + v_i)
\end{align*}

and the marginal likelihood function for panel data \( l_i \) is as follows:

\begin{align*}
l_i(\beta, \kappa, \sigma^2) &= \int_{-\infty}^{\infty} e^{-\frac{v^2}{2\sigma^2}} \left\{ \prod_{t=1}^{n_t} f(y_{it}, \kappa, X_{it}\beta + v_i) \right\} dv_i \\
&= \int_{-\infty}^{\infty} g(y_{it}, \kappa, X_{it}, v_i) \, dv_i \quad (11)
\end{align*}

Integral (11) can be approached with the M-point Gauss-Hermite quadrature as follows:

\begin{align*}
\int_{-\infty}^{\infty} e^{-x^2}h(x)\,dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)
\end{align*}

Equation (11) is equivalent to the equation:

\begin{align*}
\int_{-\infty}^{\infty} g(x)\,dx \approx \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(a^*_m)
\end{align*}

with \( w^*_m \) is a quadrature weight and \( a^*_m \) is a quadrature abscissa. The log likelihood function \( L \) is the sum of the log likelihood function for the panel data \( l_i \).

2.4. Inference on Ordinal Probit Regression Model for Panel Data

According to [11] to examine the effect of predictor variables simultaneously on the response variable used the following hypothesis:

\begin{align*}
H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0 \\
H_1: \text{There is at least one } \beta_j \neq 0 \text{; } j = 1, 2, \ldots, p
\end{align*}

to test this hypothesis used wald test statistics as follows:

\begin{align*}
W = \hat{\beta}^T V^{-1} \hat{\beta} \sim \chi^2_p \quad (12)
\end{align*}

with

\begin{align*}
V^{-1} = I(\hat{\beta}) = - H(\hat{\beta}) = \frac{\delta^2}{\hat{\beta}^T \delta \hat{\beta}^T}
\end{align*}

Wald test statistics in (12) distributed Chi-Square with \( p \) degrees of freedom. The test criteria are \( H_0 \) rejected if \( W > \chi^2_{p, \alpha} \).

If testing is simultaneously significant, then proceed with individual testing for each predictor variable in the ordinal probit regression model for panel data with the hypothesis:

\begin{align*}
H_0: \beta_j = 0 \\
H_1: \beta_j \neq 0
\end{align*}
To test the hypothesis, the standard normal test statistics are used as follows:

\[ Z_j = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \]  

(13)

\( s(\hat{\beta}_j) \) is the standard deviation of the \( \hat{\beta}_j \) estimator. The \( Z_j \) test statistic in equation (13) has a standard normal distribution. A critical area to test the hypothesis with a significance level of \( \alpha \) is to reject \( H_0 \) if \( |Z_j| > Z_{\alpha/2} \).

The suitability test of the ordinal probit regression model for panel data with random effects was used to compare with the standard ordinal probit regression model. The hypothesis used to test the suitability of the ordinal probit regression model for panel data with random effects is

\( H_0 : \) Ordinal probit regression model for panel data with random effect is not appropriate.

\( H_1 : \) Ordinal probit regression model for panel data with random effect is appropriate

To test the hypothesis is used the Likelihood Ratio Test (LRT) as follows:

\[ \Lambda = -2 \ln \left( \frac{L_{H_0}}{L_{H_1}} \right) \]  

(14)

with \( L_{H_0} \) is the maximum value of the likelihood function standard ordinal probit regression model, \( L_{H_1} \) is the maximum value of likelihood function ordinal probit regression model for panel data with random effects. According to [12], LRT statistics are Chi-Square distribution with 1 degree of freedom. The decision is that \( H_0 \) is rejected if \( \Lambda > \chi^2_{(1, \ \alpha)} \).

3. Result and Discussion

The data used in this research are data on the development of the Environmental Quality Index (EQI) of 33 provinces in Indonesia and the factors that are thought to influence it, namely the Gross Regional Domestic Product (GRDP) \( (X_1) \) and the Human Development Index (HDI) \( (X_2) \) during the period 2011-2016. EQI is grouped into three categories, namely 1 (not good), 2 (good), and 3 (very good).

3.1. EQI Modeling

Ordinal probit regression modeling for panel data with random effects begins by selecting the best quadrature point using the smallest Akaike Information Criteria (AIC). Quadrature points are used starting from values 4 to 45. The estimation results of the EQI model obtained the smallest AIC value of 201.9371 in the 7 quadrature point. After obtaining the best quadrature point, the analysis performed is inference on the ordinal probit regression model with random effect.

The results of parameter testing simultaneously using the Wald test with the help of Stata 14 software obtained a p-value of 0.0007. Because the p-value is less than \( \alpha = 5\% \), the decision rejects \( H_0 \) and the conclusion is that there is at least one predictor variable that affects EQI. After getting the conclusion that there are variables that have a significant effect simultaneously, then the analysis continues on individual testing. Testing results of individuals using standard normal test statistics with the help of Stata 14 software obtained the following results:

| Variable | Coefficient | std error | Z      | p-val |
|----------|-------------|-----------|--------|-------|
| \( X_1 \) | -0.001655   | 0.000975  | -1.70  | 0.090 |
| \( X_2 \) | -0.187092   | 0.071514  | -2.62  | 0.009 |
Based on Table 3.1 obtained variables that have a significant effect on EQI is HDI with p-value of 0.009 smaller than α = 5% then the decision rejects $H_0$ and the conclusion is the HDI variable has a significant effect on EQI. Next GRDP variable is not significant with p-value of 0.090 more than 5%.

Based on Table 3.1 obtained ordinal probit regression model for EQI panel data with random effect as follows:

$$\hat{y}_{it}^* = -0.001655X_{1it} - 0.187092X_{2it} \quad (15)$$

$\hat{y}_{it}^*$ is the estimation of the EQI latent response variable and the cutpoint estimation results with the help of Stata 14 software, the cut 1 value is -14.3536 and the cut 2 is -9.612621. Based on the cut point value, then from (15) the category of EQI response variable is obtained as follows:

$$y_{it} = \begin{cases} 
1 & \text{if } \hat{y}_{it}^* \leq -14.3536 \\
2 & \text{if } -14.3536 < \hat{y}_{it}^* \leq -9.612621 \\
3 & \text{if } -9.612621 < \hat{y}_{it}^* 
\end{cases} \quad (16)$$

The results of the model suitability test obtained LRT statistical value of 61.99 with p-value of 0.000 and the decision was to reject $H_0$ and concluded that the ordinal probit regression model for panel data with random effects was appropriate. APPER calculations are done to see the accuracy of classification of the EQI response variables before and after the estimation. The calculation with the help of the program on OSS-R software obtained the accuracy value of the classification as 70.202%.

### 3.2. Interpretation of the EQI Model

The lowest EQI value of 38.69 in DKI Jakarta Province in 2016 had a GRDP of 1540.078 trillion rupiah and an HDI of 79.6. If the GRDP and HDI values are substituted to (15), then the predicted EQI value in DKI Jakarta Province in the 11th observation in 2016 is $\hat{y}_{(11)6}^* = -17.44135$. Based on (16) the predicted value of EQI is included in the not good category. Categorizing the predicted value of EQI in DKI Jakarta Province in 2016 did not experience a change compared to the observation value, which remained in the not good category. The value of GRDP is very large and a relatively high HDI does not make EQI in DKI Jakarta Province better.

Based on Table 3.1, the probability of DKI Jakarta Province in 2016 with the GRDP of 1540.078 trillion rupiah and HDI of 79.6 having EQI in the not good category is

$$\hat{P}_{(11)61} = \Phi(\hat{\kappa}_1 - X_{(11)\bar{b}}) - \Phi(\hat{\kappa}_0 - X_{(11)\bar{b}})$$

$$= \Phi[-14.3536 + 17.44135] - 0$$

$$= \Phi(3.08775)$$

$$= 0.9989$$

Probability of DKI Jakarta Province in 2016 with a GRDP of 1540.078 trillion rupiah and HDI of 79.6 having EQI in the good category are

$$\hat{P}_{(11)62} = \Phi(\hat{\kappa}_2 - X_{(11)\bar{b}}) - \Phi(\hat{\kappa}_1 - X_{(11)\bar{b}})$$

$$= \Phi[-9.612621 - (-17.44135)] - \Phi(3.08775)$$

$$= \Phi(7.828729) - 0.9989$$

$$= 1 - 0.9989$$

$$= 0.0011$$

Therefore the probability of DKI Jakarta Province in 2016 with the GRDP of 1540.078 trillion rupiah and the HDI of 79.6 has EQI in the very good category of 0.
The HDI predictor variable has a significant effect on EQI. It must be a concern for policy makers in order to create a better environmental quality target. The results of the EQI modeling obtained information that the increase in HDI actually resulted in a decrease in the quality of the environment. This is the focus of attention, because the two variables that are targeted to be better are both contradictory. The increasing value of HDI causes an increase in the index of knowledge, life expectancy and standard of living. The higher human knowledge makes the technology created even higher. Technology is used by humans to facilitate their activities, but the technology created is not environmentally friendly. Plastics are one of the simplest examples of human-made technology that is used for various things, activities become easier, but it becomes a problem when there is excessive accumulation of plastic production and causes environmental pollution. The average plastic bag is only used for 25 minutes, but to be destroyed and decomposed in nature it takes up to 500 years. Just as plastic and styrofoam waste results from human activities produce microplastic particles which increasingly threaten DKI Jakarta. So many new technologies created by humans to meet their standard of living, but unfortunately they do not think about their impact on the environment in the long run. Technological sophistication in this digital era does not give birth to a culture of protecting and preserving the environment.

4. Conclusions

EQI modeling results in Indonesia based on ordinal probit regression approach for panel data with random effect obtained the best quadrature point value of 7 with the smallest AIC, which is equal to 201.937. The EQI prediction model Province in Indonesia is as follows:

\[
\hat{y}_{it}^{*} = -0.001655X_{1it} - 0.187092X_{2it}
\]

EQI prediction results based on cut point estimates are categorized as follows:

\[
y_{it} = \begin{cases} 
1 & \text{if} \quad \hat{y}_{it}^{*} \leq -14.3536 \\
2 & \text{if} \quad -14.3536 < \hat{y}_{it}^{*} \leq -9.612621 \\
3 & \text{if} \quad -9.612621 < \hat{y}_{it}^{*}
\end{cases}
\]

The classification accuracy of the modeling is 70.202%.

Based on EQI modeling in Indonesia using ordinal probit regression approach for panel data with random effect it is known that the predictor variables GRDP and HDI have a negative effect on EQI. For \( \alpha = 5\% \) the GRDP variable does not have a significant effect on EQI, while the HDI variable has a significant effect on EQI. Every 1 unit increase in HDI resulted in a decrease in EQI value of 18.71\%. Probability of Province DKI Jakarta in 2016 with a GRDP of 1540.078 trillion rupiah and a HDI of 79.6 having EQI in the not good category of 99.89\%, good category of 0.11\% and very good category of 0\%. This shows that in 2016 Province DKI Jakarta had EQI in the not good poor category is very high and for the very good category it was impossible.

Acknowledgments
The author would like to thank the Faculty of Science and Technology of Airlangga University for funding the participation of the International Conference of Mathematics, Science and Computer Science.

References
[1] Hartanto A 2017 Menteri LHK Sebut Indonesia Sedang Hadapi Masalah Lingkungan [https://tirto.id], 24 Mei 2018
[2] MaskoerI J 1988 Ilmu Alamiah Dasar untuk Perguruan Tinggi dan Umum (Surabaya:Bina Ilmu.)
[3] Anonim 2014 Indeks Kualitas Lingkungan Hidup Indonesia 2014 (Jakarta:Kementrian Lingkungan Hidup dan Kehutanan Republik Indonesia).
[4] Yakin A 1997 Ekonomi Sumberdaya Dan Lingkungan: Teori Dan Kebijakan Pembangunan Berkelanjutan (Jakarta:Academi Presindo)
[5] Rahajeng A 2014 Pertumbuhan Ekonomi dan Kualitas Lingkungan Hidup Indonesia 2014:
Modifikasi Kurva Kutznet, *Prosiding Seminar Nasional Teknologi Terapan* (Yogyakarta:Sekolah Vokasi Universitas Gadjah Mada)

[6] Prasurya R A 2016 Analisis Pengaruh PDRB Terhadap Kualitas Lingkungan Hidup di Pulau Sumatera Tahun 2011-2014 *Skripsi* (Universitas Lampung)

[7] Hsiao C 2003 *Analysis of Panel Data 2nd edition* (Cambridge:Cambridge University Press)

[8] Pfarr C, Schmid A and Schneider U 2011 Estimating Ordered Categorical Variables Using Panel Data: a Generalized Ordered Probit Model with an Autofit Procedure *Journal of Economics and Econometrics* 54(1) pp 7-23

[9] Agresti A 2002 Categorical Data Analysis, 2nd edition (New York:John Wiley & Sons,Inc)

[10] Naylor J C and Smith A F M 1982 Applications of a Method for the Efficient Computation of Posterior Distributions *Journal of the Royal Statistical Society* 46 pp 214-25

[11] Harrell F E 2015 Regression Modeling Strategies, 2nd edition, (Switzerland : Springer)

[12] Hogg R V, McKeans J W and Craig A T 2004 Introduction to Mathematical Statistics, Sixth Edition (New Jersey:Prentice Hall)