ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 5, Issue 1, April 2025, pp. 56-66

A Zero-Inflated Ordered Probit Approach to Modeling Household Poverty Levels

Nidya Putri Yudhani [a,1], Vita Ratnasari [a,2,*], Santi Puteri Rahayu [a,3]

[a] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 61111, Indonesia
[1] 6003232026@student.its.ac.id; [2] vita_ratna@its.ac.id*; [3] santi_pr@statistika.its.ac.id
* Corresponding author

ARTICLE INFO

Keywords
Binary Probit
Household Poverty
Latent Variable
Ordinal Probit
ZIOP

ABSTRACT

This research addressed the limitations of the ordered probit (OP) regression model in handling data that contains an excessive number of zero responses. The zero-inflated ordered probit (ZIOP) model was employed to overcome this issue. This model separates the estimation of structural zeros and ordinal outcomes through two distinct components: a binary probit for zero inflation and an OP for ordered categories. Due to the absence of closed-form solutions, parameter estimation was conducted using the maximum likelihood estimation (MLE) method with the Berndt-Hall-Hall-Hausman (BHHH) iterative algorithm. The analysis was based on 4,067 household-level observations from Indonesia's National Socio-Economic Survey, incorporating indicators of health, education, and standard of living derived from the multidimensional poverty index (MPI) framework. The result of the Vuong test (4.56) confirmed that the ZIOP model significantly outperformed the conventional OP model for zero-inflated ordinal data. Therefore, the ZIOP model is considered more appropriate for analyzing household poverty classifications with a high prevalence of zero observations.

1. Introduction

The conventional ordered probit (OP) model has limitations in explaining the excess of zero observations. A proposal has been made to extend the OP model to zero-inflated ordered probit (ZIOP) to address this issue [1]. As part of the ZIOP model, the null observations are split into two parts: an OP model to guess the response variable's ordinal category and a binary probit model to find the null status. The maximum likelihood estimation (MLE) method and Berndt-Hall-Hall-Hausman (BHHH) numerical iteration are used to guess the parameters in the ZIOP model [2]. This model has an inflation variable that affects the probability of zero inflation; the proportion of zeros is higher than expected [3].

The ZIOP method can be applied to poverty cases in Indonesia, a complex and multidimensional problem that is a priority for development [4]. Poverty is considered one of the measures of developmental success listed in the sustainable development goals (SDGs), which aim to end poverty in all forms globally. In 2024, D.I. Yogyakarta Province has the most impoverished population in Java [5]. According to data, the number of impoverished people in D.I. Yogyakarta was 488,530 in March 2017 and reduced to 475,720 in March 2020. However, this number rose to 506,450 in March

2021 due to the impact of COVID-19. In March 2024, the population of impoverished people was recorded at 445,550, reflecting a decrease of 2,900 compared to March 2023. The majority of individuals living in poverty were in urban areas, with 319,400 people in urban areas and 126,150 in rural areas.

The Statistics Indonesia (Badan Pusat Statistik, BPS) estimates poverty rates based on data from the National Socio-Economic Survey using a basic needs approach. Individuals are classified as poor when their average monthly per capita expenditure is below the poverty line, which reflects the minimum requirements for food and nonfood consumption [6]. In addition, the United Nations Development Programme (UNDP) and BPS assess poverty using the multidimensional poverty index (MPI), which includes indicators across health, education, and living standards [7]. This study adopted predictor variables from these three dimensions to investigate the factors influencing household poverty in D.I. Yogyakarta in 2024.

Several studies have applied zero-inflated models to address data with an excess of zero outcomes, particularly in fields such as health, economics, and social sciences [1], [2], [8]–[12]. For instance, zero-inflated count models are commonly used to analyze over dispersed count data, while the ZIOP model has been introduced more recently to handle ordinal outcomes with structural zeros. However, applications of the ZIOP model in the context of multidimensional poverty analysis remain limited. Existing literature tends to focus on income-based poverty measures or utilize standard ordered response models that do not account for zero inflation in poverty categorization.

This study contributes to the literature by applying the ZIOP model to household-level poverty data in D.I. Yogyakarta, Indonesia, using indicators aligned with the MPI. Households are categorized into five groups: nonpoor, other vulnerable poor, near poor, poor, and very poor. The ZIOP approach allows for a dual modeling structure: a binary probit component to distinguish between structurally nonpoor and potentially poor households, and an OP component to capture the severity of poverty among the latter. By integrating these components, the study aims to provide a more accurate and nuanced understanding of poverty dynamics and their predictors in a developing region.

2. Material and Method

2.1. Data Variables

This analysis utilized secondary data sourced from the March 2024 National Socio-Economic Survey and the March 2024 poverty threshold specific to D.I. Yogyakarta Province. The sample size of 4,067 residences was the unit of analysis in this study, which was a household. The variables employed in this investigation are illustrated in Table 1. The binary model in ZIOP consisted of five predictor variables and an ordinal response variable.

Table 1. Research Variable Tendency of Household Poverty Levels

Research Variables	Variable Name	Data Scale
Y	Household poverty levels	Ordinal
X_1	Can read/write Latin/alphabet letters	Nominal
X_2	Highest education	Ordinal
X_3	Has a health complaint	Nominal
X_4	PKH housing assistance recipients	Nominal
X_5	Health insurance recipient	Nominal

In addition, Table 2 describes the ordinal response variable and one predictor variable for the ordinal model in ZIOP that describes the level of the household poverty. This variable was derived from a number of the binary probit predictor variables that explain the propensity poverty of households.

Table 2. Research Variable of Household Poverty Levels

Research Variables	Variable Name	Data Scale
Y	Household poverty levels	Ordinal
$\boldsymbol{Z_1}$	Highest education	Ordinal

2.2. Multicollinearity Test

A multiculturality test is implemented to determine the existence of any association among the predictor variables within the regression model [13]. The variance inflation factor (VIF) value was calculated using (1) to determine the correlation between predicator variables.

$$VIF = \frac{1}{1-R^2} \tag{1}$$

where R^2 is the coefficient of determination of X. The multicollinearity occurs if the VIF value is >10 [14].

2.3. Biner Probit Model

The normal cumulative distribution function (CDF) method is used by the probit model. This method can be used to get around the problems with linear probability models [15]. In the binary probit model, categorical data serves as the response variable. A value of 1 signifies the presence of a feature, while a value of 0 signifies its absence. The binary probit regression model is described in (2).

$$y_1 = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta} + \varepsilon \tag{2}$$

let $\boldsymbol{x} = \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_p \end{bmatrix}^T$ represent a predictor vector with dimensions $(p+1) \times 1$, where p denotes the number of predictor variables. Similarly, the coefficient vector is given by $\boldsymbol{\beta} = \begin{bmatrix} 1 & \beta_1 & \beta_2 & \dots & \beta_p \end{bmatrix}^T$, also of size $(p+1) \times 1$, and contains the parameters to be estimated. The error term (ε) is assumed to follow a standard normal distribution, that is, $\varepsilon \sim N(0,1)$ [16].

2.4. Ordered Probit Model

The OP regression is a sort of regression model that can be used to show the relationship between response variables and predictor factors. The response variable is a random variable that is sorted by order, and the predictor variable can be either a continuous fixed variable (interval or ratio scale) or a discrete fixed variable (nominal or ordinal scale). OP regression modeling, as in (3).

$$y_2 = \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{\gamma} + \delta \tag{3}$$

here, $\boldsymbol{z} = \begin{bmatrix} 1 & z_1 & z_2 & \dots & z_p \end{bmatrix}^T$ denotes the predictor vector of dimension $(p+1) \times 1$, where p is the number of predictors. The vector $\boldsymbol{\gamma} = \begin{bmatrix} 1 & \gamma_1 & \gamma_2 & \dots & \gamma_p \end{bmatrix}^T$ contains the parameters to be estimated, also of size $(p+1) \times 1$. The error term (δ) is assumed to follow a standard normal distribution with a mean of 0 and a variance of 1, denoted as $\delta \sim N(0,1)$ [16].

2.5. Zero-inflated Ordered Probit Model

The ZIOP model is an extension of the OP model developed to accommodate data characterized by excess zeros. The ZIOP model performs split population by dividing zero observations into binary and OP models to overcome this problem [1]. The ZIOP model has an inflated variable where the inflated variable in ZIOP is a binary variable that describes whether the observation will have the possibility of zero-inflation or not. By modelling the inflated variable as a binary variable, the ZIOP model can overcome the zero-inflated nature of the data and provide more accurate estimates [2].

It is assumed that s denotes an indicator variable that indicates zero-inflated observations $(s = 0)$ and non-zero-inflated observations $(s = 1)$. The mapping between s and y_1^* is as in (4)–(5).

$$s = 0, \text{ if } y_1^* \leq 0 \tag{4}$$

$$s = 1, \text{ if } y_1^* > 0 \tag{5}$$

The latent variable y_1^* can be obtained with the following by (6).

p-ISSN 2798-253X
e-ISSN 2798-3153

$$y_1^* = \boldsymbol{x}^T\boldsymbol{\beta} + \varepsilon \tag{6}$$

The likelihood of an observation not belonging to the zero-inflated component is expressed in (7).

$$P(s = 1|x) = P(y_1^* > 0|x) = \Phi(\boldsymbol{x}^T\boldsymbol{\beta}) \tag{7}$$

Here, $\Phi(\boldsymbol{x}^T\boldsymbol{\beta}) = \Phi(.)$ denotes the CDF of the standard normal distribution. The OP regression function in (8) connects the latent variable y_2^* to the observed ordinal levels $r = \{r_0, r_1, \dots, r_j, \dots, r_{k-1}\}$ if $s = 1$.

$$y_2^* = \boldsymbol{z}^T\boldsymbol{\gamma} + \delta \tag{8}$$

where, in ZIOP regression, error(ε, δ) in binary and ordinal latent variables are assumed to be independent.

The variable y_2^* cannot be observed, so in categorizing the response variable y_2^*, a certain threshold is used, for example μ. The categorization of OP levels in ZIOP is denoted by r.

for $y_2^* \leq \mu_0$ is categorized with $r = 0$,
for $\mu_0 < y_2^* \leq \mu_1$ is categorized with $r = 1$,
for $\mu_1 < y_2^* \leq \mu_2$ is categorized with $r = 2$,
$\vdots$
for $y_2^* \geq \mu_k$ is categorized with $r = k$,

where k represents the highest category in the OP, and μ_{k-1} denoted the threshold value for each category, adhering to the constraint $\mu_0 < \mu_1 < \mu_2 < \cdots < \mu_{k-1}$. The mapping between r and y_2^* is generally written as in the following (9).

$$r(y_2^*) = \begin{cases} 0, & \text{if } y_2^* \leq \mu_0 \\ j, & \text{if } \mu_{j-1} < y_2^* \leq \mu_j \ ; j = 1, \dots, k-1 \\ k, & \text{if } \mu_{k-1} < y_2^* \end{cases} \tag{9}$$

The variable y_2^* has the same probability density function (PDF) as shown in (10).

$$f(y_2^*) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2^* - \boldsymbol{z}^T\boldsymbol{\gamma})^2\right) \text{ for } -\infty < y_2^* < \infty \tag{10}$$

The OP regression model is generally written in (11).

$$P(r) = \begin{cases} P(r = 0) = \Phi(\mu_0 - \boldsymbol{z}^T\boldsymbol{\gamma}) \\ P(r = j) = \Phi(\mu_j - \boldsymbol{z}^T\boldsymbol{\gamma}) - \Phi(\mu_{j-1} - \boldsymbol{z}^T\boldsymbol{\gamma}); j = 1, \dots, k-1 \\ P(r = k) = 1 - \Phi(\mu_{k-1} - \boldsymbol{z}^T\boldsymbol{\gamma}) \end{cases} \tag{11}$$

s and r cannot be observed individually in relation to zero. The observed response variable in the ZIOP model is $y = s \times r$. Thus, the observation of zero in ZIOP ($y = 0$) occurs when $s = 0$ or $s = 1$, and $r = 0$. To observe a positive y, requires both $s = 1$ and $r > 0$. The relationship of s and r in relation to zero can be written in (12).

$$y = s \times r = \begin{cases} 0, & \text{if } s = 0 \text{ or } r = 0 \\ 1, & \text{if } s = 1 \text{ and } r = 1 \\ \vdots \\ k, & \text{if } s = 1 \text{ and } r = k \end{cases} \tag{12}$$

The general expression for the probability of the ZIOP model is presented in (13).

$$P(y) = \begin{cases} P(y = 0|x, z) = P(s = 0|x) + P(s = 1|x)P(r = 0|z, s = 1) \\ P(y = j|x, z) = P(s = 1|x)P(r = j|z, s = 1) \qquad ; j = 1, \dots, k-1 \\ P(y = k|x, z) = P(s = 1|x)P(r = k|z, s = 1) \end{cases} \tag{13}$$

Thus, the ZIOP model is obtained which is written as (14).

p-ISSN 2798-253X
e-ISSN 2798-3153

$$P(y) = \begin{cases} P(y=0|x,z) = (1-\Phi(\boldsymbol{x^T\beta})) + \Phi(\boldsymbol{x^T\beta})\Phi(\mu_0 - \boldsymbol{z^T\gamma}) \\ P(y=j|x,z) = \Phi(\boldsymbol{x^T\beta})\Phi(\mu_j - \boldsymbol{z^T\gamma}) - \Phi(\mu_{j-1} - \boldsymbol{z^T\gamma}) & ; j=1,\dots,k-1 \\ P(y=k|x,z) = \Phi(\boldsymbol{x^T\beta})(1-\Phi(\mu_{k-1} - \boldsymbol{z^T\gamma})) \end{cases} \quad (14)$$

In the ZIOP model, $\Phi(\boldsymbol{x^T\beta})$ represents the CDF of the standard normal distribution in the binary probit component, while $\Phi(\mu_j - \boldsymbol{z^T\gamma})$ denotes the CDF of the standard normal distribution applied in the OP component.

2.6. Estimate the ZIOP Model

The binary probit regression component manages the prevalence of excess zero observations by enabling the model to distinguish between structural and ordinal outcomes. Parameter estimation is conducted using the MLE approach [14]. The corresponding log-likelihood function is defined in (15).

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=0}^{k} h_{ij} \ln[P(y=j|\boldsymbol{x},\boldsymbol{z},\boldsymbol{\theta})] \quad (15)$$

where $\boldsymbol{\theta}$ denotes the vector of parameters to be estimated.

2.7. Vuong Test

The Vuong test is employed to compare non-nested models, particularly when evaluating models from different regression families, where it often provides more reliable results than the Akaike information criterion (AIC) [17]. This test is especially recommended for distinguishing between zero-inflated models and their conventional counterparts [11]. The Vuong statistic (v) evaluates whether the predictive capabilities of the two models differ significantly in explaining the ordinal response variable. The formula used to calculate the Vuong statistics is provided in (16).

$$v = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}} \quad (16)$$

where n is the sample size and v follows an asymptotic standard normal distribution. The null hypothesis H_0 is rejected when $v > Z_{\alpha/2}$, suggesting that the ZIOP model performs significantly better than the traditional OP model.

2.8. Marginal Effect of the ZIOP Model

The marginal effects for categorical predictors are calculated by examining the change in probabilities of each outcome y as predictor variables change [16]. In the ZIOP model, these effects are calculated as differences in category probabilities, controlling for other variables, as shown in (17).

$$\underset{P(y=j)}{ME} = P(y=j|x_i,z_i) - P(y=j|x_i,z_i) \quad (17)$$

2.9. Modeling Procedure

The modeling procedure in this study was carried out through a series of structured steps. First, household level microdata was obtained from the March 2024 National Socio-Economic Survey for the D.I. Yogyakarta Province. The response and predictor variables were then identified, and the response variable was defined as an ordinal variable based on the official poverty line established by BPS. Subsequently, a descriptive statistical analysis was conducted to provide an overview of the data characteristics. A multicollinearity diagnostic was performed to ensure the reliability of the predictors. Following this step, the household poverty status was modeled using the ZIOP approach, in accordance with the formulation presented in (15). The model parameters were then tested through hypothesis testing procedures. A comparison between the ZIOP model and the conventional ordered probit model was conducted using the Vuong test, as described in (16). Finally, the best-fitting model was interpreted to determine the significant predictor variables and to assess the severity of poverty across household categories.

 p-ISSN 2798-253X
e-ISSN 2798-3153

3. Results and Discussion

The objective of this study was to examine the distribution of household poverty status, which was classified into five groups: nonpoor, other vulnerable poor, near poor, poor, and very poor. A household is considered nonpoor if its monthly per capita expenditure exceeds 1.6 times the official poverty threshold. The poverty line represents the minimum financial requirement to fulfill daily food consumption of 2,100 kilocalories per person, and essential nonfood needs. According to BPS, in March 2024, the poverty line for D.I. Yogyakarta was IDR602,437.00 per capita per month. Therefore, households with monthly per capita expenditures equal to or below this value are classified as poor.

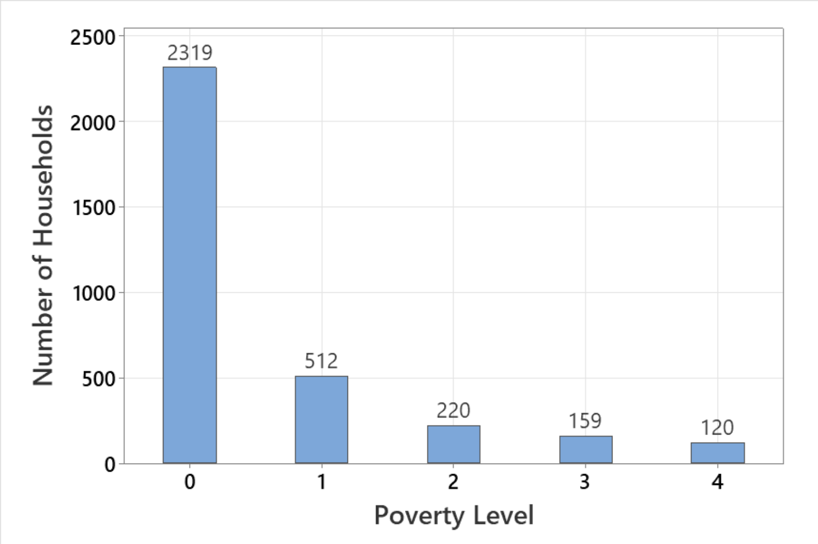

Fig. 1 Frequency of households by poverty level.

As shown in Fig. 1, 72.8% of households (2,961 out of 4,067) in the dataset fall into the nonpoor category. The high proportion of households classified as nonpoor highlights the need for a model that can effectively handle ordinal response variables with a significant number of zero outcomes. The ZIOP model was therefore applied, as it can simultaneously address excess zeros and the ordinal nature of the poverty level classifications.

3.1. Multicollinearity Test

Prior to the model estimation process, multicollinearity diagnostics were performed by computing the VIF for each explanatory variable to ensure the independence of predictors. The results, presented in Table 3, indicate that all VIF values are below the commonly accepted threshold of 10. This confirms the absence of multicollinearity, ensuring the reliability of the estimated coefficients in subsequent modeling stages.

Table 3. Multicollinearity Test of Predictor Variables

Predictor Variables	VIF
Can read/write Latin/alphabet letters (X_1)	1.27
Highest education (X_2)	1.13
Has a health complaint (X_3)	1.17
PKH housing assistance recipients (X_4)	1.09
Health insurance recipient (X_5)	1.55

3.2. Zero-inflated Ordered Probit Model

The ZIOP model was estimated using five predictor variables, with the results shown in Table 4. The model comprises two components: the binary probit part, which models the inflation of zeros, and the ordered probit part, which captures the ordinal structure of poverty levels.

Table 4. Research Variable Level of Poor Households

Predictor Variable		**Coef**	**Std Error**	***p*-value**
Poverty levels				
Highest education (Z_1)	Junior high school	-0.212	0.118	0.073
	Senior high school	-0.204	0.127	0.109
	Diploma	-1.546	0.609	0.011
	Undergraduate	-1.009	0.560	0.072
	Postgraduate	-2.319	2.220	0.296
Inflate				
Can read/write Latin/alphabet letters (X_1)	No	0.503	0.356	0.158
Highest education (X_2)	Junior high school	-0.059	0.128	0.647
	Senior high school	-0.535	0.119	0.000
	Diploma	-0.202	1.275	0.874
	Undergraduate	-1.193	0.423	0.005
	Postgraduate	-0.786	3.566	0.825
Has a health complaint (X_3)	No	0.465	0.197	0.018
PKH housing assistance recipients (X_4)	No	-0.563	0.417	0.177
Health insurance recipient (X_5)	No	-0.288	0.168	0.086
Cons		0.559	0.931	0.548
Cons ($Y = 0$)		-0.611	0.892	
Cons ($Y = 1$)		0.303	0.420	
Cons ($Y = 2$)		0.780	0.315	
Cons ($Y = 3$)		1.308	0.244	

From the results in Table 4, several model equations of opportunity were constructed as in (18).

$$P(y) = \begin{cases} P(y=0) = [1-\Phi(B)] + \Phi(B)\Phi(-0.611 - A) \\ P(y=1) = \Phi(B)[\Phi(0.303 - A) - \Phi(-0.611 - A)] \\ P(y=2) = \Phi(B)[\Phi(0.780 - A) - \Phi(0.303 - A)] \\ P(y=3) = \Phi(B)[\Phi(1.308 - A) - \Phi(0.780 - A)] \\ P(y=4) = \Phi(B)[1 - \Phi(1.308 - A)] \end{cases} \tag{18}$$

where A is the OP function on ZIOP shown in (19).

$$A = -0.212Z_{1.2} - 0.204Z_{1.3} - 1.546Z_{1.4} - 1.009Z_{1.5} - 2.319Z_{1.6} \tag{19}$$

The binary probit function in ZIOP is denoted by B shown in (20).

$$\begin{aligned} B = 0.559 + 0.503X_{1.2} - 0.056X_{2.2} - 0.535X_{2.3} - 0.202X_{2.4} - 1.193X_{2.5} - 0.786X_{2.6} \\ + 0.465X_{3.2} - 0.563X_{4.2} - 0.288X_{5.2} \end{aligned} \tag{20}$$

The formulation of the probit regression model is aligned with the number of categories present in the response variable (y). In this context, the five equations represent the ZIOP model structure corresponding to each category of the response variable, where $y = 0$ denotes the lowest category and $y = 4$ indicates the highest.

The interpretation of the ZIOP model commenced with identifying statistically significant variables. For example, higher education levels were associated with lower probabilities of being classified as poor, with postgraduate education showing a relatively large (although statistically insignificant) negative coefficient. This suggested a strong protective effect of higher education against poverty, albeit with limited sample size for that subgroup. Likewise, within the inflation component, attainment of an undergraduate degree notably decreased the likelihood of a household belonging to the excess-zero group, suggesting that these households had a minimal chance of falling into the most impoverished category.

3.2.1. Simultaneous Testing of the ZIOP Model

Simultaneous parameter hypothesis testing is conducted to determine whether the predictor variables collectively have a statistically significant effect on the response variable. The hypotheses for this test are formulated as follows:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ or $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$

H_1: At least one $\beta_u \neq 0$ and $\gamma_t \neq 0$, where $u = 1,2, \dots, q$ and $t = 1,2, \dots, p$

The likelihood ratio test results are derived based on the following calculation procedure.

$$G^2 = 2\left[\ln L(\widehat{\mathbf{\Omega}}) - \ln L(\widehat{\boldsymbol{\omega}})\right] = 15.44$$

The likelihood ratio test yielded a G^2 statistics of 15.44, computed using Stata 16. Based on the test results, the associated p-value was 0.000. Given that this p-value is less than the specified significance level of $\alpha = 0.10$, the null hypothesis is rejected. This finding implies that, collectively, at least one predictor variable within the model has a statistically significant influence on the classification of household poverty levels. Accordingly, the model is deemed appropriate for subsequent analysis.

3.2.2. Partial Testing of the ZIOP Model

Partial hypothesis testing aims to evaluate the significance of each predictor variable individually in influencing the response variable. The hypotheses for each parameter are formulated as follows:

$H_0: \beta_u = 0$ or $\gamma_t = 0$
$H_1: \beta_u \neq 0$ or $\gamma_t \neq 0$

where $u = 1,2, \dots, q$ and $t = 1,2, \dots, p$, representing the parameters in the ordinal and binary components of the ZIOP model, respectively.

The estimation results of the ZIOP model using Stata 16, as presented in Table 4, provide coefficient estimates, standard errors, and p-values for each parameter. Based on these results, several predictor variables were identified as statistically significant. In the ordinal component (poverty levels), the variable highest education (X_2), specifically the diploma category was statistically significant at the 5% level (p-value = 0.011), indicating that higher educational attainment significantly reduces the probability of a household falling into deeper poverty levels.

Meanwhile, in the binary component of the model, which captures the zero-inflation process, three predictors show significant effects. Highest education (X_2), in both the senior high school (p-value = 0.000) and undergraduate (p-value = 0.005) categories, significantly decreased the probability of a household belonging to the structurally nonpoor group. The presence of health complaint (X_3) (p-value = 0.018) positively influenced the likelihood of being in the nonpoor category, possibly reflecting the impact of health vulnerability on poverty classification. Furthermore, the health insurance recipient (X_5) demonstrated a marginal level of significance (p-value = 0.086), implying a potential protective role of health insurance coverage against structural poverty.

These findings highlight that education, health status, and access to social protection mechanisms play a critical role in distinguishing household poverty status. The significant variables can inform targeted interventions to reduce poverty by focusing on these key household characteristics.

3.2.3. Vuong Test

The Vuong test was used to determine which model performed better: the ZIOP or the conventional probit ordinal. The hypothesis are as follows:

H_0: The ZIOP model is equivalent in quality to the OP model.
H_1: The ZIOP model is different from the OP model

The results of the Vuong statistics are calculated as follows.

p-ISSN 2798-253X
e-ISSN 2798-3153

$$v = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}} = 4.56$$

The outcome is H_0 because the value $v > Z_{\alpha/2}$ (4.56 > 1.64), demonstrating that the ZIOP model surpasses the OP model.

3.2.4. Marginal Effect

Marginal effects quantify the impact of variations in a predictor variable on the response variable, assuming all other factors are held constant. This includes the marginal effect of the primary business sector variable of the head of household, illustrated with a household example, head of household with unable to read/write/alphabet letters ($X_{1.2} = 1$), last education elementary school ($X_{2.2} = 0$, $X_{2.3} = 0$, $X_{2.4} = 0$, $X_{2.5} = 0$, $X_{2.6} = 0$), have no health complaints ($X_{3.2} = 1$), Not a recipient of the Family Hope Program (Program Keluarga Harapan, PKH) housing assistance ($X_{4.2} = 1$), and not a recipient of health insurance ($X_{5.2} = 1$). It is important to note that the variable highest education, while conceptually the same across both model components, is represented by different symbols in the ZIOP framework: X_2 is used in the binary probit component and Z_1 in the OP component.

The marginal effect for the highest education variable, which is a categorical variable, was calculated from the difference in probability values across each education category and each response category. Based on the previously derived probability model equations, the following are the calculated probability values for the elementary school category, while the remaining results can be found in Table 5.

$$A = -0.212(0) - 0.204(0) - 1.546(0) - 1.009(0) - 2.319(0) = 0$$

$$B = 0.559 + 0.503(1) - 0.056(0) - 0.535(0) - 0.202(0) - 1.193(0) - 0.786(0) + 0.465(1) - 0.563(1) - 0.288(1) = 0.676$$

The probability for each category is as follows.

$$P(y = 0) = [1 - \Phi(0.676)] + \Phi(0.676)\Phi(-0.611 - 0) = 0.453$$

$$P(y = 1) = \Phi(0.676)[\Phi(0.303 - 0) - \Phi(-0.611 - 0)] = 0.262$$

$$P(y = 2) = \Phi(0.676)[\Phi(0.780 - 0) - \Phi(0.303 - 0)] = 0.123$$

$$P(y = 3) = \Phi(0.676)[\Phi(1.308 - 0) - \Phi(0.780 - 0)] = 0.092$$

$$P(y = 4) = \Phi(0.676)[1 - \Phi(1.308 - 0)] = 0.072$$

The results obtained in Table 5 indicate that households where the head of the household has a postgraduate education have a probability of 0.527 (52.7%) of being classified as nonpoor, compared to households where the head of the household has only completed elementary school. This finding suggests that as the education level of the head of the household increases, the likelihood of being classified as poor or extremely poor decreases. Conversely, the probability of being classified as nonpoor increases with higher levels of education. This demonstrates a positive relationship between educational attainment and economic status, underscoring the crucial role of access to education in reducing poverty levels within society.

Table 5. Probability Value of Highest Education Variable

Highest Education	$P(y=0)$	$P(y=1)$	$P(y=2)$	$P(y=3)$	$P(y=4)$
Elementary school	0.453	0.262	0.123	0.092	0.072
Junior high school	0.520	0.258	0.104	0.070	0.047
Senior high school	0.634	0.196	0.080	0.054	0.036
Diploma	0.881	0.097	0.015	0.005	0.002
Undergraduate	0.896	0.076	0.017	0.008	0.003
Postgraduate	0.980	0.018	0.002	0.000	0.000

p-ISSN 2798-253X
e-ISSN 2798-3153

Highest Education	$P(y=0)$	$P(y=1)$	$P(y=2)$	$P(y=3)$	$P(y=4)$
The difference (junior high school- elementary school)	0.067	-0.004	-0.019	-0.022	-0.025
The difference (senior high school- elementary school)	0.181	-0.066	-0.043	-0.038	-0.036
The difference (diploma- elementary school)	0.428	-0.165	-0.108	-0.087	-0.070
The difference (undergraduate- elementary school)	0.443	-0.186	-0.106	-0.084	-0.069
The difference (postgraduate- elementary school)	0.527	-0.244	-0.121	-0.092	-0.072

4. Conclusion

The ZIOP model provides a more nuanced and effective analytical framework for classifying household poverty levels, particularly in settings characterized by excess zero values and ordinal responses. The results clearly establish that this model outperforms the conventional OP model, making it more suitable for addressing the complexities of poverty data. This study emphasizes that household poverty is a multidimensional issue shaped by distinct socioeconomic characteristics. Among these, educational attainment stands out as the most influential factor in determining both the likelihood and severity of poverty. Additionally, health-related conditions and access to social protection, especially health insurance, play a significant role in shaping household vulnerability. These findings demonstrate that policy responses to poverty must go beyond income metrics and instead target structural disadvantages related to education, health, and welfare access. The ZIOP model thus proves to be not only statistically appropriate but also practically valuable in informing poverty alleviation strategies that are both targeted and evidence driven.

References

[1] M.N. Harris and X. Zhao, “A zero-inflated ordered probit model, with an application to modelling tobacco consumption,” *J. Econom.*, vol. 141, no. 2, pp. 1073–1099, Dec. 2007, doi: 10.1016/j.jeconom.2007.01.002.

[2] N. Rejeki, V. Ratnasari, and M. Ahsan, “Modelling of poor household in East Kalimantan using zero inflated ordered probit (ZIOP) approach,” *Procedia Comput. Sci.*, vol. 234, 2024, pp. 278–285, doi: 10.1016/j.procs.2024.03.002.

[3] B.E. Bagozzi, D.W. Hill, W.H. Moore, and B. Mukherjee, “Modeling two types of peace: The zero-inflated ordered probit (ZIOP) model in conflict research,” *J. Confl. Resolut.*, vol. 59, no. 4, pp. 728–752, Jun. 2015, doi: 10.1177/0022002713520530.

[4] S. Alkire, F. Kövesdi, E. Scheja, and F. Vollmer, “Moderate multidimensional poverty index: Paving the way out of poverty,” *Soc. Indic. Res.*, vol. 168, pp. 409–445, Aug. 2023, doi: 10.1007/s11205-023-03134-5.

[5] Badan Pusat Statistik, “Profil kemiskinan D.I. Yogyakarta Maret 2024,” 2024. [Online]. Available: https://yogyakarta.bps.go.id

[6] Badan Pusat Statistik, “Perhitungan dan analisis kemiskinan makro Indonesia.” 2021. [Online]. Available: https://www.bps.go.id/id/publication/2021/11/30/9c24f43365d1e41c8619dfe4/penghitungan-dan-analisis-kemiskinan-makro-indonesia-tahun-2021.html

[7] UNDP (United Nations Development Programme), “Global multidimensional poverty index 2023: unstacking global poverty: Data for high impact action,” 2023. [Online]. Available: https://hdr.undp.org/system/files/documents/hdp-document/2023mpireporten.pdf

[8] J. Wu, W. Fan, and W. Wang, “A zero-inflated ordered probit model to analyze hazmat truck drivers’ violation behavior and associated risk factors,” *IEEE Access*, vol. 8, pp. 110974–110985, 2020, doi: 10.1109/ACCESS.2020.3001165.

[9] P. Downward, F. Lera-Lopez, and S. Rasciute, “The zero-inflated ordered probit approach to modelling sports participation,” *Econ. Model.*, vol. 28, no. 6, pp. 2469–2477, Nov. 2011, doi: 10.1016/j.econmod.2011.06.024.

[10] H. Wang, Z. Liu, X. Wang, D. Huang, L. Cao, and J. Wang, “Analysis of the injury-severity outcomes of maritime accidents using a zero-inflated ordered probit model,” *Ocean Eng.*, vol. 258, Aug. 2022, doi: 10.1016/j.oceaneng.2022.111796.

[11] C. Xu, S. Xu, C. Wang, and J. Li, “Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression,” *Physica A*, vol. 524, pp. 121–129, Jun. 2019, doi: 10.1016/j.physa.2019.03.036.

[12] X. Jiang, B. Huang, R.L. Zaretzki, S. Richards, X. Yan, and H. Zhang, “Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models,” *Accid. Anal. Prev.*, vol. 57, pp. 55–66, Aug. 2013, Art. no 23628942, doi: 10.1016/j.aap.2013.03.018.

[13] S.R. Ajija, D.W. Sari, R. Setianto, and M. Primanthi, *Cara Cerdas Menguasai Eviews*. Jakarta, Indonesia: Salemba Empat, 2011.

[14] I. Ghozali, *Aplikasi Analisis Multivariate SPSS 23*. Semarang, Indonesia: Badan Penerbit Universitas Diponegoro, 2016.

[15] T.A. Wicaksono, “Determinan Pemekaran Wilayah di Indonesia: Study Kasus Kabupaten/Kota 2001-2004,” Undergraduate thesis, Fak. Ekon. Bisnis, Univ. Indonesia, Jawa Barat, Indonesia, 2008.

[16] W.H. Greene, *Econometric Analysis*, 5th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2003.

[17] Q.H. Vuong, “Likelihood ratio tests for model selection and non-nested hypotheses,” *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989, doi: 10.2307/1912557.