Article

A Logit Model for Bivariate Binary Responses

Purhadi Purhadi [1] and M. Fathurahman [2],*

1 Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia; purhadi@statistika.its.ac.id
2 Department of Statistics, Mulawarman University, Samarinda 75123, Indonesia
* Correspondence: fathur@fmipa.unmul.ac.id

Abstract: This article provides a bivariate binary logit model and statistical inference procedures for parameter estimation and hypothesis testing. The bivariate binary logit (BBL) model is an extension of the binary logit model that has two correlated binary responses. The BBL model responses were formed using a 2×2 contingency table, which follows a multinomial distribution. The maximum likelihood and Berndt–Hall–Hall–Hausman (BHHH) methods were used to obtain the BBL model. Hypothesis testing of the BBL model contains the simultaneous test and the partial test. The test statistics of the simultaneous test and the partial test were determined using the maximum likelihood ratio test method. The likelihood ratio statistics of the simultaneous test and the partial test were approximately asymptotically chi-square distributed with $3p$ degrees of freedom. The BBL model was applied to a real dataset, and the BBL model with the single covariate was better than the BBL model with multiple covariates.

Keywords: logit model; bivariate binary responses; maximum likelihood; BHHH; maximum likelihood ratio test

1. Introduction

The logit model is a model that is often used for modeling categorical data in various research fields. Several studies have recently developed logit models for multiple correlated responses. McCullagh and Nelder [1] introduced a multivariate logistic transform used to construct logit models with two or more correlated responses. A multivariate logistic transform, including the numerical optimization methods, has been proposed in [2–6]. Lipsitz, Laird, and Harrington [7] examined the maximum likelihood (ML) method for binary data models, which connects the probability of success at each time point to a set of covariates. Liang, Zeger, and Qaqish [8] discussed the regression modeling of the marginal means of the responses using the generalized estimating equation approach, wherein there are dependencies between responses. An alternating logit model for jointly regressing the responses on covariates and modeling the dependencies among responses in the framework of pairwise odds ratios was proposed by [9]. Cessie and Houwelingen [10] modeled regression for correlated binary responses, in which the form of marginal response probabilities is the logit link function. Lang and Agresti [11] considered the model-fitting methods for analyzing the parameters simultaneously and parsimoniously. The ML estimator properties of the kappa coefficient in the bivariate binary logistic model using the small and moderate sized samples through Monte Carlo simulation were investigated by [12,13]. Molenberghs and Lesaffre [14] presented a simple generalized linear model formulation for marginal and association modeling of multivariate categorical data. The specifications of the association models in [15], in which the dependence ratios contrast with other models for a multivariate binary response that is specified by odds ratios or correlation coefficients, were employed by [16].

Studies have also proposed both the conditional and marginal models. A flexible conditional model for a multivariate binary response vector with covariates was examined

by [17]. Islam, Chowdhury, and Briollais [18] developed a new simple procedure to construct the conditional and marginal models for the bivariate binary responses. The marginal and conditional probabilities of the responses were expressed as functions of covariates. El-Sayed, Islam, and Alzaid [19] provided the estimation and test procedures for association measures in the correlated binary data. A generalized approach using both the conditional and marginal models was demonstrated by [20,21]. The responses of the model have a bivariate Bernoulli distribution. The ML and Newton–Raphson methods were used to estimate the model's parameters, whereas the likelihood ratio test method was used to test the parameters' significance. The properties of the ML estimators of the regression parameters have also been investigated.

Other models have also been developed. Sinha, Laird, and Fitzmaurice [22] extended the univariate logit model in [23] to the case of a logit model for multivariate-correlated responses with missing covariates and observed auxiliary information. A robust model for misclassified correlated binary responses was described by [24]. O'Brien and Dunson [25] provided an exact Bayesian analysis of a marginal logistic model. The multivariate logistic regression model in a framework of the geographically-weighted regression was proposed by [26,27].

Corresponding to the previous studies, in this study we constructed a logit model, namely, the bivariate binary logit (BBL) model, which has two correlated binary responses. Following [1,2], the BBL model's responses follow a multinomial distribution. Therefore, the ML method can be used to estimate the BBL model's parameters. The ML estimator is not closed-form, and it needs an iterative procedure using a numerical optimization method. We used the Berndt–Hall–Hall–Hausman (BHHH) iterative method [28]. However, the BHHH method has not been used in previous studies. On the other hand, the BHHH method can be used as an alternative to the numerical optimization method when the elements of the Hessian matrix are unavailable. Following [29], the maximum likelihood ratio test (MLRT) method was used to test the significance of parameters both simultaneously and partially. The performance of the BBL model was evaluated using an empirical study.

This article is organized as follows. In Section 2, we describe the BBL model specifically. Section 3 investigates the estimation of the BBL model's parameters using the ML and BHHH methods. Hypothesis testing of the BBL model is discussed in Section 4. Section 5 demonstrates an application of the BBL model to real data. The conclusions are given in Section 6.

2. Bivariate Binary Logit Model

Bivariate binary logit (BBL) models are one of the families of multivariate logit models and are used to model the relationships between two correlated binary responses with one or more covariates. Let Y_1 and Y_2 be two bivariate binary responses and $\boldsymbol{y} = \begin{bmatrix} Y_{11} & Y_{10} & Y_{01} & Y_{00} \end{bmatrix}^T$ be a vector of responses. The elements of $\boldsymbol{y}$ have the probabilities of γ_{11}, γ_{10}, γ_{01}, and γ_{00}, respectively, which are presented in Table 1.

Table 1. Probabilities for the responses.

Y_1	Y_2		Total
	$Y_2=1$	$Y_2=0$	
$Y_1 = 1$	γ_{11}	γ_{10}	γ_1
$Y_1 = 0$	γ_{01}	γ_{00}	$1 - \gamma_1$
Total	γ_2	$1 - \gamma_2$	1

According to Fathurahman, Purhadi, Sutikno, and Ratnasari [27], the BBL model responses in Table 1 follow a multinomial distribution. Therefore, the joint probability function of the responses can be defined as follows:

$$P(Y_{11} = y_{11}, Y_{10} = y_{10}, Y_{01} = y_{01}, Y_{00} = y_{00}) = \prod_{q=0}^{1} \prod_{r=0}^{1} \gamma_{qr}^{y_{qr}}, \tag{1}$$

where $0 < \gamma_{qr} < 1$; $q, r = 0, 1$; $y_{qr} = 0, 1$; $y_{00} = 1 - y_{11} - y_{10} - y_{01}$; and $\gamma_{00} = 1 - \gamma_{11} - \gamma_{10} - \gamma_{01}$. q and r are the values of the responses. y_{qr} is the value of Y_{qr}, which represents the elements of the vector of responses. $\gamma_{qr} = P(Y_1 = q, Y_2 = r)$ is the joint probability of the responses. $\gamma_1 = P(Y_1 = 1)$ and $\gamma_2 = P(Y_2 = 1)$ are the marginal probabilities of Y_1 and Y_2, respectively.

Let $\boldsymbol{x} = \begin{bmatrix} 1 & X_1 & X_2 & \cdots & X_p \end{bmatrix}^T$ be the vector of covariates, which is $(p + 1)$-dimensional. Then the BBL model is expressed as follows:

$$\begin{aligned}
\tau_1(\boldsymbol{x}) &= \text{logit}[\gamma_1(\boldsymbol{x})] = \log\left[\frac{\gamma_1(\boldsymbol{x})}{1-\gamma_1(\boldsymbol{x})}\right] = \boldsymbol{x}^T\boldsymbol{\theta}_1, \\
\tau_2(\boldsymbol{x}) &= \text{logit}[\gamma_2(\boldsymbol{x})] = \log\left[\frac{\gamma_2(\boldsymbol{x})}{1-\gamma_2(\boldsymbol{x})}\right] = \boldsymbol{x}^T\boldsymbol{\theta}_2, \\
\tau_3(\boldsymbol{x}) &= \log[\psi(\boldsymbol{x})] = \log\left[\frac{\gamma_{11}(\boldsymbol{x})\gamma_{00}(\boldsymbol{x})}{\gamma_{10}(\boldsymbol{x})\gamma_{01}(\boldsymbol{x})}\right] = \boldsymbol{x}^T\boldsymbol{\theta}_3,
\end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$ are vectors of parameters, $\gamma_1(\boldsymbol{x})$ and $\gamma_2(\boldsymbol{x})$ are marginal probabilities of responses, and $\psi(\boldsymbol{x})$ is the odds ratio of responses depending on covariates, which shows that the responses are correlated.

The vectors of parameters are symbolized by

$$\begin{aligned}
\boldsymbol{\theta}_1 &= \begin{bmatrix} \theta_{01} & \theta_{11} & \theta_{21} & \cdots & \theta_{p1} \end{bmatrix}^T, \\
\boldsymbol{\theta}_2 &= \begin{bmatrix} \theta_{02} & \theta_{12} & \theta_{22} & \cdots & \theta_{p2} \end{bmatrix}^T, \\
\boldsymbol{\theta}_3 &= \begin{bmatrix} \theta_{03} & \theta_{13} & \theta_{23} & \cdots & \theta_{p3} \end{bmatrix}^T.
\end{aligned} \tag{3}$$

The marginal probabilities of responses are defined as follows:

$$\begin{aligned}
P(Y_1 = 1 | \boldsymbol{x}) &= \gamma_1(\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^T\boldsymbol{\theta}_1)}{1+\exp(\boldsymbol{x}^T\boldsymbol{\theta}_1)}, \\
P(Y_2 = 1 | \boldsymbol{x}) &= \gamma_2(\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^T\boldsymbol{\theta}_2)}{1+\exp(\boldsymbol{x}^T\boldsymbol{\theta}_2)}.
\end{aligned} \tag{4}$$

The joint probability of $\gamma_{11}(\boldsymbol{x})$ in Equation (2) is defined by

$$P(Y_1 = 1, Y_2 = 1 | \boldsymbol{x}) = \gamma_{11}(\boldsymbol{x}) = \begin{cases} \frac{1}{2}(\psi(\boldsymbol{x})-1)^{-1}\left(a - \sqrt{a^2 + b}\right), & \psi(\boldsymbol{x}) \neq 1 \\ \gamma_1(\boldsymbol{x})\gamma_2(\boldsymbol{x}), & \psi(\boldsymbol{x}) \neq 1 \end{cases}, \tag{5}$$

where $a = 1 + (\gamma_1(\boldsymbol{x}) + \gamma_2(\boldsymbol{x}))(\psi(\boldsymbol{x}) - 1)$ and $b = 4\psi(\boldsymbol{x})(1 - \psi(\boldsymbol{x}))\gamma_1(\boldsymbol{x})\gamma_2(\boldsymbol{x})$. If $\psi(\boldsymbol{x}) = 1$, then the responses are independent [30].

Based on Table 1 and Equation (5), the probabilities of $\gamma_{10}(\boldsymbol{x})$, $\gamma_{01}(\boldsymbol{x})$, and $\gamma_{00}(\boldsymbol{x})$ in Equation (2) are as follows:

$$\begin{aligned}
\gamma_{10}(\boldsymbol{x}) &= \gamma_1(\boldsymbol{x}) - \gamma_{11}(\boldsymbol{x}), \\
\gamma_{01}(\boldsymbol{x}) &= \gamma_2(\boldsymbol{x}) - \gamma_{11}(\boldsymbol{x}), \\
\gamma_{00}(\boldsymbol{x}) &= 1 - \gamma_{11}(\boldsymbol{x}) - \gamma_{10}(\boldsymbol{x}) - \gamma_{01}(\boldsymbol{x}) = 1 - \gamma_1(\boldsymbol{x}) - \gamma_2(\boldsymbol{x}) + \gamma_{11}(\boldsymbol{x}).
\end{aligned} \tag{6}$$

3. Estimation of the BBL Model

The estimation of the BBL model's parameters is one of the main results of this study. The BBL model in Equation (2) has $3(p + 1)$ parameters, where $(p + 1)$ parameters show

the dependencies among responses, and $2(p+1)$ parameters describe the relationships between responses and covariates. The BBL model's parameters are denoted by $\boldsymbol{\theta}$ and expressed as

$$\boldsymbol{\theta} = \left[\begin{array}{ccc} \boldsymbol{\theta}_1^T & \boldsymbol{\theta}_2^T & \boldsymbol{\theta}_3^T \end{array}\right]^T, \tag{7}$$

where $\boldsymbol{\theta}_1^T$, $\boldsymbol{\theta}_2^T$, and $\boldsymbol{\theta}_3^T$ are given by Equation (3).

To obtain the parameters estimator of the BBL model in Equation (7), the ML method was employed. Based on the ML method, the estimator of $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$, maximized by the likelihood function and the log-likelihood function. The ML estimator can be obtained by determining the first partial derivatives of the log-likelihood function, then equating them to zero.

Based on Equation (2), the likelihood equation contains the interdependence equations, which have a non-explicit form. Therefore, the ML estimator of the BBL model's parameters was not obtained analytically. The ML estimator was approximated by the likelihood equation's roots, which were obtained via an iterative process using the BHHH method. Determining the ML estimator of the BBL model's parameters using the BHHH method needs the gradient vector and the Hessian matrix. In the following, we present Lemmas 1 and 2 for the gradient vector and Hessian matrix, respectively.

Lemma 1. *Let* $\boldsymbol{y}_i = \left[\begin{array}{cc} Y_{1i} & Y_{2i} \end{array}\right]^T = \left[\begin{array}{cccc} Y_{11i} & Y_{10i} & Y_{01i} & Y_{00i} \end{array}\right]^T$, $i = 1, 2, \dots, n$ *be a random vector sample that is mutually independent and identical with a multinomial distribution denoted by* $\boldsymbol{y}_i \sim MULT(1; \gamma_{11}(\boldsymbol{x}_i), \gamma_{10}(\boldsymbol{x}_i), \gamma_{01}(\boldsymbol{x}_i), \gamma_{00}(\boldsymbol{x}_i))$, *where* $\gamma_{11}(\boldsymbol{x}_i)$, $\gamma_{10}(\boldsymbol{x}_i)$, $\gamma_{01}(\boldsymbol{x}_i)$, *and* $\gamma_{00}(\boldsymbol{x}_i)$ *are probabilities of the random variables of* Y_{11i}, Y_{10i}, Y_{01i}, *and* Y_{00i} *that contain the parameter* $\boldsymbol{\theta}$. *If the likelihood function of the BBL model is denoted by* $\mathcal{L}(\boldsymbol{\theta})$, *where* $\boldsymbol{\theta}$ *is as in Equation (7), then the gradient vector is*

$$\boldsymbol{g}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1}\right]^T & \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right]^T & \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_3}\right]^T \end{array}\right]^T, \tag{8}$$

where

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} = \sum_{i=1}^{n} \frac{1}{\Delta_{1i}} \boldsymbol{x}_i \left[y_{11i}\left(\frac{\gamma_{01i}}{\gamma_{2i}}\right) + y_{10i}\left(\frac{\gamma_{00i}}{1-\gamma_{2i}}\right) - y_{01i}\left(\frac{\gamma_{11i}}{\gamma_{2i}}\right) - y_{00i}\left(\frac{\gamma_{10i}}{1-\gamma_{2i}}\right) \right],$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} = \sum_{i=1}^{n} \frac{1}{\Delta_{1i}} \boldsymbol{x}_i \left[y_{11i}\left(\frac{\gamma_{10i}}{\gamma_{1i}}\right) - y_{10i}\left(\frac{\gamma_{11i}}{\gamma_{1i}}\right) + y_{01i}\left(\frac{\gamma_{00i}}{1-\gamma_{1i}}\right) - y_{00i}\left(\frac{\gamma_{01i}}{1-\gamma_{1i}}\right) \right],$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_3} = \sum_{i=1}^{n} \Delta_{2i} \boldsymbol{x}_i \left[\frac{y_{11i}}{\gamma_{11i}} - \frac{y_{10i}}{\gamma_{10i}} - \frac{y_{01i}}{\gamma_{01i}} + \frac{y_{00i}}{\gamma_{00i}} \right],$$

$$\Delta_{1i} = \frac{\gamma_{11i}\gamma_{10i}\gamma_{01i}\gamma_{00i}}{\gamma_{1i}(1-\gamma_{1i})\gamma_{2i}(1-\gamma_{2i})\Delta_{2i}},$$

$$\Delta_{2i} = \left(\frac{1}{\gamma_{11i}} + \frac{1}{\gamma_{10i}} + \frac{1}{\gamma_{01i}} + \frac{1}{\gamma_{00i}} \right)^{-1}.$$

Proof of Lemma 1. Suppose that $(Y_{1i}, Y_{2i}) = (Y_{11i}, Y_{10i}, Y_{01i}, Y_{00i})$ is a vector of the random sample that is independently and identically multinomial distributed; then the joint probability is defined by

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{\theta}) &= P(Y_{11i} = y_{11i}, Y_{10i} = y_{10i}, Y_{01i} = y_{01i}, Y_{00i} = y_{00i}) \\ &= \gamma_{11i}^{y_{11i}}(\boldsymbol{x}_i)\gamma_{10i}^{y_{10i}}(\boldsymbol{x}_i)\gamma_{01i}^{y_{01i}}(\boldsymbol{x}_i)\gamma_{00i}^{y_{00i}}(\boldsymbol{x}_i). \end{aligned} \tag{9}$$

As in Equation (9), the likelihood function is as follows:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) &= \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} P(Y_{11i} = y_{11i}, Y_{10i} = y_{10i}, Y_{01i} = y_{01i}, Y_{00i} = y_{00i}) \\
&= \prod_{i=1}^{n} \gamma_{11i}^{y_{11i}}(\boldsymbol{x}_i) \gamma_{10i}^{y_{10i}}(\boldsymbol{x}_i) \gamma_{01i}^{y_{01i}}(\boldsymbol{x}_i) \gamma_{00i}^{y_{00i}}(\boldsymbol{x}_i).
\end{aligned}
\tag{10}
$$

For simplicity, let $\gamma_{qr}^{y_{qr}}(\boldsymbol{x}_i) = \gamma_{qri}^{y_{qri}}$ for $q, r = 0, 1$; then the likelihood function in Equation (10) can be rewritten as

$$
\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left(\gamma_{11i}^{y_{11i}} \gamma_{10i}^{y_{10i}} \gamma_{01i}^{y_{01i}} \gamma_{00i}^{y_{00i}} \right).
\tag{11}
$$

To obtain the log-likelihood function of the BBL model, both sides of the likelihood function in Equation (11) were transformed by the natural logarithm, which gives

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) &= \log \mathcal{L}(\boldsymbol{\theta}) \\
&= \sum_{i=1}^{n} (y_{11i} \log \gamma_{11i} + y_{10i} \log \gamma_{10i} + y_{01i} \log \gamma_{01i} + y_{00i} \log \gamma_{00i}).
\end{aligned}
\tag{12}
$$

The log-likelihood function in Equation (12) is that the vector of $\boldsymbol{\theta}$ has $3(p+1)-$ dimensions. Following the definition in Greene [31], the gradient vector of the log-likelihood function in Equation (12) is

$$
\boldsymbol{g}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1}\right]^T & \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2}\right]^T & \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_3}\right]^T \end{array} \right]^T,
\tag{13}
$$

where the vector of $\boldsymbol{\theta}$ is given by Equation (7).

Regarding the BBL model in Equation (2), we define the vector of $\boldsymbol{\tau}$, which is denoted by $\boldsymbol{\tau} = \left[\begin{array}{ccc} \tau_1 & \tau_2 & \tau_3 \end{array} \right]^T$, where $\tau_1 = \tau_1(\boldsymbol{x})$, $\tau_2 = \tau_2(\boldsymbol{x})$, and $\tau_3 = \tau_3(\boldsymbol{x})$. The vector of the joint probability of $\boldsymbol{y}$ is defined by $\boldsymbol{\gamma} = \left[\begin{array}{cccc} \gamma_{11} & \gamma_{10} & \gamma_{01} & \gamma_{00} \end{array} \right]^T$. Furthermore, the derivative of $\boldsymbol{\tau}$ with respect to $\boldsymbol{\gamma}$ is denoted by $\partial \boldsymbol{\tau}/\partial \boldsymbol{\gamma}$. To get a symmetrical matrix of $\partial \boldsymbol{\tau}/\partial \boldsymbol{\gamma}$, suppose $\tau_0 = \log \gamma_{..}$ with $\gamma_{..} = \sum_{q=0}^{1} \sum_{r=0}^{1} \gamma_{qr}$; then the vector of $\boldsymbol{\tau}$ is $\boldsymbol{\tau} = \left[\begin{array}{cccc} \tau_0 & \tau_1 & \tau_2 & \tau_3 \end{array} \right]^T$. Thus, the matrix of $\partial \boldsymbol{\tau}/\partial \boldsymbol{\gamma}$ is

$$
\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\gamma}} = \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \frac{1}{\gamma_1} & \frac{1}{\gamma_1} & -\frac{1}{1-\gamma_1} & -\frac{1}{1-\gamma_1} \\ \frac{1}{\gamma_2} & -\frac{1}{1-\gamma_2} & \frac{1}{\gamma_2} & -\frac{1}{1-\gamma_2} \\ \frac{1}{\gamma_{11}} & -\frac{1}{\gamma_{10}} & -\frac{1}{\gamma_{01}} & \frac{1}{\gamma_{00}} \end{array} \right].
\tag{14}
$$

The inverse matrix of $\partial \boldsymbol{\tau}/\partial \boldsymbol{\gamma}$ in Equation (14) is as follows:

$$
\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\gamma}}^{-1} = \left[\begin{array}{cccc} \gamma_{11} & \frac{\gamma_{11}\gamma_{01}}{\gamma_2 \Delta_1} & \frac{\gamma_{11}\gamma_{10}}{\gamma_1 \Delta_1} & \Delta_2 \\ \gamma_{10} & \frac{\gamma_{10}\gamma_{00}}{(1-\gamma_2)\Delta_1} & -\frac{\gamma_{11}\gamma_{10}}{\gamma_1 \Delta_1} & -\Delta_2 \\ \gamma_{01} & -\frac{\gamma_{11}\gamma_{01}}{\gamma_2 \Delta_1} & \frac{\gamma_{01}\gamma_{00}}{(1-\gamma_1)\Delta_1} & -\Delta_2 \\ \gamma_{00} & -\frac{\gamma_{10}\gamma_{00}}{(1-\gamma_2)\Delta_1} & \frac{\gamma_{01}\gamma_{00}}{(1-\gamma_1)\Delta_1} & \Delta_2 \end{array} \right],
\tag{15}
$$

where

$$
\Delta_1 = \frac{\gamma_{11}\gamma_{10}\gamma_{01}\gamma_{00}}{\gamma_1(1-\gamma_1)\gamma_2(1-\gamma_2)\Delta_2}
$$

and

$$\Delta_2 = \left(\frac{1}{\gamma_{11}} + \frac{1}{\gamma_{10}} + \frac{1}{\gamma_{01}} + \frac{1}{\gamma_{00}} \right)^{-1}.$$

The gradient vector of the log-likelihood function in Equation (12) can be written as

$$g(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{16}$$

In relation to Equations (13)–(15) and the chain rule of derivatives, the elements of the gradient vector in Equation (16) can be obtained as follows:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} &= \sum_{i=1}^{n} \left(\frac{y_{11i}}{\gamma_{11i}} \left(\frac{\partial \gamma_{11i}}{\partial \boldsymbol{\theta}_1} \right) + \frac{y_{10i}}{\gamma_{10i}} \left(\frac{\partial \gamma_{10i}}{\partial \boldsymbol{\theta}_1} \right) + \frac{y_{01i}}{\gamma_{01i}} \left(\frac{\partial \gamma_{01i}}{\partial \boldsymbol{\theta}_1} \right) + \frac{y_{00i}}{\gamma_{00i}} \left(\frac{\partial \gamma_{00i}}{\partial \boldsymbol{\theta}_1} \right) \right) \\ &= \sum_{i=1}^{n} \frac{1}{\Delta_{1i}} x_i \left(y_{11i} \left(\frac{\gamma_{01i}}{\gamma_{2i}} \right) + y_{10i} \left(\frac{\gamma_{00i}}{1 - \gamma_{2i}} \right) - y_{01i} \left(\frac{\gamma_{11i}}{\gamma_{2i}} \right) - y_{00i} \left(\frac{\gamma_{10i}}{1 - \gamma_{2i}} \right) \right); \end{aligned} \tag{17}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} &= \sum_{i=1}^{n} \left(\frac{y_{11i}}{\gamma_{11i}} \left(\frac{\partial \gamma_{11i}}{\partial \boldsymbol{\theta}_2} \right) + \frac{y_{10i}}{\gamma_{10i}} \left(\frac{\partial \gamma_{10i}}{\partial \boldsymbol{\theta}_2} \right) + \frac{y_{01i}}{\gamma_{01i}} \left(\frac{\partial \gamma_{01i}}{\partial \boldsymbol{\theta}_2} \right) + \frac{y_{00i}}{\gamma_{00i}} \left(\frac{\partial \gamma_{00i}}{\partial \boldsymbol{\theta}_2} \right) \right) \\ &= \sum_{i=1}^{n} \frac{1}{\Delta_{1i}} x_i \left(y_{11i} \left(\frac{\gamma_{10i}}{\gamma_{1i}} \right) - y_{10i} \left(\frac{\gamma_{11i}}{\gamma_{1i}} \right) + y_{01i} \left(\frac{\gamma_{00i}}{1 - \gamma_{1i}} \right) - y_{00i} \left(\frac{\gamma_{01i}}{1 - \gamma_{1i}} \right) \right); \end{aligned} \tag{18}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_3} &= \sum_{i=1}^{n} \left(\frac{y_{11i}}{\gamma_{11i}} \left(\frac{\partial \gamma_{11i}}{\partial \boldsymbol{\theta}_3} \right) + \frac{y_{10i}}{\gamma_{10i}} \left(\frac{\partial \gamma_{10i}}{\partial \boldsymbol{\theta}_3} \right) + \frac{y_{01i}}{\gamma_{01i}} \left(\frac{\partial \gamma_{01i}}{\partial \boldsymbol{\theta}_3} \right) + \frac{y_{00i}}{\gamma_{00i}} \left(\frac{\partial \gamma_{00i}}{\partial \boldsymbol{\theta}_3} \right) \right) \\ &= \sum_{i=1}^{n} \Delta_{2i} x_i \left(\frac{y_{11i}}{\gamma_{11i}} - \frac{y_{10i}}{\gamma_{10i}} - \frac{y_{01i}}{\gamma_{01i}} + \frac{y_{00i}}{\gamma_{00i}} \right), \end{aligned} \tag{19}$$

where Δ_{1i} and Δ_{2i}, for $i = 1, 2, \ldots, n$, given in Equation (15). □

Lemma 2. *If the log-likelihood function of the BBL model is $\ell(\boldsymbol{\theta})$ and the vector of $\boldsymbol{\theta}$ is the BBL model's parameters, then the Hessian matrix of $\ell(\boldsymbol{\theta})$ is*

$$H(\boldsymbol{\theta}) = -\frac{1}{n} \Big[g^T(\boldsymbol{\theta}) g(\boldsymbol{\theta}) \Big], \tag{20}$$

where n is the sample size.

Proof of Lemma 2. The BBL model's parameters ($\boldsymbol{\theta}$) and the log-likelihood function ($\ell(\boldsymbol{\theta})$) were given in Equations (7) and (12), respectively. Based on Lemma 1, the gradient vector of the log-likelihood function $\ell(\boldsymbol{\theta})$ is $g(\boldsymbol{\theta})$. According to Greene [31], the Hessian matrix can be obtained by the Berndt–Hall–Hall–Hausman (BHHH) method. On the other hand, the Hessian matrix depends on the gradient vector [31], which is shown below:

$$E[g(\boldsymbol{\theta})] = 0, Var[g(\boldsymbol{\theta})] = E\Big[g^T(\boldsymbol{\theta}) g(\boldsymbol{\theta}) \Big]. \tag{21}$$

Meanwhile, the gradient vector and the Hessian matrix associated with the information matrix and can be expressed by

$$I(\boldsymbol{\theta}) = -H(\boldsymbol{\theta}), Var[g(\boldsymbol{\theta})] = nI(\boldsymbol{\theta}). \tag{22}$$

The information matrix in Equation (22) is also referred to as the Fisher information matrix [32]. Based on Equations (21) and (22), the Hessian matrix is

$$H(\boldsymbol{\theta}) = -\frac{1}{n} \Big[g^T(\boldsymbol{\theta}) g(\boldsymbol{\theta}) \Big]. \tag{23}$$

Regarding Lemmas 1 and 2, an iteration process can be carried out using the BHHH method. Following [33], the BHHH algorithm in this study is as follows:

- Determine the initial value for $\hat{\boldsymbol{\theta}}^{(0)} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1^{T(0)} & \hat{\boldsymbol{\theta}}_2^{T(0)} & \hat{\boldsymbol{\theta}}_3^{T(0)} \end{bmatrix}^T$.
- Determine the tolerance value (ε) for the BHHH iteration process stopping.
- Start the BHHH iteration process using the formula:

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \hat{\boldsymbol{\theta}}^{(t)} - \boldsymbol{H}^{-1}\left(\hat{\boldsymbol{\theta}}^{(t)}\right)\boldsymbol{g}\left(\hat{\boldsymbol{\theta}}^{(t)}\right), \ t = 0, 1, 2, \ldots, T. \tag{24}$$

- The iteration stops at the *T-th* iteration if the condition of convergence is satisfied, which is $\hat{\boldsymbol{\theta}}^{(T+1)} - \hat{\boldsymbol{\theta}}^{(T)} \leq \varepsilon$. The estimator values of the parameters are obtained in the last iteration.

Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) determine the best model in this study. The AIC and BIC values can be obtained by

$$\text{AIC} = -2\ell\left(\hat{\boldsymbol{\theta}}\right) + (p+1), \tag{25}$$

$$\text{BIC} = -2\ell\left(\hat{\boldsymbol{\theta}}\right) + \log(n)(p+1), \tag{26}$$

where $\ell\left(\hat{\boldsymbol{\theta}}\right)$ is the log-likelihood value of the parameter's estimate, p is the number of covariates, and n is the sample size. The best model is the BBL model, which has the smallest values of AIC and BIC. $\square$

4. Hypothesis Testing of the BBL Model

Hypothesis testing of the BBL model contains the simultaneous test and the partial test. The simultaneous test and the partial test obtain the significance of the BBL model's parameters jointly and individually, respectively. The simultaneous test and the partial test in this study were done using the maximum likelihood ratio test (MLRT) method.

The hypotheses of the simultaneous test are as follows:

$$\begin{aligned} H_0 &: \theta_{1h} = \theta_{2h} = \cdots = \theta_{ph} = 0, \ \ h = 1, 2, 3 \\ H_1 &: \text{at least one of } \theta_{gh} \neq 0, \ \ g = 1, 2, \ldots, p. \end{aligned} \tag{27}$$

In the following we present a lemma used to determine the likelihood ratio (LR) statistic, the distribution of the LR statistic, and the rejection region of the simultaneous test.

Lemma 3. *If $\boldsymbol{\theta}$ is the BBL model's parameter and $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta}$, then:*

a) *The LR statistic of the simultaneous test is $G_1^2 = 2\left(L\left(\hat{\boldsymbol{\theta}}\right) - L\left(\hat{\boldsymbol{\theta}}^*\right)\right)$, where $\hat{\boldsymbol{\theta}}^*$ is the ML estimator of the parameter space under the null hypothesis and $\hat{\boldsymbol{\theta}}$ is the ML estimator of the parameter space under the population.*

b) *The distribution of the LR statistic follows an asymptotic chi-square distribution, which is $G_1^2 = 2\left(L\left(\hat{\boldsymbol{\theta}}\right) - L\left(\hat{\boldsymbol{\theta}}^*\right)\right) \xrightarrow{d} \chi_{v_1}^2, \ n \to \infty.$*

c) *The rejection region at the significance level of α is $G_1^2 > \chi_{(\alpha, v_1)}^2.$*

Proof of Lemma 3. (a) The first step to obtain the statistical test of the hypothesis in Equation (27) is to define the parameter space under the null hypothesis, denoted by $\Omega_{01} = \{\theta_{01}, \theta_{02}, \theta_{03}\}$. Furthermore, we determine the likelihood function formulated by

$$L(\Omega_{01}) = \prod_{i=1}^{n}\left((\gamma_{11i}^*)^{y_{11i}}(\gamma_{10i}^*)^{y_{10i}}(\gamma_{01i}^*)^{y_{01i}}(\gamma_{00i}^*)^{y_{00i}}\right). \tag{28}$$

Let $\hat{\boldsymbol{\theta}}^* = \begin{bmatrix} \hat{\theta}_{01} & \hat{\theta}_{02} & \hat{\theta}_{03} \end{bmatrix}^T$ be the ML estimator that maximizes the likelihood function of $L(\Omega_{01})$. The ML estimator of $\boldsymbol{\theta}^*$ can be obtained by using Lemmas 1 and 2. Therefore, the maximum likelihood function of $L(\Omega_{01})$ is

$$L(\hat{\Omega}_{01}) = \prod_{i=1}^{n} \left((\hat{\gamma}_{11i}^*)^{y_{11i}} (\hat{\gamma}_{10i}^*)^{y_{10i}} (\hat{\gamma}_{01i}^*)^{y_{01i}} (\hat{\gamma}_{00i}^*)^{y_{00i}} \right). \tag{29}$$

where:

$$\hat{\gamma}_{11i}^* = \begin{cases} \frac{1}{2}(\psi_{1i}-1)^{-1}\left(a_{1i} - \sqrt{a_{1i}^2 + b_{1i}}\right), & \psi_{1i} \neq 1 \\ \hat{\gamma}_{1i}^* \hat{\gamma}_{2i}^*, & \psi_{1i} \neq 1 \end{cases}$$
$$a_{1i} = 1 + (\hat{\gamma}_{1i}^* + \hat{\gamma}_{2i}^*)(\psi_{1i}-1)$$
$$b_{1i} = 4\psi_{1i}(1-\psi_{1i})\hat{\gamma}_{1i}^* \hat{\gamma}_{2i}^*$$
$$\psi_{1i} = \hat{\gamma}_{11i}^* \hat{\gamma}_{00i}^* / (\hat{\gamma}_{10i}^* \hat{\gamma}_{01i}^*)$$
$$\hat{\gamma}_{1i}^* = \exp(\hat{\theta}_{01}) / (1 + \exp(\hat{\theta}_{01}))$$
$$\hat{\gamma}_{2i}^* = \exp(\hat{\theta}_{02}) / (1 + \exp(\hat{\theta}_{02}));$$
$$\hat{\gamma}_{10i}^* = \hat{\gamma}_{1i}^* - \hat{\gamma}_{11i}^*;$$
$$\hat{\gamma}_{01i}^* = \hat{\gamma}_{2i}^* - \hat{\gamma}_{11i}^*;$$
$$\hat{\gamma}_{00i}^* = 1 - \hat{\gamma}_{1i}^* - \hat{\gamma}_{2i}^* + \hat{\gamma}_{11i}^*.$$

The parameters under the population are $\Omega_{11} = \left\{ \theta_{gh}, \ g = 0, 1, \ldots, p; h = 1, 2, 3 \right\}$, and the likelihood function is

$$L(\Omega_{11}) = \prod_{i=1}^{n} \left(\gamma_{11i}^{y_{11i}} \gamma_{10i}^{y_{10i}} \gamma_{01i}^{y_{01i}} \gamma_{00i}^{y_{00i}} \right). \tag{30}$$

Suppose the ML estimator that maximizes the likelihood function of $L(\Omega_{11})$ is $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1^T & \hat{\boldsymbol{\theta}}_2^T & \hat{\boldsymbol{\theta}}_3^T \end{bmatrix}^T$. However, the ML estimator of $\boldsymbol{\theta}$ was obtained using Lemmas 1 and 2. Therefore, the maximum likelihood function of $L(\Omega_{11})$ is

$$L(\hat{\Omega}_{11}) = \prod_{i=1}^{n} \left(\hat{\gamma}_{11i}^{y_{11i}} \hat{\gamma}_{10i}^{y_{10i}} \hat{\gamma}_{01i}^{y_{01i}} \hat{\gamma}_{00i}^{y_{00i}} \right). \tag{31}$$

where:

$$\hat{\gamma}_{11i} = \begin{cases} \frac{1}{2}(\psi_{2i}-1)^{-1}\left(a_{2i} - \sqrt{a_{2i}^2 + b_{2i}}\right), & \psi_{2i} \neq 1 \\ \hat{\gamma}_{1i}\hat{\gamma}_{2i}, & \psi_{2i} \neq 1 \end{cases},$$

with $a_{2i} = 1 + (\hat{\gamma}_{1i} + \hat{\gamma}_{2i})(\psi_{2i}-1)$, $b_{2i} = 4\psi_{2i}(1-\psi_{2i})\hat{\gamma}_{1i}\hat{\gamma}_{2i}$, $\psi_{2i} = \hat{\gamma}_{11i}\hat{\gamma}_{00i}/(\hat{\gamma}_{10i}\hat{\gamma}_{01i})$, $\hat{\gamma}_{1i} = \exp(\boldsymbol{x}_i^T \hat{\boldsymbol{\theta}}_1)/(1 + \exp(\boldsymbol{x}_i^T \hat{\boldsymbol{\theta}}_1))$, $\hat{\gamma}_{2i} = \exp(\boldsymbol{x}_i^T \hat{\boldsymbol{\theta}}_2)/(1 + \exp(\boldsymbol{x}_i^T \hat{\boldsymbol{\theta}}_2))$; $\hat{\gamma}_{10i} = \hat{\gamma}_{1i} - \hat{\gamma}_{11i}$; $\hat{\gamma}_{01i} = \hat{\gamma}_{2i} - \hat{\gamma}_{11i}$; and $\hat{\gamma}_{00i} = 1 - \hat{\gamma}_{1i} - \hat{\gamma}_{2i} + \hat{\gamma}_{11i}$.

The likelihood ratio (LR) statistic of the hypothesis in Equation (27) is

$$\Lambda = \frac{L(\hat{\Omega}_{01})}{L(\hat{\Omega}_{11})}. \tag{32}$$

With regard to Equations (29) and (31), the LR statistic in Equation (32) can be written as

$$\Lambda = \frac{\prod_{i=1}^{n} \left((\hat{\gamma}_{11i}^*)^{y_{11i}} (\hat{\gamma}_{10i}^*)^{y_{10i}} (\hat{\gamma}_{01i}^*)^{y_{01i}} (\hat{\gamma}_{00i}^*)^{y_{00i}} \right)}{\prod_{i=1}^{n} \left(\hat{\gamma}_{11i}^{y_{11i}} \hat{\gamma}_{10i}^{y_{10i}} \hat{\gamma}_{01i}^{y_{01i}} \hat{\gamma}_{00i}^{y_{00i}} \right)}. \tag{33}$$

However, the form of the LR statistic in Equation (33) is complicated, and we cannot do the calculation analytically. Therefore, to simplify the calculation, the LR statistic is transformed in a form equivalent to

$$\Lambda^{-2} = \left[\frac{L(\hat{\Omega}_{01})}{L(\hat{\Omega}_{11})} \right]^{-2} = \left[\frac{L(\hat{\Omega}_{11})}{L(\hat{\Omega}_{01})} \right]^2. \tag{34}$$

The LR statistic in Equation (34) is also transformed using the natural logarithm. Thus, the formula of the LR statistic is

$$G_1^2 = -2 \log \Lambda = -2 \log \left[\frac{L(\hat{\Omega}_{01})}{L(\hat{\Omega}_{11})} \right] = 2(L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}^*)), \tag{35}$$

where $L(\hat{\boldsymbol{\theta}}) = \log L(\hat{\Omega}_{11})$ and $L(\hat{\boldsymbol{\theta}}^*) = \log L(\hat{\Omega}_{01})$.

(b) Suppose the ML estimator under the population is partitioned by

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{11}^T & \hat{\boldsymbol{\theta}}_{12}^T \end{bmatrix}^T,$$

where $\hat{\boldsymbol{\theta}}_{11} = \begin{bmatrix} \hat{\theta}_{1h} & \hat{\theta}_{2h} & \cdots & \hat{\theta}_{ph} \end{bmatrix}^T$, $h = 1, 2, 3$, and $\hat{\boldsymbol{\theta}}_{12} = \begin{bmatrix} \hat{\theta}_{01} & \hat{\theta}_{02} & \hat{\theta}_{03} \end{bmatrix}^T$.

The ML estimator and the known parameter in the null hypothesis are partitioned by

$$\hat{\boldsymbol{\theta}}^* = \begin{bmatrix} \boldsymbol{\theta}_{01}^{*T} & \hat{\boldsymbol{\theta}}_{02}^{*T} \end{bmatrix}^T,$$

where $\boldsymbol{\theta}_{01(3p \times 1)}^* = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T$ and $\hat{\boldsymbol{\theta}}_{02}^* = \begin{bmatrix} \hat{\theta}_{001} & \hat{\theta}_{002} & \hat{\theta}_{003} \end{bmatrix}^T$, with the true parameter partitioned as $\boldsymbol{\theta}^* = \begin{bmatrix} \boldsymbol{\theta}_{01}^{*T} & \boldsymbol{\theta}_{12}^T \end{bmatrix}^T$.

The hypothesis in Equation (27) can be rewritten as

$$H_0 : \boldsymbol{\theta}_{11} = \boldsymbol{\theta}_{01}^* \quad H_1 : \boldsymbol{\theta}_{11} \neq \boldsymbol{\theta}_{01}^*. \tag{36}$$

The LR statistic in Equation (35) can be specified by

$$G_1^2 = 2(L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}^*)) = 2(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)) - 2(L(\hat{\boldsymbol{\theta}}^*) - L(\boldsymbol{\theta}^*)). \tag{37}$$

The function of $L(\boldsymbol{\theta}^*)$ is approximated using Taylor's second-order expansion around $\hat{\boldsymbol{\theta}}$; we have

$$L(\boldsymbol{\theta}^*) \approx L(\hat{\boldsymbol{\theta}}) + \boldsymbol{g}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}})^T \boldsymbol{I}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}),$$

where

$$\boldsymbol{g}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}'}$$
$$\boldsymbol{I}(\hat{\boldsymbol{\theta}}) = -\left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}.$$

Since $\boldsymbol{g}(\hat{\boldsymbol{\theta}}) = 0$, we have

$$2(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)) \approx (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*). \tag{38}$$

Analogously, in the previous step, the function of $L(\boldsymbol{\theta}^*)$ was approximated using Taylor's second-order expansion around $\hat{\boldsymbol{\theta}}^*$, which is

$$L(\boldsymbol{\theta}^*) \approx L(\hat{\boldsymbol{\theta}}^*) + \boldsymbol{g}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^*) - \frac{1}{2}(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^*)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^*),$$

or it can be written as

$$2(L(\hat{\boldsymbol{\theta}}^*) - L(\boldsymbol{\theta}^*)) \approx (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*). \tag{39}$$

Following Equations (38) and (39), the LR statistic in Equation (37) can be rewritten as

$$G_1^2 \approx \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) - \left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right). \tag{40}$$

Suppose the forms of the partition of the Fisher information matrix and its inverse are as follows.

$$\boldsymbol{I}(\hat{\boldsymbol{\theta}})_{(3p+3)\times(3p+3)} = \begin{bmatrix} \boldsymbol{I}_{11(3p\times 3p)} & \boldsymbol{I}_{12(1\times(3p+3))} \\ \boldsymbol{I}_{21((3p+3)\times 1)} & \boldsymbol{I}_{22(3\times 3)} \end{bmatrix},$$

and

$$\left[\boldsymbol{I}(\hat{\boldsymbol{\theta}})\right]^{-1}_{(3p+3)\times(3p+3)} = \begin{bmatrix} \boldsymbol{I}^{11(3p\times 3p)} & \boldsymbol{I}^{12(1\times(3p+3))} \\ \boldsymbol{I}^{21((3p+3)\times 1)} & \boldsymbol{I}^{22(3\times 3)} \end{bmatrix}.$$

Following the concept of the conditional distribution, given as $\boldsymbol{\theta}_{11} = \boldsymbol{\theta}_{01}^*$, $\hat{\boldsymbol{\theta}}_{11}$, and $\hat{\boldsymbol{\theta}}_{12}$, we have

$$\hat{\boldsymbol{\theta}}_{02}^* = \hat{\boldsymbol{\theta}}_{12} - \boldsymbol{I}^{21}\boldsymbol{I}_{11}^{-1}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right). \tag{41}$$

Simplifying the form of Equation (41) using a simple manipulation of the partitioned matrix of the Fisher information gets

$$\hat{\boldsymbol{\theta}}_{02}^* = \hat{\boldsymbol{\theta}}_{12} + \boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right). \tag{42}$$

Since $\left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right) = \left(\boldsymbol{0}, \hat{\boldsymbol{\theta}}_{02}^* - \boldsymbol{\theta}_{12}\right)$ and $\boldsymbol{\theta}_{02}^* - \boldsymbol{\theta}_{12} = \hat{\boldsymbol{\theta}}_{12} - \boldsymbol{\theta}_{12} + \boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right)$, we have

$$\begin{aligned} \left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right)^T \boldsymbol{I}(\hat{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\right) &= \left(\hat{\boldsymbol{\theta}}_{02}^* - \boldsymbol{\theta}_{12}\right)^T \boldsymbol{I}_{22}\left(\hat{\boldsymbol{\theta}}_{02}^* - \boldsymbol{\theta}_{12}\right) \\ &= \begin{bmatrix} \hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^* \\ \hat{\boldsymbol{\theta}}_{12} - \boldsymbol{\theta}_{12} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{I}_{12}\boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^* \\ \hat{\boldsymbol{\theta}}_{12} - \boldsymbol{\theta}_{12} \end{bmatrix} \end{aligned}$$

The LR statistic in Equation (40) can be formulated as

$$G_1^2 \approx \left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right)^T \left(\boldsymbol{I}_{11} - \boldsymbol{I}_{12}\boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21}\right)\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right) = \left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right)^T \boldsymbol{I}^{11^{-1}}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right). \tag{43}$$

When considering the normality properties for ML under the regularity conditions [31], the distribution of the partitioned matrix is

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^* \\ \hat{\boldsymbol{\theta}}_{12} - \boldsymbol{\theta}_{12} \end{bmatrix} \xrightarrow{d} N\left(\boldsymbol{0}, [\boldsymbol{I}(\boldsymbol{\theta})]^{-1} \equiv \begin{bmatrix} \boldsymbol{I}^{11} & \boldsymbol{I}^{12} \\ \boldsymbol{I}^{21} & \boldsymbol{I}^{22} \end{bmatrix}\right), \ n \to \infty.$$

Therefore, the LR statistic is obtained as follows:

$$\begin{aligned} \left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{11}\right) &\xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}^{11}), \ n \to \infty, \\ \left[\boldsymbol{I}^{11}\right]^{-1/2}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{11}\right) &\xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_{v_1}), \ n \to \infty, \\ G_1^2 = \left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right)^T \boldsymbol{I}^{11^{-1}}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right) &\xrightarrow{d} \chi_{v_1}^2, \ n \to \infty. \end{aligned} \tag{44}$$

The LR statistic in Equation (44) has an asymptotic chi-square distribution with v_1 degrees of freedom. v_1 is the difference of the parameter sets under the population and the null hypothesis.

(c) Determining the rejection region or the critical region of the null hypothesis in Equation (27) requires the MLRT method, where the null hypothesis is rejected when $\Lambda < c, 0 < c \le 1$, and where Λ is given in Equation (32) and c is a constant. Let α be the significance level for $0 < \alpha < 1$, and $0 < c_\alpha \le 1$ be a constant; the c_α value depends on

the significance level of α and satisfies $P_{\boldsymbol{\theta}\epsilon\Omega_{01}}(\Lambda < c_\alpha) = \alpha$. Based on the definition of the significance level of α, we have

$$
\begin{aligned}
\alpha \quad &= P_{\boldsymbol{\theta}\epsilon\Omega_{01}}(\Lambda < c_\alpha) \\
&= P(-2\log\Lambda > -2\log c_\alpha) \\
&= P\left(2\left[L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}^*)\right] > c\right) \\
&= P\left(G_1^2 > c\right),
\end{aligned}
\tag{45}
$$

where G_1^2 is the LR statistic, which has an asymptotic chi-square distribution with v_1 degrees of freedom. The value of the constant c in Equation (45) is $\chi^2_{(\alpha,v_1)}$ and $\int_c^\infty f(w)dw = \alpha$, where $f(w) = w^{(v_1/2)-1}e^{-w/2}/\left(\Gamma(v_1/2)2^{v_1/2}\right)$ is the probability density function of the chi-square distribution with v_1 degrees of freedom. Therefore, the rejection region at the significance level of α is $G_1^2 > \chi^2_{(\alpha,v_1)}$.

The last one is a partial test. This test aims to obtain covariates that have a significant effect on the responses individually. The procedures of the partial test in this study follow the simultaneous test. The hypothesis of the partial test is

$$
\begin{aligned}
H_0 \;:\; & \theta_{11} = \theta_{12} = \theta_{13} = 0, \\
H_1 \;:\; & \text{at least one of } \theta_{1h} \neq 0, \;\; h = 1,2,3.
\end{aligned}
\tag{46}
$$

The parameter set under the null hypothesis for each covariate is $\Omega_{02} = \{\theta_{01}, \theta_{02}, \theta_{03}\}$. The parameter set under the population for each covariate is $\Omega_{12} = \{\theta_{01}, \theta_{11}, \theta_{02}, \theta_{12}, \theta_{03}, \theta_{13}\}$. Analogously, in the proof of Theorem 1, the LR statistic for the hypothesis in Equation (46) is

$$
G_2^2 = \left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right)^T I_{11}^{-1}\left(\hat{\boldsymbol{\theta}}_{11} - \boldsymbol{\theta}_{01}^*\right) \xrightarrow{d} \chi^2_{v_2} \;, n \to \infty.
\tag{47}
$$

The LR statistic in Equation (47) has an asymptotic chi-square distribution with v_2 degrees of freedom. The rejection region at the significance level α is $G_2^2 > \chi^2_{(\alpha,v_2)}$. $\square$

5. Application

The BBL model was applied to model the factors influencing the status of the human development index (HDI) and public health development index (PHDI) of regencies/municipalities in Kalimantan, Indonesia, in 2018. The HDI is an index measured from four components of the essential dimensions of human development: life expectancy, the average length of schooling, expected length of schooling, and adjusted per-capita income. Life expectancy represents an indicator of health, the average length of schooling and the expected length of schooling represent educational indicators, and adjusted per capita income represents an economic indicator [34]. The PHDI is an index that measures the health of the regencies/municipalities and provinces in the Republic of Indonesia [35].

The HDI status data and covariates' data were collected from the National Bureau of Statistics of the Republic of Indonesia, whereas the PHDI data were collected from the Republic of Indonesia's Ministry of Health. The variables in this study consist of two responses and five covariates. The responses are the HDI status and the PHDI status of regencies/municipalities, denoted by Y_1 and Y_2. The covariates are the economic growth (X_1), the net enrollment rate of the junior high school (X_2), the percentage of people that have the minimum level of education in junior high school (X_3), the number of doctors per 1000 people (X_4), and the number of public health centers (X_5). The regencies and municipalities' HDI status has four categories: low HDI, medium HDI, high HDI, and very high HDI [34]. Regencies/municipalities in Kalimantan, Indonesia, in 2018, had HDI in the medium and high categories. Therefore, the HDI status (Y_1) has two categories: the medium HDI coded by 0 and the high HDI coded by 1.

Meanwhile, the Ministry of Health of the Republic of Indonesia classifies regencies/municipalities' health status based on the PHDI into two categories. Regencies/municipalities

with a low PHDI have health problems, and vice versa [36]. Therefore, the PHDI status (Y_2) has two categories: the regencies/municipalities with low PHDI values are coded by 0, and the regencies/municipalities with high PHDI values are coded by 1. This study's observation unit is the regency/municipality. Five provinces in Kalimantan, Indonesia were used (2018 data), including 47 regencies and nine municipalities. Therefore, the sample size is 56.

The descriptive statistics of the responses HDI status (Y_1) and PHDI status (Y_2), consisting of observed frequencies, are presented in Table 2.

Table 2. The observed frequencies of the responses.

Y_1	Y_2		Total
	$Y_2=1$	$Y_2=0$	
$Y_1 = 1$	20	3	23
$Y_1 = 0$	6	27	33
Total	26	30	56

Table 2 shows that 20 regencies/municipalities had high HDI and PHDI, and six regencies/municipalities had high HDI and low PHDI. We also see that three regencies/municipalities had medium HDI and high PHDI. Finally, 27 regencies/municipalities had medium HDI and low PHDI. The HDI status (Y_1) and PHDI status (Y_2) of regencies/municipalities are displayed in Figure 1.
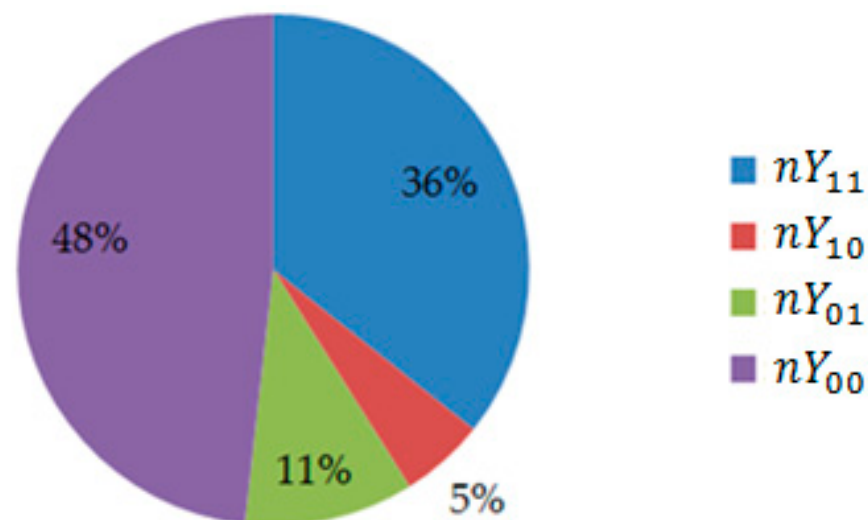
Figure 1. The proportions of the responses with certain HDI status (Y_1) and PHDI status (Y_2). nY_{00} is the number of regencies/municipalities that had medium HDI and low PHDI; nY_{01} is the number of regencies/municipalities that had medium HDI and high PHDI; nY_{10} is the number of regencies/municipalities that had high HDI and low PHDI; nY_{11} is the number of regencies/municipalities that had high HDI and PHDI.

The descriptive statistics of the responses show that the majority of regencies/municipalities in Kalimantan, Indonesia, in 2018, had medium HDI and low PHDI. The descriptive statistics of the covariates are summarized in Table 3.

Table 3. The summary of descriptive statistics for the covariates.

Covariates	Minimum	Maximum	Mean	Standard Deviation
X_1	−4.10	7.99	5.08	1.83
X_2	68.37	98.82	81.19	8.14
X_3	35.58	81.37	54.25	11.05
X_4	1.00	46.40	10.23	9.47
X_5	5.00	33.00	17.57	6.93

The HDI status (Y_1) and PHDI status (Y_2) are correlated. Based on the observed frequencies in Table 2, the odds ratio (OR) value of HDI status (Y_1) and PHDI status (Y_2) was 30 with a 95% confidence interval of $6.6826 \leq OR \leq 134.6783$. This result indicates that the responses are highly positively correlated. Meanwhile, we also employed the dependence test of the responses HDI status (Y_1) and PHDI status (Y_2), provided in Table 4.

Table 4. Statistical test values of the dependence test of responses.

Statistical Tests	χ^2	df	p-Value
Pearson	25.7750	1	3.8370×10^{-7}
Pearson with Yates' continuity correction	23.0840	1	1.5510×10^{-6}
Likelihood ratio	28.2420	1	1.0708×10^{-7}

Three statistical tests demonstrated a dependence test of HDI status (Y_1) and PHDI status (Y_2). The result in Table 4 shows that all of the statistical test values had greater than the chi-square table value (i.e., $\chi^2_{(0.05,1)} = 3.8415$) and p-values less than the significance level value (i.e., $\alpha = 0.05$). Therefore, the conclusion was to reject the null hypothesis (H_0), and the HDI status (Y_1) and PHDI status (Y_2) are dependencies. Based on the OR value and the dependence test, the HDI status (Y_1) and PHDI status (Y_2) are appropriate for the BBL model.

The variance inflation factor detected the multicollinearity of the covariates. The variance inflation factor values of all covariates in Table 5 are less than ten, which indicates that the covariates are independent of each other (i.e., no multicollinearity). Therefore, all covariates can be used in the BBL model.

Table 5. Variance inflation factor (VIF) values of the covariates.

Covariates	VIF
X_1	1.0404
X_2	1.4948
X_3	2.3002
X_4	2.5617
X_5	1.5165

The estimation of the BBL model's parameters using the ML and BHHH methods was employed. Table 6 provides the bias values and the numbers of BHHH iterations of the parameter estimation process for the BBL model with the single and multiple covariates.

Table 6. The bias values and the numbers of BHHH iterations for the BBL model with single and multiple covariates.

Covariates	Bias	Iteration
X_1	6.2266×10^{-5}	1000
X_2	1.9227×10^{-8} *	23
X_3	2.0283×10^{-8} *	25
X_4	3.8014×10^{-6} *	9
X_5	5.6987×10^{-5}	1000
$X_2 X_3 X_4$	8.5186×10^{-8} *	146

* Convergent at tolerance limit value (i.e., $\varepsilon = 1 \times 10^{-5}$).

The BBL model with the single covariate of economic growth (X_1) and public health centers (X_5) in Table 6 was not convergent. Therefore, both covariates, economic growth (X_1) and public health centers (X_5), were not used in the BBL model. Based on Table 6, the BBL model for modeling the factors that affect the HDI status and PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018 was obtained.

Table 7 displays the ML estimates of the BBL model with multiple covariates (i.e., X_2, X_3, X_4), giving the parameter estimates, the LR statistic of the simultaneous test (G_1^2), the degrees of freedom (df), and the *p*-value.

Table 7. Parameter estimates and the statistical test value of the simultaneous test for the BBL model with the multiple covariates.

Parameter	Estimation	G_1^2	df	*p*-Value
θ_{01}	-0.0034			
θ_{11}	-0.1916			
θ_{21}	0.0449			
θ_{31}	0.0013			
θ_{02}	-0.0016			
θ_{12}	-0.1403			
θ_{22}	0.0395	99.7390	9	1.7685×10^{-21}
θ_{32}	0.0011			
θ_{03}	-0.0023			
θ_{13}	-0.1440			
θ_{23}	-0.1071			
θ_{33}	0.0002			

The LR statistic value in Table 7 is 99.739, and the *p*-value is 1.7685×10^{-21} ($p < 0.001$). Meanwhile, the chi-square table's value with nine degrees of freedom and a 5% significance level was 16.919. The LR statistic value is greater than the chi-square table's value, and the *p*-value is less than the 5% significance level. Therefore, the null hypothesis was rejected, and we conclude that the net enrollment rate of the junior high school, the percentage of people that have the minimum level of education in junior high school, and the number of doctors per 1000 people were jointly significantly affecting the HDI status and the PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018. The BBL model for the HDI status and the PHDI status of regencies/municipalities can be written as follows:

$$\hat{\tau}_1(\mathbf{x}) = -0.0034 - 0.1916 X_2 + 0.0449 X_3 + 0.0013 X_4,$$
$$\hat{\tau}_2(\mathbf{x}) = -0.0016 - 0.1403 X_2 + 0.0395 X_3 + 0.0011 X_4,$$
$$\hat{\tau}_3(\mathbf{x}) = -0.0023 - 0.1440 X_2 - 0.1071 X_3 + 0.0002 X_4.$$

The partial test using the MLRT method was used to obtain the covariates that individually affect the HDI status and the PHDI status of regencies/municipalities. Table 8 describes the BBL model with the single covariate, which covers the parameter estimates, the LR statistic value (G_2^2), the degrees of freedom (df), and the *p*-value.

Table 8. Parameter estimates and the LR statistic value of the partial test for the BBL model with the single covariate.

Covariate	Parameter	Estimation	G_2^2	df	*p*-Value
X_2	θ_{01}	-0.0010	90.2309	3	1.9542×10^{-19}
	θ_{11}	-0.0446			
	θ_{02}	-0.0026			
	θ_{12}	-0.2251			
	θ_{03}	0.0008			
	θ_{13}	0.0931			
X_3	θ_{01}	-0.0104	111.4570	3	5.3304×10^{-24}
	θ_{11}	-0.1630			
	θ_{02}	-0.0079			
	θ_{12}	-0.2145			
	θ_{03}	0.0004			
	θ_{13}	0.0483			
X_4	θ_{01}	-15.3744	174.2092	3	1.5700×10^{-37}
	θ_{11}	12.5080			
	θ_{02}	-17.0178			
	θ_{12}	9.3722			
	θ_{03}	4.2464			
	θ_{13}	4.1612			

The LR statistic's value of the estimated parameter for each covariate (the net enrollment rate of the junior high school, the percentage of people that have the minimum level of education in junior high school, and the number of doctors per 1000 people; Table 8) was greater than the chi-square table's value; the chi-square table's value with three degrees of freedom and 5% significance level was 7.8147. Meanwhile, the *p*-value of each covariate was less than the 5% significance level. Therefore, we concluded that the net enrollment rate of the junior high school, the percentage of people that have the minimum level of education in junior high school, and the number of doctors per 1000 people individually significantly influenced the HDI status and the PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018.

The BBL model with a single covariate (e.g., X_4) for the HDI status and the PHDI status of regencies/municipalities can be expressed as follows:

$$\hat{\tau}_1(\mathbf{x}) = -15.3744 + 12.5080 X_4,$$
$$\hat{\tau}_2(\mathbf{x}) = -17.0178 + 9.3722 X_4,$$
$$\hat{\tau}_3(\mathbf{x}) = 4.2464 + 4.1612 X_4.$$

The AIC and BIC methods in Equations (25) and (26) were used for the evaluation of the BBL model's performance. The AIC and BIC values of the BBL models are shown in Table 9.

Table 9. The AIC and BIC values of the BBL model are the single and multiple covariates.

Covariates	AIC	BIC
X_2	512.9107	525.0628
X_3	546.2378	558.3899
X_4	546.1944	558.3465
$X_2 X_3 X_4$	552.6286	576.9328

The BBL model with the single covariate in Table 9 has the smallest AIC and BIC values compared to the BBL model with the multiple covariates. Therefore, the BBL model with the single covariate is the best model for modeling the relationships between the responses (i.e., the HDI status and the PHDI status) and the covariates (i.e., the net enrollment rate of the

junior high school, the percentage of people that have the minimum level of education in junior high school, and the number of doctors per 1000 people) of regencies/municipalities in Kalimantan, Indonesia, in 2018. Furthermore, the net enrollment rate of the junior high school, the percentage of people that have the minimum level of education in junior high school, and the number of doctors per 1000 people individually significantly affected the HDI status and the PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018.

However, some recommendations and future research from this work are possible. Firstly, the logit models in this research are limited to two responses. The BBL model, with more than two responses, should be considered for future research. Secondly, other numerical optimization methods that improve the performance of the BBL model should also be considered for future research.

6. Conclusions

The BBL model is the development of the binary logit model. It was constructed using the multinomial distribution. The ML method was applied to get the BBL model's parameter estimator. The ML estimator of the BBL model's parameters does not have a closed-form, and it needs an iterative numerical procedure. The BHHH iterative method was used. Hypothesis testing of the BBL model includes the simultaneous test and the partial test. The simultaneous and partial tests were done by using the MLRT method. The LR statistics of the simultaneous test and the partial test were asymptotically chi-square distributed. The BBL model was applied to model the factors affecting the HDI status and the PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018. The BBL model with the single covariate performed better than the BBL model with multiple covariates. The factors significantly affecting the HDI status and the PHDI status of regencies/municipalities in Kalimantan, Indonesia, in 2018, were the net enrollment rate of junior high schools, the percentage of people who have the minimum level of education in junior high school, and the number of doctors per 1000 people.

Abbreviations

AIC	Akaike's information criterion
BHHH	Berndt–Hall–Hall–Hausman
BIC	Bayesian information criterion
HDI	Human development index
LR	Likelihood ratio

ML	Maximum likelihood
MLRT	Maximum likelihood ratio test
MULT	Multinomial distribution
PHDI	Public health development index
VIF	Variance inflation factor

References

1. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman and Hall: London, UK, 1989.
2. Glonek, G.F.V.; McCullagh, P. Multivariate logistic models. *J. R. Stat. Soc. B* **1995**, *57*, 533–546. [CrossRef]
3. Kauermann, G. A note on multivariate logistic models for contingency table. *Austral. J. Stat.* **1997**, *39*, 261–276. [CrossRef]
4. Glonek, G.F.V. A class of regression models for multivariate categorical responses. *Biometrika* **1996**, *83*, 15–28. [CrossRef]
5. Bergsma, W.P.; Rudas, T. Marginal models for categorical data. *Ann. Stat.* **2002**, *30*, 140–159.
6. Qaqish, B.F.; Ivanova, A. Multivariate logistic models. *Biometrika* **2006**, *93*, 1011–1017.
7. Lipsitz, S.R.; Laird, N.M.; Harrington, D.P. Maximum likelihood regression methods for paired binary data. *Stat. Med.* **1990**, *9*, 1517–1525. [CrossRef] [PubMed]
8. Liang, K.Y.; Zeger, S.L.; Qaqish, B. Multivariate regression analyses for categorical data (with discussion). *J. R. Stat. Soc. B* **1992**, *54*, 3–40.
9. Carey, V.; Zeger, S.L.; Diggle, P. Modelling multivariate binary data with alternating logistic regressions. *Biometika* **1993**, *80*, 517–526. [CrossRef]
10. Cessie, S.L.; Houwelingen, J.C. Logistic regression for correlated binary data. *Appl. Stat.* **1994**, *43*, 95–108. [CrossRef]
11. Lang, J.B.; Agresti, A. Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *J. Am. Stat. Assoc.* **1994**, *89*, 625–632. [CrossRef]
12. Shoukri, M.M.; Martin, S.W.; Mian, I.U.H. Maximum likelihood estimation of the kappa coefficient from models of matched binary responses. *Stat. Med.* **1995**, *14*, 83–99. [CrossRef] [PubMed]
13. Shoukri, M.M.; Mian, I.U.H. Maximum likelihood estimation of the kappa coefficient from bivariate logistic regression. *Stat. Med.* **1996**, *15*, 1409–1419. [CrossRef]
14. Molenberghs, G.; Lesaffre, E. Marginal modelling of multivariate categorical data. *Stat. Med.* **1999**, *18*, 2237–2255. [CrossRef]
15. Ekholm, A.; Smith, P.W.F.; McDonal, J.W. Marginal regression analysis of a multivariate binary response. *Biometrika* **1995**, *82*, 847–854. [CrossRef]
16. Ekholm, A.; McDonald, J.W.; Smith, P.W.F. Association models for a multivariate binary response. *Biometrics* **2000**, *56*, 712–718. [CrossRef] [PubMed]
17. Joe, H.; Liu, Y. A model for a multivariate binary response with covariates based on compatible conditionally specified logistic regressions. *Stat. Probab. Lett.* **1996**, *31*, 113–120. [CrossRef]
18. Islam, M.A.; Chowdhury, R.I.; Briollais, L. A bivariate binary model for testing dependence in outcomes. *Bull. Malays. Math. Sci. Soc.* **2012**, *35*, 845–858.
19. El-Sayed, A.M.; Islam, M.A.; Alzaid, A.A. Estimation and test of measures of association for correlated binary data. *Bull. Malays. Math. Sci. Soc.* **2013**, *36*, 985–1008.
20. Islam, M.A.; Alzaid, A.A.; Chowdhury, R.I.; Sultan, K.S. A generalized bivariate Bernoulli model with covariate dependence. *J. Appl. Stat.* **2013**, *40*, 1064–1075. [CrossRef]
21. Bhuyan, M.J.; Islam, M.A.; Rahman, M.S. A bivariate Bernoulli model for analyzing malnutrition. *Health Serv. Outcomes Res. Method* **2018**, *18*, 109–127. [CrossRef]
22. Sinha, S.S.; Laird, N.M.; Fitzmaurice, G.M. Multivariate logistic regression with incomplete covariate and auxiliary information. *J. Multivar. Anal.* **2010**, *101*, 2389–2397. [CrossRef]
23. Horton, N.J.; Laird, N.M. Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics* **2001**, *57*, 34–42. [CrossRef]
24. Chen, Z.; Yi, G.Y.; Wu, C. Marginal methods for correlated binary data with misclassified responses. *Biometrika* **2011**, *98*, 647–662. [CrossRef]
25. O'Brien, S.M.; Dunson, D.B. Bayesian multivariate logistic regression. *Biometrics* **2004**, *60*, 739–746. [CrossRef]
26. Fathurahman, M.; Purhadi; Sutikno; Ratnasari, V. Hypothesis testing of geographically weighted bivariate logistic regression. *J. Phys. Conf. Ser.* **2019**, *1417*, 012008. [CrossRef]
27. Fathurahman, M.; Purhadi; Sutikno; Ratnasari, V. Geographically Weighted Multivariate Logistic Regression Model and Its Application. *Abstr. Appl. Anal.* **2020**, *2020*, 8353481. [CrossRef]
28. Berndt, E.K.; Hall, B.H.; Hall, R.E.; Hausman, J.A. Estimation and inference in nonlinear structural models. *Ann. Econ. Soc. Meas.* **1974**, *3*, 653–665.
29. Mardalena, S.; Purhadi, P.; Purnomo, J.D.T.; Prastyo, D.D. Parameter estimation and hypothesis testing of multivariate Poisson inverse Gaussian regression. *Symmetry* **2020**, *12*, 1738. [CrossRef]
30. Dale, J.R. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **1986**, *42*, 909–917. [CrossRef] [PubMed]
31. Greene, W.H. *Econometric Analysis*, 6th ed.; Pearson Education: Cranbury, NJ, USA, 2008.
32. Pawitan, Y. *All Likelihood: Statistical Modelling and Inference Using Likelihood*, 1st ed.; Clarendon Press: Oxford, UK, 2001.

33. Rahayu, A.; Purhadi; Sutikno; Prastyo, D.D. Multivariate gamma regression: Parameter estimation, hypothesis testing, and its application. *Symmetry* **2020**, *12*, 813. [CrossRef]
34. National Bureau of Statistics. *Human Development Index 2018*; BPS: Jakarta, Indonesia, 2019.
35. Ministry of Health. *Public Health Development Index*; LPB: Jakarta, Indonesia, 2019.
36. Ministry of Health. *General Guidelines for Dealing with Areas with Health Problems*; LPB: Jakarta, Indonesia, 2010.