J. Basic. Appl. Sci. Res., 1(3)178-188, 2011

ISSN 2090-424X
Journal of Basic and Applied Scientific Research
www.textroad.com

Estimation and Test Statistic in Bivariate Probit Model $(r \times c)$

Vita Ratnasari[1], Purhadi[2], Ismaini[2], Suhartono[2]

[1]Doctorate Program in Statistics, Faculty of Mathematics and Natural Science, Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia
[2]Lecturer at Statistics Department, Faculty of Mathematics and Natural Science, Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia

ABSTRACT

One of statistics models that could be used to analyze categorical response data is probit model. This paper focuses to discuss about estimation and test statistic in bivariate probit model $(r \times c)$. Bivariate probit model $(r \times c)$ is a probit model which involves two response variables, i.e. the first variable has r category and the second has c category. Both response variables are correlated. In this model, the estimation of model parameters is done by Maximum Likelihood Estimation method with the Newton-Raphson iteration. Furthermore, test statistic to evaluate the significance of model parameters is obtained by using Maximum Likelihood Ratio Test, particularly the G^2 test for overall test and *t* test for partial test.

KEY WORDS: Bivariate probit, Maximum Likelihood Estimation, Maximum Likelihood Ratio Test.

INTRODUCTION

Probit model is a statistics model that can explain the relationship between the discrete response variable and continuous, discrete or mix predictor variables. There are many researches about probit model, particularly univariate probit model such as (Aitchison and Silvey,1957), (Boes and Winkelmann, (2005), (Jackman, 2000), (McKelvey and Zavoina, 1975), (Ronning and Kukuk, 1996), and (Snapinn and Small, 1986). In fact, the response variables in many cases could be more than one and correlated each others.

Up to now, research on bivariate probit model mostly focused on the application, such as (Bokosi 2007), (Mozumder, *et al.,* 2008) and (Yamamoto and Shankar, 2004). Therefore, this paper will discuss about theory of probit models with two response variables, especially the estimation and test statistic. Bivariate probit model is a probit model which involves two response variables, i.e. the first variable has r category and the second has c category. In this case, both response variables are correlated.

In general, the initial step for building bivariate probit model is to estimate the model parameters and determine the test statistic to validate significance of the parameters. One of estimation methods that frequently used for estimating the parameters of bivariate probit model is Maximum Likelihood Estimation (MLE). Furthermore, the test statistic that usually be used to evaluate significance of model parameters is Maximum Likelihood Ratio Test. Thus, this paper will focus to discuss further about estimation and test statistic in bivariate probit model $(r \times c)$.

***Corresponding Author:** Dr. Vita Ratnasari, Doctorate Program in Statistics, Faculty of Mathematics and Natural Science, Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia.

METHODS

The probit model r category is built from a latent regression in the same man-ner as the binomial probit model. We begin with $y_i^* = \boldsymbol{\beta}^T \mathbf{x}_i + \boldsymbol{\varepsilon}_i$, where $\mathbf{x}_i$ is a vector of predictor variable for the ith observation and $\boldsymbol{\beta}^*$ is the unknown parameter. As usual, y^* is unobserved variable, that follow as: (Greene, 2008).

$$y = 0 \;\text{ if } \gamma_0 < y^* \le \gamma_1$$
$$y = 1 \;\text{ if } \gamma_1 < y^* \le \gamma_2$$
$$\vdots$$
$$y = r-1 \;\text{ if } \gamma_{r-1} < y^* \le \gamma_r$$

The probability for each observed respon has r category, i.e:

$$\mathrm{P}(y=0) = P(\gamma_0 < y^* \le \gamma_1) = \Phi(\gamma_1 - \boldsymbol{\beta}^T\mathbf{x}) - \Phi(\gamma_0 - \boldsymbol{\beta}^T\mathbf{x})$$
$$\mathrm{P}(y=1) = P(\gamma_1 < y^* \le \gamma_2) = \Phi(\gamma_2 - \boldsymbol{\beta}^T\mathbf{x}) - \Phi(\gamma_1 - \boldsymbol{\beta}^T\mathbf{x})$$
$$\vdots$$
$$\mathrm{P}(y=r-1) = P(\gamma_{r-1} < y^* \le \gamma_r) = \Phi(\gamma_r - \boldsymbol{\beta}^T\mathbf{x}) - \Phi(\gamma_{r-1} - \boldsymbol{\beta}^T\mathbf{x})$$

Bivariate probit model $(r \times c)$ is a probit model which involves two response variables, i.e. $y_1^* = \boldsymbol{\beta}_1^T\mathbf{x} + \boldsymbol{\varepsilon}_1$ and $y_2^* = \boldsymbol{\beta}_2^T\mathbf{x} + \boldsymbol{\varepsilon}_2$. The first variable has r category that is

$$y_1 = 0 \;\text{ if } \gamma_0 < y_1^* \le \gamma_1$$
$$y_1 = 1 \;\text{ if } \gamma_1 < y_1^* \le \gamma_2$$
$$\vdots$$
$$y_1 = r-1 \;\text{ if } \gamma_{r-1} < y_1^* \le \gamma_r.$$

Whereas, the second has c category, that is

$$y_2 = 0 \;\text{ if } \delta_0 < y_2^* \le \delta_1$$
$$y_2 = 1 \;\text{ if } \delta_1 < y_2^* \le \delta_2$$
$$\vdots$$
$$y_3 = c-1 \;\text{ if } \delta_{c-1} < y_2^* \le \delta_c.$$

Variables (y_1^*, y_2^*) that satisfy normal bivariate distribution can be written as $(y_1^*, y_2^*) \sim N(\boldsymbol{\mu}, \Sigma)$. For example, there is p predictor variables $x_1, x_2, \cdots, x_p$, with $E(y_1^*) = \boldsymbol{\beta}_1^T\mathbf{x}$ and $E(y_2^*) = \boldsymbol{\beta}_2^T\mathbf{x}$. So, it can be written $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\beta}_1^T\mathbf{x} \\ \boldsymbol{\beta}_2^T\mathbf{x} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

Bivariate normal density function (y_1^*, y_2^*) is:

$$f(y_1^*, y_2^*) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \begin{pmatrix} \mathbf{y}_1^* - \boldsymbol{\beta}_1^T \mathbf{x} \\ \mathbf{y}_2^* - \boldsymbol{\beta}_2^T \mathbf{x} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \mathbf{y}_1^* - \boldsymbol{\beta}_1^T \mathbf{x} \\ \mathbf{y}_2^* - \boldsymbol{\beta}_2^T \mathbf{x} \end{pmatrix} \right)$$

The probability of bivariate normal density function (y_1^*, y_2^*) with thresholds γ and δ is as follows:

$$P(y_1^* < \gamma, y_2^* < \delta) = \int_{-\infty}^{\gamma} \int_{-\infty}^{\delta} f(y_1^*, y_2^*) dy_1^* dy_2^* .$$

Suppose $z_1 = \dfrac{y_1^* - E(y_1^*)}{\sigma_{y_1^*}}$, if $\sigma_{y_1^*}^2 = 1$ so $z_1 = y_1^* - \boldsymbol{\beta}_1^T \mathbf{x}$, and $z_2 = \dfrac{y_2^* - E(y_2^*)}{\sigma_{y_2^*}}$ if $\sigma_{y_2^*}^2 = 1$ so $z_2 = y_2^* - \boldsymbol{\beta}_2^T \mathbf{x}$. Therefore, bivariate normal density function (z_1, z_2) is as follows:

$$P(y_1^* < \gamma, y_2^* < \delta) = P(z_1 < z_{ji}, z_2 < z_{ki}) = \int_{-\infty}^{z_{j1}} \int_{-\infty}^{z_{k2}} \phi(z_1, z_2) dz_1 dz_2$$

where $\phi(z_1, z_2, \rho) = \dfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\dfrac{1}{2} \dfrac{1}{\left[1-\rho^2\right]} \left[Z_1^2 - 2\rho Z_1 Z_2 + Z_2^2 \right] \right]$,

$j = 0, 1, 2, \cdots, r-1$ and $k = 0, 1, 2, \cdots, c-1$,

$z_{01} = \gamma_0 - \boldsymbol{\beta}_1{}^T \mathbf{x}$, $z_{11} = \gamma_1 - \boldsymbol{\beta}_1{}^T \mathbf{x}$, $z_{21} = \gamma_2 - \boldsymbol{\beta}_1{}^T \mathbf{x}$, ..., $z_{(r-1)1} = \gamma_{(r-1)} - \boldsymbol{\beta}_1{}^T \mathbf{x}$ and

$z_{02} = \delta_1 - \boldsymbol{\beta}_2{}^T \mathbf{x}$, $z_{12} = \delta_1 - \boldsymbol{\beta}_2{}^T \mathbf{x}$, $z_{22} = \delta_2 - \boldsymbol{\beta}_2{}^T \mathbf{x}$, ..., $z_{(c-1)2} = \delta_{(c-1)} - \boldsymbol{\beta}_2{}^T \mathbf{x}$,

$\boldsymbol{\beta}_1^T = \left[\beta_{10}\ \beta_{11}\ \beta_{12}\ \cdots \beta_{1p} \right]$ and $\boldsymbol{\beta}_2^T = \left[\beta_{20}\ \beta_{21}\ \beta_{22}\ \cdots \beta_{2p} \right]$

$\mathbf{x}^T = \left[1\ x_1\ x_2\ \cdots x_p \right]$.

Value $\gamma_0, \gamma_1, \cdots, \gamma_{(r-1)}$ is threshold on the first response variable with $(r-1)$ category, and value $\delta_0, \delta_1, \cdots, \delta_{(c-1)}$ is threshold on the second response variable with $(c-1)$ category. Both response variables are formed in contingency table $(r \times c)$ as in Table 1. Table 1 will have $(r \times c)$ categories, i.e. $Y_{00}, Y_{10}, Y_{01}, \cdots, Y_{(r-1)\times(c-1)}$ which value 0 or 1. Y_{jk} is an event that happens on the first response variable of j category and the second response variable of k category, where $Y_{00} = 1 - \sum\limits_{\substack{j=1 \\ j \neq k = 0}}^{r-1} \sum\limits_{k=0}^{c-1} Y_{jk}$.

Y_{00} is an event that happens in the area of $-\infty < y_1^* < \gamma_0$ and $-\infty < y_2^* < \delta_0$

Y_{01} is an event that happens in the area of $-\infty < y_1^* < \gamma_0$ and $\delta_0 < y_2^* < \delta_1$

Y_{10} is an event that happens in the area of $\gamma_0 < y_1^* < \gamma_1$ and $-\infty < y_2^* < \delta_0$

Y_{11} is an event that happens in the area of $\gamma_0 < y_1^* < \gamma_1$ and $\delta_0 < y_2^* < \delta_1$

$\vdots$

$Y_{(r-1)(c-1)}$ is an event that happens in the area of $\gamma_{r-1} < y_1^* < \infty$ and $\delta_{c-1} < y_2^* < \infty$.

The event on Table 1 follows Multinomial distribution, i.e.

$(Y_{10}, Y_{01}, Y_{11}, \cdots, Y_{(r-1)(c-1)}) \sim M(1; P_{10}, P_{01}, P_{11}, \cdots, P_{(r-1)(c-1)})$.

The density function of Multinomial distribution, is as follows:

$$P(Y_{10} = y_{10}, Y_{01} = y_{01}, Y_{11} = y_{11}, ..., Y_{(r-1)(c-1)} = y_{(r-1)(c-1)}) = \prod_{j=0}^{r-1}\prod_{k=0}^{c-1} P_{jk}^{jk} .$$

Table 1: Table of Frequency Contingency and Probability of Two Response Variables $(r \times c)$

$\mathbf{Y_1}$	$\mathbf{Y_2}$				
	0	1	$\cdots$	$(c-1)$	**Total**
0	$Y_{00}; P_{00}$	$Y_{01}; P_{01}$	$\cdots$	$Y_{0(c-1)}; P_{0(c-1)}$	P_{0+}
1	$Y_{10}; P_{10}$	$Y_{11}; P_{11}$	$\cdots$	$Y_{1(c-1)}; P_{1(c-1)}$	P_{1+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$(r-1)$	$Y_{(r-1)0}; P_{(r-1)0}$	$Y_{(r-1)1}; P_{(r-1)1}$	$\cdots$	$Y_{(r-1)(c-1)}; P_{(r-1)(c-1)}$	$P_{(r-1)+}$
Total	P_{+0}	P_{+1}	$\cdots$	$P_{+(c-1)}$	$P_{++} = 1$

In details, probability value on Table 1 is as follows.

$$P_{00} = \int_{-\infty}^{z_{02}}\int_{-\infty}^{z_{01}} \phi(z_1, z_2, \rho)\, dz_1\, dz_2 = \Phi(z_{01}, z_{02})$$

$$P_{01} = \int_{z_{02}}^{z_{12}}\int_{-\infty}^{z_{01}} \phi(z_1, z_2, \rho)\, dz_1\, dz_2 = \Phi(z_{01}, z_{12}) - \Phi(z_{01}, z_{02})$$

$$P_{10} = \int_{-\infty}^{z_{02}}\int_{z_{01}}^{z_{11}} \phi(z_1, z_2, \rho)\, dz_1\, dz_2 = \Phi(z_{11}, z_{02}) - \Phi(z_{01}, z_{02})$$

$$P_{11} = \int_{z_{02}}^{z_{12}}\int_{z_{01}}^{z_{11}} \phi(z_1, z_2, \rho)\, dz_1\, dz_2 = \Phi(z_{11}, z_{12}) - \Phi(z_{01}, z_{12}) - \Phi(z_{11}, z_{02}) + \Phi(z_{01}, z_{02})$$

$\vdots$

$$P_{jk} = \int_{z_{(k-1)2}}^{z_{k2}}\int_{z_{(j-1)1}}^{z_{j1}} \phi(z_1, z_2, \rho)\, dz_1\, dz_2$$

$$= \Phi(z_{j1}, z_{k2}) - \Phi(z_{(j-1)1}, z_{k2}) - \Phi(z_{j1}, z_{(k-1)2}) + \Phi(z_{(j-1)1}, z_{(k-1)2})$$

$$\vdots$$

$$P_{(r-1)(c-1)} = \int_{z_{(c-1)2}}^{\infty} \int_{z_{(r-1)1}}^{\infty} \phi(z_1, z_2, \rho)\, dz_1\, dz_2$$

$$= 1 - \Phi(z_{(r-1)1}) - \Phi(z_{(c-1)2}) + \Phi(z_{(r-1)1}, z_{(c-1)2})$$

RESULTS AND DISCUSSION

The following is the discussion of parameter estimation and test statistic of bivariate binary probit model $(r \times c)$.

Parameter Estimation

Table 1 shows that the event on each respondent will have Multinomial distribution, i.e. $(Y_{10}, Y_{01}, Y_{11}, \cdots, Y_{(r-1)(c-1)}) \sim M(1; P_{10}, P_{01}, P_{11}, \cdots, P_{(r-1)(c-1)})$. Based on (Greene, 2008) and (Gujarati, 2003), the parameters on probit model could be estimated by using MLE method. The initial step to gain parameter estimation by MLE method is by taking n sampel randomly, i.e.

$$\left(Y_{00i}, Y_{01i}, Y_{10i}, Y_{11i}, \cdots, Y_{(r-1)(c-1)i} \,\middle|\, X_{1i}, X_{2i}, \cdots, X_{pi}\right), \text{ where } i = 1, 2, ..., n.$$

Function of the likelihood can be written as follows.

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} P\left(Y_{00i} = y_{00i}, Y_{11i} = y_{10i}, Y_{01i} = y_{01i}, \cdots,, Y_{(r-1)(c-1)i} = y_{(r-1)(c-1)i}\right)$$

or

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{j=0}^{r-1} \prod_{k=-0}^{c-1} P_{jk}^{Y_{jki}}.$$

Function ln of the likelihood is available on the equation (1).

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{j=0}^{r-1} \sum_{k=0}^{c-1} \left[y_{jki} \ln P_{jki} \right] \tag{1}$$

Probability on the equation (1) contains parameter $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T & \boldsymbol{\beta}_2^T \end{bmatrix}^T$. Then, $\ln L(\boldsymbol{\beta})$ is derived to its parameter $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, i.e.

$$\frac{\partial \ln L(\cdot)}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial}{\partial \boldsymbol{\beta}_1^T} \left[\sum_{i=1}^{n} \sum_{j=0}^{r-1} \sum_{k=0}^{c-1} \left[y_{jki} \ln P_{jki} \right] \right] = \sum_{i=1}^{n} \sum_{j=0}^{r-1} \sum_{k=0}^{c-1} \left[y_{jki} \frac{1}{P_{jki}} \frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T} \right] \tag{2}$$

and

$$\frac{\partial \ln L(\cdot)}{\partial \boldsymbol{\beta}_2^T} = \frac{\partial}{\partial \boldsymbol{\beta}_2^T}\left[\sum_{i=1}^{n}\sum_{j=0}^{r-1}\sum_{k=0}^{c-1}\left[y_{jki}\ln P_{jki}\right]\right] = \sum_{i=1}^{n}\sum_{j=0}^{r-1}\sum_{k=0}^{c-1}\left[y_{jki}\frac{1}{P_{jki}}\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2^T}\right] \tag{3}$$

where:

$$P_{jk} = \Phi(z_{j1}, z_{k2}) - \Phi(z_{(j-1)1}, z_{k2}) - \Phi(z_{j1}, z_{(k-1)2}) + \Phi(z_{(j-1)1}, z_{(k-1)2})$$

$$\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1^T} - \frac{\partial \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1^T} - \frac{\partial \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1^T} + \frac{\partial \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1^T} \tag{4}$$

$$\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2^T} = \frac{\partial \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2^T} - \frac{\partial \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2^T} - \frac{\partial \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2^T} + \frac{\partial \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2^T}. \tag{5}$$

Equations (4) and (5) is obtained by substituting:

$$\frac{\partial \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial}{\partial \boldsymbol{\beta}_1^T}\int_{-\infty}^{z_{j1i}}\int_{-\infty}^{z_{k2i}} \phi(z_{1i}, z_{2i})dz_1 dz_2$$

$$\frac{\partial \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial}{\partial \boldsymbol{\beta}_1^T}\int_{-\infty}^{z_{(j-1)1i}}\int_{-\infty}^{z_{k2i}} \phi(z_{1i}, z_{2i})dz_1 dz_2$$

$$\frac{\partial \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial}{\partial \boldsymbol{\beta}_1^T}\int_{-\infty}^{z_{j1i}}\int_{-\infty}^{z_{(k-1)2i}} \phi(z_{1i}, z_{2i})dz_1 dz_2$$

$$\frac{\partial \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1^T} = \frac{\partial}{\partial \boldsymbol{\beta}_1^T}\int_{-\infty}^{z_{(j-1)1i}}\int_{-\infty}^{z_{(k-1)2i}} \phi(z_{1i}, z_{2i})dz_1 dz_2$$

The estimates obtained by Equations (2) and (3) are not close form, then one of numerical approaches that can be used to find the estimates is Newton-Raphson method. Through the process of Newton-Raphson iteration, the maximum likelihood estimator can be obtained for $\boldsymbol{\beta}$, where $\boldsymbol{\beta}^{(m)}$ is the parameter estimation at iteration m. Newton Raphson iteration process is the need of the vector $\mathbf{g}(\boldsymbol{\beta})$ and the Hessian matrix. Vector $\mathbf{g}(\boldsymbol{\beta})$ is the first derivative of the function ln likelihood to its parameters. Hessian Matrix elements $\mathbf{H}(\boldsymbol{\beta})$ are the second derivatives of their parameter. Vector component $\mathbf{g}(\boldsymbol{\beta})$ which sizes $[2(p+1)\times 1]$ is as follows.

$$\mathbf{g}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1^T} \\ \dfrac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2^T} \end{bmatrix}_{[2(p+1)\times 1]}$$

The vector components $\mathbf{g}(\boldsymbol{\beta})$ could be calculated in equation (2) and (3). Hessian matrix sizes is $[2(p+1)\times 2(p+1)]$, i.e:

$$\mathbf{H}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} & \dfrac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \\ \dfrac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} & \dfrac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} \end{bmatrix}_{[2(p+1)\times 2(p+1)]}$$

In details, matrix components of Hessian could be found on equations (6), (7), and (8). The equation (6) is derived from equation (2) to $\boldsymbol{\beta}_1$, i.e.

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} = \sum_{k=0}^{c-1}\sum_{j=0}^{r-1}\sum_{i=1}^{n}\left[y_{jki} \frac{\partial}{\partial \boldsymbol{\beta}_1}\left(\frac{1}{P_{jki}}\right)\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T} + y_{rci}\left(\frac{1}{P_{jki}}\right)\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right] \tag{6}$$

where:

$$\frac{\partial}{\partial \boldsymbol{\beta}_1}\left(\frac{1}{P_{jki}}\right) = -\frac{1}{P_{jki}^2}\left(\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1}\right).$$

The value of $\dfrac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T}$ is calculated from equation (4) and

$$\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} = \frac{\partial^2 \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} - \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} - \frac{\partial^2 \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}.$$

Equation (7) is obtained by deriving equation (3) to $\boldsymbol{\beta}_2$, i.e.

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} = \sum_{k=0}^{c-1}\sum_{j=0}^{r-1}\sum_{i=1}^{n}\left[y_{jki} \frac{\partial}{\partial \boldsymbol{\beta}_2}\left(\frac{1}{P_{jki}}\right)\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2^T} + y_{jki}\left(\frac{1}{P_{jki}}\right)\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right] \tag{7}$$

where:

$$\frac{\partial}{\partial \boldsymbol{\beta}_2}\left(\frac{1}{P_{jki}}\right) = -\frac{1}{P_{jki}^2}\left(\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2}\right).$$

Whereas, $\dfrac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2^T}$ is obtained from (5) and

$$\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} = \frac{\partial^2 \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} - \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} - \frac{\partial^2 \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}.$$

Equation (8) is found from equation (2) to $\boldsymbol{\beta}_2$, i.e.

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} = \sum_{k=0}^{c-1}\sum_{j=0}^{r-1}\sum_{i=1}^{n}\left[y_{jki} \frac{\partial}{\partial \boldsymbol{\beta}_2}\left(\frac{1}{P_{jki}}\right)\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T} + y_{jki}\left(\frac{1}{P_{jki}}\right)\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T}\right] \tag{8}$$

where:

$$\frac{\partial}{\partial \boldsymbol{\beta}_2}\left(\frac{1}{P_{jki}}\right) = -\frac{1}{P_{jki}^2}\left(\frac{\partial P_{jki}}{\partial \boldsymbol{\beta}_2}\right).$$

Then, $\dfrac{\partial P_{jki}}{\partial \boldsymbol{\beta}_1^T}$ is derived from equation (4) and

$$\frac{\partial^2 P_{jki}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} = \frac{\partial^2 \Phi(z_{j1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} - \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{k2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} - \frac{\partial^2 \Phi(z_{j1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} + \frac{\partial^2 \Phi(z_{(j-1)1i}, z_{(k-1)2i})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T}.$$

Parameter Significance Testing

To test the goodness of fit of the model, the parameters are evaluated. It is intended to determine whether the predictor variables contained in the model have a significant effect or not. Testing of the model parameters is carried out either at simultaneously and individual. The method used to obtain the test statistic is MLRT.

Simultaneously hypothesis is a hypothesis that state whether the variable $x_1, x_2, ..., x_p$ has a significant effect on the response variable P_{jk}. The hypotheses are:

$$H_0 : \beta_{11} = \beta_{12} = \cdots = \beta_{1p} = \beta_{21} = \beta_{22} = \cdots = \beta_{2p} = 0$$

H_1: at least has one $\beta_{st} \neq 0$, where $s = 1, 2$ and $t = 1, 2, ..., p$.

The parameter under population is $\Omega = \{ \beta_{10}, \beta_{11}, \ldots, \beta_{1p}, \beta_{20}, \beta_{21}, \ldots, \beta_{2p} \}$, whereas the parameter under H$_0$ is $\omega = \{ \beta_{10}, \beta_{20} \}$.

Test Statistic is obtained by making ratios of $L(\hat{\omega})$ and $L(\hat{\Omega})$, $\Lambda = \dfrac{L(\hat{\omega})}{L(\hat{\Omega})}$. H$_0$ is rejected if $\Lambda = \dfrac{L(\hat{\omega})}{L(\hat{\Omega})} < \lambda_0$, where $0 < \lambda_0 < 1$. Thus, $G^2 = -2\ln\Lambda$ is obtained and according to (Agresti, 2002), G^2 follows χ^2 distribution, i.e.

$$G^2 = -2\ln\Lambda = -2\ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2\ln L(\hat{\Omega}) - 2\ln L(\hat{\omega}).$$

In details, $L(\hat{\Omega})$ is:

$$L(\hat{\Omega}) = \left[\Phi(z_{01}, z_{02})\right]^{y_{00}} \times \left[\Phi(z_{11}, z_{02}) - \Phi(z_{01}, z_{02})\right]^{y_{01}} \times \left[\Phi(z_{11}, z_{02}) - \Phi(z_{01}, z_{02})\right]^{y_{10}} \times$$

$$\times\left[\Phi(z_{j1}, z_{k2}) - \Phi(z_{(j-1)1}, z_{k2}) - \Phi(z_{j1}, z_{(k-1)2}) + \Phi(z_{(j-1)1}, z_{(k-1)2})\right]^{y_{jk}} \times \cdots \times$$

$$\times\left[1 - \Phi(z_{(r-1)1}) - \Phi(z_{(c-1)2}) + \Phi(z_{(r-1)1}, z_{(c-1)2})\right]^{y_{(r-1)(c-1)}}$$

or

$$L(\hat{\Omega}) = \left[\Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{02} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x})\right]^{y_{00}} \times \left[\Phi(\gamma_{11} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{02} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) - \Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{02} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x})\right]^{y_{01}} \times$$

$$\times \left[\Phi(\gamma_{11} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{02} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) - \Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{02} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x})\right]^{y_{10}} \times$$

$$\times \left[\Phi(\gamma_{j1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{k2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) - \Phi(\gamma_{(j-1)1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{k2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) - \Phi(\gamma_{j1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{(k-1)2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) + \right.$$

$$\left. + \Phi(\gamma_{(j-1)1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{(k-1)2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x})\right]^{y_{jk}} \times \cdots \times$$

$$\times \left[1 - \Phi(\gamma_{(r-1)1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}) - \Phi(\delta_{(c-1)2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x}) + \Phi(\gamma_{(r-1)1} - \hat{\boldsymbol{\beta}}_1^T \mathbf{x}, \delta_{(c-1)2} - \hat{\boldsymbol{\beta}}_2^T \mathbf{x})\right]^{y_{(r-1)(c-1)}}.$$

The values of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ could be obtained by using equations (2) and (3). Moreover, the value of $L(\hat{\omega})$ is as follows:

$$L(\hat{\omega}) = \left[\Phi(\gamma_{01} - \hat{\beta}_{10}, \delta_{02} - \hat{\beta}_{20})\right]^{y_{00}} \times \left[\Phi(\gamma_{11} - \hat{\beta}_{10}, \delta_{02} - \hat{\beta}_{20}) - \Phi(\gamma_{01} - \hat{\beta}_{10}, \delta_{02} - \hat{\beta}_{20})\right]^{y_{01}} \times$$

$$\times \left[\Phi(\gamma_{11} - \hat{\beta}_{10}, \delta_{02} - \hat{\beta}_{20}) - \Phi(\gamma_{01} - \hat{\beta}_{10}, \delta_{02} - \hat{\beta}_{20})\right]^{y_{10}} \times$$

$$\times \left[\Phi(\gamma_{j1} - \hat{\beta}_{10}, \delta_{k2} - \hat{\beta}_{20}) - \Phi(\gamma_{(j-1)1} - \hat{\beta}_{10}, \delta_{k2} - \hat{\beta}_{20}) - \Phi(\gamma_{j1} - \hat{\beta}_{10}, \delta_{(k-1)2} - \hat{\beta}_{20}) + \right.$$

$$\left. + \Phi(\gamma_{(j-1)1} - \hat{\beta}_{10}, \delta_{(k-1)2} - \hat{\beta}_{20})\right]^{y_{jk}} \times \cdots \times$$

$$\times \left[1 - \Phi(\gamma_{(r-1)1} - \hat{\beta}_{10}) - \Phi(\delta_{(c-1)2} - \hat{\beta}_{20}) + \Phi(\gamma_{(r-1)1} - \hat{\beta}_{10}, \delta_{(c-1)2} - \hat{\beta}_{20})\right]^{y_{(r-1)(c-1)}}.$$

Then, the values $\hat{\beta}_{10}$ and $\hat{\beta}_{20}$ could be found by using equations (9) and (10). Derivative of ln likelihood to β_{10} is

$$\frac{\partial \ln L(\cdot)}{\partial \beta_{10}} = \sum_{i=1}^{n} \sum_{j=0}^{r-1} \sum_{k=0}^{c-1} \left[y_{jki} \frac{1}{P_{jki}^*} \frac{\partial P_{jki}^*}{\partial \beta_{10}} \right]. \tag{9}$$

Thus, the result of $\dfrac{\partial P_{jki}^*}{\partial \beta_{10}}$ is

$$\frac{\partial P_{jki}^*}{\partial \beta_{10}} = \frac{\partial \Phi(z_{j1i}^*, z_{k2i}^*)}{\partial \beta_{10}} - \frac{\partial \Phi(z_{(j-1)1i}^*, z_{k2i}^*)}{\partial \beta_{10}} - \frac{\partial \Phi(z_{j1i}^*, z_{(k-1)2i}^*)}{\partial \beta_{10}} + \frac{\partial \Phi(z_{(j-1)1i}^*, z_{(k-1)2i}^*)}{\partial \beta_{10}}$$

where $z_{j1i}^* = \gamma_{j1i} - \beta_{10}$, $z_{(j-1)1i}^* = \gamma_{(j-1)1i} - \beta_{10}$ and $z_{k2i}^* = \delta_{k2i} - \beta_{20}$, $z_{(k-1)2i}^* = \delta_{(k-1)2i} - \beta_{20}$.

While, the first derivative of $\ln(\boldsymbol{\beta})$ to β_{20} is

$$\frac{\partial \ln L(\cdot)}{\partial \beta_{20}} = \sum_{i=1}^{n}\sum_{j=0}^{r-1}\sum_{k=0}^{c-1}\left[y_{jki} \frac{1}{P^*_{jki}} \frac{\partial P^*_{jki}}{\partial \beta_{20}} \right] \tag{10}$$

where the derivative of probability P^*_{2i} and P^*_{01i} to β_{20} is

$$\frac{\partial P^*_{jki}}{\partial \beta_{20}} = \frac{\partial \Phi(z^*_{j1i}, z^*_{k2i})}{\partial \beta_{20}} - \frac{\partial \Phi(z^*_{(j-1)1i}, z^*_{k2i})}{\partial \beta_{20}} - \frac{\partial \Phi(z^*_{j1i}, z^*_{(k-1)2i})}{\partial \beta_{20}} + \frac{\partial \Phi(z^*_{(j-1)1i}, z^*_{(k-1)2i})}{\partial \beta_{20}} .$$

If $G^2 > \chi^2_{\alpha,v}$ then H_0 is rejected, where degree of freedom (v) is the number of model parameters under population subtracted by the number of model parameters under H_0, with a large population. After conducting the test of parameter significance simultaneously, the next step is the partial testing. In this testing, it is desirable to know the contribution of each predictor variable. The hypothesis of individual testing on bivariate binary probit model is

H_0 : $\beta_{st} = 0$

H_1 : $\beta_{st} \neq 0$, where $s = 1,2$ and $t = 0,1,2,...,p$.

Set of the parameters (ω) under the null hypothesis is

$$\omega = \left\{ \beta_{s^*t^*}, s^* = 1,2; t^* = 0,1,...,p; s^* \neq s, t^* \neq t \right\}$$

where:

$$L(\hat{\Omega}) = \left[\Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{02} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) \right]^{y_{00}} \times \left[\Phi(\gamma_{11} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{02} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) - \Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{02} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) \right]^{y_{01}} \times$$

$$\times \left[\Phi(\gamma_{11} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{02} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) - \Phi(\gamma_{01} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{02} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) \right]^{y_{10}} \times$$

$$\times \left[\Phi(\gamma_{j1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{k2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) - \Phi(\gamma_{(j-1)1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{k2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) - \Phi(\gamma_{j1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{(k-1)2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) + \right.$$

$$\left. + \Phi(\gamma_{(j-1)1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{(k-1)2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) \right]^{y_{jk}} \times \cdots \times$$

$$\times \left[1 - \Phi(\gamma_{(r-1)1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*) - \Phi(\delta_{(c-1)2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) + \Phi(\gamma_{(r-1)1} - \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{x}^*, \delta_{(c-1)2} - \hat{\boldsymbol{\beta}}_2^{*T}\mathbf{x}^*) \right]^{y_{(r-1)(c-1)}}$$

and $\mathbf{x}^* = \{1, x_1, x_2, ..., x_{(t-1)}, x_{(t+1)}, ..., x_p\}$

$\hat{\boldsymbol{\beta}}_1^* = \{\beta_{10}, \beta_{11}, ..., \beta_{1(t-1)}, \beta_{1(t+1)}, ..., \beta_{1p}\}$ and $\hat{\boldsymbol{\beta}}_2^* = \{\beta_{20}, \beta_{21}, ..., \beta_{2p}\}$ or

$\hat{\boldsymbol{\beta}}_1^* = \{\beta_{10}, \beta_{11}, ..., \beta_{1p}\}$ and $\hat{\boldsymbol{\beta}}_2^* = \{\beta_{20}, \beta_{21}, ..., \beta_{2(t-1)}, \beta_{2(t+1)}, ..., \beta_{2p}\}$.

The partial test statistic was done by using MLRT method as in simultaneous testing, so test statistic t was gained, that is $t = \frac{\hat{\beta}_{rs}}{SE(\hat{\beta}_{rs})}$. For large sample, t follows Normal standard distribution, i.e. $t \sim N(0,1)$. If $|t| > Z_{\alpha/2}$, then the null hypothesis is rejected.

Conclusion

This paper already discussed about theoretical part of bivariate probit model, particularly about estimation method and test statistic. The results showed that estimation of model parameters by using MLE yielded not closed form solution. Then, these estimated parameters could be obtained by Newton-Raphson iteration. Furthermore, two test statistics to validate significance of model parameters could be constructed by MLRT method based on asymptotic properties of these estimators, i.e. G^2 for overall test and t for partial test.

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Aitchison, J. & Silvey, S.D. (1957). The Generalization of Probit Analysis to the Case of Multiple Responses. *Biometrika* 44(2), 131-140.

Boes, S. & Winkelmann, R. (2005), Ordered Response Models, *Technical Report*, Socioeconomic Institute, University of Zurich.

Bokosi, F.K. (2007). Household Poverty Dinamics in Malawi: A Bivariate Probit Analysis. *Journal of Applied Sciences*. 7(2), 258-262.

Greene, W.H. (2008), *Econometrics Analysis*, Fourth Edition, Prentice Hall, Englewood Cliffs, New Jersey.

Gujarati, D.N.(2003). *Basic Econometric.* Fourth Edition. McGraw Hill, New York.

Jackman, S. (2000), Models for Ordered Outcomes, *Technical Report*, Stanford University.

McKelvey, R.D. & Zavoina, W. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology* 4, 103-120.

Mozumder, P., Raheem, N., Talberth, J., & Berrens, R.P. (2008). Investigating intended evacuation from wildfires in the wildland–urban interface: Application of a bivariate probit model. *Forest Policy and Economics* 10, 415-423.

Ronning, G & Kukuk, M. (1996). Efficient Estimation of Ordered Probit Models. *Journal of the American Statistical Association*. 91(435), 1120-1129.

Snapinn, S.M. & Small, R.D. (1986). Test of Significance Using Regression Models for Ordered Categorical Data. *Biometrics* 42, 583-592.

Yamamoto, T. & Shankar,V.N. (2004). Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. *Accident Analysis & Prevention* 36, 869-876.