IOP Conference Series: Materials
Science and Engineering

PAPER • **OPEN ACCESS**

Modeling of Parity Status of The Mother and Basic Immunization Giving to Infants with Semiparametric Bivariate Probit (Case Study: North Kalimantan Province in 2017)

To cite this article: Rahmi Amelia *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052004

IOP Conf. Series: Materials Science and Engineering **546** (2019) 052004 doi:10.1088/1757-899X/546/5/052004

Modeling of Parity Status of The Mother and Basic Immunization Giving to Infants with Semiparametric Bivariate Probit (Case Study: North Kalimantan Province in 2017)

Rahmi Amelia[1], Muhammad Mashuri,[1] Vita Ratnasari, M.Si [1*]

[1]Institut Teknologi Sepuluh Nopember (ITS), Surabaya 60111, Indonesia

*corresponding author: vitaratna70@gmail.com

Abstract. The bivariate probit regression model is a probit regression model consisting of two response variables with errors between the two variables correlate each other. The correlation between the two response variables can occur as a result of the presence of endogeneity, a condition in which a response variable becomes an exogenous variable in another response variable. Besides, the important issue that cannot be underestimated is undetectable nonlinear relationships between response variables and predictors, especially discrete or continuous predictor variables. The bivariate probit regression that does not ignore endogeneity cannot detect the nonlinear relationships between response variables and predictors, so one of the regression models that can overcome the problem is bivariate probit regression model with a semiparametric approach. The first step in semiparametric bivariate probit modeling is testing the hypothesis of exogeneity to determine whether there is a case of endogeneity or not. The exogenous test used in this study is the Lagrange Multiplier (LM) and Likelihood Ratio (LR) test. The data used in this study consisted of two binary categorical response variables, they are parity status of the mother and basic immunization giving to infants in North Kalimantan Province in 2017. The results of the exogenous test using the LM test and LR test stated that there was a significant correlation between response variables. The AIC value of the semiparametric bivariate probit model is 1301.602, while the bivariate probit model produces AIC of 1316.789, so it can be concluded that the semiparametric bivariate probit model provides better modeling results than the bivariate probit model.

1. Introduction

Dependent variables or response variables encountered in everyday life are often in the form of categorical data. Categorical data is often referred to as qualitative data to distinguish from quantitative variables such as weight, age, income, and a number of children in a family [1].

The suitable regression analysis for categorical response variables assumed to follow binomial distribution is logistic regression or probit regression. Both are included in the Generalized Linear Model (GLM). Logistic regression uses the link function while probit regression uses the standard normal cumulative function. One advantage of using probit regression is that the values obtained from fitting models can be directly converted into probabilities by using values from the standard normal distribution table.

Categorical response variables are used to have only two categories, namely "yes" or "no". Probit regression which response variables only have two categories is called binary probit regression and called multinomial probit regression if there are more than two categories.

Probit regression which consists of one response variable is called univariate probit regression. The univariate probit regression then develops into a bivariate probit regression, namely probit regression which consists of two correlated response variables. Chen and Tsurumi [2] implemented a bivariate probit model to analyze the factors that influence differences in formal sector employment between men and women in urban areas in China while Coates, et.al. [3] implemented a bivariate probit model in evaluating the effect of external funding on financial problems and volunteering at a nonprofit sports organization in Germany. Ratnasari,et.al. [4] examined parameter estimates and test statistics from the bivariate probit regression model using Maximum Likelihood Estimation (MLE) and Maximum Likelihood Ratio Test (MLRT). According to Ratnasari [5], the first step in obtaining the parameter estimates of the bivariate probit model is to form a contingency table and calculate the probability of each cell.

The correlation between the two binary response variables can also be caused due to an endogenous case [6]. The endogeneity can occur if there are unobserved variables that are related to binary predictors and binary responses. If the endogeneity is ignored in modeling, it can lead to biased and inconsistent estimation results. Another thing that is of concern is the possibility of a nonlinear relationship between response variables and predictors, especially continuous predictor variables. The non-detected nonlinearity will also affect the estimation results.

Probit regression which does not overlook endogenous issues and allows for flexible dependencies of the response variables in continuous predictor variables is a bivariate probit model with a semiparametric approach [6]. Ieva, et.al. [7] analyzed the relationship of hospital mortality with the effectiveness of patients affected by myocardial ST-elevation (STEMI) using semiparametric bivariate probit regression. The same thing was done by Marra and Radice [6] in the survey data in Botswana regarding the relationship between education and female fertility rates. Umami, et.al. [8] applied the semiparametric bivariate probit model for working the mother status and exclusive breastfeeding in Bangkalan Regency.

The semiparametric bivariate probit model uses the smooth function to overcome endogeneity and anticipate nonlinearity. Hastie and Tibshirani [9] introduced generalized additive models (GAM) that replace linear shapes with smooth functions. Marra and Radice [10] using the penalized likelihood approach to GAM with spline regression. The method is applied to cross section data which aims to analyze the relationship between BC plasma concentrations and personal characteristics. The results showed that estimation using penalized regression splines that allowed nonlinearity to provide results that were more informative than if only using parametric components. Marra and Radice [6] using fisher scoring iterations as the optimization method that have similarities to the Newton Raphson method. According to Okeh and Oyeka [11], Fisher scoring iterations are considered to have better performance in overcoming the possibility of non-convergence.

The problem of endogeneity found in the semiparametric bivariate probit model can be identified by conducting a test called the exogenous test. Marra, Radice, and Filippou [12] discussed the exogenous test in the use of recursive spline regression and sample selection of bivariate probit models. There are four exogeneity tests discussed, including Lagrange Multiplier, Wald, Likelihood Ratio, and Gradient Test. The four tests were applied to survey data regarding the impact of ownership of health insurance on the use of health facilities in the USA. The results showed that only the gradient test failed to reject Ho. Monfardini and Radice [13] had conducted a Monte Carlo experiment which aims to test the exogenity of a number of samples of bivariate probit models with several alternative tests before. There are four tests used, among others, Likelihood Ratio, t-test, Lagrange Multiplier, and Conditional Moment (CM).

Infectious diseases are still the main cause of death throughout the world. According to the Health Law Number 36 of 2009, immunization is the top priority of the Ministry of Health in order to prevent infectious diseases and as a concrete form of commitment from the government to reduce mortality in children. The immunization program that has existed since 1956 has also proven effective and efficient. One of the real positive impacts is that since 1974 Indonesia was declared free of smallpox.

However, there are things that are quite concerning in the territory of the Unitary Republic of Indonesia, there are still children who have not received complete immunization and some have never received any immunizations at all. The Ministry of Health states that there are at least 1.7 million children who have not received immunizations or have not yet completed their immunization status.

The complete basic immunization coverage for a total of seven types of vaccines nationally reaches 85 percent. In each type of vaccine, in some provinces and certain years, coverage of complete basic immunization is only in the range of 50-60 percent. High national achievement figures do not necessarily make the achievement of an immunization evenly across Indonesia. North Kalimantan Province is one of the provinces that must be considered. Complete basic immunization coverage for infants in this province has always been low in recent years. There is no coverage that exceeds 75 percent.

Sustainability and increasing the coverage of government programs are certainly inseparable from the role of parents, especially the mothers. The mothers with multiparas classified as having more knowledge about immunization because they are considered more experienced than primiparous the mothers. In addition, the age of the mother's first marriage, education, employment, participation in family planning, and ownership of health insurance are also suspected to have an impact on the completeness of basic immunization giving.

Previous research examining the relationship between parity and basic immunization has been done by Silviana [14] which states that there is a significant relationship between parity and the level of knowledge of the mothers regarding basic immunization in infants at the Umbulharjo I Health Center in Yogyakarta. The study of Dwiyanti [15] also stated that parity had a significant effect on the completeness of basic immunization in infants at Gribig Malang Health Center.

Factors influencing the completeness of basic immunization giving to infants can be known through regression analysis. Regression analysis is a method used to assess dependencies between response variables to one or more predictor variables in order to estimate the value of response variables based on the value of known predictor variables [16].

Based on the background described, the method that is able to analyze the effect of parity status of the mother on complete basic immunization giving to infants in North Kalimantan Province in 2017 is semiparametric bivariate probit which is then compared to bivariate probit that does not ignore the presence of endogeneity based on AIC values.

2. Theoretical Review

In this section, we will review some of the theories used.

2.1. Semiparametric Bivariate Probit Model

Endogeneity is a condition where the response variable becomes a predictor variable in another response variable. If the case of endogeneity is ignored, it can cause the estimation results to be biased and inconsistent [7]. The bivariate probit regression model that does not ignore the endogenous case are as follows

$$y_{1i}^{*} = \mathbf{x}_{1i}^{T}\boldsymbol{\alpha}_1 + \varepsilon_{1i} \tag{1}$$

$$y_{2i}^{*} = \beta y_{1i} + \mathbf{x}_{2i}^{T}\boldsymbol{\alpha}_2 + \varepsilon_{2i} \quad i = 1,\ldots,n \tag{2}$$

y_{1i}^{*} and y_{2i}^{*} are continues latent variables

$y_{vi} = 1_{\{y_{v1}^{*}>0\}},\ v = 1,2$

$\mathbf{x}_{1i}^{T} = \left(1, \mathbf{x}_{12i}, \ldots, \mathbf{x}_{1p_1 i}\right)_{n\,\mathrm{x}P_1}$

$\boldsymbol{\alpha}_1$ = parameters vector

$\mathbf{x}_{2i}^T = \left(1, \mathbf{x}_{21i}, \ldots, \mathbf{x}_{2p_1 i}\right)_{n\,\mathrm{x}P_2}$
$\boldsymbol{\alpha}_2$ = coefficient vector
β = parameter of binary endogenous variable y_{1i}

Semiparametric bivariate probit regression is bivariate probit regression that does not ignore the endogeneity problem and included in Generalized Additive Models (GAM). The advantage of using the semiparametric bivariate probit regression model is being able to handle the undetectable nonlinearity of the continuous predictor variable. The undetectable nonlinearity can lead to errors in modeling which also affect the accuracy of the estimation results. The model of semiparametric bivariate probit regression is expressed in the following equation

$$y_{1i}^* = \mathbf{x}_{1i}^T \boldsymbol{\delta}_1 + \sum_{k_1=1}^{K_1} s_{1k_1}\left(\mathbf{x}_{1k_1 i}\right) + \varepsilon_{1i} \tag{3}$$

$$y_{2i}^* = \beta y_{1i} + \mathbf{x}_{2i}^T \boldsymbol{\delta}_2 + \sum_{k_2=1}^{K_2} s_{2k_2}\left(\mathbf{x}_{2k_2 i}\right) + \varepsilon_{2i} \quad i = 1,\ldots,n \tag{4}$$

Where $i = 1, \ldots, n$, $j = 1,2$

y_{1i}^* and y_{2i}^* are continues latent variables
$y_{vi} = 1_{\{y_{v1}^* > 0\}}$, $v = 1,2$
$\mathbf{x}_{ji}^T = \left(1, \mathbf{x}_{j2i}, \ldots, \mathbf{x}_{jQ_j i}\right)_{n\,\mathrm{x}Q_j}$ as parametric model components
$\boldsymbol{\delta}_1$ = corresponding parameter vector of $\mathbf{x}_{1i}^T$
$\boldsymbol{\delta}_2$ = coefficient vector
s_{jk_j} = smooth function of K_j continuous covariates $\mathbf{x}_{jk_j i}$
β = parameter of binary endogenous variable y_{1i}
ε_{1i} and ε_{2i} is assumed to follow normal distribution $\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \theta_3\\ \theta_3 & 1\end{bmatrix}\right)$
Based on equation (3) dan (4) if the errors are correlated so it is recommended to do parameter estimation simultaneously. So that linear predictors can be defined as follows:

$$\eta_{1i} = \mathbf{x}_{1i}^T \theta_1 \tag{5}$$

$$\eta_{2i} = \mathbf{x}_{2i}^T \boldsymbol{\theta}_2 \tag{6}$$

Both binary response variables y_{1i} dan y_{2i} of probit bivariate model can be formed as follows:
$\mathrm{P}(y_{1i} = 1, y_{2i} = 1 | \mathrm{x}_{1i}, \mathrm{x}_{2i}) = \Phi_2(\eta_{1i}, \eta_{2i}; \boldsymbol{\theta}_3)$
Φ_2 is bivariate normal distribution function with zero means, diagonal of variance-covariance matrix is 1 and correlation $\boldsymbol{\theta}_3$. Because of $\boldsymbol{\theta}_3$ is limited between [-1,1], so that $\boldsymbol{\theta}_3$ is transformed to $\boldsymbol{\theta}_3^* = \tanh^{-1}(\boldsymbol{\theta}_3)$. Writing $P(y_{1i} = m_1, y_{2i} = m_2)$ as $p_{m_1 m_2 i}$ where $m_v = 0,1$ and $v = 1,2$
The categorizing can be defined as:
$y_{m_1} y_{m_2} i = \begin{cases} 1, \text{if } (y_{1i} = m_1 \text{ dan } y_{2i} = m_2) \\ 0, \text{others} \end{cases}$

2.2. Parameter Estimation of Semiparametric Bivariate Probit

The method used to estimate the parameters of bivariate probit model is Maximum Likelihod Estimation while the method used to estimate the parameters of bivariate semiparametric probit model is Penalized Maximum Likelihood Estimation using the equation

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\theta}^T \boldsymbol{S}_\lambda \boldsymbol{\theta} \tag{7}$$

where the log-likelihood function is

$$\boldsymbol{\ell}(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_{11i} \log (p_{11i}) + y_{01i}\log(p_{01i}) + y_{10i}\log(p_{01i}) + y_{00i}\log(p_{00i})) \tag{8}$$

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1 \boldsymbol{\theta}_2, \theta_3)$

The equation obtained in parameters estimate of probit bivariate semiparametric model is not closed form, hence, Fisher Scoring iteration is used to solve the equation based on following equation

$$\hat{\boldsymbol{\theta}}^{|a+1|} = \hat{\boldsymbol{\theta}}^{[a]} + (\mathfrak{I}^{[a]} + \boldsymbol{S}_\lambda)^{-1} (\mathbf{g}^{[a]} \boldsymbol{S}_\lambda \hat{\boldsymbol{\theta}}^{[a]}) \tag{9}$$

Equation (9) is also can be solved by implementing decomposition of $\mathfrak{I}$ in every iteration using *trust region* algorithm

$$\hat{\boldsymbol{\theta}}^{|a+1|} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i + \boldsymbol{S}_\lambda\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i \mathbf{z}_i\right) \tag{10}$$

Minimize the prediction error criterion as follows

$$\boldsymbol{V}_u^w(\boldsymbol{\lambda}) = \frac{1}{n}\left\|\sqrt{\mathbf{W}}(\boldsymbol{z} - \mathbf{X}\boldsymbol{\theta})\right\|^2 - 1 + \frac{2}{n}\gamma \operatorname{tr}(\mathbf{A}) \tag{11}$$

$\mathbf{A} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X} + \boldsymbol{S}_\lambda)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}$ is a hat matrix

2.3. Multicollinearity Test

One of the assumptions that must be fulfilled in probit regression modeling is the absence of multicollinearity between predictor variables. One indicator that can be used to identify the presence of multicollinearity is to see how much correlation between these variables. Correlation values can be calculated using Rank-Spearman correlation.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{12}$$

According to Gujarati and Porter [17], if the correlation between variables is more than 0.8, it can be concluded that multicollinearity occurs in both of these variables.

2.4. The Hypothesis Testing of Exogeneity

Hypothesis testing of this exogenous variable is used to determine the existence of exogenous variables in the model stated in ρ. If $\rho = 0$, ε_{1i} and ε_{2i} is not correlated so it can be identified that there is no endogeneity. The hypothesis testing of exogeneity use *Lagrange Multiplier* (LM) and Likelihood Ratio (LR) test with the following hypothesis :

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

and the test statistics is defined as:

$$LM = \left\{\mathbf{g}_{\hat{\delta}_{H0}} - \bar{S}_{\lambda_{Ho}} \hat{\delta}_{H0}\right\}^T \mathbf{I}^{-1} \left\{\mathbf{g}_{\hat{\delta}_{H0}} - \bar{S}_{\lambda_{Ho}} \hat{\delta}_{H0}\right\} \tag{13}$$

$$\begin{aligned} LR &= 2\left\{\ell\left(\hat{\boldsymbol{\delta}}_{H_1}\right) - \ell\left(\hat{\boldsymbol{\delta}}_{H_0}\right)\right\} \\ &= 2\left\{\ell\left(\hat{\boldsymbol{\delta}}_{H_1}\right) - \left[\ell_{M_1}\left(\hat{\boldsymbol{\delta}}_{H_0,1}\right) + \ell_{M_2}\left(\hat{\boldsymbol{\delta}}_{H_0,2}\right)\right]\right\} \end{aligned} \tag{14}$$

$$LR \xrightarrow{d} \chi^2_{dfH_1 - dfH_0}$$

H_0 rejected if $LM^* > \chi^2_{(1,\alpha)}$ or $LR > \chi^2_{\alpha,\left(\chi^2_{dfH_1 - dfH_0}\right)}$

2.5. The Goodness of Fit Semiparametric Bivariate Probit Model
In semiparametric bivariate probit, the goodness of fit criteria that can be used is Akaike Information Criteria or AIC [10]. AIC value is calculated based on the formula

$$AIC = D + 2tr(\mathbf{A}) \tag{15}$$

2.6. Parity Status and Basic Immunization
According to Varney, Kriebs, and Gegor [18] parity itself consists of two categories:

1. Primipara

Primipara is a woman who has had one pregnancy in which the fetus or fetuses reached the point of viability.

2. Multipara

Multipara is a woman who has had more than one pregnancy in which the fetus or fetuses reached the point of viability.

Immunization is one of the preventions of infectious diseases by giving certain bacterial or viral antigens that have been weakened or turned off in order to stimulate the body's immune system or form antibodies with the aim of increasing immunity so as to prevent or reduce the transmission of disease [19].

Each country has a different immunization program, depending on the priorities and health conditions in each country. The determination of this type of immunization is based on expert studies and epidemiological analysis of emerging diseases. In Indonesia, the immunization program requires every baby (ages 0-11 months) to get a complete basic immunization consisting of Hepatitis B, BCG, DPT-HB-Hib, Polio, and Measles.

3. Result and Discussion

3.1. Data
The data used in this study is secondary data taken from the Survey Sosial Ekonomi Nasional (SUSENAS) in North Kalimantan Province carried out by the Badan Pusat Statistik (BPS) in 2017. The response variables used in this study were maternal parity status and giving basic immunization. The response variables are categorized as follows:

1. Parity Status of the mother (Y_1) :
   Code 0 : primipara
   Code 1 : multipara
2. Basic Immunization Giving to Infants (Y_2) :
   Code 0 : not given complete basic immunization
   Code 1 : given complete basic immunization

This study uses four predictor variables categorized as follows:

1. Age of the mother's first marriage (X_1)
2. Education of The mother (X_2)
   Code 1 : Less than or equal Elementary School
   Code 2 : Junior High School
   Code 3 : Senior High School
   Code 4 : College
3. Occupation of The mother (X_3)
   Code 0 : Not working
   There are 19 categories of the mother's occupation sectors that are classified into 4 strata
   Code 1 : 1st strata (agriculture, horticulture, plantation, fisheries, animal husbandry, forestry, mining)
   Code 2 : 2nd strata (processing industries, construction, finance, and insurance)
   Code 3 : 3rd strata (trade, education services, health services, community services)

Code 4: 4th strata (electricity and gas, hotels and restaurants, transportation and warehousing, information and communication, others)

4. Participation in family planning (Keluarga Berencana/KB) (X_4)
   Code 1 : Ever underwent
   Kode 2 : Undergo
   Kode 5 : Not undergo
5. Ownership of health insurance (X_5)
   Code 0: Do not have any health insurance
   Code 1: Have at least one health insurance

The research consists of 672 households that have children aged 13-59 months.

Table 1. Percentage Based on Parity Status of The mother and Basic Immunization Giving

		Basic Immunization Giving		
		Not Complete	Complete	Total
Parity Status of The mother	Primipara	14	4	18
	Multipara	66	16	82
	Total	80	20	100

Table 1 shows that multiparous the mothers provide complete basic immunization of 12% greater than primiparous the mothers. It indicates that primiparous the mothers need to get more attention regarding the importance of basic immunization in infants.

The existence of multicollinearity among predictor variables can be identified by looking at the Spearman correlation coefficient. The processing results show that the correlation between predictor variables does not exceed 0.8, so it can be concluded that there is no multicollinearity between predictor variables.

3.2. Hypothesis Testing of Exogeneity

Parity status of the mother and basic immunization giving are thought to correlate with endogenous problems. The result of the Lagrange Multiplier test concludes that there is a correlation between the parity status of the mother and basic immunization giving that indicates the presence of endogeneity among both variables. It can be seen from the p-value = 0.02713663 (p-value less than α=5 %). This is also supported by the results of the Likelihood Ratio test which produces p-value = 0.02336232.

3.3. Semiparametric Bivariate Probit Model

The data in this study were processed by R studio program version 1.1.463.

Table 2. Coefficient and p-value values in each parameter of the Semiparametric Bivariate Probit Model (SBP) and Bivariate Probit (BP)

SBP				BP			
Equation 1							
	Estimate	Pr(>|z|)			Estimate	Pr(>|z|)	
(Intercept)	1.02303	1.30E-12	**	(Intercept)	1.78304	1.97E-07	**
x_{22}	-0.38421	0.0206	**	x_{22}	-0.3971	0.0171	**
x_{23}	-0.18294	0.2065		x_{23}	-0.23031	0.1138	
x_{24}	-0.06023	0.7918		x_{24}	-0.07479	0.7497	
x_{31}	0.18624	0.2187		x_{31}	0.1904	0.2173	
x_{32}	0.32799	0.256		x_{32}	0.32815	0.2438	
x_{33}	0.20889	0.398		x_{33}	0.23567	0.3443	
x_{34}	0.32791	0.2884		x_{34}	0.33597	0.2718	
x_{42}	0.0557	0.6362		x_{42}	0.10971	0.3962	
x_{45}	-0.22307	0.0592		x_{45}	-0.13843	0.317	
	edf	p-value					
$s(x_1)$	1.754	0.0125	**	x_1	-0.03835	0.0143	**
Equation 2							
	Estimate	Pr(>|z|)			Estimate	Pr(>|z|)	
(Intercept)	-1.82682	<2e-16	**	(Intercept)	-2.3745	3.92E-16	**
y_{11}	1.352889	<2e-16	**	y_{11}	1.369721	< 2e-16	**
x_{22}	0.318565	0.0222	**	x_{22}	0.326201	0.0188	**
x_{23}	0.247345	0.0461	**	x_{23}	0.224401	0.0661	
x_{24}	-0.01595	0.9365		x_{24}	-0.00746	0.9698	
x_{31}	-0.26925	0.0388	**	x_{31}	-0.28772	0.0264	**
x_{32}	-0.17296	0.4633		x_{32}	-0.19618	0.4001	
x_{33}	-0.47193	0.0303	**	x_{33}	-0.4521	0.0338	**
x_{34}	-0.10726	0.6639		x_{34}	-0.10245	0.677	
x_{51}	0.007243	0.9323		x_{51}	-0.00739	0.9293	
	edf	p-value					
$s(x_1)$	3.783	0.00265	**	x_1	0.02704	0.039	**

**) significance at level α=5 %

Based on the result shown in Table 2 (x_{21}, x_{30}, x_{41}, and x_{50} are set to reference categories), the models formed are :

1. Semiparametric Bivariate Probit

$$y_{1i}^* = 1.02303 - 0.38421x_{22} - 0.18294x_{23} - 0.06023x_{24} + 0.18624x_{31} + 0.32799x_{32} + 0.20889x_{33} + 0.32791x_{34} + 0.0557x_{42} - 0.22307x_{45} + f_1(sx_1)$$

$$f_1(sx_1) = 0.00008sx_1^3 - 0.0016sx_1^2 - 0.0981sx_1 + 1.9743$$

$$y_{2i}^* = -1.82682 + 1.352889y_{11} + 0.318565x_{22} + 0.247345x_{23} - 0.01595x_{24} - 0.26925x_{31} - 0.17296x_{32} - 0.47193x_{33} - 0.10726x_{34} - 0.007243x_{51} + f_2(sx_1)$$

IOP Conf. Series: Materials Science and Engineering **546** (2019) 052004 doi:10.1088/1757-899X/546/5/052004

$$f_2(sx_1) = 0.000002sx_1^6 - 0.0002sx_1^5 + 0.0118\, sx_1^4 - 0.3362sx_1^3 + 5.1622sx_1^2 - 40.317sx_1 + 124.02$$

2. Bivariate Probit

$$y_{1i}^* = 1.78304 - 0.03835x_1 - 0.3971x_{22} - 0.23031x_{23} - 0.07479x_{24} + 0.1904x_{31} + 0.32815x_{32} + 0.23567x_{33} + 0.33597x_{34} + 0.10971x_{42} - 0.13843x_{45}$$

$$y_{2i}^* = -2.3745 + 0.02704x_1 + 1.369721y_{11} + 0.326201x_{22} + 0.224401x_{23} - 0.00746x_{24} - 0.28772x_{31} - 0.19618x_{32} - 0.4521x_{33} - 0.10245x_{34} - 0.00739x_{51}$$

Semiparametric bivariate probit (SBP) yields significant variables as many as bivariate probit that does not ignore the presence of endogeneity (BP), they are the age of the mother's first marriage, occupation of the mother, and education of the mother. The regression coefficient of parity status of the mother as an exogenous variable in complete basic immunization giving has a positive sign, meaning that the relationship between the two variables is unidirectional or in other words, a multiparous mother tends to give more complete basic immunization to her baby compared to a primiparous mother.

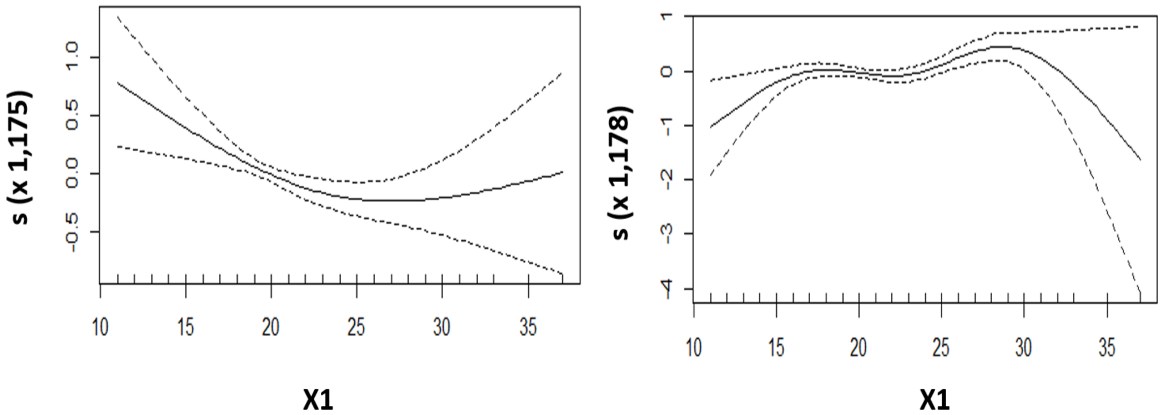

Figure 1. Smooth Function of the age of the mother's first marriage (X_1)

Figure 1 is an estimate of the age of the mother's first marriage (X_1) smooth function based on spline regression in parity status of the mother (left) and complete basic immunization giving (bottom). Estimation of degrees of freedom (right) in each equation indicated by the value contained in brackets. The figure 1 shows that there is a nonlinear relationship between the age of the first marriage of the mother and the mother's parity status. This is supported by the value of edf which is equal to 1.75. The same thing occurs in the relationship between the age of first marriage of the mother with basic immunization giving that produces edf of 3.78. Both images are in accordance with the theory that if the edf is equal to one it will form a straight or linear line.

3.4. Model Selection

Both the semiparametric bivariate probit model and bivariate probit which does not ignore the presence of endogeneity produce significant parameters with the same amount as the marks of each parameter also tend to have similarities. The semiparametric bivariate probit model produces an AIC of 1301,602 while the bivariate probit model that does not ignore the case of endogeneity but without using the smooth function produces AIC of 1316.789. Based on the AIC value of the two models, the best model is the semiparametric bivariate probit model because it has a smaller AIC value.

4. Conclusion

There is a case of endogeneity between the parity status of the mother and basic immunization giving based on Lagrange Multiplier and Likelihood Ratio tests. The significant variables obtained from the two models are the age of the mother's first marriage, occupation of the mother, and education of the mother. The best model based on AIC criterion for parity status of the mother and basic immunization giving in North Kalimantan Province in 2017 is semiparametric bivariate probit model.

References

[1] Agresti, A. (2007). *An Introduction To Categorical Data Analysis.* New Jersey: John Wiley & Sons.

[2] Chen, G., & Tsurumi, H. (2010). Probit and Logit Model Selection. *Communication in Statistics-Theory and Methods*, 159-175.

[3] Coates, D., Wicker, P., Feiler, S., & Breuer, C. (2014). A Bivariate Probit Examination of Financial and Volunteer Problems of Non Profit Sport Clubs. *International Journal of Sport Finance* , 230-248.

[4] Ratnasari, V., Purhadi, Zain, I., & Suhartono. (2011). Estimation and Test Statistics in Bivariate Probit Model. *Journal of Basic and Applied Scientific Research*, 178-188.

[5] Ratnasari, V. (2012). *Estimation and Significance Test in Bivariate Probit Models.* Surabaya: Institut Teknologi Sepuluh Nopember.

[6] Marra, G., & Radice, R. (2011). Estimation of Semiparametric Recursive Bivariate Probit Model in The Presence of Endogeneity. *The Canadian Journal of Statistics, 39*, 259-279.

[7] Ieva, F., Marra, G., Paganoni, M., & Radice, R. (2014). A Semiparametric Bivariate Probit Model for Joint Modeling of Outcomes in STEMI Patients. *The Journal of Computational and Mathematical Method in Medicine*, 1-7.

[8] Umami, A., Ratnasari, V., & Zain, I. (2018). Semiparametric Bivariate Probit Model in Data Working The mother Status and Exclusive Breastfeeding. *Journal of Physics: Conference Series*.

[9] Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science, 1*, 297-310.

[10] Marra, G., & Radice, R. (2010). Penalized Regression Splines: Theory and Application to Medical Research. *Statistical Methods in Medical Research*, 107-125.

[11] Okeh, U.M, & Oyeka, I. (2013). Estimating the Fisher Scoring Matrix Formula From Logistic Model. *American Journal of Theorical and Applied Statistics*, 221-227.

[12] Marra, G., Radice, R., & Filippou, P. (2015). Regression Spline Bivariate Probit Models: A Practical Approach to Testing for Exogeneity. *Communication in Statistics-Simulation and Computation*.

[13] Monfardini, C., & Radice, R. (2008). Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study. *Oxford Bulletin of Economics and Statistics* , 271-282.

[14] Silviana, N. (2013). *Relationship of Parity with the Level of Knowledge of The mother About Basic Immunization in Babies at the Umbulharjo I Puskesmas in Yogyakarta.* Yogyakarta: Sekolah Tinggi Ilmu Kesehatan Aisyiyah.

[15] Dwiyanti, F. (2010). *Relationship Between Maternal Parity and Completeness of Giving Basic Immunization to Infants at the Gribig Puskesmas.* Malang: Universitas Muhammadiyah Malang.

[16] Gujarati, D., & Porter, D. (2009). *Basic Econometric.* New York: McGraw-Hill.

[17] Gujarati, D., & Porter, D. (2013). *Dasar-Dasar Ekonometrika.* Jakarta: Salemba Empat.

[18] Varney, H., Kriebs, J., & Gegor, C. (2004). *Varney's Midwifery.* Massachussets: Jones and Bartlett Publishers.

[19] Kementerian Kesehatan RI. (2016). *Situasi Imunisasi di Indonesia.* Jakarta: Pusat Data dan Informasi Kementerian Kesehatan RI.