PAPER • **OPEN ACCESS**

Hypothesis testing of Geographically weighted bivariate logistic regression

To cite this article: M Fathurahman *et al* 2019 *J. Phys.: Conf. Ser.* **1417** 012008

Hypothesis testing of Geographically weighted bivariate logistic regression

M Fathurahman[1,2], Purhadi[3], Sutikno[4], and V Ratnasari[5]

[1] Department of Statistics, Mulawarman University, Samarinda, Indonesia
[2] Doctoral Student at the Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3,4,5] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

e-mail: fathur@fmipa.unmul.ac.id

Abstract. In this study, the hypothesis testing of geographically weighted bivariate logistic regression (GWBLR) procedure is proposed. The GWBLR model is a bivariate logistic regression (BLR) model which all of the regression parameters depend on the geographical location in the study area. The geographical location is expressed as a point coordinate in two-dimensional geographic space (longitude and latitude). The response variables of BLR model are constructed from a (2 □ 2) contingency table and it follows the multinomial distribution. The purpose of this study is to test the GWBLR model parameters. There are three hypothesis tests. The first is a parameters similarity test using the Vuong test method. The test is to obtain a significant difference between GWBLR and BLR. The second is a simultaneous test using the likelihood ratio test method. The simultaneous test is to obtain the simultaneous significance of the regression parameters. The last is a partial test using Wald test method. The result showed that the Vuong statistic and Wald statistic have an asymptotic standard normal distribution, whereas the likelihood ratio statistic has an asymptotic chi-squared distribution.

1. Introduction

Geographically weighted regression (GWR) is an effective technique for modelling spatial non-stationary data [1, 2]. Hypothesis testing is one of the statistical inference tools that have an important contributed in GWR modelling and it was developed [3, 4, 5]. The response variable of the GWR model in previous studies is quantitative data and normally distributed. However, in fact, not only the response variable is quantitative data and normally distributed, but also is qualitative (categorical) data and follows other distributions, such as Bernoulli, binomial, and multinomial.

Recently, the GWR models with response variable are categorical data have been proposed. GWR and a logistic regression model are combined to form a geographically weighted logistic regression (GWLR) [6]. The proposed model has two categories of the response variable which follows Bernoulli distribution and it was applied to model the spatial variation in the relation between erosion (presence or absence) and several controlling variables for the Afon Dyfi in West Wales. Furthermore, a geographically weighted logistic model (GWLM) was employed and it has shown that the GWLM is an efficient tool to account for spatial heterogeneity, showing better fit and residual properties compared to the logistic regression model (LRM) and logistic mixed model (LMM) in modeling the occurrence of cloud cover, using the spatial data derived from satellite imagery and GIS [7]. On the other hand, the GWLM and global logistic model (GLM) were used to model and analyze the spatial variation in the human factors associated with forest fires [8]. GWLM techniques have shown a high predictive potential for human-caused wildfire occurrence modelling, surpassing classical regression techniques like GLM and allowing the detection of non-stationary relationships between response and independent variables. However, the number of categories in the response variable of the GWR model for categorical data is also more than two categories. There are several studies have been developed.

The geographically weighted multinomial logistic regression (GWMLR) model was introduced and it has the number of category of the response variable is more than two categories [9, 10]. In the GWMLR model, each category is unordered and follows the multinomial distribution. As a follow-up, previous studies, the geographically weighted ordinal logistic regression (GWOLR) and geographically weighted ordinal logistic regression semiparametric (GWOLRS) model were proposed [11, 12]. Like the GWMLR model, the response variable of GWOLR and GWOLRS model follows the multinomial distribution and has more than two categories, but each category is ordered. In the GWOLRS model, some of the independent variables are global and the other variables are local [12].

The previous studies extend the GWR models for categorical data that have only one response variable (univariate). However, in many fields of research, several cases have two response variables (bivariate). Therefore, a new spatial regression model is proposed in this study, namely the geographically weighted bivariate logistic regression (GWBLR). The GWBLR model is to focus on two variable responses which each variable has two categories. This study aims to test the GWBLR model parameters. There are three kinds of hypothesis tests. The first test is a parameters similarity test by using the Vuong test method. This test is used to obtain if there is a significant difference between GWBLR model and BLR model. The second test is a simultaneous test by using a likelihood ratio test (LRT) method. This test aimed to obtain the simultaneous significance of the regression parameters. The last test is a partial test by using the Wald test method. This test is used to obtain the significance of each parameter in the regression model.

2. Bivariate logistic regression

The bivariate logistic regression (BLR) is an extension of the univariate logistic regression when there are two categorical data of response variables and they are correlated to each other. In this study, each of the response variables has two categories. Let Y_1 and Y_2 are response variables of the BLR model. The response variables can be shown in Table 1 and the joint probability of the response variables in Table 1 is presented in Table 2.

Table 1. The (2×2) contingency table of the response variables.

Y_1	Y_2		Total
	$Y_2 = 1$	$Y_2 = 0$	
$Y_1 = 1$	Y_{11}	Y_{10}	Y_{1+}
$Y_1 = 0$	Y_{01}	Y_{00}	$Y_{0+} = n - Y_{1+}$
Total	Y_{+1}	$Y_{+0} = n - Y_{+1}$	$Y_{++} = n$

Table 2. The joint probability of the response variables.

Y_1	Y_2		Total
	$Y_2 = 1$	$Y_2 = 0$	
$Y_1 = 1$	p_{11}	p_{10}	p_{1+}
$Y_1 = 0$	p_{01}	p_{00}	$p_{0+} = 1 - p_{1+}$
Total	p_{+1}	$p_{+0} = 1 - p_{+1}$	$p_{++} = 1$

Based on Table 1 and Table 2, the random variables Y_{11}, Y_{10}, Y_{01}, and Y_{00} are follows the multinomial distribution with joint probability function defined by

$$P(Y_{11} = y_{11}, Y_{10} = y_{10}, Y_{01} = y_{01}) = \prod_{g=0}^{1} \prod_{h=0}^{1} \frac{p_{gh}^{y_{gh}}}{y_{gh}!}, 0 < p_{gh} < 1,$$

where $g, h = 0{,}1$; $y_{gh} = 0{,}1$; $y_{00} = 1 - y_{11} - y_{10} - y_{01}$; and $p_{00} = 1 - p_{11} - p_{10} - p_{01}$.

Furthermore, the BLR model can be written as follows

$$g_1(x) = logit\ (p_1(x)) = \beta_1^T x,$$
$$g_2(x) = logit\ (p_2(x)) = \beta_2^T x,$$
$$g_3(x) = ln\ ln\ \psi_1(x) \ \ = \beta_3^T x,$$

where $x = [1\ x_1\ x_2 \cdots x_k]^T$ is a vector of the covariate, $\beta_1^T = [\beta_{01}\ \beta_{11}\ \beta_{21}\ \cdots\ \beta_{k1}]$, $\beta_2^T = [\beta_{02}\ \beta_{12}\ \beta_{22}\ \cdots\ \beta_{k2}]$, and $\beta_3^T = [\beta_{03}\ \beta_{13}\ \beta_{23}\ \cdots\ \beta_{k3}]$ are the parameters, $p_1(x)$ is the marginal probability function of Y_1, $p_2(x)$ is the marginal probability function of Y_2, and $\psi_1(x)$ is the odds ratio that is shown an association between Y_1 and Y_2. The marginal probability function and the odds ratio are defined by

$$p_1(x) = p_{1+}(x) = \frac{e^{\beta_1^T x}}{1+e^{\beta_1^T x}},$$
$$p_2(x) = p_{+1}(x) = \frac{e^{\beta_2^T x}}{1+e^{\beta_2^T x}},$$
$$\psi_1(x) = \frac{p_{11}(x)p_{00}(x)}{p_{10}(x)p_{01}(x)}.$$

According to [13], the joint probability of $p_{11}(x)$ in Equation can be obtained as follows

$$p_{11}(x) = \{\frac{a_1 - \sqrt{a_1^2 + b_1}}{2(\psi_1(x) - 1)}, \psi_1(x) \neq 1 \ \ p_1(x)p_2(x), \psi_1(x) = 1$$

where $a_1 = 1 + (\psi_1(x) - 1)(p_1(x) + p_2(x))$, $b_1 = -4\psi_1(x)(\psi_1(x) - 1)p_1(x)p_2(x)$ with $p_1(x)$ and $p_2(x)$ in Equations and. Furthermore, based on Table 2 and Equation, the joint probabilities of $p_{10}(x)$, $p_{01}(x)$, and $p_{00}(x)$ are obtained as follows

$$p_{10}(x) = p_1(x) - p_{11}(x),$$
$$p_{01}(x) = p_2(x) - p_{11}(x),$$
$$p_{00}(x) = 1 - p_{11}(x) - p_{10}(x) - p_{01}(x)$$
$$= 1 - p_1(x) - p_2(x) + p_{11}(x).$$

3. Geographically weighted bivariate logistic regression

The geographically weighted bivariate logistic regression (GWBLR) is developed of the BLR model which all of the regression parameters depend on the geographical location in the study area. The geographical location is expressed as a point coordinate in two-dimensional geographic space (longitude and latitude) [14].

Let $u_i = (u_{1i}, u_{2i})$ denotes a vector of point coordinate for i^{th} location where the data is observed for $i = 1,2, \dots, n$, with u_{1i} is latitude and u_{2i} is longitude, then from BLR model in Equations -, it can be developed to the new BLR model with all of the parameters depend on the geographical location which is called GWBLR model. Here, the parameters are assumed to be functions of the location on which the observations are obtained. Suppose, all of the parameters of the BLR model in Equations - depend on the geographical location, then GWBLR model at i^{th} location with coordinate u_i has an expression as follows

$$h_1(x_i) = logit\ (\pi_1(x_i)) = \beta_1^T(u_i)x_i, i = 1,2, \dots, n$$
$$h_2(x_i) = logit\ (\pi_2(x_i)) = \beta_2^T(u_i)x_i,$$
$$h_3(x_i) = ln\ ln\ \psi_2(x_i) \ \ = \beta_3^T(u_i)x_i,$$

where

$x_i = [1\ x_{1i}\ x_{2i} \cdots x_{ki}]^T$ is a vector of independent variables at i^{th} location,
$\beta_1^T(u_i) = [\beta_{01}(u_i)\ \beta_{11}(u_i)\ \beta_{21}(u_i)\ \cdots\ \beta_{k1}(u_i)]$, $\beta_2^T(u_i) = [\beta_{02}(u_i)\ \beta_{12}(u_i)\ \beta_{22}(u_i)\ \cdots\ \beta_{k2}(u_i)]$, and $\beta_3^T(u_i) = [\beta_{03}(u_i)\ \beta_{13}(u_i)\ \beta_{23}(u_i)\ \cdots\ \beta_{k3}(u_i)]$ are the parameters at i^{th} location,
$\pi_1(x_i)$ is the marginal probability function of Y_1 at i^{th} location, $\pi_2(x_i)$ is the marginal probability function of Y_2 at i^{th} location, and $\psi_2(x_i)$ is the odds ratio that is shown an association between Y_1 and

Y_2 at i^{th} location. The marginal probability function of response variables and the odds ratio at i^{th} location is defined by

$$\pi_1(x_i) = \pi_{1+}(x_i) = \frac{expexp\ (\beta_1^T(u_i)x_i)}{1+expexp\ (\beta_1^T(u_i)x_i)},$$
$$\pi_2(x_i) = \pi_{+1}(x_i) = \frac{expexp\ (\beta_2^T(u_i)x_i)}{1+expexp\ (\beta_2^T(u_i)x_i)},$$
$$\psi_2(x_i) = \frac{\pi_{11}(x_i)\pi_{00}(x_i)}{\pi_{10}(x_i)\pi_{01}(x_i)}.$$

Based on the joint probability function in Equation, the joint probability of $\pi_{11}(x_i)$ in Equation can be determined by

$$\pi_{11}(x_i) = \{\frac{a_2 - \sqrt{a_2^2 + b_2}}{2(\psi_2(x_i)-1)}, \psi_2(x_i) \neq 1 \quad \pi_1(x_i)\pi_2(x_i), \psi_2(x_i) = 1$$

where $a_2 = 1 + (\psi_2(x_i) - 1)(\pi_1(x_i) + \pi_2(x_i))$, $b_2 = -4\psi_2(x_i)(\psi_2(x_i) - 1)\pi_1(x_i)\pi_2(x_i)$ with $\pi_1(x_i)$, $\pi_2(x_i)$, and $\psi_2(x_i)$ in Equations -. The joint probabilities of $\pi_{10}(x_i)$, $\pi_{01}(x_i)$, and $\pi_{00}(x_i)$ are obtained as follows

$$\pi_{10}(x_i) = \pi_1(x_i) - \pi_{11}(x_i),$$
$$\pi_{01}(x_i) = \pi_2(x_i) - \pi_{11}(x_i),$$
$$\pi_{00}(x_i) = 1 - \pi_1(x_i) - \pi_2(x_i) + \pi_{11}(x_i).$$

4. Hypothesis testing of the GWBLR model parameters

Hypothesis testing consists of three kinds of tests. The first test is a parameter similarity test between GWBLR and BLR. The second test is a simultaneous test of GWBLR parameters. The last test is a partial test of GWBLR parameters.

4.1. Parameter similarity test between GWBLR and BLR

The aim of this test is used to obtain a significant difference between GWBLR model and BLR model. Parameter similarity test is also called the goodness of fit test [15]. The null hypothesis (H_0) and the alternative hypothesis (H_1) are as follows

$$H_0: \beta_{rs}(u_i) = \beta_{rs}, \quad r = 1, 2, \dots, k; s = 1, 2, 3; \ i = 1, 2, \dots, n,$$
H_1: at least one of $\beta_{rs}(u_i) \neq \beta_{rs}$.

where r is the index of the covariate, k is the number of covariates, s is the index of parameters, i is the index of research sample (the study area), and n is the number of the study area.

The test statistic for hypotheses in Equation can be determined by using the Vuong test method. The part of this method can be determined by using the likelihood ratio test (LRT) procedure [15]. Suppose the parameters set under H_0 is $\omega = \{\beta_{rs}, r = 0, 1, 2, \dots, k; s = 1, 2, 3\}$. Thus, to determine the likelihood function under H_0

$$L(\omega) = \prod_{i=1}^{n}\left[p_{11i}^{y_{11i}}(x_i)p_{10i}^{y_{10i}}(x_i)p_{01i}^{y_{01i}}(x_i)p_{00i}^{y_{00i}}(x_i)\right].$$

Let $p_{ghi}(x_i) = p_{ghi}$, for $g, h = 0{,}1$ and $i = 1, 2, \dots, n$. The likelihood function in Equation can be rewritten as

$$L(\omega) = \prod_{i=1}^{n}\left[p_{11i}^{y_{11i}}p_{10i}^{y_{10i}}p_{01i}^{y_{01i}}p_{00i}^{y_{00i}}\right].$$

Based on Equation to determine maximum log-likelihood function under H_0

$$ln\ ln\ L(\widehat{\omega}) = \sum_{i=1}^{n}(y_{11i}\ ln\ ln\ \hat{p}_{11i} + y_{10i}\ ln\ ln\ \hat{p}_{10i} + y_{01i}\ ln\ ln\ \hat{p}_{01i} + y_{00i}\ ln\ ln\ \hat{p}_{00i}),$$

where $\hat{p}_{11i}$, $\hat{p}_{10i}$, $\hat{p}_{01i}$, and $\hat{p}_{00i}$ for $i = 1, 2, \dots, n$ are formulated as follows

$$\hat{p}_{11i} = \{\frac{a_3 - \sqrt{a_3^2 + b_3}}{2(\hat{\psi}_3 - 1)}, \hat{\psi}_3 \neq 1 \quad \hat{p}_{1i}\hat{p}_{2i}, \hat{\psi}_3 = 1$$

where $a_3 = 1 + (\hat{\psi}_3 - 1)(\hat{p}_{1i} + \hat{p}_{2i})$, $b_3 = -4\hat{\psi}_3(\hat{\psi}_3 - 1)\hat{p}_{1i}\hat{p}_{2i}$, $\hat{p}_{1i} = \frac{e^{\hat{\beta}_1^T x_i}}{(1+e^{\hat{\beta}_1^T x_i}}),$ $\hat{p}_{2i} = \frac{e^{\hat{\beta}_2^T x_i}}{(1+e^{\hat{\beta}_2^T x_i}})$, and $\hat{\psi}_3 = \frac{\hat{p}_{11i}\hat{p}_{00i}}{\hat{p}_{10i}\hat{p}_{01i}}$, with $\hat{\beta}_1^T = [1\ \hat{\beta}_{01}\ \hat{\beta}_{11}\ \hat{\beta}_{21} \cdots\ \hat{\beta}_{k1}]$

and $\hat{\beta}_1^T = [1\ \hat{\beta}_{01}\ \hat{\beta}_{11}\ \hat{\beta}_{21} \cdots\ \hat{\beta}_{k1}]$. Thus, $\hat{\beta}_1^T$ and $\hat{\beta}_2^T$ are maximum likelihood estimators (MLEs) for β_1 and β_2.

$$\hat{p}_{10i} = \hat{p}_{1i} - \hat{p}_{11i}.$$
$$\hat{p}_{01i} = \hat{p}_{2i} - \hat{p}_{11i}.$$
$$\hat{p}_{00i} = 1 - \hat{p}_{1i} - \hat{p}_{2i} + \hat{p}_{11i}.$$

Furthermore, to obtain parameters set under population is

$$\Omega = \{\beta_{rs}(u_i), r = 0,1,2,\dots,k; s = 1,2,3; i = 1,2,\dots,n\}.$$

Thus, the likelihood function and maximum log-likelihood function under population are given by

$$L(\Omega) = \prod_{i=1}^{n}\left[\pi_{11i}^{y_{11i}}\pi_{10i}^{y_{10i}}\pi_{01i}^{y_{01i}}\pi_{00i}^{y_{00i}}\right],$$
$$ln\ ln\ L(\hat{\Omega}) = \sum_{i=1}^{n}(y_{11i}\ ln\ ln\ \hat{p}_{11i} + y_{10i}\ ln\ ln\ \hat{p}_{10i} + y_{01i}\ ln\ ln\ \hat{p}_{01i} + y_{00i}\ ln\ ln\ \hat{p}_{00i}),$$

where $\hat{\pi}_{11i}$, $\hat{\pi}_{10i}$, $\hat{\pi}_{01i}$, and $\hat{\pi}_{00i}$ for $i = 1,2,\dots,n$ are formulated as follows

$$\hat{\pi}_{11i} = \{\frac{a_4 - \sqrt{a_4^2 + b_4}}{2(\hat{\psi}_4 - 1)}, \hat{\psi}_4 \neq 1\ \ \hat{\pi}_{1i}\hat{\pi}_{2i}, \hat{\psi}_4 = 1$$

where $a_4 = 1 + (\hat{\psi}_4 - 1)(\hat{\pi}_{1i} + \hat{\pi}_{2i})$, $b_4 = -4\hat{\psi}_4(\hat{\psi}_4 - 1)\hat{\pi}_{1i}\hat{\pi}_{2i}$, $\hat{\pi}_{1i} = \frac{e^{\hat{\beta}_1^T(u_i)x_i}}{[1+e^{\hat{\beta}_1^T(u_i)x_i}]}$, $\hat{\pi}_{2i} = \frac{e^{\hat{\beta}_2^T(u_i)x_i}}{[1+e^{\hat{\beta}_2^T(u_i)x_i}]}$, $\hat{\beta}_1^T(u_i) = [1\ \hat{\beta}_{01}(u_i)\ \hat{\beta}_{11}(u_i)\ \hat{\beta}_{21}(u_i) \cdots\ \hat{\beta}_{k1}(u_i)]$, $\hat{\beta}_2^T(u_i) = [1\ \hat{\beta}_{02}(u_i)\ \hat{\beta}_{12}(u_i)\ \hat{\beta}_{22}(u_i) \cdots\ \hat{\beta}_{k2}(u_i)]$, and $\hat{\psi}_4 = \frac{\hat{\pi}_{11i}\hat{\pi}_{00i}}{\hat{\pi}_{10i}\hat{\pi}_{01i}}$ with $\hat{\beta}_1^T(u_i)$ and $\hat{\beta}_2^T(u_i)$ are maximum likelihood estimators (MLEs) for $\beta_1(u_i)$ and $\beta_2(u_i)$.

$$\hat{\pi}_{10i} = \hat{\pi}_{1i} - \hat{\pi}_{11i}.$$
$$\hat{\pi}_{01i} = \hat{\pi}_{2i} - \hat{\pi}_{11i}.$$
$$\hat{\pi}_{00i} = 1 - \hat{\pi}_{1i} - \hat{\pi}_{2i} + \hat{\pi}_{11i}.$$

The test statistic for testing hypotheses in Equation can be obtained by

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

where $m_i = ln\ ln\ L(\hat{\Omega}) - ln\ ln\ L(\hat{\omega})$, $\underline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$, and n is sample size. The $ln\ ln\ L(\hat{\omega})$ and $ln\ ln\ L(\hat{\Omega})$ are obtained in Equations and.

The Vuong statistic in Equation was asymptotically standard normal distributed. Therefore, reject H_0 if $|V| > Z_{1-\alpha/2}$, where the critical value $Z_{1-\alpha/2}$ is $1 - \alpha/2$ quantile from a standard normal distribution.

4.2. Simultaneous test

This test is used to determine the simultaneous significance of the regression parameters. Consider the hypotheses

$$H_0: \beta_{1s}(u_i) = \beta_{2s}(u_i) = \cdots = \beta_{ks}(u_i) = 0, s = 1,2,3;\ i = 1,2,\dots,n$$
$$H_1: \text{at least one of the } \beta_{rs}(u_i) \neq 0, r = 1,2,\dots,k.$$

The test statistic for hypotheses in Equation can be determined by using the LRT method [15]. Suppose the parameters set under H_0 is $\omega^* = \{\beta_{01}(u_i), \beta_{02}(u_i), \beta_{03}(u_i), i = 1,2,\dots,n\}$. Thus, to determine the likelihood function under H_0

$$L(\omega^*) = \prod_{i=1}^{n}\left[\left(\pi_{11i}^*(x_i)\right)^{y_{11i}}\left(\pi_{10i}^*(x_i)\right)^{y_{10i}}\left(\pi_{01i}^*(x_i)\right)^{y_{01i}}\left(\pi_{00i}^*(x_i)\right)^{y_{00i}}\right].$$

Let $\pi_{ghi}^*(x_i) = \pi_{ghi}^*$, for $g,h = 0,1$ and $i = 1,2,\dots,n$. The likelihood function in Equation can be rewritten as

$$L(\omega^*) = \prod_{i=1}^{n}[(\pi_{11i}^*)^{y_{11i}}(\pi_{10i}^*)^{y_{10i}}(\pi_{01i}^*)^{y_{01i}}(\pi_{00i}^*)^{y_{00i}}].$$

Based on Equation to determine maximum log-likelihood function under H_0

$$ln\ ln\ L(\hat{\omega}^*) = \sum_{i=1}^{n}(y_{11i}\ ln\ ln\ \hat{\pi}_{11i}^* + y_{10i}\ ln\ ln\ \hat{\pi}_{11i}^* + y_{01i}\ ln\ ln\ \hat{\pi}_{11i}^* + y_{00i}\ ln\ ln\ \hat{\pi}_{11i}^*),$$

where $\hat{\pi}_{11i}^*$, $\hat{\pi}_{10i}^*$, $\hat{\pi}_{01i}^*$, and $\hat{\pi}_{00i}^*$ for $i = 1,2,\dots,n$ are formulated as follows

$$\hat{\pi}_{11i}^{*} = \{\frac{a_5 - \sqrt{a_5^2 + b_5}}{2(\hat{\psi}_5 - 1)}, \hat{\psi}_5 \neq 1 \quad \hat{\pi}_{1i}^{*}\hat{\pi}_{2i}^{*}, \hat{\psi}_5 = 1$$

where $a_5 = 1 + (\hat{\psi}_5 - 1)(\hat{\pi}_{1i}^{*} + \hat{\pi}_{2i}^{*})$, $b_5 = -4\hat{\psi}_5(\hat{\psi}_5 - 1)\hat{\pi}_{1i}^{*}\hat{\pi}_{2i}^{*}$, $\hat{\pi}_{1i}^{*} = \frac{e^{\hat{\beta}_{01}(u_i)x_i}}{[1+e^{\hat{\beta}_{01}(u_i)x_i}]}$, $\hat{\pi}_{2i}^{*} = \frac{e^{\hat{\beta}_{02}(u_i)x_i}}{[1+e^{\hat{\beta}_{02}(u_i)x_i}]}$, and $\hat{\psi}_5 = \frac{\hat{\pi}_{11i}^{*}\hat{\pi}_{00i}^{*}}{\hat{\pi}_{10i}^{*}\hat{\pi}_{01i}^{*}}$ with $\hat{\beta}_{01}(u_i)$ and $\hat{\beta}_{02}(u_i)$ are maximum likelihood estimators (MLEs) for $\beta_{01}(u_i)$ and $\beta_{02}(u_i)$.

$$\hat{\pi}_{10i}^{*} = \hat{\pi}_{1i}^{*} - \hat{\pi}_{11i}^{*}.$$
$$\hat{\pi}_{01i}^{*} = \hat{\pi}_{2i}^{*} - \hat{\pi}_{11i}^{*}.$$
$$\hat{\pi}_{00i}^{*} = 1 - \hat{\pi}_{1i}^{*} - \hat{\pi}_{2i}^{*} + \hat{\pi}_{11i}^{*}.$$

Furthermore, to obtain parameters set under population. The parameters set underpopulation are similar to the parameter similarity test in Section 4.1. Therefore, the likelihood function and maximum log-likelihood function underpopulation are obtained by Equations and.

The test statistic for hypotheses in Equation is given by

$$G^2 = -2[ln\ ln\ L(\hat{\omega}^{*})\ - ln\ ln\ L(\hat{\Omega})\ ],$$

where $L(\hat{\Omega})$ and $L(\hat{\omega}^{*})$ are obtained in Equations and the test statistic in Equation can be rewritten as follows

$$G^2 = 2[\sum_{i=1}^{n}(y_{11i}\ ln\ ln\ \hat{\pi}_{11i}\ + y_{10i}\ ln\ ln\ \hat{\pi}_{10i}\ + y_{01i}\ ln\ ln\ \hat{\pi}_{01i}\ + y_{00i}\ ln\ ln\ \hat{\pi}_{00i}\ ) + \\ -\sum_{i=1}^{n}(y_{11i}\ ln\ ln\ \hat{\pi}_{11i}^{*}\ + y_{10i}\ ln\ ln\ \hat{\pi}_{10i}^{*}\ + y_{01i}\ ln\ ln\ \hat{\pi}_{01i}^{*}\ + y_{00i}\ ln\ ln\ \hat{\pi}_{00i}^{*}\ )].$$

The likelihood ratio (LR) statistic is also referred to as the *Wilk's lambda* statistic [16]. The LR statistic in Equation has an asymptotic chi-squared distribution with v degree of freedom, where v is the difference between the effective number of parameters in GWBLR model without independent variables (the reduced model) and GWBLR model with independent variables (the full model). Therefore, reject H_0 when $G^2 > \chi^2_{(v,1-\alpha)}$, where $\chi^2_{(v,1-\alpha)}$ is the $1-\alpha$ quantile from a chi-square distribution (χ^2) with v degree of freedom.

4.3. *Partial test*

The last test on hypothesis testing of the GWBLR model parameters is a partial test. This test is used to determine the significance of each parameter in the regression model. The form of hypotheses can be expressed as

$$H_0: \beta_{rs}(u_i) = 0, r = 1, 2, \dots, k; s = 1, 2, 3;\ i = 1, 2, \dots, n,$$
$$H_1: \beta_{rs}(u_i) \neq 0.$$

The test statistic for hypotheses in Equation is done by using the Wald test method. The basic idea of this method is following properties of the MLEs, particularly asymptotically normally distributed. Therefore, the Wald statistic is formulated by

$$W = \frac{\hat{\beta}_{rs}(u_i)}{\sqrt{\widehat{Var}(\hat{\beta}_{rs}(u_i))}} \sim N(0,1),$$

where $\widehat{Var}(\hat{\beta}_{rs}(u_i))$ is the diagonal elements of the matrix $[I(\beta(u_i))]^{-1}$ and $I(\beta(u_i))$ is a Fisher information matrix [17]. The $[I(\beta(u_i))]^{-1}$ the matrix can be determined by

$$[I(\beta(u_i))]^{-1} = \left(-E\left[\frac{\partial^2 ln ln\ L(\beta(u_i))}{\partial^2\beta(u_i)}\right]\right)^{-1} = \\ \left(E\left[\left(\frac{\partial ln ln\ L(\beta(u_i))}{\partial\beta(u_i)}\right)\left(\frac{\partial ln ln\ L(\beta(u_i))}{\partial\beta(u_i)}\right)^T\right]\right)^{-1}.$$

The Wald statistic in Equation was asymptotically standard normal distributed [17]. Thus, reject H_0 if $|W| > Z_{1-\alpha/2}$, where the critical value $Z_{1-\alpha/2}$ is $1-\alpha/2$ quantile from a standard normal distribution.

5. Conclusion

GWBLR model is a local form of BLR which all of the regression parameters depend on the geographical location. Hypothesis testing on the GWBLR model is based on the GWR method, that is, hypothesis testing is done locally, dependent on the spatial weighting function in the study area. There are three kinds of hypothesis tests. The first test is a parameter similarity test by using the Vuong test method. The test is used to determine a significant difference between GWBLR and BLR. The Vuong test statistic was asymptotically standard normal distributed. The second test is a simultaneous test using the LRT method. The simultaneous test is used to determine the simultaneous significance of the regression parameters. The LR statistic has an asymptotic chi-squared distribution with the degree of freedom is the difference between the effective number of parameters in the reduced model and the full model. The last test is a partial test using the Wald test method and it was applied to determine the significance of each parameter in the regression model. The Wald test statistic was asymptotically standard normal distributed.

References

[1] Brunsdon C, Fotheringham A S and Charlton M 1996 Geographically weighted regression: a method for exploring spatial nonstationarity *Geographical Analysis* **28**(4) 281-298

[2] Brunsdon C, Fotheringham S and Charlton M 1998 Geographically weighted regression: modelling spatial non-stationarity *Journal of the Royal Statistical Society Series D (The Statistician)* **47**(3) 431-443

[3] Brunsdon C, Fotheringham A S and Charlton M 1999 Some notes on parametric significance tests for geographically weighted regression *Journal of Regional Science* **39**(3) 497-524

[4] Leung Y, Mei C-L and Zhang W-X 2000a Statistical tests for spatial nonstationarity based on the geographically weighted regression model *Environment and Planning A* **32** 9-32

[5] Leung Y, Mei C-L and Zhang W-X 2000b Testing for spatial autocorrelation among the residuals of the geographically weighted regression *Environment and Planning A* **32** 871-890

[6] Atkinson P M, German S E, Sear D A and Clark M J 2003 Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression *Geographical Analysis* **35**(1) 58-82

[7] Wu W and Zhang L 2013 Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in North-eastern Puerto Rico *Applied Geography* **37** 52-62

[8] Rodrigues M, de la Riva J and Fotheringham S 2014 Modeling the spatial variation of the explanatory factors of human-caused wildfires in spain using geographically weighted logistic regression *Applied Geography* **48** 52-63

[9] Luo J and Kanala N K 2008 Modeling urban growth with geographically weighted multinomial logistic regression *Proc. of SPIE* **7144** 71440M-1-71440M-2

[10] Wang Y, Kockelman K M and Wang X 2011 Anticipation of land use change through use of geographically weighted regression models for discrete response *Transportation Research Record: Journal of the Transportation Research Board* No. 2245 111-123

[11] Purhadi, Rifada M and Wulandari S P 2012 Geographically weighted ordinal logistic regression model *International Journal of Mathematics and Computation* **16**(3) 116-126

[12] Wardhani N W S, Pramoedyo H and Dianati Y N 2014 Food security and vulnerability modeling of East Java Province based on Geographically Weighted Ordinal Logistic Regression Semiparametric (GWOLRS) Model *Journal of Degraded and Mining Lands Management* **2**(1) 231-234

[13] El-Sayed A M, Islam M A and Alzaid A A 2013 Estimation and test of measures of association for correlated binary data *Bulletin of the Malaysian Mathematical Sciences Society* **36**(4) 985-1008

[14] Fotheringham A S, Charlton M and Brunsdon C 2002 *Geographically Weighted Regression: The analysis of spatially varying relationships* (Chichester: John Wiley & Sons)

[15] Triyanto, Purhadi, Otok B W and Purnami S W 2016 Hypothesis testing of geographically weighted multivariate poisson regression *Far East Journal of Mathematical Sciences* **100**(5) 747-762
[16] Harini S, Purhadi, Mashuri M and Sunaryo S 2012 Statistical test for multivariate geographically weighted regression model using The Method of Maximum Likelihood Ratio Test *International Journal of Applied Mathematics & Statistics* **29**(5) 110-115
[17] Greene W H 2003 *Econometric Analysis* (New Jersey: Pearson Education)