Article

Parameter Estimation and Hypothesis Testing of Geographically Weighted Multivariate Generalized Poisson Regression

Sarni Maniar Berliana [1,2,†] ⓘ**, Purhadi** [1,*,†] ⓘ**, Sutikno** [1,†] **and Santi Puteri Rahayu** [1,†]

[1]   Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia; sarni@stis.ac.id (S.M.B.); sutikno@statistika.its.ac.id (S.S.); santi_pr@statistika.its.ac.id (S.P.R.)

[2]   Department of Statistics, Politeknik Statistika STIS, Jakarta 13330, Indonesia

*   Correspondence: purhadi@statistika.its.ac.id

†   These authors contributed equally to this work.

check for updates

Abstract: We introduce a new multivariate regression model based on the generalized Poisson distribution, which we called geographically-weighted multivariate generalized Poisson regression (GWMGPR) model, and we present a maximum likelihood step-by-step procedure to obtain parameters for it. We use the maximum likelihood ratio test to examine the significance of the regression parameters and to define their critical region.

1. Introduction

Poisson distribution is one of the most widely utilized univariate and multivariate regression models for count data analysis. The univariate regression model is comprised of one response variable, while the multivariate regression model is made up of two or more response variables that are correlated. There are several multivariate count data models based on Poisson distribution, namely multivariate Poisson regression model, multivariate Poisson-gamma mixture model, multivariate Poisson log-normal model, and multivariate generalized Poisson regression model [1,2].

Poisson distribution is characterized by its restrictive assumption on equidispersion, a condition where the sample variance is equal to the mean. However, when the sample mean is less (greater) than the variance, it is called overdispersion (underdispersion). The multivariate Poisson regression model assumes equidispersion and allows for nonnegative correlations, while the multivariate Poisson-gamma mixture model handles overdispersion and nonnegative correlations. Furthermore, the multivariate Poisson-log-normal model allows overdispersion and a more flexible correlation structure. The multivariate generalized Poisson regression (MGPR) model is the most flexible amongst all because it allows any type of dispersion and correlation.

The MGPR model developed by Famoye [2] is a regression model based on multivariate generalized Poisson distribution, which was obtained by the same procedure in constructing bivariate generalized Poisson distribution, which is a product of generalized Poisson marginal with a multiplicative factor [3,4]. To date, bivariate and multivariate generalized Poisson regression models have been applied to real data with units of analysis such as elderly, women, patients, soccer teams, or households [2,5,6]. These models are global regression models that assume the relationships between variables are spatially constant.

When the units of analysis used are locations, spatial effects need to be taken into account in modeling the relationships between the explanatory variables and the response variables. Two possible issues arise when the sample data contain spatial elements, namely spatial dependency and spatial heterogeneity [7]. Out of these two, spatial dependency or spatial autocorrelation is best known and most acknowledged. Spatial dependency occurs when the intensity of an event at one location influences the intensity at surrounding locations. Meanwhile, spatial heterogeneity describes the distribution of an event and the distribution of a relationship across a landscape. This means that what matters in a socio-economic relationship may not be homogenously distributed across space, such as infant mortality rates or unemployment rates at different locations. Both of these issues tend to arise when data are collected by location units such as provinces, districts/cities, sub-districts, villages, etc.

One statistical technique commonly used to address spatial effects is geographically weighted regression (GWR) model [8]. The GWR model allows the relationships between response variables and explanatory variables vary over space so the regression coefficients systematically vary across space. With this method, separate regression models are estimated for each location unit. All observations, i.e., location units, are included in each regression but with different weights. The farther a spatial unit from the location for which the regression is fitted, the smaller its data are weighted. This is in line with Tobler's First Law, which stated that everything is related to everything else, but near things are more related than distant things [9].

The GWR models for univariate response applied in previous studies were based on normal distribution [10–13]. Another study used Poisson distribution to develop a geographically weighted Poisson regression model and semi-parametric geographically weighted Poisson regression [14]. Meanwhile, Rodrigues et al. [15] and Chen et al. [16] used the GWR model based on logistic distribution and negative binomial distribution, respectively. The development of the GWR model for multivariate responses based on several different distributions has been carried out, such as based on normal distribution [17], Poisson distribution [18], Weibull distribution [19], Student-*t* distribution [20], negative binomial distribution [21], and logistic distribution [22].

Furthermore, two GWR models for count data which use multivariate responses have been developed, based on the Poisson distribution and negative binomial distribution. The first model is called geographically weighted multivariate Poisson regression (GWMPR) model [18], while the second model is called geographically weighted negative binomial regression (GWMNBR) model [21]. The GWMPR model assumes equidispersion, whereas the GWMNBR model only handles cases of overdispersion. In this paper, we propose an alternative multivariate regression model that can handle any type of dispersion by taking into account spatial effects in measuring the relationships between the response variables and the explanatory variables. The proposed model is called a geographically weighted multivariate generalized Poisson regression (GWMGPR) model.

The following discussion in this paper is divided into several main topics. In Section 2, we define the GWMGPR model followed by discussions on the parameter estimation using the maximum likelihood estimation (MLE) method in Section 3. We present simultaneous hypothesis testing in Section 4 to test the significance of all parameters together. In this section, we define the distribution of the test statistic and the critical region for the corresponding hypothesis by applying the maximum likelihood ratio test. We give conclusions in Section 5.

2. GWMGPR Model

The GWMGPR model is an extension of existing model of MGPR developed by Famoye [2] as follows:

$$P\left(y_1,\ldots,y_g\right) = \prod_{l=1}^{g}\left[\left(\frac{\mu_l}{1+\varphi_l\mu_l}\right)^{y_l}\frac{(1+\varphi_l y_l)^{y_l-1}}{y_l!}\exp\left[-\frac{\mu_l}{1+\varphi_l\mu_l}\left(1+\varphi_l y_l\right)\right]\right]$$

$$\times\left[1+\sum_{l<m}^{g}\gamma_{lm}\left(e^{-y_l}-z_l\right)\left(e^{-y_m}-z_m\right)\right], \tag{1}$$

where $y_1, \ldots, y_g = 0, 1, 2, \ldots$ with $z_l = \exp\left(\frac{\mu_l(r_l-1)}{1+\varphi_l\mu_l}\right)$; $l = 1, 2, \ldots, g$, and $\ln(r_l) - \frac{\varphi_l\mu_l(r_l-1)}{1+\varphi_l\mu_l} + 1 = 0$, $l = 1, 2, \ldots, g$, where y_l is the response variable $(l = 1, 2, \ldots, g)$, φ_l is the dispersion parameter of y_l, and γ_{lm} shows the correlation between y_l and y_m $(l, m = 1, 2, \ldots, g; l < m)$. When $\varphi_l = 0$, the MGPR model is reduced to multivariate Poisson regression model. Meanwhile, if $\varphi_l < 0$ or $\varphi_l > 0$, then the MGPR model experiences underdispersion or overdispersion, respectively. The variable y_l and y_m are negatively or positively correlated when $\gamma_{lm} < 0$ or $\gamma_{lm} > 0$, respectively.

We use the usual log-linear relationship between the marginal Y_{il} and the covariates $\mathbf{x}_i$ as follows:

$$\mu_{il} = q_{il} \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_l\right), \tag{2}$$

where q_{il} is an exposure measure for observation i of the response variable l, $\mathbf{x}_i$ is a vector of the explanatory variables, and $\boldsymbol{\beta}_l$ is a parameter vector of the MGPR model for the response variable l,

$$\mathbf{x}_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix}^T_{(p+1)\times 1}, \quad i = 1, 2, \ldots, n,$$

$$\boldsymbol{\beta}_l = \begin{bmatrix} \beta_{l0} & \beta_{l1} & \beta_{l2} & \cdots & \beta_{lp} \end{bmatrix}^T_{(p+1)\times 1}, \quad l = 1, 2, \ldots, g.$$

Equation (2) produces one regression model with the same regression coefficients for all n observations (locations), which we called the global model. Meanwhile, the local regression model produces n regression models; therefore, each location has its own regression coefficients. This is the underlying idea of the GWMGPR model, which is to build a model that takes into account spatial variations in relationships that allows the regression coefficients to vary across geographical locations. In this case, the location-specific regression coefficients are functions of longitude and latitude coordinates denoted by $\mathbf{u}_i = (u_{1i}, u_{2i})$, a vector of two-dimensional coordinates of location i. This means that the response variables $Y_{1i}, Y_{2i}, \ldots, Y_{il}$ are predicted by explanatory variables $\mathbf{x}_i$, each of which has a regression coefficient depending on the location where the data were observed.

Let n random samples $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, 2, \ldots, n$ from the random variables $\mathbf{y}_i \sim MGP$ $(\boldsymbol{\mu}(\mathbf{u}_i), \boldsymbol{\varphi}(\mathbf{u}_i), \boldsymbol{\gamma}(\mathbf{u}_i))$ be given, where $\mathbf{u}_i$ is a two-dimensional coordinate vector of location i; then, the GWMGPR model based on MGPR in (2) is written as

$$\mu_{il}(\mathbf{u}_i) = q_{il} \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_l(\mathbf{u}_i)\right), \tag{3}$$

where q_{il} is the exposure for location i of the response variable l with

$$\mathbf{x}_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix}^T_{(p+1)\times 1}, \quad i = 1, 2, \ldots, n,$$

$$\boldsymbol{\beta}_l(\mathbf{u}_i) = \begin{bmatrix} \beta_{l0}(\mathbf{u}_i) & \beta_{l1}(\mathbf{u}_i) & \beta_{l2}(\mathbf{u}_i) & \cdots & \beta_{lp}(\mathbf{u}_i) \end{bmatrix}^T_{(p+1)\times 1}, \quad l = 1, 2, \ldots, g; \; i = 1, 2, \ldots, n.$$

3. Parameter Estimation

As stated earlier, we use the MLE method to obtain the estimators of the GWMGPR model. The MLE method is widely used to find estimators from observed data. This method requires that the estimators, say $\hat{\boldsymbol{\theta}}$, maximizes the log-likelihood function by solving the gradient function to zero. For a more detailed discussion about MLE, including its principles and properties, see Pawitan [23].

Supposing n random samples $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, 2, \ldots, n$ are taken independently from random variables $\mathbf{y}_i \sim MGP(\boldsymbol{\mu}(\mathbf{u}_i), \boldsymbol{\varphi}(\mathbf{u}_i), \boldsymbol{\gamma}(\mathbf{u}_i))$, then the likelihood function of location i that corresponds to (1) is

$$L\left(\Omega\right) = L\left(\mu_l\left(\mathbf{u}_i\right), \varphi_l\left(\mathbf{u}_i\right), \gamma_{lm}\left(\mathbf{u}_i\right)\right), \quad (i = 1, 2, \ldots, n; \ l, m = 1, 2 \ldots, g; \ l < m),$$

$$= \prod_{i=1}^{n} \left[\prod_{l=1}^{g} \left(\left(\frac{\mu_{il}\left(\mathbf{u}_i\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} \right)^{y_{il}} \frac{\left(1 + \varphi_l\left(\mathbf{u}_i\right)y_{il}\right)^{y_{il}-1}}{y_{il}!} \right. \right.$$

$$\left. \exp\left(-\frac{\mu_{il}\left(\mathbf{u}_i\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} \left(1 + \varphi_l\left(\mathbf{u}_i\right)y_{il}\right) \right) \right)$$

$$\left. \times \left(1 + \sum_{l<m}^{g} \gamma_{lm}\left(\mathbf{u}_i\right) \left(e^{-y_{il}} - z_{il}\right)\left(e^{-y_{im}} - z_{im}\right) \right) \right], \tag{4}$$

where

$$z_{il} = \exp\left(\frac{\mu_{il}\left(\mathbf{u}_i\right)\left(r_l - 1\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} \right), \tag{5}$$

$$\ln\left(r_l\right) - \frac{\varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)\left(r_l - 1\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} + 1 = 0, \tag{6}$$

and $\mu_{il}\left(\mathbf{u}_i\right)$ is stated in (3). Hence, the parameter vector of location i of the corresponding GWMGPR model is written as

$$\theta\left(\mathbf{u}_i\right) = \left[\boldsymbol{\beta}_1^T\left(\mathbf{u}_i\right) \quad \cdots \quad \boldsymbol{\beta}_g^T\left(\mathbf{u}_i\right) \quad \varphi_1\left(\mathbf{u}_i\right) \quad \cdots \quad \varphi_g\left(\mathbf{u}_i\right) \quad \gamma_{12}\left(\mathbf{u}_i\right) \quad \cdots \quad \gamma_{(g-1)g}\left(\mathbf{u}_i\right) \right]^T. \tag{7}$$

Let Q_1 be the log-likelihood function of (4) such that

$$Q_1 = \ln\left(L\left(\mu_l\left(\mathbf{u}_i\right), \varphi_l\left(\mathbf{u}_i\right), \gamma_{lm}\left(\mathbf{u}_i\right)\right)\right), \quad (i = 1, 2, \ldots, n; \ l, m = 1, 2 \ldots, g; \ l < m),$$

$$= \sum_{i=1}^{n} \left[\sum_{l=1}^{g} \left(y_{il} \ln\left(\frac{\mu_{il}\left(\mathbf{u}_i\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} \right) + \left(y_{il} - 1\right)\ln\left(1 + \varphi_l\left(\mathbf{u}_i\right)y_{il}\right) \right. \right.$$

$$\left. \left. - \ln\left(y_{il}!\right) - \frac{\mu_{il}\left(\mathbf{u}_i\right)\left(1 + \varphi_l\left(\mathbf{u}_i\right)y_{il}\right)}{1 + \varphi_l\left(\mathbf{u}_i\right)\mu_{il}\left(\mathbf{u}_i\right)} \right) + \ln\left(1 + Q_2\right) \right], \tag{8}$$

where

$$Q_2 = \sum_{l<m}^{g} \gamma_{lm}\left(\mathbf{u}_i\right)\left(e^{-y_{il}} - z_{il}\right)\left(e^{-y_{im}} - z_{im}\right), \tag{9}$$

and z_{il} is defined in (5).

To estimate the parameters for location i, we can approximate Equation (2) by Equation (3). The regression model is built on using a subset or the entire data points by giving a weight based on the distance of the data points to location i. We give more weight to data points that are closer to location i than other data points that are farther away. Thus, the weights vary according to the proximity of the data points to location i. This means that we need to compute a weighting matrix for each location where the weights represent the proximity of each data point to location i with closer points having more weight in the estimation of the parameters for location i.

The above explanation means that the local parameters of $\hat{\theta}$ are estimated by a weighting function, w_{ij}. The weights $w_{ij}, i, j = 1, 2, \ldots, n$ at each location $\mathbf{u}_i$ are defined by a function of distance d_{ij} between the center of location i and other locations. Several weighting functions (kernels) are presented in Fotheringham et al. [8]. The kernels generally define weights such that locations close to the regression point will make more contributions in estimating the parameters than locations farther away. The most common kernels are the Gaussian function and bi-square function accompany with fixed or adaptive bandwidth.

Fixed Gaussian weighting function is defined as

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2 \right),$$

and fixed bi-square weighting function is

$$
w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2; & \text{if } d_{ij} < b \\ 0 & ; \quad \text{otherwise,} \end{cases}
$$

where d_{ij} is Euclidean distance between location i and location j

$$
d_{ij} = \sqrt{\left(u_{1i} - u_{1j}\right)^2 + \left(u_{2i} - u_{2j}\right)^2},
$$

and b is the bandwidth that controls the degree of distance-decay. The bandwidth is a measure of the distance-decay in the weighting function and indicates the extent to which the local model should be smoothed [8]. When b is large (small), it gives relative higher (lower) weight for more remote locations. This means that a fixed weighting functions may lead to a problem of undersmoothing in areas with only small observations and oversmoothing in areas with a high density of observations.

An alternative to the fixed weighting functions is adaptive kernels or spatially varying kernels, which give smaller bandwidth in areas with more data points and greater bandwidth in areas with less data points. Two types of adaptive kernel are adaptive Gaussian and adaptive bi-square. Adaptive Gaussian weighting function is defined as

$$
w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b_i}\right)^2\right),
$$

and adaptive bi-square weighting function is stated as

$$
w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b_i}\right)^2\right)^2; & \text{if } d_{ij} \leq b_i \\ 0 & ; \text{ otherwise,} \end{cases}
$$

where b_i is the bandwidth of location i.

Therefore, the implementation of GWMGPR requires a choice of specific function: a bisquare or a Gaussian function and the choice of bandwidth. The results of GWMGPR are more sensitive to the bandwidth of the particular weighting function chosen rather than to the selection of the weighting function. This means that we need to determine optimum bandwidth so there are sufficient data points within the kernel to get a reliable estimation in terms of bias and standard error but not too large of bandwidth so that there are too many data points such that the heterogeneity is washed away or not too high of bandwidth that the kernel only includes one data point for each location, then the system breaks down. The greater the bandwidth, the lower standard error of the coefficient regression but the higher bias, and vice versa.

There are several techniques used to obtain optimal bandwidth in terms of optimum trade-off between bias and standard error, such as the generalized cross-validation (GCV), which is used to determine the best combination of the bias and standard error. The optimum bandwidth is indicated by the minimum value of GCV, which is defined as follows [8]:

$$
GCV = \frac{n \sum\limits_{i=1}^{n} \left(\mathbf{y}_i - \hat{\mathbf{y}}_i\left(b\right)\right)^T \left(\mathbf{y}_i - \hat{\mathbf{y}}_i\left(b\right)\right)}{\left(n - v_1\right)^2},
$$

where $\hat{\mathbf{y}}_i\left(b\right)$ is the prediction of $\mathbf{y}_i$ for bandwidth b, and v_1 is the number of effective parameters in the GWMGPR model.

Once the weights w_{ij} are calculated, the parameter estimation is determined. To obtain the estimator of location i, we need information of location j by applying a weighting function w_{ij} to the

log-likelihood function. When the optimal weighting function is obtained, the log-likelihood function of location i in (8) is written as a weighted log-likelihood function as follows:

$$\mathbb{Q}_1 = \ln \left(L \left(\mu_l \left(\mathbf{u}_i \right), \varphi_l \left(\mathbf{u}_i \right), \gamma_{lm} \left(\mathbf{u}_i \right) \right) \right) w_{ij}, \quad (l, m = 1, 2 \ldots, g; \; l < m),$$

$$\mathbb{Q}_1 = \sum_{j=1}^{n} \left[\sum_{l=1}^{g} \left(y_{jl} \ln \left(\frac{\mu_{jl} \left(\mathbf{u}_i \right)}{1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right)} \right) + \left(y_{jl} - 1 \right) \ln \left(1 + \varphi_l \left(\mathbf{u}_i \right) y_{jl} \right) \right. \right.$$
$$\left. \left. - \ln \left(y_{jl}! \right) - \frac{\mu_{jl} \left(\mathbf{u}_i \right) \left(1 + \varphi_l \left(\mathbf{u}_i \right) y_{jl} \right)}{1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right)} \right) + \ln \left(1 + \mathbb{Q}_2 \right) \right] w_{ij}, \tag{10}$$

where

$$\mathbb{Q}_2 = \sum_{l<m}^{g} \gamma_{lm} \left(\mathbf{u}_i \right) \left(e^{-y_{jl}} - z_{jl} \right) \left(e^{-y_{jm}} - z_{jm} \right), \tag{11}$$

$$z_{jl} = \exp \left(\frac{\mu_{jl} \left(\mathbf{u}_i \right) \left(r_l - 1 \right)}{1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right)} \right), \tag{12}$$

$$\ln \left(r_l \right) - \frac{\varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right) \left(r_l - 1 \right)}{1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right)} + 1 = 0, \tag{13}$$

and

$$\mu_{jl} \left(\mathbf{u}_i \right) = q_{jl} \; \exp \left(\mathbf{x}_j^T \boldsymbol{\beta}_l \left(\mathbf{u}_i \right) \right). \tag{14}$$

The MLE method is applied to the parameter estimation; therefore, the weighted log-likelihood function in (10) is maximized over the vector parameter $\boldsymbol{\theta} \left(\mathbf{u}_i \right)$ in (7). The first partial derivatives of $\mathbb{Q}_1$ with respect to $\boldsymbol{\theta} \left(\mathbf{u}_i \right)$ are shown as follows:

$$\frac{\partial \mathbb{Q}_1}{\partial \boldsymbol{\beta}_l^T \left(\mathbf{u}_i \right)} = \sum_{j=1}^{n} \left(\frac{\left(y_{jl} - \mu_{jl} \left(\mathbf{u}_i \right) \right) \mathbf{x}_j}{\left(1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right) \right)^2} + \frac{\partial \mathbb{Q}_2 / \partial \boldsymbol{\beta}_l^T \left(\mathbf{u}_i \right)}{1 + \mathbb{Q}_2} \right) w_{ij}, \tag{15}$$

$$\frac{\partial \mathbb{Q}_1}{\partial \varphi_l \left(\mathbf{u}_i \right)} = \sum_{j=1}^{n} \left(-\frac{\mu_{jl} \left(\mathbf{u}_i \right) y_{jl}}{1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \left(\mathbf{u}_i \right)} + \frac{\left(y_{jl} - 1 \right) y_{jl}}{\left(1 + \varphi_l \left(\mathbf{u}_i \right) y_{jl} \left(\mathbf{u}_i \right) \right)} \right.$$
$$\left. - \frac{\mu_{jl} \left(\mathbf{u}_i \right) \left(y_{jl} - \mu_{jl} \left(\mathbf{u}_i \right) \right)}{\left(1 + \varphi_l \left(\mathbf{u}_i \right) \mu_{jl} \right)^2} + \frac{\partial \mathbb{Q}_2 / \partial \varphi_l \left(\mathbf{u}_i \right)}{1 + \mathbb{Q}_2} \right) w_{ij}, \tag{16}$$

$$\frac{\partial \mathbb{Q}_1}{\partial \gamma_{lm} \left(\mathbf{u}_i \right)} = \sum_{j=1}^{n} \left(\frac{1}{1 + \mathbb{Q}_2} \frac{\partial \mathbb{Q}_2}{\partial \gamma_{lm} \left(\mathbf{u}_i \right)} \right) w_{ij} \tag{17}$$

with $\mathbb{Q}_2$ is stated in (11), where

$$\frac{\partial \mathbb{Q}_2}{\partial \boldsymbol{\beta}_l^T \left(\mathbf{u}_i \right)} = -\sum_{l<m}^{g} \gamma_{lm} \left(\mathbf{u}_i \right) \left(e^{-y_{jm}} - z_{jm} \right) \frac{\partial z_{jl}}{\partial \boldsymbol{\beta}_l^T \left(\mathbf{u}_i \right)},$$

$$\frac{\partial \mathbb{Q}_2}{\partial \varphi_l \left(\mathbf{u}_i \right)} = -\sum_{l<m}^{g} \gamma_{lm} \left(e^{-y_{jm}} - z_{jm} \right) \frac{\partial z_{jl}}{\partial \varphi_l \left(\mathbf{u}_i \right)},$$

$$\frac{\partial \mathbb{Q}_2}{\partial \gamma_{lm} \left(\mathbf{u}_i \right)} = \sum_{l<m}^{g} \left(\frac{\partial}{\partial \gamma_{lm} \left(\mathbf{u}_i \right)} \gamma_{jlm} \left(e^{-y_{jl}} - z_{jl} \right) \left(e^{-y_{jm}} - z_{jm} \right) \right)$$
$$= \left(e^{-y_{jl}} - z_{jl} \right) \left(e^{-y_{jm}} - z_{jm} \right),$$

and

$$\frac{\partial z_{jl}}{\partial \boldsymbol{\beta}_l^T (\mathbf{u}_i)} = \frac{z_{jl} \; \mathbf{x}_j \; \mu_{jl} \, (r_l - 1)}{\left(1 + \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i)\right) \left(1 + \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i) \, - \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i) \; r_l\right)},$$

$$\frac{\partial z_{jl}}{\partial \varphi_l (\mathbf{u}_i)} = \frac{z_{jl} \; \mu_{jl}^2 (\mathbf{u}_i) \, (r_l - 1)^2}{\left(1 + \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i)\right) \left(1 + \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i) \, - \varphi_l (\mathbf{u}_i) \, \mu_{jl} (\mathbf{u}_i) \; r_l\right)},$$

with z_{jl} and r_l are stated in (12) and (13), respectively.

From Equations (15) to (17), the total number of parameters of GWMGPR model, notated by a_1, is $(g + 2p + 3) \, gn/2$. The first partial derivatives of $\mathbb{Q}_1$ with respect to each parameter in (15) through (17) are not closed-form and the MLE does not give analytical solutions. A numerical optimization, such as the Newton–Raphson algorithm, as the simplest and the most common optimization method, is needed when there is no analytical solution available to obtain the maximum likelihood estimators. The Newton–Raphson algorithm is

$$\hat{\boldsymbol{\theta}}^{(h+1)} (\mathbf{u}_i) = \hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i) - \mathbf{H}^{-1} \left(\hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i)\right) \mathbf{g} \left(\hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i)\right),$$

where $\hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i)$ is the vector of parameter of location i at the *hth* iteration, $\mathbf{H}^{-1} \left(\hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i)\right)$ is the inverse of Hessian matrix of location i at the *hth* iteration, and $\mathbf{g} \left(\hat{\boldsymbol{\theta}}^{(h)} (\mathbf{u}_i)\right)$ is the gradient vector of location i at the *hth* iteration. For a more detailed discussion about numerical optimization methods, see Cameron and Trivedi [24].

4. Simultaneous Hypothesis Testing

Simultaneous hypothesis testing aims to test more than one parameter, for example to test all the parameters $\beta_{lk} (\mathbf{u}_i)$, $(l = 1, 2, \ldots g; \; i = 1, 2, \ldots, n)$ together. This is used to determine the significance of the regression parameters in the model jointly. The hypothesis for simultaneous testing of parameter $\beta_{lk} (\mathbf{u}_i)$ is stated as follows:

$$H_0 : \beta_{l1} (\mathbf{u}_i) = \beta_{l2} (\mathbf{u}_i) = \ldots = \beta_{lp} (\mathbf{u}_i) = 0; \; (i = 1, 2, \ldots, n; \; l, m = 1, 2, \ldots, g),$$
$$H_1 : \text{at least one } \beta_{lk} (\mathbf{u}_i) \neq 0. \tag{18}$$

If we cannot reject H_0, then it means that none of the regression parameter values are significant. In case of rejection, it means that one or more of $\beta_{lk} (\mathbf{u}_i)$'s have a significant effect in the model.

The critical region of hypothesis testing in (18) is obtained using the likelihood ratio test technique by calculating the ratio between the maximum value of the likelihood function under H_0, $L (\omega)$, with the maximum value of the likelihood function under the population, $L (\Omega)$. The maximum value of $L (\Omega)$ is the maximum value in (4). Meanwhile, the calculation of $L (\omega)$ is described below.

The first step to obtain the maximum $L (\omega)$ in determining the critical region for the hypothesis in (18) is to define the parameter space under H_0, $\omega = \{\beta_{\omega l0} (\mathbf{u}_i), \varphi_{\omega l} (\mathbf{u}_i), \gamma_{\omega lm} (\mathbf{u}_i)\}$, $(l, m = 1, 2, \ldots, g; \; l < m)$. Let $\boldsymbol{\theta}_{\omega 2} (\mathbf{u}_i)$ be the vector of parameters under H_0

$$\boldsymbol{\theta}_{\omega 2} (\mathbf{u}_i) = \begin{bmatrix} \beta_{\omega 10} (\mathbf{u}_i) & \cdots & \beta_{\omega g0} (\mathbf{u}_i) & \varphi_{\omega 1} (\mathbf{u}_i) & \cdots & \varphi_{\omega g} (\mathbf{u}_i) & \gamma_{\omega 12} (\mathbf{u}_i) & \cdots & \gamma_{\omega (g-1)g} (\mathbf{u}_i) \end{bmatrix}^T,$$

and the GWMGPR model of location i under H_0 be

$$\mu_{\omega il} (\mathbf{u}_i) = q_{il} \; \exp \left(\beta_{\omega l0} (\mathbf{u}_i)\right). \tag{19}$$

The likelihood function in space ω is

$$L\left(\omega\right) = \prod_{i=1}^{n} P\left(\mathbf{y}_i \,|\beta_{\omega l0}\left(\mathbf{u}_i\right), \varphi_{\omega l}\left(\mathbf{u}_i\right), \gamma_{\omega lm}\left(\mathbf{u}_i\right)\right), \quad (l, m = 1, 2, ..., g; \; l < m). \tag{20}$$

Let $Q_{1\omega} = \ln L\left(\omega\right)$, then the log-likelihood function under H_0 is written as

$$
\begin{aligned}
Q_{1\omega} &= \sum_{i=1}^{n} P\left(\mathbf{y}_i \,|\beta_{\omega l0}\left(\mathbf{u}_i\right), \varphi_{\omega l}\left(\mathbf{u}_i\right), \gamma_{\omega lm}\left(\mathbf{u}_i\right)\right), \quad (l, m = 1, 2, ..., g; \; l < m), \\
&= \sum_{i=1}^{n} \left[\sum_{l=1}^{g} \left(y_{il} \ln \left(\frac{\mu_{\omega il}\left(\mathbf{u}_i\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega il}\left(\mathbf{u}_i\right)} \right) + (y_{il} - 1) \ln \left(1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) y_{il}\right) \right. \right. \\
&\quad \left. \left. - \ln\left(y_{il}!\right) - \frac{\mu_{\omega il}\left(\mathbf{u}_i\right)\left(1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) y_{il}\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega il}\left(\mathbf{u}_i\right)} \right) + \ln\left(1 + Q_{2\omega}\right) \right],
\end{aligned}
\tag{21}
$$

where

$$Q_{2\omega} = \sum_{l<m}^{g} \gamma_{\omega lm}\left(\mathbf{u}_i\right)\left(e^{-y_{il}} - z_{\omega il}\right)\left(e^{-y_{im}} - z_{\omega im}\right),$$

$$z_{\omega il} = \exp\left(\frac{\mu_{\omega il}\left(\mathbf{u}_i\right)\left(r_{\omega l} - 1\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega il}\left(\mathbf{u}_i\right)} \right),$$

and

$$\ln\left(r_{\omega l}\right) - \frac{\varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega il}\left(\mathbf{u}_i\right)\left(r_{\omega l} - 1\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega il}\left(\mathbf{u}_i\right)} + 1 = 0.$$

By carrying out the same procedure as described previously, we need the information of location j to obtain the estimator of location i. The contribution of each location depends on the weighting function w_{ij}, such that the weighted log-likelihood function corresponding to (21) is written as

$$
\begin{aligned}
\mathbb{Q}_{1\omega} &= w_{ij} \ln L\left(\omega\right) \\
\mathbb{Q}_{1\omega} &= \sum_{j=1}^{n} \left[\sum_{l=1}^{g} \left(y_{jl} \ln \left(\frac{\mu_{\omega jl}\left(\mathbf{u}_i\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega jl}\left(\mathbf{u}_i\right)} \right) + \left(y_{jl} - 1\right) \ln \left(1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) y_{jl}\right) \right. \right. \\
&\quad \left. \left. - \ln\left(y_{jl}!\right) - \frac{\mu_{\omega jl}\left(\mathbf{u}_i\right)\left(1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) y_{jl}\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega jl}\left(\mathbf{u}_i\right)} \right) + \ln\left(1 + \mathbb{Q}_{2\omega}\right) \right] w_{ij},
\end{aligned}
\tag{22}
$$

where

$$\mathbb{Q}_{2\omega} = \sum_{l<m}^{g} \gamma_{\omega lm}\left(\mathbf{u}_i\right)\left(e^{-y_{jl}} - z_{\omega jl}\right)\left(e^{-y_{jm}} - z_{\omega jm}\right), \tag{23}$$

$$z_{\omega jl} = \exp\left(\frac{\mu_{\omega jl}\left(\mathbf{u}_i\right)\left(r_l - 1\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega jl}\left(\mathbf{u}_i\right)} \right),$$

$$\ln\left(r_l\right) - \frac{\varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega jl}\left(\mathbf{u}_i\right)\left(r_l - 1\right)}{1 + \varphi_{\omega l}\left(\mathbf{u}_i\right) \mu_{\omega jl}\left(\mathbf{u}_i\right)} + 1 = 0,$$

and $\mu_{\omega jl}\left(\mathbf{u}_i\right)$ is defined as

$$\mu_{\omega jl}\left(\mathbf{u}_i\right) = q_{jl}\;\exp\left(\beta_{\omega l0}\left(\mathbf{u}_i\right)\right).$$

The first partial derivatives of $\mathbb{Q}_{1\omega}$ with respect to each parameter in the parameter vector $\theta_{\omega 2}(\mathbf{u}_i)$ are

$$\frac{\partial \mathbb{Q}_{1\omega}}{\partial \beta_{\omega l 0}(\mathbf{u}_i)} = \sum_{j=1}^{n} \left(\frac{\left(y_{jl} - \mu_{\omega jl}(\mathbf{u}_i) \right)}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \right)^2} + \frac{\partial \mathbb{Q}_{2\omega}/\partial \beta_{\omega l 0}(\mathbf{u}_i)}{1 + \mathbb{Q}_{2\omega_2}} \right) w_{ij}, \tag{24}$$

$$\frac{\partial \mathbb{Q}_{1\omega}}{\partial \varphi_{\omega l}(\mathbf{u}_i)} = \sum_{j=1}^{n} \left(-\frac{\mu_{\omega jl}(\mathbf{u}_i) \, y_{jl}}{1 + \varphi_{\omega l}(\mathbf{u}_i) \, \mu_{\omega jl}(\mathbf{u}_i)} + \frac{\left(y_{jl} - 1 \right) y_{jl}}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \, y_{jl} \right)} \right.$$
$$\left. -\frac{\mu_{\omega jl}(\mathbf{u}_i) \left(y_{il} - \mu_{\omega jl}(\mathbf{u}_i) \right)}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \, \mu_{\omega jl}(\mathbf{u}_i) \right)^2} + \frac{\partial \mathbb{Q}_{2\omega}/\partial \varphi_{\omega l}(\mathbf{u}_i)}{1 + \mathbb{Q}_{2\omega}} \right) w_{ij}, \tag{25}$$

$$\frac{\partial \mathbb{Q}_{1\omega}}{\partial \gamma_{\omega l m}(\mathbf{u}_i)} = \sum_{j=1}^{n} \left(\frac{1}{1 + \mathbb{Q}_{2\omega}} \frac{\partial \mathbb{Q}_{2\omega}}{\partial \gamma_{\omega l m}(\mathbf{u}_i)} \right) w_{ij}, \tag{26}$$

where $\mathbb{Q}_{2\omega}$ is defined in (23) with

$$\frac{\partial \mathbb{Q}_{2\omega}}{\partial \beta_{\omega l 0}(\mathbf{u}_i)} = - \sum_{l<m}^{g} \gamma_{\omega l m}(\mathbf{u}_i) \left(e^{-y_{jm}} - z_{\omega jm} \right) \frac{\partial z_{\omega jl}}{\partial \beta_{\omega l 0}(\mathbf{u}_i)},$$

$$\frac{\partial \mathbb{Q}_{2\omega}}{\partial \varphi_{\omega l}(\mathbf{u}_i)} = - \sum_{l<m}^{g} \gamma_{\omega l m}(\mathbf{u}_i) \left(e^{-y_{jm}} - z_{\omega jm} \right) \frac{\partial z_{\omega jl}}{\partial \varphi_{\omega l}(\mathbf{u}_i)},$$

$$\frac{\partial \mathbb{Q}_{2\omega}}{\partial \gamma_{\omega l m}(\mathbf{u}_i)} = \sum_{l<m}^{g} \left(\frac{\partial}{\partial \gamma_{\omega l m}} \gamma_{\omega l m}(\mathbf{u}_i) \left(e^{-y_{jl}} - z_{\omega jl} \right) \left(e^{-y_{jm}} - z_{\omega jm} \right) \right)$$
$$= \left(e^{-y_{jl}} - z_{\omega jl} \right) \left(e^{-y_{im}} - z_{\omega jm} \right),$$

and

$$\frac{\partial z_{\omega jl}}{\partial \beta_{\omega l 0}(\mathbf{u}_i)} = \frac{z_{\omega jl} \, \mu_{\omega jl}(\mathbf{u}_i)}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \, \mu_{\omega jl}(\mathbf{u}_i) \right)}$$
$$\times \frac{\left(r_{\omega l} - 1 \right)}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \ - \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \ r_{\omega l} \right)},$$

$$\frac{\partial z_{\omega jl}}{\partial \varphi_{\omega l}(\mathbf{u}_i)} = \frac{z_{\omega jl} \, \mu_{\omega jl}^2(\mathbf{u}_i)}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \right)}$$
$$\times \frac{\left(r_{\omega l} - 1 \right)^2}{\left(1 + \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \ - \varphi_{\omega l}(\mathbf{u}_i) \ \mu_{\omega jl}(\mathbf{u}_i) \ r_{\omega l} \right)}.$$

Equations (24) through (26) are not closed-form, so a numerical iteration is needed to solve the equations. The total number of parameters in the parameter space ω, notated by a_2, is $(g+3) \, gn/2$.

Using a significance level $\alpha \in (0, 1)$, the critical region of the likelihood ratio test for the GWMGPR model with regard to the hypothesis in (18) is

$$\Lambda = \frac{\max \ L(\omega)}{\max \ L(\Omega)} < k_{\alpha}, \tag{27}$$

where k_α is a constant that depends on α with $0 \le k_\alpha \le 1$, $L(\Omega)$ and $L(\omega)$ are stated in (4) and (20), respectively. Equation (27) can be written as

$$G^2 = -\ln\left(\Lambda^2\right) = -2\ln(\Lambda)$$
$$= 2\left(\ln L\left(\hat{\Omega}\right) - \ln L\left(\hat{\omega}\right)\right). \tag{28}$$

The distribution of the test statistic G^2 in (28) is written in the following Theorem 1.

Theorem 1. *Supposing the maximum likelihood estimator under H_0 and the population regarding to hypothesis testing stated in (18) are*

$$\hat{\boldsymbol{\theta}}_{\omega 2} \atop {(g+3)gn/2 \times 1} = \left[\hat{\beta}_{\omega 10}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\beta}_{\omega g0}\left(\mathbf{u}_i\right) \quad \hat{\varphi}_{\omega 1}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\varphi}_{\omega g}\left(\mathbf{u}_i\right) \quad \hat{\gamma}_{\omega 12}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\gamma}_{\omega(g-1)g}\left(\mathbf{u}_i\right)\right]^T,$$

$$\hat{\boldsymbol{\theta}} \atop {(g+2p+3)gn/2 \times 1} = \left[\hat{\boldsymbol{\beta}}_1^T\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\boldsymbol{\beta}}_g^T\left(\mathbf{u}_i\right) \quad \hat{\varphi}_1\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\varphi}_g\left(\mathbf{u}_i\right) \quad \hat{\gamma}_{12}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\gamma}_{(g-1)g}\left(\mathbf{u}_i\right)\right]^T,$$

respectively, then the test statistic G^2 in (28) follows a chi-square distribution with gpn degrees of freedom.

Proof of Theorem 1. Suppose the maximum likelihood estimator under the population related to the hypothesis stated in (18) is partitioned as follows:

$$\hat{\boldsymbol{\theta}} \atop {(g+2p+3)gn/2 \times 1} = \left[\hat{\boldsymbol{\theta}}_1 \quad \hat{\boldsymbol{\theta}}_2\right]^T,$$

where

$$\hat{\boldsymbol{\theta}}_1 = \left[\hat{\beta}_{l1}\left(\mathbf{u}_i\right) \quad \hat{\beta}_{l2}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\beta}_{lp}\left(\mathbf{u}_i\right)\right]^T, \quad (l = 1, 2, \ldots, g),$$

and

$$\hat{\boldsymbol{\theta}}_2 = \left[\hat{\beta}_{10}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\beta}_{g0}\left(\mathbf{u}_i\right) \quad \hat{\varphi}_1\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\varphi}_g\left(\mathbf{u}_i\right) \quad \hat{\gamma}_{12}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\gamma}_{(g-1)g}\left(\mathbf{u}_i\right)\right]^T.$$

The maximum likelihood estimators and the known parameters in H_0 are partitioned as

$$\hat{\boldsymbol{\theta}}_\omega = \left[\boldsymbol{\theta}_{\omega 1} \quad \hat{\boldsymbol{\theta}}_{\omega 2}\right]^T,$$

where

$$\boldsymbol{\theta}_{\omega 1} \atop {gpn \times 1} = \left[0 \quad 0 \quad \cdots \quad 0\right]^T,$$

and

$$\hat{\boldsymbol{\theta}}_{\omega 2} \atop {(g+3)gn/2 \times 1} = \left[\hat{\beta}_{\omega 10}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\beta}_{\omega g0}\left(\mathbf{u}_i\right) \quad \hat{\varphi}_{\omega 1}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\varphi}_{\omega g}\left(\mathbf{u}_i\right) \quad \hat{\gamma}_{\omega 12}\left(\mathbf{u}_i\right) \quad \cdots \quad \hat{\gamma}_{\omega(g-1)g}\left(\mathbf{u}_i\right)\right]^T,$$

with the true parameter is partitioned as

$$\boldsymbol{\theta}_\omega = \left[\boldsymbol{\theta}_{\omega 1} \quad \boldsymbol{\theta}_2\right]^T.$$

Therefore, the hypothesis in (18) can be written as

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{\omega 1}$$
$$H_1 : \boldsymbol{\theta}_1 \ne \boldsymbol{\theta}_{\omega 1}.$$

The test statistic G^2 stated in (28) can be described as follows:

$$
\begin{aligned}
G^2 &= 2 \left[\ln L \left(\hat{\boldsymbol{\theta}} \right) - \ln L \left(\hat{\boldsymbol{\theta}}_\omega \right) \right] \\
&= 2 \left[\ln L \left(\hat{\boldsymbol{\theta}} \right) - \ln L \left(\boldsymbol{\theta}_\omega \right) \right] - 2 \left[\ln L \left(\hat{\boldsymbol{\theta}}_\omega \right) - \ln L \left(\boldsymbol{\theta}_\omega \right) \right].
\end{aligned}
\tag{29}
$$

Let $\ell \left(\boldsymbol{\theta}_\omega \right) = \ln L \left(\boldsymbol{\theta}_\omega \right)$, using Taylor's second order expansion around $\hat{\boldsymbol{\theta}}$, we get

$$
\ell \left(\boldsymbol{\theta}_\omega \right) \approx \ell \left(\hat{\boldsymbol{\theta}} \right) + \mathbf{g} \left(\hat{\boldsymbol{\theta}} \right) \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}} \right) - 1/2 \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}} \right)^T \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}} \right),
$$

where

$$
\mathbf{g} \left(\hat{\boldsymbol{\theta}} \right) = \left. \frac{\partial \ell \left(\boldsymbol{\theta} \right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}},
$$

and

$$
\mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) = - \left. \frac{\partial^2 \ell \left(\boldsymbol{\theta} \right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}.
$$

Since $\mathbf{g} \left(\hat{\boldsymbol{\theta}} \right) = \mathbf{0}$, then

$$
2 \left[\ell \left(\hat{\boldsymbol{\theta}} \right) - \ell \left(\boldsymbol{\theta}_\omega \right) \right] \approx \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\omega \right)^T \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\omega \right).
\tag{30}
$$

Meanwhile, when the function $\ell \left(\boldsymbol{\theta}_\omega \right)$ is represented via second order approximation around $\hat{\boldsymbol{\theta}}_\omega$, the following is obtained:

$$
\ell \left(\boldsymbol{\theta}_\omega \right) \approx \ell \left(\hat{\boldsymbol{\theta}}_\omega \right) + \mathbf{g} \left(\hat{\boldsymbol{\theta}} \right) \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}}_\omega \right) - 1/2 \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}}_\omega \right) \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\boldsymbol{\theta}_\omega - \hat{\boldsymbol{\theta}}_\omega \right),
$$

or it can be written as

$$
2 \left[\ell \left(\hat{\boldsymbol{\theta}}_\omega \right) - \ell \left(\boldsymbol{\theta}_\omega \right) \right] \approx \left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right)^T \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right).
\tag{31}
$$

Following (30) and (31), the test statistic in (29) can be written as

$$
G^2 \approx \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\omega \right)^T \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\omega \right) - \left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right)^T \mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right).
\tag{32}
$$

Let the partition of Fisher information matrix $\mathcal{I} \left(\hat{\boldsymbol{\theta}} \right)$ and its inverse $\mathcal{I}^{-1} \left(\hat{\boldsymbol{\theta}} \right)$ be

$$
\mathcal{I} \left(\hat{\boldsymbol{\theta}} \right) = \begin{bmatrix} \mathcal{I}_{11[gpn \times gpn]} & \mathcal{I}_{12[gpn \times (g+3)gn/2]} \\ \mathcal{I}_{21[(g+3)gn/2 \times gpn]} & \mathcal{I}_{22[(g+3)gn/2 \times (g+3)gn/2]} \end{bmatrix},
$$

and

$$
\mathcal{I}^{-1} \left(\hat{\boldsymbol{\theta}} \right) = \begin{bmatrix} \mathcal{I}^{11}_{gpn \times gpn} & \mathcal{I}^{12}_{gpn \times (g+3)gn/2} \\ \mathcal{I}^{21}_{(g+3)gn/2 \times gpn} & \mathcal{I}^{22}_{(g+3)gn/2 \times (g+3)gn/2} \end{bmatrix}.
$$

Based on the nature of the conditional distribution, given as $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{\omega 1}$, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$, we get

$$
\hat{\boldsymbol{\theta}}_{\omega 2} = \hat{\boldsymbol{\theta}}_2 - \mathcal{I}^{21} \left(\mathcal{I}^{11} \right)^{-1} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right).
\tag{33}
$$

A simpler form of (33) using a simple manipulation of the partitioned matrix of the Fisher information is

$$
\hat{\boldsymbol{\theta}}_{\omega 2} = \hat{\boldsymbol{\theta}}_2 + \mathcal{I}_{22}^{-1} \mathcal{I}_{21} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right).
$$

Since $\left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right) = \left(\mathbf{0}, \hat{\boldsymbol{\theta}}_{\omega 2} - \boldsymbol{\theta}_2 \right)$ and $\boldsymbol{\theta}_{\omega 2} - \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 + \mathcal{I}_{22}^{-1}\mathcal{I}_{21}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right)$, then

$$
\left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right)^T \mathcal{I}\left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}_\omega \right) = \left(\hat{\boldsymbol{\theta}}_{\omega 2} - \boldsymbol{\theta}_2 \right)^T \mathcal{I}_{22} \left(\hat{\boldsymbol{\theta}}_{\omega 2} - \boldsymbol{\theta}_2 \right)
$$

$$
= \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{bmatrix}^T \begin{bmatrix} \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{bmatrix}.
$$

The statistic G^2 in (32) can then be written as

$$
G^2 \approx \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\omega 2} \right)^T \mathcal{I}\left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\omega 2} \right) - \left(\hat{\boldsymbol{\theta}}_{\omega 2} - \boldsymbol{\theta}_{\omega 2} \right)^T \mathcal{I}\left(\hat{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}_{\omega 2} - \boldsymbol{\theta}_{\omega 2} \right)
$$

$$
= \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right)^T \left(\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21} \right) \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right)
$$

$$
= \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right)^T \left(\mathcal{I}^{11} \right)^{-1} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right).
$$

With regard to the normality for MLE under the regularity conditions (see Cameron and Trivedi [25]), the distribution of the partitioned matrix is

$$
\begin{bmatrix} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 \\ \hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \end{bmatrix} \overset{a}{\sim} N\left(\mathbf{0}, \mathcal{I}^{-1}\left(\boldsymbol{\theta} \right) \equiv \begin{bmatrix} \mathcal{I}^{11} & \mathcal{I}^{12} \\ \mathcal{I}^{21} & \mathcal{I}^{22} \end{bmatrix} \right),
$$

such that

$$
\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 \right) \overset{a}{\sim} N\left(\mathbf{0}, \mathcal{I}^{11} \right),
$$

$$
\left(\mathcal{I}^{11} \right)^{1/2} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 \right) \overset{a}{\sim} N\left(\mathbf{0}, \mathbf{I}_{gpn} \right),
$$

$$
G^2 = \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right)^T \left(\mathcal{I}^{11} \right)^{-1} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{\omega 1} \right) \overset{a}{\sim} \chi^2_{gpn}.
$$

$\square$

Regarding Equation (27), the critical region of the test statistic G^2 is

$$
\begin{aligned}
\alpha &= P\left(\Lambda < k_\alpha; \omega \right) \\
&= P\left(-2\ln \Lambda > -2\ln k_\alpha; \omega \right) \\
&= P\left(G^2 > c; \omega \right), \text{ where } c = -2\ln k_\alpha \\
&= P\left(G^2 > \chi^2_{\alpha, gpn} \right).
\end{aligned}
$$

Therefore, H_0 in (18) is rejected when $G^2 > \chi^2_{\alpha, gpn}$.

5. Conclusions

We developed a local regression model, called a geographically weighted multivariate generalized Poisson regression (GWMGPR) model. We have shown the step-by-step procedure to obtain the estimation of the parameters and the test statistic for hypothesis testing of the GWMGPR model. It is easy to see that the same procedure of developing the GWMGPR model presented in this article can be applied in developing other GWR models based on different multivariate distributions.

The distribution of the test statistic developed in this paper is based on an asymptotic approach. Therefore, further study is needed for development of exact distributions of the test statistics. Despite the advantages of presenting regression coefficients for each location, GWMGPR is a complex model that needs a challenging coding program. Furthermore, simulation and application of the

GWMGPR model using real data to validate the theoretical results in this paper that are being prepared for our next manuscript.

Author Contributions: Conceptualization, S.M.B., P.P., and S.S.; methodology, S.M.B., P.P., S.S. and S.P.R.; writing—original draft preparation, S.M.B.; writing—review and editing, S.M.B. and P.P. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GWMGPR	Geographically weighted multivariate generalized Poisson regression
GWMNBR	Geographically weighted multivariate negative binomial regression
GWMPR	Geographically weighted multivariate Poisson regression
GWR	Geographically weighted regression
MGPR	Multivariate generalized Poisson regression
MLE	Maximum likelihood estimation

References

1. Winkelmann, R. *Econometric Analysis of Count Data*, 5th ed.; Springer: Berlin, Germany, 2008.
2. Famoye, F. A Multivariate Generalized Poisson Regression Model. *Commun. Stat. Theory Methods* **2015**, *44*, 497–511. [CrossRef]
3. Famoye, F.; Okafor, R.; Adamu, M.O. A multivariate Generalized Poisson Distribution. *J. Stat. Theory Appl.* **2011**, *10*, 519–532.
4. Famoye, F. A New Bivariate Generalized Poisson Distribution. *Stat. Neerl.* **2010**, *64*, 112–124. [CrossRef]
5. Famoye, F. Comparisons of Some Bivariate Regression Models. *J. Stat. Comput. Simul.* **2012**, *82*, 937–949. [CrossRef]
6. Zamani, H.; Faroughi, P.; Ismail, N. Bivariate Generalized Poisson Regression Model: Applications on Health Care Data. *Empir. Econ.* **2016**, *51*, 1607–1621. [CrossRef]
7. Anselin, L. *Spatial Econometrics: Methods and Models*; Springer: Dordrecht, The Netherlands, 1988.
8. Fotheringham, A.S.; Brundson, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*; John Wiley & Son Ltd.: Chichester, UK, 2002.
9. Tobler, W.R. Cellular Geography. In *Theory and Decision Library (An International Series in the Philosophy and Methodology of the Social and Behavioral Sciences)*; Gale, S., Olsson, G., Eds.; Springer: Dordrecht, The Netherlands, 1979; pp. 379–386.
10. Bascuñán, M.M.; Quezada, C.R. Geographically Weighted Regression for Modelling the Accessibility to the Public Hospital Network in Concepción Metropolitan Area, Chile. *Geospat. Health* **2016**, *11*, 263–273.
11. Fotheringham, A.S.; Kelly, M.H.; Charlton, M. The Demographic Impacts of the Irish Famine: Towards a Greater Geographical Understanding. *Trans. Inst. Br. Geogr.* **2012**, *38*, 221–237. [CrossRef]
12. Lin, C.H.; Wen, T.H. Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue. *Int. J. Environ. Res. Public Health* **2011**, *8*, 2798–2815. [CrossRef] [PubMed]
13. Zhao, F.; Park, N. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transp. Res. Rec. J. Transp. Res. Board* **2004**, *1879*, 99–107. [CrossRef]
14. Nakaya, T.; Fotheringham, A.S.; Brunsdon, C.; Charlton, M. Geographhically Weighted Poisson Regression for Disease Association Mapping. *Stat. Med.* **2005**, *24*, 2695–2717. [CrossRef] [PubMed]
15. Rodrigues, M.; de la Riva, J.; Fotheringham, S. Modeling the Spatial Variation of the Explanatory Factors of Human-caused Wildfires in Spain Using Geographically Weighted Logistic Regression. *Appl. Geogr.* **2014**, *48*, 52–63. [CrossRef]
16. Chen, J.; Liu, L.; Xiao, L.; Xu, C.; Long, D. Integrative Analysis of Spatial Heterogeneity and Overdispersion of Crime with a Geographically Weighted Negative Binomial Model. *Int. J. Geo-Inf.* **2020**, *9*, 60. [CrossRef]

17. Harini, S.; Purhadi, Mashuri, M.; Sunaryo, S. Statistical Test for Multivariate Geographically Weighted Regression Model Using the Method of Maximum Likelihood Ratio Test. *Int. J. Appl. Math. Stat.* **2012**, *29*, 110–115.

18. Purhadi, B.W.O.; Purnami, S.W. Parameter Estimation of Geographically Weigthed Multivariate Poisson Regression. *Appl. Math. Sci.* **2015**, *9*, 4081–4093.

19. Suyitno; Purhadi; Sutikno; Irhamah. Parameter Estimation of Geographically Weighted Trivariate Weibull Regression Model. *Appl. Math. Sci.* **2016**, *10*, 861–878.

20. Sugiarti, H.; Purhadi; Sutikno; Purnami, S.W. Parameter Estimation of Geographically Weighted Multivariate t Regression Model. *J. Theor. Appl. Inf. Technol.* **2016**, *92*, 45–51.

21. Dewi, Y.S.; Purhadi; Sutikno; Purnami, S.W. Comparison of Nelder Mead and BFGS Algorithms on Geographically Weighted Multivariate Negative Binomial. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2019**, *9*, 979–987. [CrossRef]

22. Fathurahman, M.; Purhadi; Sutikno; Ratnasari, V. Geographically Weighted Multivariate Logistic Regression Model and Its Application. *Abstr. Appl. Anal.* **2020**, *2020*, 8353481. [CrossRef]

23. Pawitan, Y. *In All Likelihood*; Clarendon Press: Oxford, UK, 2001.

24. Cameron, A.C.; Trivedi, P.K. *Microeconometrics Methods and Applications*; Cambridge University Press: New York, NY, USA, 2005.

25. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2013.