

PARAMETER ESTIMATION OF GEOGRAPHICALLY WEIGHTED MULTIVARIATE t REGRESSION MODEL

¹HARMI SUGIARTI, ²PURHADI, ³SUTIKNO, ⁴SANTI WULAN PURNAMI

¹Department of Statistics, Universitas Terbuka, Tangerang Selatan, Indonesia

^{2,3,4}Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: ¹harmi.sugiarti@gmail.com, ²purhadi@statistika.its.ac.id, ³sutikno@statistika.its.ac.id, ⁴santiwulan08@gmail.com

ABSTRACT

The use of ordinary linear regression model in spatial heterogeneity data often does not suitable within the data points, especially the relationship between response variable and explanatory variables. Therefore, the geographically weighted t regression (GWtR) is used to overcome spatial heterogeneity term. The model is an extension of geographically weighted regression (GWR) which the response variable follows multivariate t distribution. The aim of this study is to obtain the estimator of geographically weighted multivariate t regression (GWMtR) model with known degrees of freedom. The maximum likelihood estimation (MLE) method will be applied to maximize a weighted logarithm likelihood function. Based on the EM algorithm, the estimator of geographically weighted multivariate t regression model can be determined.

Keywords: *Maximum Likelihood Estimation (MLE), EM Algorithm, Geographically Weighted Regression, Multivariate t Model*

1. INTRODUCTION

The linear regression model are often used to describe the relationship between the response variable and the independent variable. The estimation of linear regression model by the ordinary least squares (OLS) method can be used when the error has normal distribution. A number of studies of multivariate t linear regression were performed. A study of univariate linear regression and multivariate linear regression are developed to obtain a robust linear regression model when data follow the heavy tail distribution. The maximum likelihood estimation (MLE) method can be applied to find the mean and covariance matrix of estimator robust. In addition, it is suggested to treat the degrees of freedom of t distribution as a known parameter when the sample size is small and to estimate it when the sample size is large [1].

Although, MLE method and Bayesian method have a weakness when the data have null probability under sampling model assumption, the inference of multivariate t linear regression with unknown degrees of freedom was developed. Fernandez and Steel [2] suggested that Bayesian

analysis is based on a set of observations with the accuracy of the initial data. MLE method is not recommended without further study of the properties of the local maxima when it is used to find the estimator of model with errors multivariate t distribution and unknown degrees of freedom.

Liu and Rubin [3] developed a maximum likelihood method using EM algorithm to estimate the parameters of the multivariate t regression model with known and unknown degrees of freedom. The algorithm has analytically quite simple and has stable monotone convergence to a local maximum likelihood estimate.

When the relationship between the response variable and the independent variables is explained by observing geographic factors for each location, the use of spatial regression models is more suitable than the ordinary linear regression model which is constrained by the assumption of spatial data [4]. Spatial data is the measurement data that contains a location information, so the observation at one location depends on the observation at other nearby location. Therefore, the use of the GWR model serves to overcome the spatial heterogeneity.

Some studies on models of classical GWR model has been discussed, where an error follows a normal distribution. Although Lu, *et al.* [5] proposed non-Euclidean distance (non-ED) metrics calibrated with Euclidean distance (ED), road network distance and travel time metrics in GWR model, but the error still follows a normal distribution.

An early review of the GWtR model has been done by Sugiarti, *et al.* [6]. The study indicates that the parameters of the model GWtR can be estimated by maximum likelihood method. The method maximize the weighted logarithm likelihood function of the response variable and get the value of the estimator using the Newton-Raphson iteration. In this paper we propose the maximum likelihood estimator of geographically weighted multivariate t regression model with known degrees of freedom using EM algorithm.

2. MULTIVARIATE t DISTRIBUTION

Base on Nadarajah and Dey [7], vector y_i is said to have the q -variate t distribution with degrees of freedom ν , mean vector μ , and scale matrix Ψ , if its probability density function (pdf) is given by:

$$f(y_i) = \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{(\pi\nu)^{\frac{q}{2}} \Gamma\left(\frac{\nu}{2}\right) |\Psi|^{\frac{1}{2}}} \left[1 + \frac{D_i(\theta)}{\nu}\right]^{-\frac{\nu+q}{2}} \quad (1)$$

where $D_i(\theta) = (y_i - \mu)^T \Psi^{-1} (y_i - \mu)$, $-\infty < y_i < \infty$, $-\infty < \mu < \infty$, $|\Psi| > 0$, and $\nu > 0$.

An illustration of bivariate t distribution with $\nu=2$, $\mu = 0$, and $\Psi = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ is shown in Figure 1.

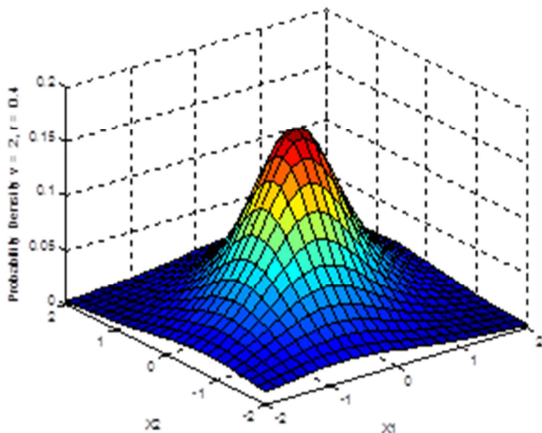


Figure 1: Bivariate t Distribution with $\nu=2$

The bivariate t distribution has heavier tail than Normal distribution. It can be shown when degrees of freedom $\nu=30$ is increased as in Figure 2.

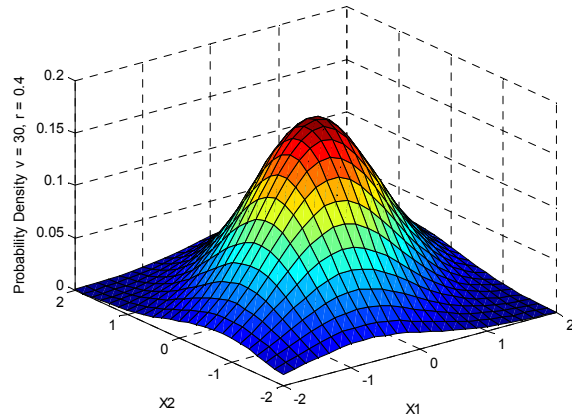


Figure 2: Bivariate t Distribution with $\nu=30$

The expectation and variance of vector y_i can be expressed by:

$$E(y_i) = \mu ; \text{var}(y_i) = \left(\frac{\nu}{\nu-2}\right) \Psi ; \quad \nu > 2$$

3. MULTIVARIATE t REGRESSION

Multivariate t regression model is developed based on multivariate regression model with response variables that follow multivariate t distribution. Suppose q response variables (Y_1, Y_2, \dots, Y_q) associated with p independent variables (X_1, X_2, \dots, X_p) can be written as:

$$y_i = \mathbf{B}^T x_i + \epsilon_i ; i=1,2,\dots,n \quad (2)$$

$$y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iq} \end{pmatrix} ; \mathbf{B}^T = \begin{pmatrix} \beta_{01} & \beta_{11} & \dots & \beta_{p1} \\ \beta_{02} & \beta_{12} & \dots & \beta_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0q} & \beta_{1q} & \dots & \beta_{pq} \end{pmatrix} ;$$

$$x_i = \begin{pmatrix} 1 \\ X_{i1} \\ \vdots \\ X_{ip} \end{pmatrix} ; \epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iq} \end{pmatrix}$$

Vector y_i follows the q -variate t distribution with degrees of freedom ν , mean vector $\mathbf{B}^T x_i$ and scale

matrix Ψ . The probability density function (pdf) of vector y_i is given by:

$$f(y_i) = \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{(\pi\nu)^{\frac{q}{2}} \Gamma\left(\frac{\nu}{2}\right) |\Psi|^{\frac{1}{2}}} \left[1 + \frac{D_i(\theta)}{\nu}\right]^{-\frac{\nu+q}{2}} \quad (3)$$

where $D_i(\theta) = (y_i - B^T x_i)^T \Psi^{-1} (y_i - B^T x_i)$ is Mahala-nobis distance. The expectation and variance of vector y_i can be denoted by:

$$E(y_i) = B^T x_i; \text{var}(y_i) = \left(\frac{\nu}{\nu-2}\right) \Psi; \nu > 2$$

Based on equation (3), the likelihood function can be stated as

$$\begin{aligned} \ell(\theta) &= \prod_{i=1}^n f(y_i) \\ &= \frac{\left[\Gamma\left(\frac{\nu+q}{2}\right)\right]^n}{(\pi\nu)^{\frac{qn}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^n |\Psi|^{\frac{n}{2}}} \prod_{i=1}^n \left[1 + \frac{D_i(\theta)}{\nu}\right]^{-\frac{\nu+q}{2}} \end{aligned} \quad (4)$$

The maximum likelihood estimation of parameter multivariate t regression can be obtained by the maximized of logarithm likelihood function

$$\begin{aligned} \ln \ell(\theta) &= n \ln \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{(\pi\nu)^{\frac{q}{2}} \Gamma\left(\frac{\nu}{2}\right)} - \frac{n}{2} \ln |\Psi| \\ &\quad - \frac{\nu+q}{2} \sum_{i=1}^n \ln \left[1 + \frac{D_i(\theta)}{\nu}\right] \end{aligned} \quad (5)$$

Furthermore, assuming that the likelihood function is differentiable, the estimator can be found by solving the simultaneous equations below.

$$\frac{\partial \ln \ell(\theta)}{\partial B} = 0 \quad \text{and} \quad \frac{\partial \ln \ell(\theta)}{\partial \Psi} = 0 \quad (6)$$

Differentiating with respect to B and Ψ , respectively is given by

$$\begin{aligned} \frac{\partial \ln \ell(\theta)}{\partial B} &= \nu + q \sum_{i=1}^n \frac{x_i (y_i - B^T x_i)^T \Psi^{-1}}{[\nu + D_i(\theta)]} \\ &= \sum_{i=1}^n \frac{(\nu + q) x_i (y_i - B^T x_i)^T \Psi^{-1}}{[\nu + D_i(\theta)]} \\ &= \sum_{i=1}^n \omega_i x_i (y_i - B^T x_i)^T \Psi^{-1} \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln \ell(\theta)}{\partial \Psi} &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Psi^{-1} \frac{\partial \Psi}{\partial \sigma_{ij}} \right) \\ &\quad - \frac{\nu+q}{2} \sum_{i=1}^n \frac{\varepsilon_i^T \left[-\Psi^{-1} \frac{\partial \Psi}{\partial \sigma_{ij}} \Psi^{-1} \right] \varepsilon_i}{1 + \frac{D_i(\theta)}{\nu}} \\ &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Psi^{-1} \frac{\partial \Psi}{\partial \sigma_{ij}} \right) \\ &\quad + \frac{1}{2} \sum_{i=1}^n \omega_i \varepsilon_i^T \left[\Psi^{-1} \frac{\partial \Psi}{\partial \sigma_{ij}} \Psi^{-1} \right] \varepsilon_i \end{aligned}$$

where $\omega_i = \frac{\nu+q}{\nu + D_i(\theta)}$, $\varepsilon_i = y_i - B^T x_i$, and $\frac{\partial \Psi}{\partial \sigma_{ij}}$ is the symmetric $q \times q$ matrix that has ones in row i column j and row j column i, and zeros elsewhere.

Since the equation (6) does not result in closed form solution, the estimator of B and Ψ can be determined by the iteratively process. Liu and Rubin [3] used the EM algorithm to obtain the maximum likelihood estimator that consists of E-step (Estimation step) followed by M-step (Maximization step). The E-step of the EM algorithm aims to obtain the conditional expectation of the complete data sufficient statistics when given the observed values. The M-step involves weighted least squares estimation of B and Ψ . Thus, the EM algorithm iterates successively until the convergence is reached.

At iteration $(r+1)$ with input $B^{(r)}$ and $\Psi^{(r)}$, the E-step of EM algorithm calculate the expectation of ln likelihood complete and the expected sufficient statistics, respectively as follows:

$$\begin{aligned} \omega_i^{(r+1)} &= \frac{\nu+q}{\nu + D_i^{(r)}(\theta)} \quad \text{and} \quad \hat{y}^{(r)} = X \hat{B}^{(r)} \\ S_{\omega XX}^{(r+1)} &= \sum_{i=1}^n \omega_i^{(r+1)} x_i x_i^T \\ S_{\omega XY}^{(r+1)} &= \sum_{i=1}^n \omega_i^{(r+1)} x_i (\hat{y}_i^{(r)})^T \\ S_{\omega YY}^{(r+1)} &= \sum_{i=1}^n \omega_i^{(r+1)} \hat{y}_i^{(r)} (\hat{y}_i^{(r)})^T + \frac{1}{n} \sum_{i=1}^n \Psi_i^{(r)} \end{aligned} \quad (7)$$

The M-step of EM algorithm will calculate

$$\begin{aligned} \hat{B}^{(r+1)} &= \left(S_{\omega XX}^{(r+1)} \right)^{-1} S_{\omega XY}^{(r+1)} \\ \hat{\Psi}_i^{(r+1)} &= \frac{1}{n} \left[S_{\omega YY}^{(r+1)} - \left(S_{\omega XY}^{(r+1)} \right)^T \left(S_{\omega XX}^{(r+1)} \right)^{-1} \left(S_{\omega XY}^{(r+1)} \right) \right] \end{aligned} \quad (8)$$

4. GEOGRAPHICALLY WEIGHTED MULTIVARIATE t REGRESSION

Geographically weighted multivariate t regression (GWMtR) model is developed based on geographically weighted t regression (GWtR) model which proposed by Sugiarti *et al.* [6]. The model describes relationship between q response variables y_i and p independent variables x_i by considering the location factor that expressed as vector coordinate in two dimensional of geographic space. Hence, the estimator obtained in GWMtR model is a local estimator for each point of observation. Let $u_i = (u_{i1}, u_{i2})$ is a coordinate in two dimensional geographic space (latitude and longitude), the expectation of GWMtR model can be defined as follows:

$$E(y_i) = B^T(u_i) x_i ; i = 1, 2, \dots, n \tag{9}$$

where

$$y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iq} \end{pmatrix} ; x_i = \begin{pmatrix} 1 \\ X_{i1} \\ \vdots \\ X_{ip} \end{pmatrix} ; B^T(u_i) = \begin{pmatrix} \beta_{01}(u_i) & \beta_{11}(u_i) & \dots & \beta_{p1}(u_i) \\ \beta_{02}(u_i) & \beta_{12}(u_i) & \dots & \beta_{p2}(u_i) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0q}(u_i) & \beta_{1q}(u_i) & \dots & \beta_{pq}(u_i) \end{pmatrix}$$

and vector $y_i \sim t_q(B^T(u_i)x_i, \Psi(u_i), \nu)$ have the q -variate t distribution and probability density function is

$$f(y_i) = \frac{\Gamma(\frac{\nu+q}{2})}{(\pi\nu)^{\frac{q}{2}} \Gamma(\frac{\nu}{2}) |\Psi(u_i)|^{\frac{1}{2}}} \left[1 + \frac{D_i(\theta(u_i))}{\nu} \right]^{-\frac{\nu+q}{2}} \tag{10}$$

where

$$D_i(\theta(u_i)) = (y_i - B^T(u_i)x_i)^T (\Psi(u_i))^{-1} (y_i - B^T(u_i)x_i)$$

5. PARAMETER ESTIMATION OF GEOGRAPHICALLY WEIGHTED MULTIVARIATE t REGRESSION

Maximum likelihood estimation of parameter GWMtR model can be obtained by the maximized of weighted logarithm likelihood function (10).

$$\begin{aligned} \ln \ell^*(\theta(u_{i^*})) &= \sum_{i=1}^n w_i(u_{i^*}) \ln f(y_i) \\ &= \sum_{i=1}^n w_i(u_{i^*}) \ln \left\{ \frac{\Gamma(\frac{\nu+q}{2})}{(\pi\nu)^{\frac{q}{2}} \Gamma(\frac{\nu}{2}) |\Psi|^{\frac{1}{2}}} \right\} \\ &\quad - \sum_{i=1}^n w_i(u_{i^*}) \left(\frac{\nu+q}{2} \right) \ln \left\{ 1 + \frac{D_i(\theta(u_{i^*}))}{\nu} \right\} \end{aligned} \tag{11}$$

where $w_i(u_{i^*})$ is a weighted function for location u_i and $D_i(\theta(u_{i^*})) = (y_i - B^T(u_{i^*})x_i)^T \Psi^{-1}(u_{i^*})(y_i - B^T(u_{i^*})x_i)$

There are some types of weighting functions that can be used to describe the relationship between the observations at the location i to other locations. The one obvious choice is the Gaussian function [8]

$$w_i(u_{i^*}) = \exp\left(-\frac{1}{2} \left(\frac{d_{i^*}}{h}\right)^2\right) \tag{12}$$

$$d_{i^*} = \sqrt{(u_i - u_{i^*})^2 + (v_i - v_{i^*})^2}$$

where d_{i^*} is Euclidean distance between location i^* and location i with bandwidth h .

Lu, *et al.* [5] propose non-Euclidean distance (non-ED) metrics calibrated with Euclidean distance (ED), road network distance and travel time metrics in GWR model. The results indicate that GWR calibrated with a non-Euclidean metric can not only improve model fit, but also provide additional and useful insights about relationships within data set. However, Lu, *et al.* [9] propose a back-fitting approach to calibrate a GWR model with parameter-specific distance metrics. The results show that the approach can provide both more accurate predictions and parameter estimates, than that found with standard GWR.

In order to select an appropriate bandwidth in GWR, there are a number of criteria that can be used, one of them is *generalized cross validation criterion* (GCV) which is described Matsui, *et al.* [10]

$$GCV = \frac{1}{nq} \frac{\text{trace}\{[y - \hat{y}(h)]^T [y - \hat{y}(h)]\}}{\left(1 - \frac{v_1}{nq}\right)^2}$$

$$v_1 = \text{trace}(X^T W(u_i) X)^{-1} X^T W(u_i) \tag{13}$$

where $\hat{y}(h)$ is the fitted value of $y(h)$ using a bandwidth of h , $W(u_i)$ is weighted matrix that consists weighting function in equation (12).

Estimation of GWMtR's parameter model is obtained through the following equation.

$$\frac{\partial \ln \ell^*(\theta(u_{i*}))}{\partial B(u_{i*})} = 0 \text{ and } \frac{\partial \ln \ell^*(\theta(u_{i*}))}{\partial \Psi(u_{i*})} = 0 \quad (14)$$

Differentiating with respect to $B(u_{i*})$ and $\Psi(u_{i*})$, respectively is given by

$$\begin{aligned} \frac{\partial \ln \ell^*(\theta(u_{i*}))}{\partial B(u_{i*})} &= \sum_{i=1}^n \frac{(v+q) w_i(u_{i*}) \Psi^{-1} \varepsilon_i x_i^T}{v+D_i(\theta(u_{i*}))} \\ \frac{\partial \ln \ell^*(\theta(u_{i*}))}{\partial \Psi(u_{i*})} &= -\frac{1}{2} \sum_{i=1}^n w_i(u_{i*}) \text{tr} \left(\Psi(u_{i*})^{-1} \frac{\partial \Psi(u_{i*})}{\partial \sigma_{ij}} \right) \\ &\quad - \left(\frac{1}{2} \right) \sum_{i=1}^n w_i(u_{i*}) \omega_i(u_{i*}) \varepsilon_i^T \Phi(u_{i*}) \varepsilon_i \end{aligned} \quad (15)$$

where

$$\begin{aligned} \varepsilon_i &= y_i - B^T(u_{i*}) x_i \\ \omega_i(u_{i*}) &= \frac{(v+q)}{v+D_i(\theta(u_{i*}))} \\ \Phi(u_{i*}) &= \Psi(u_{i*})^{-1} \frac{\partial \Psi(u_{i*})}{\partial \sigma_{ij}} \Psi(u_{i*})^{-1} \end{aligned}$$

Since the equation (14) does not result in closed form solution, the estimator of $B(u_{i*})$ and $\Psi(u_{i*})$ are determined by the iteratively process.

A joint iterative process to solve (14) is given by EM algorithm. The EM algorithm is developed to maximize the weighted ln likelihood observed through the weighted ln likelihood complete. Let $y_i|\tau_i$ has multivariate normal distribution and τ_i has Gamma distribution with density function, respectively as follows:

$$\begin{aligned} f(y_i|\tau_i) &= (2\pi)^{-\frac{q}{2}} \left| \frac{\Psi}{\tau_i} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} e^T(u_i) \left(\frac{\Psi}{\tau_i} \right)^{-1} e(u_i) \right\}; \\ f(\tau_i) &= \left[\Gamma \left(\frac{v}{2} \right) \right]^{-1} \left(\frac{v}{2} \right)^{\frac{v}{2}} \tau_i^{\frac{v}{2}-1} \exp \left\{ -\frac{v\tau_i}{2} \right\} \end{aligned}$$

where $e(u_i) = (y_i - B^T(u_i) x_i)$

Joint density function between $y_i|\tau_i$ and τ_i can be expressed by

$$\begin{aligned} f(y_i, \tau_i) &= \left[(2\pi)^{\frac{q}{2}} \Gamma \left(\frac{v}{2} \right) \left| \Psi \right|^{\frac{1}{2}} \right]^{-1} \left(\frac{v}{2} \right)^{\frac{v+q}{2}} \tau_i^{\frac{v+q}{2}-1} \\ &\quad \exp \left\{ -\frac{\tau_i}{2} [v+D_i(\theta(u_i))] \right\} \end{aligned} \quad (16)$$

Therefore, the $\tau_i|y_i$ follows Gamma distribution with parameter: $\left(\frac{v+q}{2}, \frac{v+D_i(\theta(u_i))}{2} \right)$ with its density function can be expressed by

$$\begin{aligned} f(\tau_i|y_i) &= \frac{\tau_i^{\frac{v+q}{2}-1} \exp \left\{ -\frac{\tau_i}{2} [v+D_i(\theta(u_i))] \right\}}{\Gamma \left(\frac{v+q}{2} \right) \left[\frac{v+D_i(\theta(u_i))}{2} \right]^{\frac{v+q}{2}}} \end{aligned} \quad (17)$$

The expectation of $\tau_i|y_i$ can be denoted as:

$$E(\tau_i|y_i) = \frac{v+q}{v+D_i(\theta(u_i))} \quad (18)$$

From (16), the weighted ln likelihood complete is given by

$$\begin{aligned} \ln \ell_c^*(\theta(u_{i*})) &= \sum_{i=1}^n w_i(u_{i*}) \ln f(y_i, \tau_i) \\ &= -\sum_{i=1}^n w_i(u_{i*}) \ln \left[(2\pi)^{\frac{q}{2}} \left| \Psi \right|^{\frac{1}{2}} \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^n w_i(u_{i*}) \tau_i [v+D_i(\theta(u_{i*}))] \\ &\quad + \sum_{i=1}^n w_i(u_{i*}) \ln \left[\left\{ \Gamma \left(\frac{v}{2} \right) \right\}^{-1} \left(\frac{v}{2} \right)^{\frac{v}{2}} \tau_i^{\frac{v+q}{2}-1} \right] \end{aligned} \quad (19)$$

Based on the conditional expectation of $\tau_i|y_i$, the E-step of EM algorithm will calculate the expectation of weighted ln likelihood complete. Because τ_i unknown, the use of $E(\tau_i|y_i)$ is easier than the of $E(\tau_i)$.

Hence, at iteration $(r+1)$ with input $\Theta^{(r)}(u_i) = (B^{(r)}(u_i), \Psi^{(r)}(u_i))$, the E-step of EM algorithm will calculate the expectation of weighted ln likelihood complete and the expected sufficient statistics, respectively as follows:

$$\begin{aligned} \omega_i^{(r+1)}(\mathbf{u}_{i^*}) &= E(\tau_i | \Theta^{(r)}(\mathbf{u}_{i^*})) \\ &= \frac{v+q}{v+D_i^{(r)}(\boldsymbol{\theta}(\mathbf{u}_{i^*}))} \\ S_{\omega Y}^{(r+1)}(\mathbf{u}_{i^*}) &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) E(\tau_i y_i y_i^T | \Theta^{(r)}(\mathbf{u}_{i^*})) \\ &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) \omega_i^{(r+1)}(\mathbf{u}_{i^*}) y_i y_i^T \\ &\quad + \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) \Psi_i^{(r)}(\mathbf{u}_{i^*}) \\ S_{\omega XY}^{(r+1)}(\mathbf{u}_{i^*}) &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) E(\tau_i \mathbf{x}_i y_i^T | \Theta^{(r)}(\mathbf{u}_{i^*})) \\ &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) \omega_i^{(r+1)}(\mathbf{u}_{i^*}) \mathbf{x}_i y_i \\ S_{\omega XX}^{(r+1)}(\mathbf{u}_{i^*}) &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) E(\tau_i \mathbf{x}_i \mathbf{x}_i^T | \Theta^{(r)}(\mathbf{u}_{i^*})) \\ &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) \omega_i^{(r+1)}(\mathbf{u}_{i^*}) \mathbf{x}_i \mathbf{x}_i^T \end{aligned} \quad (20)$$

The M-step of EM algorithm will calculate

$$\begin{aligned} \hat{\mathbf{B}}^{(r+1)}(\mathbf{u}_{i^*}) &= (S_{\omega XY}^{(r+1)}(\mathbf{u}_{i^*}))^{-1} S_{\omega Y}^{(r+1)}(\mathbf{u}_{i^*}) \\ \hat{\Psi}_i^{(r+1)}(\mathbf{u}_{i^*}) &= \frac{1}{n} [S_{\omega Y}^{(r+1)}(\mathbf{u}_{i^*}) - (S_{\omega XY}^{(r+1)}(\mathbf{u}_{i^*}))^T (S_{\omega XX}^{(r+1)}(\mathbf{u}_{i^*}))^{-1} (S_{\omega XY}^{(r+1)}(\mathbf{u}_{i^*}))] \end{aligned} \quad (21)$$

The estimator of $\mathbf{B}(\mathbf{u}_{i^*})$ and $\Psi(\mathbf{u}_{i^*})$ can be found by update the E-step and M-step iteratively until the convergence of algorithm is reached.

Lange, *et al.* [1] used the Fisher Information matrix to determine the estimator of the asymptotic variance-covariance matrix of $\mathbf{B}(\mathbf{u}_{i^*})$. The Fisher Information matrix can be expressed as

$$\mathbf{K}_{BB} = E \left\{ \frac{\partial \ln \ell^*(\boldsymbol{\theta}(\mathbf{u}_{i^*}))}{\partial \mathbf{B}(\mathbf{u}_{i^*})} \right\}^2$$

Let

$$\begin{aligned} g_q &= \frac{(v+q)}{v+\|z_i\|^2}, \quad z_i = [\Psi(\mathbf{u}_{i^*})]^{-\frac{1}{2}} \boldsymbol{\varepsilon}_i \\ \|z_i\|^2 &= \left(\sum_{j=1}^q z_{ij}^2 \right) = \boldsymbol{\varepsilon}_i^T [\Psi(\mathbf{u}_{i^*})]^{-1} \boldsymbol{\varepsilon}_i \end{aligned}$$

The equation (15) can be written as

$$\begin{aligned} \frac{\partial \ln \ell^*(\boldsymbol{\theta}(\mathbf{u}_{i^*}))}{\partial \mathbf{B}(\mathbf{u}_{i^*})} &= \sum_{i=1}^n \frac{(v+q) w_i(\mathbf{u}_{i^*}) [\Psi(\mathbf{u}_{i^*})]^{-1} \boldsymbol{\varepsilon}_i \mathbf{x}_i^T}{v+D_i(\boldsymbol{\theta}(\mathbf{u}_{i^*}))} \\ &= \sum_{i=1}^n w_i(\mathbf{u}_{i^*}) g_q [\Psi(\mathbf{u}_{i^*})]^{-\frac{1}{2}} z_i \mathbf{x}_i^T \end{aligned} \quad (22)$$

Hence, the contribution of the current observation in equation (22) to find estimator of $\mathbf{B}(\mathbf{u}_{i^*})$ is

$$\frac{\partial \ln \ell^*(\boldsymbol{\theta}(\mathbf{u}_{i^*}))}{\partial \mathbf{B}(\mathbf{u}_{i^*})} = w_i(\mathbf{u}_{i^*}) g_q [\Psi(\mathbf{u}_{i^*})]^{-\frac{1}{2}} z_i \mathbf{x}_i^T \quad (23)$$

Based on Lange, *et al.* [1], when q declared the dimension of $\Psi(\mathbf{u}_{i^*})$, then the expectation of some function can be expressed

$$\begin{aligned} E \left(\frac{z_i^T [\Psi(\mathbf{u}_{i^*})]^{-1} z_i}{\|z_i\|} \right) &= \frac{1}{q} [tr([\Psi(\mathbf{u}_{i^*})]^{-1})] \\ E \left\{ \|z_i\|^2 \left(\frac{v}{v+\|z_i\|^2} \right)^2 \right\} &= \frac{v^2 q}{(v+q+2)(v+q)} \\ E \left\{ \|z_i\|^2 \frac{2vq}{(v+\|z_i\|^2)^2} \right\} &= \frac{2vq^2}{(v+q+2)(v+q)} \\ E \left\{ \|z_i\|^2 \left(\frac{q}{v+\|z_i\|^2} \right)^2 \right\} &= \frac{q^3}{(v+q+2)(v+q)} \end{aligned}$$

Furthermore, the Fisher Information matrix can be found through

$$\begin{aligned} E \left\{ \frac{\partial \ln \ell^*(\boldsymbol{\theta}(\mathbf{u}_{i^*}))}{\partial \mathbf{B}(\mathbf{u}_{i^*})} \right\}^2 &= w_i^2(\mathbf{u}_{i^*}) E \left\{ \|z_i\|^2 g_q^2 \mathbf{x}_i \frac{z_i^T}{\|z_i\|} [\Psi(\mathbf{u}_{i^*})]^{-1} \frac{z_i}{\|z_i\|} \mathbf{x}_i^T \right\} \\ &= w_i^2(\mathbf{u}_{i^*}) E \left\{ \|z_i\|^2 g_q^2 \right\} \mathbf{x}_i \frac{1}{q} [tr([\Psi(\mathbf{u}_{i^*})]^{-1})] \mathbf{x}_i^T \\ &= w_i^2(\mathbf{u}_{i^*}) E \left\{ \|z_i\|^2 g_q^2 \right\} \frac{1}{q} [tr([\Psi(\mathbf{u}_{i^*})]^{-1})] \mathbf{x}_i \mathbf{x}_i^T \\ &= w_i^2(\mathbf{u}_{i^*}) \left(\frac{v+q}{v+q+2} \right) [tr(\Psi^{-1})] (\mathbf{x}_i \mathbf{x}_i^T) \end{aligned}$$

Hence, the Fisher Information matrix can be expressed as

$$\begin{aligned} \mathbf{K}_{BB} &= E \left\{ \frac{\partial \ln \ell^*(\boldsymbol{\theta}(\mathbf{u}_{i^*}))}{\partial \mathbf{B}(\mathbf{u}_{i^*})} \right\}^2 \\ &= \sum_{i=1}^n w_i^2(\mathbf{u}_{i^*}) \left(\frac{v+q}{v+q+2} \right) [tr(\Psi^{-1}(\mathbf{u}_{i^*}))] \mathbf{x}_i \mathbf{x}_i^T \end{aligned} \quad (24)$$

Then, the consistent estimator of the asymptotic variance-covariance matrix of $\mathbf{B}(\mathbf{u}_{i^*})$ can be expressed by



$$\begin{aligned} \text{cov}(\hat{\mathbf{B}}(\mathbf{u}_*)) &= \mathbf{K}_{BB}^{-1} \\ &= \left(\frac{\nu+q+2}{\nu+q} \right) \left[\sum_{i=1}^n w_i^2(\mathbf{u}_*) \left[\text{tr}(\hat{\Psi}_j^{-1}(\mathbf{u}_*)) \right] \mathbf{x}_i \mathbf{x}_i^T \right]^{-1} \end{aligned}$$

(25)

6. CONCLUSION

GWMtR model is an extension of geographically weighted regression (GWR) which response variables follow multivariate t distribution. In this model, the response variables will be predicted by independent variables for each location. Therefore, there are many estimator of coefficients regression that depend on the location where the data are observed. Parameter estimation of GWMtR model can be done by maximum likelihood estimation (MLE) method with EM algorithm.

REFERENCES:

- [1] K. Lange, R. Little and J. Taylor, "Robust Statistical Modeling Using The t Distribution," *Journal of the American Statistical Association*, vol. 84, pp. 881-896., 1989.
- [2] C. Fernandez and M. F. Steel, "Multivariate Student-t Regression Models: Pitfalls and Inference," The Netherlands, 1997.
- [3] C. Liu and D. B. Rubin, "ML Estimation of the t Distribution Using EM and its Extensions, ECM and ECME," *Statistica Sinica*, vol. 5, pp. 19-39, 1995.
- [4] L. Anselin, *Spatial Econometrics: Methods and Models*, Springer, 1988.
- [5] B. Lu, M. Charlton, P. Harris and A. Fotheringham, "Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data," *International Journal of Geographical*, pp. 1-22, 2014.
- [6] H. Sugiarti, Purhadi, Sutikno and S. Purnami, "Penaksir Parameter untuk Model Geographically Weighted t Regression (GWtR)," *Konferensi Nasional Matematika XVII*, Surabaya, 2014.
- [7] S. Nadarajah and D. Dey, "Multitude of multivariate t distributions," *Statistics: A Journal of Theoretical and Applied Statistics*, vol. 39, no. 2, pp. 149-181, 2005.
- [8] C. Brunsdon, A. Fotheringham and M. Charlton, "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity," *Geographical Analysis*, no. 28, pp. 281-298, 1998.
- [9] B. Lu, P. Harris, M. Charlton and C. Brunsdon, "Calibrating a Geographically Weighted Regression Model with Parameter-Specific Distance Metrics," *Procedia Environmental Sciences*, vol. 26, p. 110 – 115, 2015.
- [10] H. Matsui, Y. Araki and S. Konishi, "Multivariate Regression Modeling for Functional Data," *Journal of Data Science*, vol. 6, pp. 313-331, 2008.