Parameter Estimation of Multivariate Geographically Weighted Regression Model Using Matrix Laboratory

Sri Harini
Department of Mathematics
UIN Maulana Malik Ibrahim Malang
Jl. Gajayana 50 Malang 65145
Email : sriharini21@yahoo.co.id

Purhadi
Department of Statistics
Institut Teknologi Sepuluh Nopember Surabaya
Jl. Arif Rahman Hakim 1 Surabaya 60111
Email : purhadi@statistika.its.ac.id

Abstract— **Model of Multivariate Geographically Weighted Regression (MGWR) is the extension of multivariate spatial linear model with local observation characters for each observation location. To obtain distribution of the MGWR model, parameter estimation of** $\underset{\sim}{\beta}_h(u_i, v_i)$ **and variance covariance matrix of error** $\left(\Sigma(u_i, v_i)\right)$ **is require to be determined. Besides using mathematical approach, matrix laboratory (MATLAB) algorithm can also be used to obtain parameter estimation of model of MGWR. The MATLAB is a high level programming language base on numerical computing technique to solve problems which involves mathematical operations with array data bases using matrix and vector formulations. Compared to mathematical approach, MATLAB has some advantages which are extensible and no constraint of variable dimension.**

Keywords-MGWR; estimation; variance-covariance; algorithm; matrix laboratory.

I. Introduction

Multivariate Geographically Weighted Regression (MGWR) is a method that takes into account geographical factors as variables affecting the response variable [3]. MGWR model is widely used by researchers in analyzing spatial data in various fields, because the method of MGWR can be used to determine the effect of predictor variables in order to response variables both globally and locally by considering the elements of geography or location as a weighting in estimating the model parameters. the MGWR method is relatively easy in its calculations, it is more effective than others methods and can overcome that model is not stationary in the geographic location (space) [6].

The theory of multivariate spatial data with the Bayesian approach was developed ([1], [4]). Brown at al [1] find a vector of the correlation spatial with random respons. Haas [5] use a Bayesian procedure performed for spatial On the MGWR model, error vector $\tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_q \end{bmatrix}^T$ is a random and the distributed normal multivariate with mean zero and variance covariance matrix $\Sigma(u_i, v_i)$ of the sample size qxq for the-i[th] which $\Sigma(u_i, v_i)$ are :

prediction by multivariate cokriging methods and approaches hierarchy (hierarchical approach) to find models and obtained spatial predictions. Cokriging method is suitable for modeling multivariate spatial that have non-linear model and non-stationary spatial covarians [7]. The parameters estimation procedure use a combination of generalized least squares (GLS). The modeling and prediction of multivariate spatial data using factor analysis. procedures statistical inference are used to select many factors and make the appropriate model structure with an observation location [2] and the research of the parameters estimation [8] and statistical test for multivariate geographically weighted regression model using the method of maksimum likelihood ratio test [9].

II. Materials and Methods

MGWR models is the development of a multivariate spatial linear model with parameter estimator is local to each observation. Assumptions MGWR on the model used is a multivariate normally distributed with mean vector a zero and variance-covariance matrix at each location. In MGWR model the coordinates of the location is one important factor in determining the weighting that will be used to estimate the parameters of the model MGWR.

MGWR model is an extension of univariate spatial linear model with dependent variable (response variable) is more than one. On the MGWR model the relationship between response variables $Y_1, Y_2, ..., Y_q$ and independent variables $X_1, X_2, ..., X_p$ in the-i[th] location is :

$$Y_{hi} = \beta_{h0}(u_i, v_i) + \beta_{h1}(u_i, v_i) X_{1i} + ... + \beta_{hp}(u_i, v_i) X_{pi} + \varepsilon_{hi}$$

$$h = 1, 2, ..., q \text{ dan } i = 1, 2, ..., n \tag{1}$$

$$\Sigma(u_i, v_i) = \begin{bmatrix} \sigma_1^2(u_i, v_i) & \sigma_{12}(u_i, v_i) & \dots & \sigma_{1q}(u_i, v_i) \\ & \sigma_2^2(u_i, v_i) & \dots & \sigma_{2q}(u_i, v_i) \\ & & \ddots & \vdots \\ & & & \sigma_q^2(u_i, v_i) \end{bmatrix} \tag{2}$$

From equation (1), then look for the parameter estimation $\hat{\tilde{\beta}}_h(u_i, v_i)$ using the method of weighted least squares (WLS) is :

$$\hat{\tilde{\beta}}_h(u_i, v_i) = \left(\mathbf{X}^T \mathbf{W}(u_i, v_i)\mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i)\mathbf{Y}$$

(3)

and the estimator of $\hat{\mathbf{B}}(u_i, v_i)$ is :

$$\hat{\mathbf{B}}(u_i, v_i) = \left[\hat{\tilde{\beta}}_1(u_i, v_i) \quad \hat{\tilde{\beta}}_2(u_i, v_i) \quad \ldots \quad \hat{\tilde{\beta}}_q(u_i, v_i)\right]^T$$

(4)

Then estimation parameter variance covariance matrix $\hat{\mathbf{\Sigma}}(u_i, v_i)$ using the method of maximum likelihood estimation (MLE) is :

$$\hat{\mathbf{\Sigma}} = \frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\mathbf{Y}}{\left(\dfrac{\delta_1^2}{\delta_2}\right)}$$

(5)

with :

$$\delta_1 = tr\left((\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\right)$$

$$\delta_2 = tr\left((\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\right)^2$$

(6)

and

$$\mathbf{S} = \begin{bmatrix} X_1^T\left(\mathbf{X}^T\mathbf{W}(u_1,v_1)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(u_1,v_1) \\ X_2^T\left(\mathbf{X}^T\mathbf{W}(u_2,v_2)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(u_2,v_2) \\ \vdots \\ X_n^T\left(\mathbf{X}^T\mathbf{W}(u_n,v_n)\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}(u_n,v_n) \end{bmatrix}$$

(7)

III.   RESULTS AND DISCUSSION

Parameter Estimation Model MGWR Using Matrix Laboratory

In search of a parameter estimation MGWR model can be done in many ways one of them with Matrix Laboratory (MATLAB). MATLAB is a programming environment for algorithm development, data analysis, visualization, and numerical computation. MATLAB is a program for analysis and numerical computation and a language advanced mathematical programming which formed the basis thinking and use the properties of the matrix form.

As the parameter estimation using a mathematical approach, then in estimating the model parameters need to be compiled Matlab MGWR with the following algorithm. Programming is done in order to simplify the process of estimating the parameters of the gobal multivariate models and model MGWR.

To find a parameter estimation model MGWR with MATLAB can be done by the algorithm as follows :

Algorithm 1 : Parameter estimation beta of multivariate regression models

INPUT   : data $\mathbf{X}$ and $\mathbf{Y}$

Step 1   Set $\hat{\mathbf{B}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$

Where $\mathbf{X}^T = \begin{bmatrix} \underset{\sim}{1} & \underset{\sim}{X}_{1i} & \ldots & \underset{\sim}{X}_{pi} \end{bmatrix}$

Where $\mathbf{Y} = \begin{bmatrix} \underset{\sim}{Y}_1 & \underset{\sim}{Y}_2 & \ldots & \underset{\sim}{Y}_q \end{bmatrix}$

Step 2   Set   Output parameter estimation Beta $\left(\hat{\mathbf{B}}\right)$

STOP

This algorithm resulted in an estimate parameters Beta of the multivariate regression model.

Algorithm 2 : Statistics Test of a Multivariate Regression Model

INPUT   : data $\mathbf{X}$ and $\mathbf{Y}$

Step 1   Set $\mathbf{M} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$

Step 2   Set $\mathbf{M}_0 = \underset{\sim}{1}\left(\underset{\sim}{1}^T\underset{\sim}{1}\right)^{-1}\underset{\sim}{1}^T$

Step 3   F test statistics

Set $df_1 = (p-1)$

Set $df_2 = (n-p-1)$

Set $F^* = \dfrac{\left|\dfrac{\mathbf{Y}^T(\mathbf{M}-\mathbf{M}_0)\mathbf{Y}}{df_1}\right|}{\left|\dfrac{\left(\mathbf{Y}^T(\mathbf{I}-\mathbf{M})\mathbf{Y}\right)}{df_2}\right|}$

Step 4   Statistik Reject H$_0$ if $F^* \geq F_{(\alpha, df_1, df_2)}$.

Step 5   Set $\mathbf{V}\left(\hat{\underset{\sim}{\beta}}_h\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\hat{\sigma}_h^2$

Step 6   Set $SE\left(\hat{\beta}_{kh}\right) = \sqrt{g_{kk}}$

g_{kk} diagonal elements to $k+1$ of matrix $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\hat{\sigma}_h^2$

Step 7   t test statistics

Set $t^* = \dfrac{\hat{\beta}_{kh}}{SE\left(\hat{\beta}_{kh}\right)}$

Step 8   Statistik Reject H$_0$ if $t^* \geq t_{(\alpha, n-1)}$.

Step 9   Set $\hat{\mathbf{\Sigma}}_d = \dfrac{\mathbf{Y}^T(\mathbf{I}-\mathbf{M})^T(\mathbf{I}-\mathbf{M})\mathbf{Y}}{n}$

Step 10   Set output F test and t test
STOP

This algorithm resulted a statistics test is *F* test and *t* test of the multivariate regression model.

Algorithm 3 : Euclidean Distance

INPUT : coordinate (u_i, v_i)

Step 1 Set $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$

Step 2 Set Output d_{ij}

STOP

This algorithm resulted an euclidean distance. Results eucliden distance between the observation location is then used to search the entire weighting an observation. in spatial data analysis, weighting is important because the value of weighting can be used to determine the magnitude of the effect of observations with each other at each location.

Algorithm 4 : *Cross Validation* (CV) dan optimum bandwith (*l*)

INPUT : data $\mathbf{Y}$ and $\hat{\mathbf{Y}}$

Step 1 Set $\min CV(l) = \sum_{i=1}^{n} \left(\mathbf{Y}_i - \hat{\mathbf{Y}}_{(i)}(l)\right)^2$

Step 2 Set Output $CV(l), l$

STOP

This algorithm resulted *Cross Validation* (CV) dan optimum bandwith (*l*). the bandwidth is used to determine the best type of weighting to be selected on the model MGWR.

Algoritma 5 : Determination of weighting matrix

INPUT : data $\mathbf{X}$, $\mathbf{Y}$, coordinate (u_i, v_i) and *l*

Step 1 Set

$$\mathbf{W}(u_i, v_i) = \text{diag}\left(w_1(u_i, v_i), \cdots, w_n(u_i, v_i)\right)$$

Where $w_j(u_i, v_i) = \phi(d_{ij}/\sigma l)$

(if Kernel Gaussian)

Where $w_j(u_i, v_i) = \sqrt{\exp\left(-(d_{ij}/l)^2\right)}$

(if exponential)

Where $w_j(u_i, v_i) = \begin{cases} \left(1 - (d_{ij}/l)^2\right)^2, & \text{for } d_{ij} \leq l \\ 0, & \text{for } d_{ij} > l \end{cases}$

(if Bisquare)

Step 2 Set Outputs $\mathbf{W}(u_i, v_i)$

STOP

This result is further used for search estimating the parameters of the model MGWR.

Algorithm 6 : Parameter Estimation Model MGWR

INPUT : data $\mathbf{X}$, $\mathbf{Y}$, bandwith *l* and $\mathbf{W}(u_i, v_i)$

Step 1 Set

$$\hat{\mathbf{B}}(u_i, v_i) = \left(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y}$$

Step 2 Set

$$\mathbf{S}_{(nxn)} = \begin{bmatrix} X_1^T \left(\mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_1, v_1) \\ X_2^T \left(\mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_2, v_2) \\ \vdots \\ X_n^T \left(\mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_n, v_n) \end{bmatrix}$$

Step 3 Set $\delta_1 = tr\left((\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})\right)$

Step 4 Set $\delta_2 = tr\left[\left((\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})\right)^2\right]$

Step 5 Set $\hat{\mathbf{\Sigma}} = \dfrac{\mathbf{Y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{Y}}{\dfrac{\delta_1^2}{\delta_2}}$

Step 6 Set Output $\hat{\mathbf{B}}(u_i, v_i)$ and $\hat{\mathbf{\Sigma}}$

STOP

This algorithm resulted in an estimate of the parameters Beta dan variance-covariance matrix at each location of the model MGWR. More than, the algorithm in detail written in the form of script m-file and stored with a given the name.

The results of the algorithm MATLAB on the obtained MGWR model the parameter estimates $\hat{\mathbf{B}}(u_i, v_i)$ and $\hat{\mathbf{\Sigma}}$ is :

$$\hat{\tilde{\beta}}_h(u_i, v_i) = \left(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y}$$

and

$$\hat{\mathbf{\Sigma}} = \frac{\mathbf{Y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{Y}}{\left(\dfrac{\delta_1^2}{\delta_2}\right)}$$

where these parameters meet the assumptions of best linear unbias estimation.

Besides using mathematical approach, matrix laboratory (MATLAB) algorithm can also be used to obtain parameter estimation of model of MGWR. The MATLAB is a high level programming language base on numerical computing technique to solve problems which involves mathematical operations with array data bases using matrix and vector formulations. Compared to mathematical approach, MATLAB has some advantages which are extensible and no constraint of variable dimension.

estimation of model of MGWR. Because The MATLAB is a high level programming language base on numerical computing technique to solve problems which involves mathematical operations with array data bases using matrix and vector formulations.

REFERENCES

[1] Brown, P.J., Nhu D. Le dan Zidek, J.V, "Multivariate Spatial Interpolation and Exposure to Air Pollutants", *The Canadian Journal of Statistics*. vol. 22 (4), pp. 489-509, 1994.

[2] Christensen, *R* dan Amemiya, Y, "Modelling and Prediction for Multivariate Spatial Factor Analysis", *Journal of Statistical planning and inference.* no.115, pp. 543-564, 2002.

[3] Fotheringham, A.S., Brundson, C., dan Charlton, M, "*Geographically Weighted Regression: the analysis of spatially varying relationships*", John Wiley & Sons Ltd, England, 2002.

[4] Gamerman, D. and Moreira, R.B, " Multivariate Spatial Regression Models". *Journal of Multivariate Analysis*. vol. 9, pp. 262–281, 2004.

[5] Haas, C.T, " Multivariate Spatial Prediction in the Presence of Non-Linier trend and Covariance Non-Stationerity". *Journal Environmetrics*, vol. 7, pp. 145-165, 1996.

[6] Leung, Y, " Statistical Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model", *Journal, The Chinese University of Hong Kong,* Hong Kong, 2000.

[7] Mennis, J. and Jordan, L, "The Distribution of Environmental Equity: Exploring Spatial Nonstationarity in Multivariate Models of Air Toxic Releases", *Annals of the Association of American Geographers,* vol. 95, pp.249–268, 2005.

[8] Harini, S., Purhadi, Mashuri, M dan Sunaryo, S., " Linear Model Parameter Estimator Of Spatial Multivariate Using Restricted Maximum Likelihood Estimation", *Journal Of Mathematics And Technology*, Baku, Azerbaijan. Edisi Oktober, pp. 56-61. ISSN: 2078-0257, 2010

[9] Harini, S., Purhadi, Mashuri, M dan Sunaryo, S., " Statistical Test for Multivariate Geographically Weighted Regression Model Using the Method of Maximum Likelihood Ratio Test. *International Journal of Applied Mathematics & Statistics*, Vol. 29, Issue Number 5, pp. 110-115, 2012.