International Journal of Applied Mathematics & Statistics,
Int. J. Appl. Math. Stat.; Vol. 29; Issue No. 5; Year 2012,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Statistical Test for Multivariate Geographically Weighted Regression Model Using the Method of Maximum Likelihood Ratio Test

Sri Harini[1], **Purhadi**[2], **M. Mashuri**[3] and **S. Sunaryo**[4]

[1] PhD student of Department of Statistics
Institut Teknologi Sepuluh Nopember
[1]Department of Mathematics, FSAINTEK, UIN Maliki Malang (Indonesia);
Email : sriharini21@yahoo.co.id, sri_harini@statistika.its.ac.id

[2,3,4] Department of Statistics, Institut Teknologi Sepuluh Nopember
Jl. Arif Rahman Hakim 1 Surabaya 60111 (Indonesia);
Email : [2]purhadi@statistika.its.ac.id, [3]m_mashuri@statistika.its.ac.id,
[4]sonny_s@statistika.its.ac.id

ABSTRACT

Multivariate Geographically Weighted Regression (MGWR) model is an enhancement of the GWR model with the model parameter estimator that is local to each point or location where the data is collected. In MGWR model the vector error is a random and the distributed normal multivariate with mean zero and variance covariance $\boldsymbol{\Sigma}(u_i, v_i)$. The hypothesis test of MGWR model is done by comparing the suitability of the parameters coefficient simultaneously and partially from MGWR model. Determination of the statistical test is using the method of Maximum Likelihood Ratio Test (MLRT).

Keywords : statistical test, MGWR, variance covariance, MLRT.

Mathematics Subject Classification: 62M30, 62H12, 62H15, 62J10

1. INTRODUCTION

Geographically Weighted Regression (GWR) is a method that takes into account geographical factors as variables affecting the response variable [4]. GWR model is widely used by researchers in analyzing spatial data in various fields, because the method of GWR 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 GWR 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) [7]. Base on the theory [4], Mennis and Jordan [8] compare the global and GWR regression model to predict air pollution in New Jersey, USA. GWR results indicated a significant effect between population density, the number of industrial, transport density and environmental influences on the increase of air pollution in the area. This study followed found estimates parameters of GWR model [9]. In the field of geology Atkinson at al [1] find a relationship between the erosion that occurs along the river by using a model of (GWLR). GWLR model is considered appropriate for being able to suspect the influence of erosion an occur in upstream, midstream and downstream river.

The models above generally is a univariate spatial models, where the observations have only a single dependent variable that depends on the location of observation. Whereas in many cases there are several problems that have more than one dependent variable that depends on the location of the observation (multivariate spatial models). Multivariate spatial linear model is an extention of univariate spatial linear model which the dependent variable (response variable) is more than one.
The theory of multivariate spatial data with the Bayesian approach was developed ([2], [5]). Brown at al [2] find a vector of the correlation spatial with random respons. Haas [6] use a Bayesian procedure performed for spatial 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 covarian [10]. The parameters estimation procedure use a combination of generalized least squares (GLS). Referring to the researchers at the top and the lack of research about MGWR, so in this study will be sought form the test statistic of the MGWR model with MLRT method.

2. THEORETICAL MODEL

The realization of this work supposes the availability of the same known theoretical model. Based on theoretical model we use the methods of multivariate linear model and MGWR model.

2.1. Methods of MGWR studied

MGWR model is an extension of multivariate linear model with the unknown location information. Multivariate spatial linear model dependent variable (response variable) is more than one. On the MGWR model the relationship between response variables $Y_1, Y_2, \ldots, Y_q$ and independent variables $X_1, X_2, \ldots, X_p$ in the-ith location is :

$$Y_{hi} = \beta_{h0}(u_i, v_i) + \beta_{h1}(u_i, v_i)X_{1i} + \beta_{h2}(u_i, v_i)X_{2i} + \ldots + \beta_{hp}(u_i, v_i)X_{pi} + \varepsilon_{hi}$$

$h = 1,2,\ldots,q$ and $i = 1,2,\ldots,n$

On the MGWR model, error vector $\tilde{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_q]^T$ is a random and the distributed normal multivariate with mean zero and variance covariance matrix $\mathbf{\Sigma}(u_i, v_i)$ of the sample size qxq for the-ith

From equation MGWR model, then look for the parameter estimation $\hat{\tilde{\beta}}_h(u_i, v_i)$ and $\hat{\mathbf{\Sigma}}(u_i, v_i)$ using the method of weighted least squares (WLS) estimator thus obtained $\hat{\tilde{\beta}}_h(u_i, v_i)$ is :

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

and the estimator of $\hat{\boldsymbol{B}}(u_i, v_i)$ is $\hat{\mathbf{B}}(u_i, v_i) = \left[\hat{\tilde{\beta}}_1(u_i, v_i) \quad \hat{\tilde{\beta}}_2(u_i, v_i) \quad \ldots \quad \hat{\tilde{\beta}}_q(u_i, v_i)\right]^{\mathrm{T}}$

Parameter estimation $\sigma_h^2(u_i, v_i)$ is obtained by taking into account the location of the model spatial are

$$\hat{\sigma}_h^2(u_i, v_i) = \frac{\left(\tilde{\mathrm{Y}}_{\mathrm{h}} - \mathbf{X}\hat{\tilde{\beta}}(u_i,v_i)\right)^{\mathrm{T}} \mathbf{W}(u_i,v_i)\left(\tilde{\mathrm{Y}}_{\mathrm{h}} - \mathbf{X}\hat{\tilde{\beta}}(u_i,v_i)\right)}{\mathrm{n}}$$

then estimation parameter variance covariance matrix $\mathbf{\Sigma}$ is :

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

with $\delta_1 = tr\left((\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\right)$ and $\delta_2 = tr\left((\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\right)^2$

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

3. RESULTS

The hypothesis test in multivariate spatial linear model is done by comparing the suitability of the parameters coefficient simultaneously from multivariate linear regression model and the MGWR model. This test is similar to test whether the weighting $\mathbf{W}(u_i, v_i)$ that used in the process of estimating the parameters equal to one. Form of hypothesis test are as follows :

H_0: $\beta_{kh}(u_i, v_i) = \beta_{kh}$

(no influence of geographical factors on the model)

H_1: at least one $\beta_{kh}(u_i, v_i) \neq \beta_{kh}$

(there is influence of geographical factors on the model)

The set of parameters under H_0 of function likelihood $L(\omega)$ using a multivariate linear model equation is:

$$\mathbf{Y}_{nxq} = \mathbf{XB}_{(p+1)xq} + \varepsilon_{nxq}$$

The estimate $L(\widehat{\omega})$ as follows :

$$L(\widehat{\omega}) = L(\widehat{\mathbf{B}}, \widehat{\mathbf{\Sigma}}_{\omega}) = (2\pi)^{-nq/2}\left|\widehat{\mathbf{\Sigma}}_{\omega}\right|^{-n/2}\exp\left(-\frac{nq}{2}\right)$$

with the estimate of parameters $Vec(\widehat{\mathbf{B}})$ dan $\widehat{\Sigma}_{\omega}$ as follows :

$$Vec(\widehat{\mathbf{B}}) = \left(\mathbf{I}_q \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)\left(Vec(\mathbf{Y})\right)$$

$$\widehat{\mathbf{\Sigma}}_{\omega} = \frac{(\mathbf{Y}^T(\mathbf{I}-\mathbf{M})\mathbf{Y})}{n}$$

The set of parameters under population with likelihood $L(\Omega)$ function using a MGWR is :

$$L(\widehat{\Omega}) = L(\widehat{\mathbf{B}}(u_i, v_i), \widehat{\mathbf{\Sigma}}_{\Omega}) = (2\pi)^{-nq/2}\left|\widehat{\mathbf{\Sigma}}_{\Omega}\right|^{-n/2}\exp\left(-\frac{nq}{2}\right)$$

With the estimate of parameters $\widehat{\mathbf{B}}(u_i, v_i)$ and $\widehat{\mathbf{\Sigma}}_{\Omega}$ as follows :

$$\widehat{\mathbf{B}}(u_i, v_i) = (\mathbf{X}^T\, \mho(u_i, v_i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{Y}$$

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

Statistical test obtained by making ratio of $L(\widehat{\omega})$ with $L(\widehat{\Omega})$ so-called the likelihood ratio statistic test (Wilk's lamda statistic) is :

$$\Lambda = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \left(\frac{|\widehat{\mathbf{\Sigma}}_{\omega}|}{|\widehat{\mathbf{\Sigma}}_{\Omega}|}\right)^{-\frac{n}{2}}$$

$$(\Lambda)^{\frac{n}{2}} = \frac{|\widehat{\mathbf{\Sigma}}_{\Omega}|}{|\widehat{\mathbf{\Sigma}}_{\omega}|}$$

Ratio likelihood test is used to compare the multivariate linear model and the MGWR model based on the *F* test. The test will reject H_0 if $\Lambda < \Lambda_0 < 1$ with the value $0 < \Lambda_0 < 1$ is expressed as :

$$(\Lambda)^{\frac{n}{2}} = \frac{\left|\frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\mathbf{Y}}{\left(\frac{\delta_1^2}{\delta_2}\right)}\right|}{\left|\frac{(\mathbf{Y}^T(\mathbf{I}-\mathbf{M})\mathbf{Y})}{n}\right|} = U$$

Where is with *p* respon variabel and $m = 1$ follow :

$$\left(\frac{1-U_{p,2,n}^{\frac{1}{2}}}{U_{p,2,n}^{\frac{1}{2}}}\right)\frac{(n-1-p)}{p} \sim F_{(p,(n-1-p))}$$

with a significance level (α) then the decision taken will reject H_0 if the value $F^*_{hitung} > F_{(p,n-p-1)}$ or $P_{value} < \alpha$. It means the MGWR model it not the same as a multivariate linear model.

Further tests carried out simultaneously on the model MGWR with the hypothesis :

H_0: $\beta_{1h}(u_i, v_i) = \beta_{2h}(u_i, v_i) = \ldots = \beta_{kh}(u_i, v_i) = 0$

H_1: at least one $\beta_{kh}(u_i, v_i) \neq 0$

Testing simultan of the MGWR model was performed by searching the set of parameters under H_0 and under population. Determination of the statistic test using the method of Maximum Likelihood Ratio Test (MLRT).

The set of parameters under H_0 of function likelihood $L(\omega) = \left(\tilde{\beta}_0(u_i,v_i), \boldsymbol{\Sigma}_\omega(u_i,v_i)\right)$ using a MGWR model can be expressed as follows :

$$L(\widehat{\omega}) = L\left(\widehat{\tilde{\beta}_0}(u_i,v_i), \boldsymbol{\Sigma}_\omega(u_i,v_i)\right)$$

$$= (2\pi)^{-nq/2}\left|\widehat{\boldsymbol{\Sigma}}_\omega(u_i,v_i)\right|^{-n/2}\exp\left[-\frac{1}{2}\left(\mathbf{Y}-\tilde{1}^{\mathrm{T}}\widehat{\tilde{\beta}_0}(u_i,v_i)\right)^{\mathrm{T}}\mathbf{W}(u_i,v_i)\left(\mathbf{Y}-\tilde{1}^{\mathrm{T}}\widehat{\tilde{\beta}_0}(u_i,v_i)\right)\right]$$

$$= (2\pi)^{-nq/2}\left|\widehat{\boldsymbol{\Sigma}}_\omega(u_i,v_i)\right|^{-n/2}\exp\left(-\frac{nq}{2}\right)$$

To obtain the estimate of parameters $\widehat{\tilde{\beta}_0}(u_i,v_i)$ and $\widehat{\boldsymbol{\Sigma}}_\omega(u_i,v_i)$, then the likelihood function $L(\omega)$ *In* it is later revealed to $\tilde{\beta}_0^T(u_i,v_i)$ and $\widehat{\boldsymbol{\Sigma}}_\omega(u_i,v_i)$. From the results of the decline is obtained the parameter estimator $\widehat{\tilde{\beta}_0}(u_i,v_i)$ is :

$$\widehat{\tilde{\beta}_0^T}(u_i,v_i) = \frac{\left(1^T\mathbf{W}(u_i,v_i)\right)\mathbf{Y}}{\sum_{j=1}^{n}\mathrm{W_j}(u_i,v_i)}$$

and the estimator $\widehat{\boldsymbol{\Sigma}}_\omega$ is :

$$\widehat{\boldsymbol{\Sigma}}_\omega = \frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S}_\omega)^T(\mathbf{I}-\mathbf{S}_\omega)\mathbf{Y}}{\left(\frac{\delta_{1\omega}^2}{\delta_{2\omega}}\right)}$$

$$\mathbf{S}_\omega = \begin{bmatrix}\tilde{1}^T(1^T\mathbf{W}(u_1,v_1)1)^{-1}1^T\mathbf{W}(u_1,\ _1)\\ \tilde{1}^T(1^T\mathbf{W}(u_2,v_2)1)^{-1}1^T\mathbf{W}(u_2,v_2)\\ \vdots\\ \tilde{1}^T(1^T\mathbf{W}(u_n,v_n)1)^{-1}1^T\mathbf{W}(u_n,v_n)\end{bmatrix}$$

So the average $\delta_{1\omega}$ is :

$$\delta_{1\omega} = \frac{1}{\sigma_{hh^*}}(u_j,v_j)E(SSE)$$

and

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

Statistical test obtained by making ratio of $L(\widehat{\omega})$ with $L(\widehat{\Omega}\ )$ so it called the likelihood ratio statistic test (Wilk's lamda statistic). Likelihood ratio test is based on the F test for testing of the parameters simultaneously in MGWR model is :

$$\Lambda = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \left(\frac{\left|\frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S}_\omega)^T(\mathbf{I}-\mathbf{S}_\omega)\mathbf{Y}}{\delta_{1\omega}^2/\delta_{2\omega}}\right|}{\left|\frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\mathbf{Y}}{\delta_1^2/\delta_2}\right|}\right)^{-\frac{n}{2}}$$

Ratio likelihood test the MGWR model based on the *F* test. The test will reject H_0 if $\Lambda < \Lambda_0 < 1$ with the value $0 < \Lambda_0 < 1$ is expressed as :

$$(\Lambda)^{\frac{n}{2}} = \frac{\left|\frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\mathbf{Y}}{\left(\frac{\delta_1^2}{\delta_2}\right)}\right|}{\left|\frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{S}_\omega)^T(\mathbf{I}-\mathbf{S}_\omega)\mathbf{Y}}{\left(\frac{\delta_{1\omega}^2}{\delta_{2\omega}}\right)}\right|} \sim F_{\left(\frac{\delta_{1\omega}^2}{\delta_{2\omega}},\frac{\delta_1^2}{\delta_2}\right)}$$

with a significance level (α) then the decision taken will reject H_0 if the value $F_{hitung}^* > F_{\left(\frac{\ddot{a}_{1\grave{u}}^2}{\ddot{a}_{2\grave{u}}},\frac{\ddot{a}_1^2}{\ddot{a}_2}\right)}$ or $P_{value} < \acute{a}$. Can be meaning In other words, at least one $\beta_{kh}(u_i,v_i) \neq 0$.

Partial test on the MGWR model aims to determine any significant parameters affecting the response variable. Form the hypothesis is as follows :
H_0: $\beta_{kh}(u_i, v_i) = 0$
H_1: $\beta_{kh}(u_i, v_i) \neq 0$
for $k = 1,2,\ldots,p$, $h = 1,2,\ldots,q$ and $i = 1,2,\ldots,n$
Partial test on the MGWR model to To obtain the estimate parameter $\beta_{kh}(u_i, v_i)$ dan variance covariance matrix $(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\hat{ô}_h^2$ with the *standar error* (*SE*) $\hat{\beta}_{kh}(u_i, v_i)$ is :
$SE\left(\hat{\beta}_{kh}(\ _i, v_i)\right) = \sqrt{g_{kk}}$

g_{kk} are the diagonal elemen $k+1$ the matrix $(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\hat{ô}_h^2$.
$SE\left(\hat{\beta}_{kh}(u_i, v_i)\right)$ used to test the level of significance of each location by using the *t* test statistic under H_0 for $\hat{\beta}_{kh}(u_i, v_i) = 0$ then :
$t_{hitung} = \frac{\hat{\beta}_{kh}(u_i,v_i)}{SE\left(\hat{\beta}_{kh}(u_i,v_i)\right)}$

t_{hitung} will follow *t* distribution with degrees of freedom $\left(n - \frac{ä_{1ù}^2}{ä_{2ù}}\right)$ with a significance level (α), then the decision taken will reject H_0 if the value $|t_{hitung}| > t_{\left(n - \frac{ä_{1ù}^2}{ä_{2ù}}\right)}$.

4. DISCUSSION AND CONCLUSION

From the result a research obtained three the hypothesis test of MGWR model with MLRT methods are testing hypotheses simultaneously a comparing the multivariate linear regression model and the MGWR model. This test is similar to test whether the weighting $\mathbf{W}(u_i, v_i)$ used in the process of estimating the parameters equal to one.
Testing hypotheses simultaneously from MGWR model that aims to determine there is or no there is $â_{kh}(u_i, v_i) \neq 0$ or at least one $â_j(u_i, v_i) \neq 0$. The decision taken will reject H_0 if at least one $\beta_{kh}(u_i, v_i) \neq 0$.
Partial testing hypotheses on the MGWR model aims to determine any significant parameters affecting the response variable

5. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to DP2M DIKTI of National Education Ministry, for providing fund for this research as part of the PhD Grant Research Program.

6. REFERENCES

[1] Atkinson, P.M., German, S.E., Sear, D.A. dan M.J. Clark, 2003, Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis*. **35(1)**. The Ohio State University.

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

[3] Christensen, *R* dan Amemiya, Y., 2002, Modelling and Prediction for Multivariate Spatial Factor Analysis. *Journal of Statistical planning and inference.* **115**, 543-564.

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

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

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

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

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

[9] Mennis, J., 2006, Mapping the Results of Geographically Weighted Regression. *The Cartographic Journal.* **43(2)**, 171–179.

[10] Royle, J.A dan Berliner, L.M., 1999, A Hierarchical Approach to Multivariate Spatial Modeling and prediction *Journal of Agricultural, Biological, and Enviromental Statistics*. **4(1),** 29-56.