Best Weighted Selection in Handling Error Heterogeneity Problem on Spatial Regression Model

Sri Sulistijowati Handajani[1], Cornelia Ardiana Savita[2], Hasih Pratiwi[1], and Yuliana Susanti[1]
[1]Statistics of Study Program FMIPA, Universitas Sebelas Maret, Jl.Ir. Sutami 36 A Kentingan, Surakarta, Indonesia
[2]Mathematics of Study Program FMIPA, Universitas Sebelas Maret, Jl.Ir. Sutami 36 A Kentingan, Surakarta, Indonesia

Keywords: Spatial Regression Model, Heterogeneity in Error, Ensemble Technique, R^2, RMSE.

Abstract: Spatial regression model is a regression model that is formed because of the relationship between independent variables with dependent variable with spasial effect. This is due to a strong relationship of observation in a location with other adjacent locations. One of assumptions in spatial regression model is homogeneous of error variance, but we often find the diversity of data in several different locations. This causes the assumption is not met. One such case is the poverty case data in Central Java Province. The objective of this research is to get the best model from this data with the heterogeneity in error. Ensemble technique is done by simulating noises (m) from normal distribution with mean nol and a standard deviation σ of the spasial model error taken and adding noise to the dependent variable. The technique is done by comparing the queen weighted and the cross-correlation normalization weighted in forming the model. Furthermore, with these two weights, the results will be compared using R^2 and RMSE on the poverty case data in province of Central Java. Both of weights are calculated to determine the significant factors that give influence on poverty and to choose the best model. The results of the case study show that the spatial regression model of the SEM ensemble already does not have a variance error that is not homogeneous and the model using cross-normalization weight is better than the spatial regression model of SEM ensemble with Queen contiguity weight.

1 INTRODUCTION

Regression modeling is one form of classical modeling that is used to provide a model of the relationship between independent variables with dependent variable. Fulfillment of the assumptions of this modeling is necessary to ensure that the model is good and can be used for prediction. The problem that often arises is when the model does not meet the necessary assumptions.

The condition of data in the field that the more diverse the pattern resulted in the development of existing classical models. One of the assumptions that often is not met is the autocorrelation of errors. If the dependent variable is drawn from several adjacent areas, this often results in errors of the resulting model being correlated. The phenomena studied often show significant associations or interactions of variables in adjacent areas, as expressed by Anselin (1988). The development of a regression model by adding spatial effects can eliminate any dependencies between errors. This is in accordance with the results of Lesage (1997) which states that if the model obtained without considering the spatial effect then the conclusions obtained will be invalid.

Models with spatial effects have been proposed by Qu (2013), i.e. the Spatial Auto Regressive (SAR) model which shows the dependence of observation on the dependent variable (autoregression) between the locations. While Mac Millen (1992) discusses the model of Spatial Error Model (SEM) which indicates an error correlation between locations.

In spatial models that formed, there arose another problem that is the heterogeneity of errors that resulted in instability in the parameter estimation. The instability of the parameter prediction resulted in less valid results. This is also stated by Dimopoulus, Tsiros, Serelis and Chronopoulou (2004) regarding its application in the neural network model for various nonlinear problems. This instability is a weakness of the model that is formed.

Several approaches to the classical regression model have been done to overcome these assumption deviations. Such as the robust regression method by Chen (2002) with its emphasis on the detection of extreme observations called outliers and its resistance

Handajani, S., Savita, C., Pratiwi, H. and Susanti, Y.
Best Weighted Selection in Handling Error Heterogeneity Problem on Spatial Regression Model.
DOI: 10.5220/0008521002930299
In *Proceedings of the International Conference on Mathematics and Islam (ICMIs 2018)*, pages 293-299
ISBN: 978-989-758-407-7

to the model. Approach of regression model with M estimation has also been proposed by Montgomery and Peck (2006) and Yuliana and Susanti (2008). The model solutions with S estimation have been introduced by Rousseeuw and Yohai (1984), and MM estimation has been discussed also by Yohai (1987). The problems that arise is how to approach the regression model that error in addition to autocorrelation also there is heterogeneity in the error obtained from the classical regression model.

To overcome the errors that contain autocorrelation is by spatial regression modeling. In this research will be discussed the problem of heterogeneity of error on spatial regression model with solution which will be applied in this research is ensemble method. According to Mevik, Segtnan and Naes (2005) and Canuto, Oliveira, Junior, Santos and Abreu (2005), ensemble techniques can be used to reduce the diversity contained in predictive models and can improve prediction accuracy. This method is as one solution by combining k spatial regression model formed from the addition of noise. The approach is through non-hybrid ensemble approach and hybrid ensemble approach. The principle of non-hybrid ensemble method is to combine estimation results from simulation of one model to a final estimate. While hybrid ensemble method involves several suitable models and combine the predicted simulation results generated by each model into one final prediction. In this study we studied non-hybrid ensemble approach by comparing queen weights with spatial weights of cross-correlation normalization and using R^2 and RMSE indicators to select the best approach.

2 SPATIAL PANEL REGRESSION MODEL

In the data taken based on time and location, the analyzes were performed using panel data analysis. Due to the effect of spatial effect in panel data analysis so that the appropriate model used is spatial panel regression model. One of the spatial panel regression models is the panel spatial error model (SEM) (Lesage, 2009) yang sebelumnya dikembangkan dari model SEM yang diusulkan Anselin (2003). The SEM panel regression model is

$$Y_{it} = \alpha + X_{it}\beta + u_{it} \;\; ; \;\; u_{it} = \lambda W u_{it} + \varepsilon_{it} \qquad (1)$$

where Y_{it} being the dependent variable of the data in the i^{th} observation unit and the t-time, X_{it} is the independent variable of the data in the i-th observation unit and the t-time, W is the standardized spatial weighted row matrix, α is the intercept, β is the parameter of the independent variable, u_{it} spatial error in the i-th region of time t, and ε_{it} is the model error on the i-th observation and the th time.
The model (1) is further simplified to be

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + (\mathbf{I} - \boldsymbol{\lambda}\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

where $\mathbf{Z} = [\mathbf{I} : \mathbf{X}]$ dan $\boldsymbol{\gamma} = [\boldsymbol{\alpha} : \boldsymbol{\beta}]^{T}$.

By estimating the parameters using the maximum likelihood method, it is necessary to first form the likelihood function. Using the Jacobian transformation is obtained

$$\left|\frac{\partial\varepsilon}{\partial\mathbf{Y}}\right| = \left|\frac{\partial(\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma})(\mathbf{I}-\lambda\mathbf{W})}{\partial\mathbf{Y}}\right| = \left|\mathbf{I} - \boldsymbol{\lambda}\mathbf{W}\right|,$$

and likelihood function :

$$L(\gamma,\lambda,\sigma^2) = \frac{|\mathbf{I}-\boldsymbol{\lambda}\mathbf{W}|}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(-\frac{\left((\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})\right)^{T}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})}{2\sigma^2}\right) \qquad (2)$$

To facilitate the estimation of the parameters, the two sections of (2) are logged and the following results are obtained

$$lnL(\gamma,\lambda,\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|\boldsymbol{I} - \boldsymbol{\lambda W}|\left(-\frac{\left((\boldsymbol{Y}-\boldsymbol{Z\gamma})(\boldsymbol{I}-\boldsymbol{\lambda W})\right)^{T}(\boldsymbol{Y}-\boldsymbol{Z\gamma})(\boldsymbol{I}-\boldsymbol{\lambda W})}{2\sigma^2}\right) \qquad (3)$$

By partially deriving (3) against σ^2, γ and λ and making it equal to zero, we get the estimator for σ^2, γ and λ as follows

$$\frac{\partial lnL(\gamma,\lambda,\sigma^2)}{\partial\sigma^2} =$$

$$-\frac{n}{2\sigma^2} + \frac{\left((\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})\right)^{T}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})}{2{\sigma^2}^2} = 0$$

$$\widehat{\sigma^2} = \frac{(\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma}-\boldsymbol{\rho}\mathbf{W}\mathbf{Y})^{T}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\gamma}-\boldsymbol{\rho}\mathbf{W}\mathbf{Y})}{n}$$

$$\frac{\partial lnL(\gamma,\lambda,\sigma^2)}{\partial\gamma} = \frac{(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})\begin{pmatrix}(\mathbf{Z}(\boldsymbol{\lambda}\mathbf{WY})^{\boldsymbol{T}}) - \mathbf{Z}(\mathbf{Y})^{\boldsymbol{T}} + \\ \gamma(\mathbf{Z}^T\boldsymbol{Z} - \mathbf{Z}(\mathbf{Z}\boldsymbol{\lambda}\mathbf{W})^{\boldsymbol{T}})^T\end{pmatrix}}{\sigma^2} = 0$$

$$\hat{\gamma} = \frac{(\mathbf{Y}^T\mathbf{Z} - \mathbf{Z}(\boldsymbol{\lambda}\mathbf{WY})^{\boldsymbol{T}})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})}{(\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}(\mathbf{Z}\boldsymbol{\lambda}\mathbf{W})^{\boldsymbol{T}})(\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})}$$

$$\frac{\partial lnL(\gamma,\lambda,\sigma^2)}{\partial\lambda} = \sum_{i=1}^{n}\frac{-\omega_i}{1-\lambda\omega_i} - \frac{1}{\sigma^2}\left((\mathbf{I}-\boldsymbol{\lambda}\mathbf{W})(2(\mathbf{Z}\boldsymbol{\gamma}\mathbf{WY})^T - \mathbf{Z}\boldsymbol{\gamma}(\mathbf{Z}\boldsymbol{\gamma}\mathbf{W})^T)\right) = 0$$

$$\hat{\lambda} = \left(\frac{\sigma^2}{\mathbf{I}-\boldsymbol{\lambda}\mathbf{W}}\left(\sum_{i=1}^{n}\frac{-\omega_i}{1-\lambda\omega_i}\right)\right) + \mathbf{Z}\boldsymbol{\gamma}(\mathbf{Z}\boldsymbol{\gamma}\mathbf{W})^T - 2(\mathbf{Z}\boldsymbol{\gamma}\mathbf{WY})^T$$

3 EXPERIMENTAL DETAILS

1. Derive estimation the parameters for the SEM spatial panel regression model with maximum likelihood.
2. Looking for an example case that meets one of the spatial regression models with an area approach and contains heterogeneity in the error. An example of cases taken in this study is the case of poverty in 35 districts / cities of Central Java Province. Data taken from Central Bureau of Statistics (2008-2015) with the variables taken are percentage of poor people (Y), percentage of poor people aged more than 14 years and not finished primary school (X_1), percentage of population not illiterate age 15-55 years (X_2), percentage of poor people aged more than 14 years and unemployed (X_3), percentage of poor people aged more than 14 years and agriculture work (X_4), percentage of women using contraceptives (X_5), percentage of women with poor status aged 15-49 years whose first delivery was assisted by health personnel (X_6), the percentage of households whose houses have a per capita floor area of less than 8 m^2 (X_7), percentage of households using their own toilet / joint (X_8), the percentage of households ever buy raskin rice (X_9), and percentage of population growth rate (X_{10}).
3. Determine independent variables that significantly influence the percentage of poor people with stepwise method to form a simple linear regression model.
4. Determine the weighted matrix by using the spatial matrix of Queen contiguity and the cross-correlation normalization matrix.
5. Detect spatial effect by using Moran Index test.
6. Test the LM to determine the effect of spatial dependence.
7. Establish a corresponding spatial regression model and test its assumptions.
8. Add noise which is generated k times from the normal distribution with zero mean and model error variance of σ and giving a zero value to the negative value data, to dependent variable to generate k new data.
9. In the k new data is done spatial regression modeling as follows.
   a. Test Lagrange for spatial dependence
   b. Test the Breusch Pagan to test for spatial diversity
   c. Estimate model parameters and test their significance
   d. Measuring the goodness of the spatial regression model with R^2.
10. Establish an ensemble model which is a composite of *k* spatial regression models by calculating the average coefficients of the model.
11. Compare the spatial regression model ensemble for both weights by looking at the greatest R^2 and minimum RMSE.

4 RESULT AND DISCUSSION

In the case of poverty in Central Java province, the linear regression model begins with the selection of variables that have a significant effect on the percentage of poor people in Central Java Province in 2015. Variable selection is done by stepwise method and the obtained linear regression model is

$$\hat{Y} = 12.92927 - 0.07321X_8 + 0.12883X_9 - 13.67537X_{10}$$

To see the feasibility of the model, tested normality with statistics Kolmogorov Smirnov and obtained the test statistic value is 0.0728571. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, for multicollinearity test with VIF, the three independent variables show the VIF value less than 10 so it is concluded that the above model does not occur multicolinearity. While homogeneity test of variance is used statistic of Breusch Pagan (BP) and obtained value of BP = 7,8908>χ^2(0,05;3) = 7,815 so it is concluded that model there is heteroscedasticity. In addition, spatial correlation testing is also done between the errors by using Moran Index. The test results showed positive spatial autocorrelation with IM value = 0.24 which means there is similarity error value from adjacent locations and error value tend to group.

4.1 SEM Model

There is an indication of the spatial effect of the Moran Index so that an analysis to test for the effect is necessary. This test uses Lagrange Multiplier (LM) lag and LM error statistics. LM lag statistic value of 2.1087 and LM error value of 4,0997 compared with the value of chi square table of 3.851. The conclusion obtained shows that there is no spatial effect of lag but there is a spatial effect of error on the model so that the appropriate model is the SEM model.

Determination of SEM model parameter estimation with weighted queen obtained result with each parameter is significant is as follows.

$$\hat{Y} = 11{,}92635 - 0{,}06676X_8 + 0{,}13043X_9 - 10{,}94249X_{10} + u \ ; u = 0{,}42701Wu.$$
$$R^2 = 0{,}7045$$

To overcome the non-homogeneous variance, noise is added to the dependent variable. The noise is generated from the normal distribution having a mean of zero and the standard deviation is the standard deviation error of 2.53. Noise is simulated 100 times. The next step is modeled into the SEM model for each noise simulation result and a spatial regression model of the ensemble is obtained. The spatial model of the ensemble model is the result of the average parameter estimation p of the regression model, where p is the number of spatial regression models. The spatial regression ensemble model is expressed as

$$\hat{Y} = \frac{1}{Q}\sum_{p=1}^{Q}\hat{Y}_p$$

From the data analysis obtained by regression model of spatial error ensemble from mean of estimation result of hundredth parameter of model is

$$\hat{Y} = 11.78799 - 0.06558X_8 + 0.13113X_9 - 10.86190X_{10} + u; u = 0.38606Wu$$

with the R^2 value of 0.7271, which means that 72% of the total poor is affected by the percentage of households using their own latrines / joints, the percentage of households who have bought raskin rice, and the percentage of population growth rate. To see the feasibility of the model, tested normality with statistics Kolmogorov Smirnov and obtained the test statistic value is 0,147. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, homogeneity test of variance is used statistic of Breusch Pagan (BP) and obtained value of BP = 6:99240<χ^2(0,05;3) = 7,815 so it is concluded that the model there isn't heteroscedasticity again.

While the panel data analysis is the first done by analysis with regular regression model. The formation of the linear regression model is begun by variable selection which is significant to the model with stepwise method. In the case of poverty in Central Java Province in 2008 until 2015, the obtained linear regression model is

$$\hat{Y}_{it} = 19.0626 + 0.1378X_{4it} - 0.0963X_{6it} - 0.0532X_{8it} + 0.0515X_{9it}$$

RMSE = 3.151822

To see the feasibility of the model, tested normality with statistic Kolmogorov Smirnov and obtained the test statistic value is 0.2071 > D(0.05;280) = 0.0807. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, for multicollinearity test with VIF, from the four independent variables of the model above shows the VIF value of each less than 10 which it means that the model does not occur multicollinearity. While homogeneity test of variance used statistic Breusch Pagan (BP) and obtained value of BP = 17.0267 $>\chi^2(0{,}05; 3) = 9.48773$ so concluded that model there is heteroscedasticity. In addition, spatial correlation testing is also done between the errors by using Moran Index. The test results show that there is negative spatial autocorrelation with IM = -0.0807, which means different errors in adjacent locations and the errors tend to spread. There is an indication of the spatial effect of the Moran Index so that an analysis to test for the effect is necessary. This test uses Lagrange Multiplier (LM) lag and LM error statistics. The LM lag statistic value is 2.14705197 and the LM error value is 4.91880103 and each is compared with the chi square table value of 3.841. The conclusion obtained shows that there is no spatial effect of lag but there is a spatial effect on the model error so that the appropriate model is the SEM model.

4.2 SEM Panel Model with Queen Weight

Estimation of SEM model parameters with queen weights obtained with each significant parameter are as follows

$$\hat{Y}_{it} = 21.9543446 + 0.1253795X_{4it} - 0.1102882X_{6it} - 0.0615198X_{8it} + 0.0575619X_{9it} + u_{it},$$
$$u_{it} = -0.075165\boldsymbol{W}u_{jt}$$

with the value of R^2 is 0.725807 which means 72.58% percentage of the poor is affected by the percentage of households using their own latrines / joint, the percentage of households who have bought raskin rice, the percentage of poor people aged more than 14 years working in the sector agriculture, the percentage of poor women aged 15-49 years whose first delivery was helped by health personnel and RMSE = 3.130008.

To overcome the heteroscedasticity of errors variance is done adding noise to the dependent variable. The noise is generated from the normal distribution having a mean of zero and the standard deviation is a standard deviation error of 0.23. Noise is simulated 100 times. The next step is modeled into the SEM model for each noise and the average model of the hundredth model is searched. The ensemble model of the error spatial regression model is

$$\hat{Y}_{it} = 21.78935 + 0.131977X_{4it} - 0.11098X_{6it} - 0.05896X_{8it} + 0.056906X_{9it} + u_{it},$$
$$u_{it} = -0.07991\boldsymbol{W}u_{jt}$$

with a R^2value of 0.750338, which means 75.03% of the percentage of the poor is affected by the percentage of households using their own latrines / joints, the percentage of households who have bought raskin rice, the percentage of poor people aged over 14 who work in agriculture, the percentage of poor women aged 15-49 years whose first delivery was helped by health personnel and RMSE = 3.117259. To see the feasibility of the model, tested normality with statistics Kolmogorov Smirnov and obtained the test statistic value is 0.02438 >D(0.05;280) = 0.0807. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, homogeneity test of variance is used statistic of Breusch Pagan (BP) and obtained value of BP = 6.217639<χ^2(0,05;3) = 9,48773 so it is concluded that the model there isn't heteroscedasticity again.

4.3 SEM Panel Model with Cross-Correlation Normalization Weight

While the panel data analysis for the use of weighting cross-correlation normalization first done by regular regression model analysis. The formation of the linear regression model is begun by variable selection which is significant to the model with stepwise method. In the case of poverty in Central Java Province in 2008 until 2015, the obtained linear regression model is

$$\hat{Y}_{it} = 20.5622708 + 0.1277858X_{4it} - 0.1074335X_{6it} - 0.0634190X_{8it} + 0.0573606X_{9it}$$

RMSE = 3.151822.

To see the feasibility of the model, tested normality with statistic Kolmogorov Smirnov and obtained the test statistic value is 0.2493. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, for multicollinearity test with VIF, from the four independent variables of the model above shows the VIF value of each less than 10 which it means that the model does not occur multicollinearity. While homogeneity test of variance used statistic Breusch Pagan (BP) and obtained value of BP = 12.9205 $>\chi^2(0,05;3) = 9.48773$ so concluded that model there is heteroscedasticity. In addition, spatial correlation testing is also done between the errors by using Moran Index. The test results show that there is negative spatial autocorrelation with IM = -0.03062855, which means different errors in adjacent locations and the errors tend to spread. There is an indication of the spatial effect of the Moran Index so that an analysis to test for the effect is necessary. This test uses Lagrange Multiplier (LM) lag and LM error statistics. The LM lag statistic value is 2.635515 and the LM error value is 11.721478 and each is compared with the chi square table value of 3.841. The conclusion obtained shows that there is no spatial effect of lag but there is a spatial effect on the model error so that the appropriate model is the SEM model.

Estimation of SEM model parameters with cross correlation normalization weights obtained with each significant parameter are as follows

$$\hat{Y}_{it} = 31.8914595 + 0.1222338X_{4it} - 0.1099141X_{6it} - 0.0610961X_{8it} + 0.0557544X_{9it} + u_{it},$$
$$u_{it} = -0.68753\boldsymbol{W}u_{jt}$$

with the value of R^2 is 0.7433 which means 74.3% percentage of the poor is affected by the percentage of households using their own latrines / joint, the percentage of households who have bought raskin rice, the percentage of poor people aged more than 14 years working in the sector agriculture, the percentage of poor women aged 15-49 years whose first delivery was helped by health personnel and RMSE = 3.107759.

To overcome the heteroscedasticity of errors variance is done adding noise to the dependent variable. The noise is generated from the normal distribution having a mean of zero and the standard deviation is a standard deviation error of 0.21. Noise is simulated 100 times. The next step is modeled into the SEM model for each noise and the average model of the hundredth model is obtained. The ensemble model of the error spatial regression model is

$$\hat{Y}_{it} = 35.82514 + 0.12272X_{4it} - 0.10963X_{6it} - 0.05749X_{8it} + 0.05508X_{9it} + u_{it},$$
$$u_{it} = -0.94464\boldsymbol{W}u_{jt}$$

with a R^2value of 0.7775, which means 78% of the percentage of the poor is affected by the percentage of households using their own latrines / joints, the percentage of households who have bought raskin rice, the percentage of poor people aged over 14 who work in agriculture, the percentage of poor women aged 15-49 years whose first delivery was helped by health personnel and RMSE = 3.082789. To see the feasibility of the model, tested normality with statistics Kolmogorov Smirnov and obtained the test statistic value is 0.048286. With a significance level of 5 percent then taken the conclusion of the assumption of normality error is met. Furthermore, homogeneity test of variance is used statistic of Breusch Pagan (BP) and obtained value of BP = 5.73146<χ^2(0,05;3) = 9,48773 so it is concluded that the model there isn't heteroscedasticity again.

From the two-weighting used in the panel data, the results of the analysis concluded that the model with the normalization of cross-correlation weighting was better than the model with queen weighting. From the best model, the percentage of poor people in Central Java Province in 2015 is predicted. The prediction results obtained are then grouped into six priority zones are as follows,

Table 1: Zones of percentage the poor population.

Zone	Percentage of the poor population
1	>35 %
2	30 % - 34.99 %
3	25 % - 29.99 %
4	20 % - 24,99 %
5	15 % - 19.99 %
6	<15%

Of the six priority zones in Table 1. the percentage of poverty for districts and cities in Central Java province is only at three zones, i.e. the second, third and fourth priority zones. The results of the is shown in Table 2.

Table 2: The districts and cities of poverty in Central Java Province

Zone	The districts and cities of the poor population
1	-
2	Wonosobo regency, Banjarnegara regency, Wonogiri regency
3	Cilacap regency, Banyumas regency, Purbalingga regency, Kebumen regency, Purworejo regency, Magelang regency, Boyolali regency, Klaten regency, Sukoharjo regency, Karanganyar regency, Sragen regency, Grobogan regency, Blora regency, Rembang regency, Pati regency, Kudus regency, Jepara regency, Demak regency, Semarang regency, Temanggung regency, Batang regency, Pekalongan regency, Pemalang regency, Tegal regency, Brebes regency
4	Surakarta city, Salatiga city, Pekalongan city, Magelang city, Semarang city, Tegal city, Kendak regency
5	-
6	-

5 CONCLUSIONS

The poverty case shows that the spatial regression model of the SEM ensemble already does not have a variance error that is not homogeneous and the model using cross-normalization weight is better than the spatial regression model of SEM ensemble with Queen contiguity weight.

ACKNOWLEDGMENT

The authors would like to thank to Research Group of Applied and Inference Statistics for the discussions. We also appreciate Universitas Sebelas

Maret for providing financial support through Grant of Fundamental Research 2018.

REFERENCES

Anselin, L.,1988. *Spatial Econometrics: Methods and Models.* Academic Publishers, Dordrecht.

Anselin, L., 2003. Spatial Multipliers, and Spatial Econometrics , *International Regional Science Review,* University of Illinois.

Chen,C.,2002. Robust regression and Outlier Detection with the ROBUSTREG Procedure,Paper 265-27, *Statistics and Data Analysis*, SUGI 27,North Carolina:SAS Institute Inc.

Canuto,A.M.P., Oliveira,L.,Junior,J.C.X.,Santos,A.,Abreu, M., 2005. Perfomance and diversity evaluation in hybrid and non-hybrid structures of ensembles, *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems*.

Central Bureau of Statistics.2015. *Data dan informasi kemiskinan kabupaten/kota*, BPS, Indonesia.

Dimopoulos,L.F., Tsiros,L.X., Serelis, K. & Chronopoulou,A.,2004. Combining Neural Network Models to Predict Spatial Patterns of Airborne Pollutant Accumulation in Soils around an Industrial Point Emission Source, *Journal of the Air & Waste Management Association*, 54(12), 1506–1515.

LeSage,J.P. 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20, no 1 dan 2, pp 113-129.

LeSage, J. P.,2009. Intoduction to Spatial Econometrics, CRC Press Taylor and Francis Group, Florida.

Mevik,H.B., Segtnan,V.H., T. Naes,T., 2005. Ensemble methods and partial least squares regression, *Journal of Chemometrics18* no. 11, 498-507.

Millen,D.P, Mac. 1992. Probit with spatial autocorrelation. *Journal of Regional Science* 32, No.3 , pp. 335-348.

Montgomery, D.C., and Peck,E.A.,2006, *Introduction to Linear Regression Analysis,* John Wiley & Sons Inc., New York.

Qu, Xi. 2013. *Three Essays on the Spatial Autoregressive Model in Spatial Econometrics*, Dissertation at The Ohio State University.

Rousseeuw,P.J. and Yohai, V.J.,1984. *Robust Regression by Mean of S-Estimators,* , Springer-Verlag, Berlin.

Yohai,V.J.,1987. High Breakdown Point and High Efficiency Robust Estimates for Regression,*The Annals of Statistics*, 15, 642-656.

Yuliana dan Susanti,Y.,2008. Estimasi-M dan Sifat-sifatnya pada Regresi Linear Robust", *Jurnal Math-Info,* Vol. l No. 10.