

Saefuddin, Asep; Wigena, Aji Hamim; Nuryartono, Nunung; Kusumaningrum, Dian

**Conference Paper**

## Development And Application Of Bayesian Spatial Analysis On Poverty Data In East Java, Indonesia

53rd Congress of the European Regional Science Association: "Regional Integration: Europe, the Mediterranean and the World Economy", 27-31 August 2013, Palermo, Italy

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Saefuddin, Asep; Wigena, Aji Hamim; Nuryartono, Nunung; Kusumaningrum, Dian (2013) : Development And Application Of Bayesian Spatial Analysis On Poverty Data In East Java, Indonesia, 53rd Congress of the European Regional Science Association: "Regional Integration: Europe, the Mediterranean and the World Economy", 27-31 August 2013, Palermo, Italy, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124098>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## **DEVELOPMENT AND APPLICATION OF BAYESIAN SPATIAL ANALYSIS ON POVERTY DATA IN EAST JAVA, INDONESIA**

Asep Saefuddin<sup>1)</sup>, Aji Hamim Wigena<sup>1)</sup>, Nunung Nuryartono<sup>2)</sup>, Dian Kusumaningrum<sup>1)</sup>

<sup>1)</sup>Departement of Statistics <sup>2)</sup> Departement of Economics

Bogor Agricultural University

### **ABSTRACT**

This paper is extracted from the research educational process of M.S level at Bogor Agricultural University, Indonesia. We first created a research group named Geoinformatics on Poverty and then invited some students to join and do research related to spatial statistics and poverty. Poverty is one of the crucial problems in Indonesia which is not easy to be solved. Following a survey held by the Statistics Indonesia in March 2011 showed that there were 30.02 million people or 12.49% of total Indonesian are considered poor. From the statistical point of view, poverty is an interesting topic because there are many problems of spatial data, i.e. spatial autocorrelation, error variance heterogeneity, spatial interaction, and other statistical issues. The main objective of the geoinformatics group is to compile and develop spatial statistics applied on poverty allowing spatial effects. Those are GSM (General Spatial Model), SAR (Simultan Autoregressive), CAR (Conditional Autoregressive), SEM (Spatial Error Model), Bayesian SAR, and SAE (Small Area Estimation). The models are implemented on poverty data in East Java, Indonesia. The general results show that the poverty in East Java are related to low education, poor access to clean water, government pro poor programs (health insurance for the poor, subsidized rice for the poor, and poor letter), and improper housing. Statistically, Bayesian SAR slightly perform better than the other spatial regressions. In terms of small area estimation, spatial empirical best linear unbiased prediction (SEBLUP) and empirical Bayesian methods are more efficient than empirical BLUP (EBLUP).

**Keywords:** *poverty, Bayesian spatial analysis, East Java.*

### **Introduction and Motivation**

Poverty has long been an unresolved problem in Indonesia. Statistics Indonesia in March 2011 predicted that the number of poor people in Indonesia reached 30.02 million or 12.49 percent of the total population. Research in poverty is very important to find substantial factors that cause poverty. From the statistical point of view, poverty is an interesting topic due to autocorrelation among data, spatial autocorrelation, error variance heterogeneity, spatial interaction, and other statistical issues. We then at the Department of Statistics IPB created a geoinformatics group to analyze statistical methodology applied on poverty. The first step, the group learn some spatial regressions like SAR, CAR, SEM, Bayesian SAR as well as Small Area Estimation (SAE). The second step is to apply them on poverty data of East Java.

Poverty is the state of being without the necessities of daily living, often associated with need, hardship and lack of resources across a wide range of circumstances. A person living in the condition of poverty is said to be poor or impoverished. According to Statistics Indonesia household is categorized as poor if they have income below the poverty line. The poverty line is a measure of the amount of money a government or a society believes is necessary for a person to live at a minimum level of subsistence or standard of living, which is around 1 USD per day.

### Spatial Regression Models

Spatial Regression is related to the specifications, estimation, and diagnostic checking of regression models that accommodate spatial effects. Spatial effects can be grouped into spatial dependence and spatial diversity (spatial heterogeneity). In a simple linear regression model, spatial dependence is included as an additional regressors or structural error known as spatial lag model and spatial error model respectively.

The general equation of a spatial regression model is:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \end{aligned} \quad (1)$$

Where  $\mathbf{y}$  is a dependent variable with a  $n \times 1$  dimension,  $\mathbf{X}$  is a independent matrix with a size of  $(n \times (p+1))$ ,  $\boldsymbol{\beta}$  is the parameter regression coefficient estimate with a size of  $p \times 1$ ,  $\rho$  is the spatial lag autoregression coefficient,  $\lambda$  is the spatial error autoregression coefficient where  $|\lambda| < 1$ ,  $\mathbf{u}$  is the error vector that is assumed to have an autocorrelation and has a dimension of  $n \times 1$ ,  $\mathbf{W}$  is the spatial weight matrix with a  $n \times n$  dimension, and  $n$  is the number of observations.

If  $\rho \neq 0$  and  $\lambda = 0$ , then equation (1) will be equal to:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (2)$$

The model is called SAR (simultaneously autoregressive regression).

Modeling spatial interaction that arises in spatial data is generally done by including the spatial dependence in the covariance structure either explicitly or implicitly through autoregressive models. One of the autoregressive models commonly used is Conditional Autoregressive (CAR) (See for example Besag, 1974). Another form of autoregressive spatial models is simultaneously autoregressive regression (SAR).

The autoregressive model can be written in the form of

$$\boldsymbol{\varepsilon}_i^* = \rho \sum_{j=1}^n w_{ij} (Y_{(S_j)} - \mathbf{X}\boldsymbol{\beta}) \quad ; i = 1, \dots, n$$

SAR model observes random variables in the various locations simultaneously. The independent variables in SAR are spatially correlated.

The SEM (Spatial Error Model) occurs when equation (1) has  $\rho=0$  and  $\lambda \neq 0$  producing

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (3)$$

Spatial error model is a linear regression model where the errors are spatially correlated.

The spatial models mentioned above used ML methods and do not use any initial informations (priors) in the parameter estimation. Bayesian estimation method can be used to improve the parameter estimation. Bayes method is a parameter estimation method that utilizes initial information and sample information (Gill 2002). A simulation study by Oliviera et al. (2008) found that Bayesian approach produces a more precise parameter estimates.

The SAR model in equation 1 can also be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \quad (4)$$

with  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ , therefore  $\mathbf{y} \sim N[\mathbf{X}\boldsymbol{\beta}, ((\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{M} (\mathbf{I}_n - \mathbf{B}')^{-1})]$ ,  $\mathbf{M} = \sigma^2 \mathbf{I}_n$ , and  $\sigma^2 > 0$  is unknown, while  $(\mathbf{I}_n - \mathbf{B})$  is a nonsingular matrix,  $\mathbf{B} = \phi \mathbf{W}$  with  $\phi$  is an unknown spatial parameter, and  $\mathbf{W} = (w_{ij})_{n \times n}$  is a non-negative spatial weight matrix.

If  $y_i$  is a random variable of normal distribution, the likelihood function of  $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \phi)$  based on observed data is as the following:

$$L(\boldsymbol{\eta}|\mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} |\boldsymbol{\Sigma}_\phi^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_\phi^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (5)$$

With  $\boldsymbol{\Sigma}_\phi^{-1} = (\mathbf{I}_n - \phi \mathbf{W}') (\mathbf{I}_n - \phi \mathbf{W}) = (\mathbf{I}_n - \phi \mathbf{W})^2$

The prior information distribution of  $\boldsymbol{\eta}$  equals:

$$\pi(\boldsymbol{\eta}) \propto \frac{\pi(\phi)}{(\sigma^2)^a}, \quad \boldsymbol{\eta} \in \Omega, \quad (6)$$

with  $\Omega = R^p \times (0, \infty)$  is a parameter space that has a unique characteristic that depends on its weight,  $a \in R$  whose values are determined (Gill 2002) and  $\pi(\phi)$  is the initial marginal information from  $\phi$  in the interval  $(\lambda_n^{-1}, \lambda_1^{-1})$ ;  $\lambda_i$ , for  $i = 1, \dots, n$  is the eigen value of the weight matrix  $\mathbf{W}$ .

According to Bayes theorem the distribution of  $\boldsymbol{\eta}$  needs a requirement, that is  $0 <$

$\int_{\Omega} L(\boldsymbol{\eta}|\mathbf{y}) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} < \infty$ . To simplify the posterior function in equations (5) and (6), it can be denoted as:

$$\int_{R^p \times (0, \infty)} L(\boldsymbol{\eta}|\mathbf{y}) \pi(\boldsymbol{\eta}) d\boldsymbol{\beta} d\sigma^2 = L^I(\phi|\mathbf{y}) \pi(\phi) \quad (7)$$

with  $L^I(\phi|\mathbf{y}) \propto |\boldsymbol{\Sigma}_\phi^{-1}|^{\frac{1}{2}} |\mathbf{X}' \boldsymbol{\Sigma}_\phi^{-1} \mathbf{X}|^{-\frac{1}{2}} (S_\phi^2)^{-\left(\frac{n-p}{2} + a - 1\right)}$

$$(S_\phi^2) = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\phi)' \boldsymbol{\Sigma}_\phi^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\phi)$$

$$\widehat{\boldsymbol{\beta}}_\phi = (\mathbf{X}' \boldsymbol{\Sigma}_\phi^{-1} \mathbf{X})^{-1} (\mathbf{X}' \boldsymbol{\Sigma}_\phi^{-1} \mathbf{y})$$

$L^I(\phi|y)$  is known as the integrated likelihood function of  $\phi$  (Oliviera *et al.* (2008)). The posterior distribution  $\eta$  in equation (5) will be accurate if the integrated likelihood function in equation (7) and the prior information distribution  $\pi(\phi)$  is included in the interval  $(\lambda_n^{-1}, \lambda_1^{-1})$ .

Preliminary non-informative information used is the Jeffreys initial information and uniform or naive initial information. Uniform (naive) initial informations will give equal weights for all the spatial parameters values. While the Jereys initial information does not give a limit value between  $\lambda_1^{-1}$  dan  $\lambda_n^{-1}$ , but it is important to establish a weight for a spatial parameter that is near to the lowest or highest eigen value.

In the model parameter estimation stage, when using hierarchical Bayes method, the calculation is done usually through a multidimensional integral, an alternative that can be used to calculate the posterior through numerical integration and one of the methods used are MCMC algorithms.

### Small Area Estimation Model

Classical statistical methods commonly are based on large samples. In practice, particularly if the objective is to estimate parameters in a specific sub-population in which the sample size is usually too small to produce accurate direct estimators. The sub-population is called small area and the method is small area estimation (SAE) (Rao, 2003). Nowadays, there is an increasing demand for small area estimators due to limited resources for detailed surveys.

In standard SAE models, it is assumed that small areas are independent of each other. However, in practice this assumption does not hold because of spatial dependence as mentioned by Tobler (1970). Ignorance of spatial dependence between small areas leads to biased or inconsistent estimators. One technique to capture spatial dependence is by accomodating the dependent variable or error term as an autoregressive model. The extension of basic SAE models to include spatial dependence in the error term is given by

$$y = X^T \beta + (I - \rho W)^{-1} v + e \quad (4)$$

where  $\beta$  is the vector of regression coefficients,  $W$  is the spatial weights matrix representing the spatial structure,  $\rho$  is spatial correlation parameter,  $v$  is vector of random area effects and  $e$  is vector of sampling error.

Based on the model (4), Petrucci and Salvati (2004a, 2004b), Salvati (2004), Petrucci and Salvati (2006), Pratesi and Salvati (2008), Molina et al (2009) introduced SEBLUP (Spatial Empirical Best Linear Unbiased Prediction) which is a combination of EBLUP and a spatial autoregressive error (SAR) model. They found that estimated standard

errors from SEBLUP were smaller than those of EBLUP. EBLUP and SEBLUP was also used in this research along with Spatial Bayesian SAE to see which is the most appropriate method to use in the case of poverty data in East Java.

EBLUP estimator has been discussed by Ghosh and Rao (1994), Rao (1999), Datta and Lahiri (2000), and Rao (2003). EBLUP estimator by replacing the value  $\sigma_v^2$  with its estimators  $\hat{\sigma}_v^2$  is as follows:

$$\tilde{\theta}_i = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) x_i^T \tilde{\beta}$$

EBLUP estimator obtained by the method of ML and REML estimator is not biased  $v_i$  and  $e_i$  is normally distributed with mean 0 .

MSE of EBLUP (Rao 2003) is :

$$MSE[\tilde{\theta}_i] \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2)$$

with  $g_{3i}(\sigma_v^2) = \sigma_i^4 (\sigma_i^2 + \sigma_v^2)^{-3} \bar{V}(\hat{\sigma}_v^2)$ .  $\bar{V}(\hat{\sigma}_v^2)$  with asymptote variance of  $\hat{\sigma}_v^2$  with formula ;

$$\bar{V}(\hat{\sigma}_v^2) = [\mathfrak{I}(\sigma_v^2)]^{-1} = 2 \left[ \sum_{i=1}^m z(\sigma_v^4) / (\sigma_i^2 + \sigma_v^2)^2 \right]^{-1}$$

In this study, the model of Pratesi and Salvati (2008) is used to predict the estimators of the average expenditure per capita of 35 villages in Jember (Susenas 2008). The expression is

$$\hat{\theta} = \mathbf{X}\beta + \mathbf{Z}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} + \mathbf{e}$$

The independent variable X used is the percentage of poor families having health insurance (askeskin family recipient) in the past year. The data is from Podes (Village Potential Data 2008).

Then the SEBLUP estimation is calculated through the following formula :

$$\begin{aligned} \tilde{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho}) &= x_i \hat{\beta} + \mathbf{b}_i^T \{ \hat{\sigma}_u^2 [(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}^T)]^{-1} \} \mathbf{Z}^T \\ &\quad \times \{ \text{diag}(\hat{\sigma}_i^2) + \mathbf{Z} \hat{\sigma}_u^2 [(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}^T)]^{-1} \mathbf{Z}^T \}^{-1} (\hat{\theta} - \mathbf{X}\hat{\beta}) \end{aligned}$$

with :  $\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \hat{\theta}$

MSE $[\tilde{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho})]$  for the spatial model EBLUP with a normally distributed random effect, is :

$$MSE[\tilde{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho})] = MSE[\tilde{\theta}_i^S(\sigma_u^2, \rho)] + E[\tilde{\theta}_i^S(\hat{\sigma}_u^2, \hat{\rho}) - \tilde{\theta}_i^S(\sigma_u^2, \rho)]^2$$

Some types of spatial weighting matrix used in this research are :

1. Weighted spatial correlation matrix

$$Q_{il}^{(1)} = r_{il}$$

With value  $0 \leq |r_{il}| \leq 1$ .

2. Nearest neighbors spatial weighting matrix

$$Q_{il}^{(2)} = \begin{cases} 1, & \text{for area } -i \text{ which strongly correlated with area } -l \\ 0, & \text{otherwise} \end{cases}$$

with  $r_{il}$  is the correlation between area- $i$  and area- $l$ . the size of strong correlation between area- $i$  and area- $l$  if  $|r_{il}| \geq 0.5$ , while  $d_{il}$  Euclidean distance between area- $i$  and area- $l$ .

### Data Analysis

The data used in this research from the National Survey (Susenas 2008) and Village Potential (Podes 2008). Below is the map of cities and regencies in

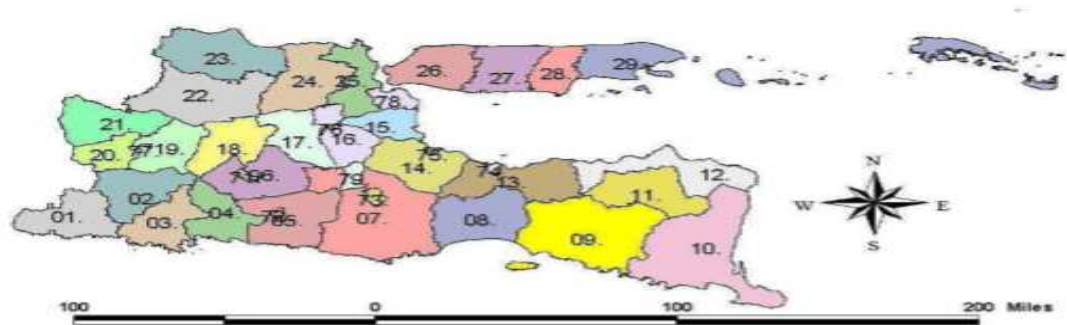


Figure 1. Administrative Map of Cities and Regencies in East Java, Indonesia

Note: code of 38 cities and regencies in East Java:

01. Pacitan	14. Pasuruan	27. Sampang
02. Ponorogo	15. Sidoarjo	28. Pamekasan
03. Trenggalek	16. Mojokerto	29. Sumenep
04. Tulungagung	17. Jombang	71. Kota Kediri
05. Blitar	18. Nganjuk	72. Kota Blitar
06. Kediri	19. Madiun	73. Kota Malang
07. Malang	20. Magetan	74. Kota Probolinggo
08. Lumajang	21. Ngawi	75. Kota Pasuruan
09. Jember	22. Bojonegoro	76. Kota Mojokerto
10. Banyuwangi	23. Tuban	77. Kota Madiun
11. Bondowoso	24. Lamongan	78. Kota Surabaya
12. Situbondo	25. Gresik	79. Kota Batu
13. Probolinggo	26. Bangkalan	

The stages of the data analysis are:

1. Preparation of response variable and independent variable
2. Taking into account the influence of spatial dependence
3. Establish a spatial weighting matrix based on each criteria required by each model.
4. Calculating spatial correlation.
5. Parameter estimation and hypothesis testing for each model.
6. Estimating the properties of parameter estimators.
7. Model Validation
8. Conclusions.

## Spatial Modeling Applications on Poverty Data in East Java

### a. GSM, SAR, and SEM

Identifying spatial effects is aimed to investigate the spatial heterogeneity and spatial dependencies. The outcome variable in these models is the percentage of poverty. Lagrange multiplier is used to detect the spatial dependencies, more specifically the spatial dependencies in the lag, error, or both (lag and error) listed in Table 1.

Table 1. Spatial Dependency Tests

Spatial Dependence Test	Value	Chi-square	p-value	conclusion
Lagrange Multiplier (SGM)	14.7199	5.99	0.0001	Reject Ho
Lagrange Multiplier (lag)	13.9483	3.84	0.0001	Reject Ho
Lagrange Multiplier (error)	19.3217	3.84	0.0000	Reject Ho

Following the Table 1 we conclude that all models are useful. The independent variables used to develop these models are listed in Table 2.

Table 2. Independent Variables used in GSM, SAR, and SEM

Sector	Variable	Information	Symbol
Education	Illiteracy	the percentage of the population who can not read at the age of 15-55 years.	X1
	Low education rate	the percentage of people who are not graduated from elementary school	X2
Housing Facility	Low access to clean water	the percentage of households that are not drinking water from mineral water, tap water, water pump, protected well or spring	X3
	Healthy house	the percentage of people who occupy a healthy home according to the Ministry of Health criteria	X4
Employment	Agriculture workers	the percentage of the population employed in agriculture	X5
	Non-Agriculture workers	the percentage of the population working in non-agricultural sectors	X6
	Formal Sector Workers	the percentage of the population employed in the formal sector	X7
	Non-Formal Sector Workers	the percentage of the population working in the informal sector	X8

The General Spatial Model (GSM), which is a combined error and lag model, is:

$y = 4.29 + 0.19Wy + 1.26 X_2 + 0.51 x_3 - 0.33x_4 + u$  with  $u = 0.54 Wu + \varepsilon$ . From the model it can be concluded that poverty in an area can be characterized by the number of people who have low education (elementary school), poor access to clean water, and inappropriate house. Compared to the regression models based on OLS (Ordinary Least Square), global



spatial model provides more precise conclusions. The parameter estimation based on SAR and SEM can be seen in Table 2 below.

Table 2. Parameter Estimations and Goodness of Fit of GSM, SAR, and SEM

Variable	GSM	SAR	SEM
Intercept	4.29**	3.3**	4.24**
X2	1.26**	1.18**	1.26**
X3	0.51**	0.57**	0.51**
X4	-0.33**	-0.61**	-0.33**
Rho	0.19**	0.27**	-
Lamda	0.54**	-	0.86**
R <sup>2</sup>	0.9330	0.9331	0.9391
AIC	129.6139	128.9857	127.5963
RMSE	5.9214	3.9849	2.3688

\*\* ) significant at  $\alpha = 5\%$  and \*) significant at  $\alpha = 10\%$

Based on Table 2 we conclude that the Spatial Error Model (SEM) is the best model and the factors that affect the increase in the percentage of poverty are the same as in GSM.

#### b. SAR and CAR

The variables used in this section are the education sector (X1 and X2), housing facility (X3 and X4), and per capita Product Domestic Regional Bruto (PDRB) (X5). In addition, we include some variables related to the government pro poor programs, i.e. health insurance (X6), which is the percentage of people who receive health care benefits that are characterized by having health insurance cards for the poor, Raskin (X7), which is the percentage of people who are allowed to buy subsidized rice at a low price, and Letter of Poor (X8), which is the percentage of people who received the letter indicating a membership of the 20 percent most bottom poor households.

Table 3. Parameter Estimations and Goodness of Fit of CAR and SAR

Variable	CAR	SAR
Intercept	3.071**	3.269**
X2	0.836**	0.849**
X3	0.146**	0.133*
X6	0.119**	0.114**
X7	0.353**	0.358**
X8	0.589**	0.357*
Rho	0.363**	0.121**
$\widehat{\sigma^2}$	0.83	0.82
AIC	112,69	111,95
$\rho$	0.157	0.121

\*\* ) significant at  $\alpha = 5\%$  and \*) significant at  $\alpha = 10\%$

We conclude that the two approaches are reciprocal.

From the SAR model it can be concluded that poverty in an area can be characterized by five significant variables, which are the number of people who just graduated maximum

in Elementary School; poor access to clean water, as well as all of the government pro poor programs (health insurance for the poor, subsidize rice for the poor, and letter for the poor).

### Poverty Analysis using SEBLUP

The model of expenditure per capita with EBLUP is  $(\hat{\theta} = 154,078.7 - 594.339x)$  and SEBLUP is  $(\hat{\theta} = 129,553.1 - 338.573x)$ . The parameter estimate of spatial correlation coefficient (*spatial autoregressive coefficient*),  $\rho$ , is 0.99 indicating a high positive correlation of spending between villages.

The goodness of the prediction of direct estimation, estimation with EBLUP method and estimation with SEBLUP method can be seen from the value of RRMSE. It can be seen in Figure 2 that the RRMSE of EBLUP estimator is slightly smaller than the one of direct estimation. While the RRMSE of SEBLUP estimate is much smaller than the RRMSE of DE and EBLUP. Hence, SEBLUP approach improves the parameter estimation significantly

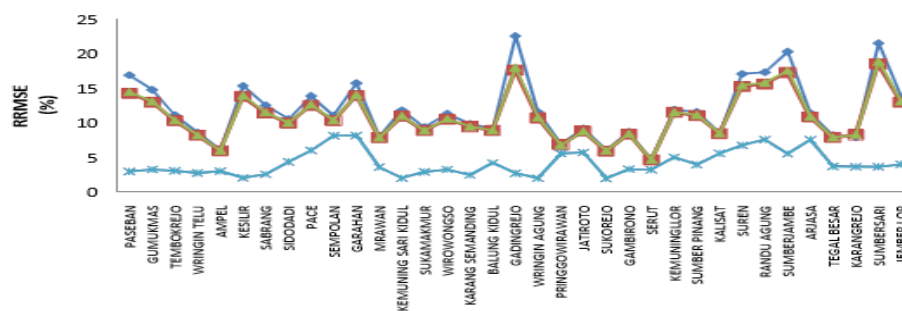


Figure 2. Comparison of RMSE of Direct Estimator ( ◆ ), EBLUP ML ( ■ ), EBLUP REML ( ▲ ), SEBLUP ML ( ✕ ) dan SEBLUP REML ( \* )

### Conclusion

In order to accommodate some regional complexities away from the rigid statistical assumption, spatial models is urgently implemented. In particular, in each model can be summed up several conclusions:

1. Spatial models is appropriate to analyze poverty data
2. SAR model is found better than the OLS to estimate poverty related factors in East Java.
3. Bayesian SAR model is slightly better than SAR indicated by the lower MSE producing a more precise parameter estimates.
4. In terms of poverty, the models find that some important factors to be improved, i.e. low education, poor access to clean water, poor access to healthy housing, and all of the government pro poor programs (health insurance for the poor, subsidize rice for the poor, and letter for the poor).

5. SEBLUP method produces much better estimates than those from the direct estimate and EBLUP of the average expenditure per capita in Jember (a district in East Java).

### **Acknowledgment**

We thank to the Directorate General of Higher Education that have provided a multi year research grant (2010-2012) for enhancing this research.

### **References**

- Besag J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems. *J of the Royal Statist Soc. Series B (Methodological)* 36 (2):192-236.
- BPS [Badan Pusat Statistik]. 2008. *Data dan Informasi Kemiskinan 2008*. Jakarta: Badan Pusat Statistik.
- Cressie, 1993. *Statistics for Spatial Data*. New York: John Wiley and Sons.
- Datta GS, Lahiri P. 2000. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Stat Sin* 10:613-627.
- Ghosh M, Rao JNK. 1994. Small area estimation: an appraisal (with discussion). *Statistical Science*, 9(1):55-93.
- LeSage JP. 2001. A Family of Geographically Weighted Regression Models. *Journal of Geographic Information Science Vol. 5, No. 2, Department of Economics University of Toledo*.
- Pratesi M, Salvati N. 2008. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications, Stat. Meth. & Appl.* 17:113-141.
- Petrucci, A. and Salvati, N. 2004a. Small Area Estimation Using Spatial Information, The Rathbun Lake Watershed Case Study. Working Paper no 2004/02, "G. Parenti" Department of Statistics, University of Florence.
- Petrucci, A. and Salvati, N. 2004b. Small Area Estimation Considering Spatially Correlated Errors : The Unit Level Random Effects Model. Working Paper no 2004/10, "G. Parenti" Department of Statistics, University of Florence.
- Petrucci, A. and Salvati, N. 2006. Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment. *Journal of Agricultural, Biological, and Environmental Statistics*, 11: 169-182.
- Rao JNK. 1999. Some recent advances in model-based small area estimation. *Survey Methodology*, 25:175-186.
- Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley and Sons.
- Tobler W., 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region". *Economic Geography*, 46(2): 234-240