Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 20 No.2, Oktober 20115, p: 39-44

available online at:
journal.ipb.ac.id/index.php/statistika

MODELING OF DENGUE HEMORRHAGIC FEVER IN BOGOR USING BAYESIAN SUR-SAR

Hilman Dwi Anggana[1], Asep Saefuddin[1], Bagus Sartono[1]

[1]Department of Statistics, Bogor Agricultural University, Indonesia
E-mail : hilmandwianggana@gmail.com[1]

ABSTRACT

The purposes of this research are (1) To develop Seemingly Unrelated Regression (SUR) system constructed by correlated Spatial Autoregressive Model (SAR) with Bayesian approach for dynamic analysis of spatial and non-spatial contributions of Dengue Hemorrhagic Fever (DHF) case in Bogor, (2) To evaluate efficiency issues on parameters estimation with SUR system. Markov Chain Monte Carlo (MCMC) sampling scheme was used to estimate all of model parameters with the number of iteration whose burn-in period was discovered. The results indicated that : there was the similar pattern of DHF spread in Bogor during 2009 – 2011, the nearby areas had a significant role to the incidence of DHF in an area in the city of Bogor, and the non-spatial contributions of DHF cases in Bogor during 2009 -2011 included in this model were dynamic. Gain efficiency of parameters estimation on modeling of DHF in Bogor with SAR for each year during 2009-2011 can be obtained if we construct all of SAR with SUR system model.

INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is one of the public health problems in Indonesia whose incidence and desease's spread are both increasing and expanding. One of the cities in Indonesia where DHF has been a serious problem is Bogor. Bogor city with high level of rainfall and located at an average altitude minimum 190 meters and maximum 350 meters above sea level is a potential area for the development of Aedes aegypti mosquitoes (Rahmat, 2014).

Preventive research to reduce the incidence of DHF should be done, such as modeling DHF cases with with statistical modeling in order to see the contributions of spatial and non-spatial contributions to the DHF (Anggana, 2012). Several studies using statistical modeling for DHF cases in Bogor have been done. Nadra (2006) applied the Auto - Gaussian regression for patterns analysis of spatial relationships from 68 villages in Bogor in 2005. Kartika (2007) used Morans index, Geary 's ratio , and Chi - squared statistics to see the spatial relationships of the number of DHF patients in the city of Bogor in 2005. Fatmawati (2011) used spatial autoregressive models (SAR) to assess the contributions of the factors that influenced the incidence of dengue in the city of Bogor in 2009. They concluded that there were significant spatial effect (spatial autocorrelation) in dengue cases in the city of Bogor.

The pattern relationship analysis of DHF cases in Bogor using spatial statistics is still limited to examining the patterns relationship of spatial and non-spatial aspects from one cross sectional data. By the fact that DHF cases continues to occur each year, more information will be obtained if modeling DHF cases using panel data. Rahmat (2014) used spatial panel data regression and spatial sequential regression to estimate the patterns relationship of DHF incidence in Bogor during 2009-2011 which the result of spatial sequential regression was better than spatial panel data regression.

Modeling that done by Rahmat (2014) still was not able to explain the dynamics of both spatial and non-spatial contributions. The dynamics of spatial and non-spatial contributions can be explained by constructing spatial regression models individually for each year. However, there is inefficient parameter estimation when we make individual regression models because the relationship among different models may occur due to the phenomenon that the spread of dengue fever remains high throughout the year as in the case of dengue fever in Bogor during 2009-2011.

One way to accommodate the relationship some individual regression models is to use Seemingly Unrelated Regression model system as proposed by Zellner (1962). Anselin (1988) proposed the concept of spatial SUR for models built by T periods , where each period has n units of the same observations . Kakamu *et al*. (2007) used spatial SUR models with a Bayesian approach for the analysis of the economic grouping during the 1991-2000 period. Bayesian approach used in SUR system because this approach is more practical in reality (Griffith , 2001) and the spatial models of this approach has several advantages (Lesage , 2005) .

There are two main objectives to be achieved in this study : 1) To develop a specific spatial regression model (SAR) as mentioned in Anggana (2012) and Fatmawati (2011) into SUR system constructed by Anselin (1988) with the Bayesian approach as proposed Kakamu *et al*. (2007) for the dynamic analysis of spatial and non-spatial contributions of DHF incidence in Bogor during 2009-2011; 2) To evaluate the efficiency of the corresponding model parameter estimation with SUR system .

MATERIALS AND METHODS

Data

Data for the analysis was from a previous research conducted by Rahmat (2014). This secondary data was obtained from Bogor Health Profile (Profil Kesehatan Bogor) and Bogor Regional Development Agency (Bappeda Bogor) during 2009-2011. Unit observations in this study are 68 villages in Bogor. As additional data, Bogor map is used to create a spatial weight matrix shown in Figure 1.

The variables used for this study were data on the number of DHF patients (Y) as dependent variable and the independent variables were population density (X_1), the population mobility (X_2), the average age of patients (X_3), and the number of clinics (X_4). The unknown coefficients ρ of the number of DHF patients (Y) weighted by standardized spatial weight matrix (**W**) represented spatial contributions while the unknown vector $\boldsymbol{\beta}$ of X_1, X_2, X_3, and X_4 described non-spatial contributions.

Figure 1 The map of Bogor city which consists 68 villages

Methodology

In general, there were many stages conducted in the modeling of DHF data in Bogor using Bayesian SUR-SAR model as follows :

1. Determined the spatial weight matrix **W** from Bogor map using queen contiguity rule. This matrix was standardized so that the total of elements of each row is one.
2. Estimated three individual Bayesian SAR models using Markov Chain Monte Carlo (MCMC) simulation as described in the previous study from Anggana (2012).
3. Checked the residuals from three individual models and then tested the significance of correlation using simple t test.
4. Specified the initial values of prior and posterior parameters from Bayesian SUR-SAR models and then performed MCMC simulation described in Kakamu *et al*. (2007).
5. Using Metropolis algorithm, generated the candidate value of spatial lag autocorrelation coefficient of t-th individual model (ρ_t^*) for t = 2009, 2010, and 2011. The process of drawing the candidate value use a standard normal distribution where the candidate value (ρ_t^*) was justified by current value (ρ_t^c) through equation(1)
$$\rho_t^* = \rho_t^c + a.N(0,1).$$
6. The candidate value (ρ_t^*) was evaluated using acceptance probability throuh

(2)

equation

$$\alpha\left(\rho_t^c, \rho_t^*\right) = \min\left(\frac{p\left(\rho_t^* \mid \rho_{-t}^c\right)}{p\left(\rho_t^c \mid \rho_{-t}^c\right)}, 1\right)$$

where $p\left(\rho_t^* \mid \rho_{-t}^c\right)$ was conditional posterior distribution of the candidate value from *t*-the individual model.

7. Using Gibbs sampler algorithm, generated $\boldsymbol{\beta}$ from conditional posterior distribution (3)
$p(\boldsymbol{\beta} \mid \rho_1, .., \rho_T, \boldsymbol{\Omega}^{-1}, \mathbf{y}, \mathbf{X}, \mathbf{W}) \sim N(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})$.
8. Using Gibbs sampler algorithm, generated $\boldsymbol{\Omega}$ from conditional posterior distribution (4)
$p(\boldsymbol{\Omega} \mid \rho_1, .., \rho_T, \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}, \mathbf{W}) \sim IW(\tilde{\boldsymbol{\Omega}}, \tilde{\upsilon})$.
9. Repeated stage 5 to stage 8 for *S* iterations. Then, discarded burn-in period which was obtained from the first *B* iterations.
10. Calculated the value of the estimated parameters by averaging the values of generated samples after burn-in period and then calculated the standard error.
11. Tested the significance of parameeter using Credible Interval (CI).
12. Evaluated the efficiency of parameters estimation using SUR-SAR model by comparing standard error of both SUR-SAR model and SAR individual models. If the ratio of SUR-SAR standard error to SAR individual models from a parameter more than one, it can be said that a parameter is more efficient to estimate using SUR-SAR model.

RESULT AND DISCUSSION

Preliminary Analysis of Contemporaneous Correlation

The first thing to be known was the presence of correlations (contemporaneous correlation) among individual SAR models during 2009-2011. If there were contemporaneous correlation among different SAR models, modeling the different SAR models in SUR system should be considered. Since the proposed SUR system model was conducted by Bayesian approach, the individual SAR models should be done by Bayesian approach too. The test was done by utilizing information of covariance errors from two individual models. The correlation of two individual model was obtained from their standardized covariance errors.

Table 1. Correlations of Bayesian SAR Errors on DHF Data

Year	2009	2010	2011
2009	1.0000000	0.3651638*	0.2662044*
2010	0.3651638*	1.0000000	0.4532906*
2011	0.2662044*	0.4532906*	1.0000000

Table 1 showed the estimated values of correlation among Bayesian SAR models error which the lowest and the highest value respectively 0.4532906 (2010 & 2011) and 0.2662044 (2009 & 2011). Based on the *T* test at 0.05 level of significance, there were significant contemporaneous correlations among Bayesian SAR models, so that the SUR system should be considered and could be performed.

SUR-SAR on the DHF cases during 2009 - 2011

Spatial autoregressive model (SAR) which is formed for each year served in a system called Seemingly Unrelated Regression (SUR). The basic form of the SUR-SAR model for DHF data in Bogor according to Anselin (1988) :

$$\begin{pmatrix}\mathbf{y}_{2009}\\ \mathbf{y}_{2010}\\ \mathbf{y}_{2011}\end{pmatrix} = \begin{pmatrix}\rho_{2009}\mathbf{W} & 0\mathbf{W} & 0\mathbf{W}\\ 0\mathbf{W} & \rho_{2010}\mathbf{W} & 0\mathbf{W}\\ 0\mathbf{W} & 0\mathbf{W} & \rho_{2011}\mathbf{W}\end{pmatrix}\begin{pmatrix}\mathbf{y}_{2009}\\ \mathbf{y}_{2010}\\ \mathbf{y}_{2011}\end{pmatrix} + \begin{pmatrix}\mathbf{X}_{2009} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{2010} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{X}_{2011}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_{2009}\\ \boldsymbol{\beta}_{2010}\\ \boldsymbol{\beta}_{2011}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\varepsilon}_{2009}\\ \boldsymbol{\varepsilon}_{2010}\\ \boldsymbol{\varepsilon}_{2011}\end{pmatrix} \quad (5)$$

In the system of vector matrix, equation (5) becomes :

$$\mathbf{y} = \left(\mathbf{D}_{\rho_t} \otimes \mathbf{W}\right)\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{, dengan } \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{I}_n) \quad (6)$$

where $\mathbf{D}_{\rho_t}$ is a diagonal matrix whose main diagonal elements are ρ_{2009}, ρ_{2010}, and ρ_{2011} and $\boldsymbol{\Omega}$ is a variance-covariance matrix of errors model. The contemporaneous correlation among individual SAR models was justified by this variance-covariance matrix.

For the *i*-th unit observation (village) at the *t*-th year, the model in equation (6) can be written as follows:

$$y_{i,t} = \rho_t \sum_{j=1}^{n} w_{ij} y_{j,t} + \beta_{0,t} + X_{1i,t}\beta_{1,t} + X_{2i,t}\beta_{2,t} + X_{3i,t}\beta_{3,t} + X_{4i,t}\beta_{4,t} + \varepsilon_{i,t} \quad (7)$$

$y_{i,t}$ is the number of DHF patients at the *i*-th (*i* = 1, 2, ..68) village and the *t*-th (*t* = 2009, 2010, 2011) year, $y_{j,t}$ is the number of DHF patients at the *j*-th village and the *t*-th year , $X_{1i,t}$ is the population density at the *i*-th village and the *t*-th year, $X_{2i,t}$ is the population mobility at the *i*-th village and the *t*-th year, $X_{3i,t}$ is the the average age of patients at the *i*-th village and the *t*-th year, $X_{4i,t}$ is the number of clinics at the *i*-th village and the *t*-th year. Based on the model in equation (7), there are 18 model parameters that must be estimated.

SUR-SAR Parameter Estimation using MCMC

SUR-SAR model with Bayesian approach was done by using MCMC sampling scheme through simulation. There were two algorithms in this simulation, Gibbs Sampler algorithm and Metropolis algorithm. All the algorithms were implemented into computer programming by open source statistical software, namely R. The initial values for the specified parameter of prior distribution were :

$$\boldsymbol{\beta}_0^{'} = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

$$\boldsymbol{\Sigma}_0 = diag(15)\times(0.1)^4\ ;\ \boldsymbol{\Omega}_0 = diag(3)\times 10^4\ ;\ \upsilon_0 = 4$$

The initial values for the specified parameter of posterior distribution were :

$$\boldsymbol{\beta}^{(0)'} = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

$$\rho_{2009}^{(0)} = 0.5\ ;\ \rho_{2010}^{(0)} = 0.5\ ;\ \rho_{2010}^{(0)} = 0.5\ ;\ c{=}0.175$$

$$\boldsymbol{\Omega}^{(0)} = \begin{pmatrix} 0.09128148 & 0.03480798 & 0.02582482 \\ 0.03480798 & 0.09776485 & 0.04921619 \\ 0.02582482 & 0.04921619 & 0.08736971 \end{pmatrix}$$

Monitoring convergence of MCMC simulation was done by the trace plots, ergodic mean plots, and autocorrelation plots. This process could determine the burn-in period which could be used as a reference point for estimating model parameters. The MCMC simulation was conducted by setting 40,000 iterations which discarded the first 20,000 as burn-in period.

Table 2. Posterior Summaries of Bayesian SUR-SAR Parameters Model on the DHF Data

Parameter	Mean	Standar error	2.5% Quantile	Median	97.5% Quantile
ρ_{2009}	0.3909247	0.1080658	0.1790822	0.3904087	0.6105364
$\beta_{0,2009}$	-2.9377247	0.5694979	-4.0486713	-2.9431045	-1.8171012
$\beta_{1,2009}$	0.7739108	0.1476283	0.4834503	0.7747688	1.0617336
$\beta_{2,2009}$	0.1704708	0.0591826	0.0539982	0.1709091	0.2857064
$\beta_{3,2009}$	0.3168774	0.1175268	0.0819740	0.3177487	0.5463437
$\beta_{4,2009}$	0.2209325	0.1433717	-0.0553308	0.2193940	0.5101589
ρ_{2010}	0.3457631	0.0934126	0.1503563	0.3460046	0.5101589
$\beta_{0,2010}$	-1.558580	0.5518064	-2.6359942	-1.5594970	-0.4712121
$\beta_{1,2010}$	0.4272585	0.1458346	0.1425126	0.4263100	0.7156373
$\beta_{2,2010}$	0.0918619	0.0532354	-0.0119283	0.0917088	0.1963527
$\beta_{3,2010}$	0.4821184	0.0832779	0.3221260	0.4808031	0.6494176
$\beta_{4,2010}$	0.2694014	0.1486674	-0.0215855	0.2684026	0.5624114
ρ_{2011}	0.5816469	0.0885843	0.3980624	0.5840148	0.7494507
$\beta_{0,2011}$	-0.5256187	0.5645437	-1.6469012	-0.5258119	0.5763252
$\beta_{1,2011}$	0.1601081	0.1465617	-0.1234518	0.1586374	0.4521945
$\beta_{2,2011}$	0.0247619	0.0529904	-0.0806109	0.0247959	0.1289564
$\beta_{3,2011}$	0.1371261	0.0704403	0.0015383	0.1367020	0.2764504
$\beta_{4,2011}$	0.5054946	0.1558145	0.2041735	0.5052070	0.8100167

Table 2 showed the estimated parameters of Bayesian SUR-SAR model obtained from mean of posterior distribution after burn-in period. Test of significance with Bayesian approach of SUR-SAR model parameters were based on 95% of credible interval (CI) that the estimated parameters were in the interval. If 2.5% quantile and 97.5% quantile had the same sign or did not include zero value between them, it could be concluded that the estimated parameter was statistically significant.

The contibutions of the population density was statistically significant in 2009 and 2010. Only in 2009, there was significant contributions of the population mobility. During 2009 – 2011, the contributions of the average age of patients was significant. However, the contributions of the number of clinics could increase the number of DHF patients significantly in 2011 was an illogical phenomenon so that it should be done for further analysis. Finally, for the spill-over effects or spatial contributions, ρ_t, all of the parameters were estimated significantly where in 2010 the values was decreased.

Model Evaluation

Model evaluation in modeling DHF case using Bayesian SUR-SAR model was conducted by two methods, evaluated either the variance-covariance matrix of errors model or the correlation matrix of errors model as well as evaluated the efficiency of parameter estimation. The variance-covariance matrix and the correlation matrix of errors model were estimated from the residuals of fitted model.

Table 3. Estimated Variance-Covariance Matrix

Year	2009	2010	2011
2009	0.08153579	).03499376	0.02768806
2010	0.03499376	).08618334	0.05513449
2011	0.02768806	).05513449	0.09272435

Table 3 showed the estimated variance-covariance matrix of errors model. From this matrix, we found that the variances were quite similar. Based on the Table 3, the covariance seemed to vary that the lowest and the highest value respectively 0.02768806 (2009 & 2011) and 0.05513449 (2010 & 2011). Thus, the covariance values were standardized by dividing with the square root of theirs variances.

Table 4. Estimated Correlation Matrix

Year	2009	2010	2011
2009	1.0000000	0.4438032	0.3385378
2010	0.4438032	1.0000000	0.6708728
2011	0.3385378	0.6708728	1.0000000

Table 4 showed the estimated correlation matrix of errors model. The values of estimated correlation matrix indicated that there were moderate relationships among individual models. The result of correlation matrix were consistent with the result of variance-covariance matrix which the lowest in (2009 & 2011) and the highest in (2010 & 2011). It showed that the serials correlation through contemporaneous correlation in SUR-SAR system played an important role in analyzing the incidence of DHF in Bogor. The same of DHF spreading pattern could be indicated that at the same village, DHF patients in this year will suffer again in the next year so that the related institutions can focus on reducing DHF in one region through the right policies.

Table 5. Standard Errors for SUR-SAR and Individual SAR

Parameter	Methods		Ratio
	SUR-SAR	Individual SAR	(SUR-SAR / Individual SAR)
ρ_{2009}	0.1080658	0.1137327	<1
$\beta_{0,2009}$	0.5694979	0.58934775	<1
$\beta_{1,2009}$	0.1476283	0.15406355	<1
$\beta_{2,2009}$	0.0591826	0.06168485	<1
$\beta_{3,2009}$	0.1175268	0.12435645	<1
$\beta_{4,2009}$	0.1433717	0.14322395	>1
ρ_{2010}	0.0934126	0.11312714	<1
$\beta_{0,2010}$	0.5518064	0.57384955	<1
$\beta_{1,2010}$	0.1458346	0.15328510	<1
$\beta_{2,2010}$	0.0532354	0.06817655	<1
$\beta_{3,2010}$	0.0832779	0.11013954	<1
$\beta_{4,2010}$	0.1486674	0.15075761	<1
ρ_{2011}	0.0885843	0.10003417	<1
$\beta_{0,2011}$	0.5645437	0.55901350	>1
$\beta_{1,2011}$	0.1465617	0.14456376	>1
$\beta_{2,2011}$	0.0529904	0.06431355	<1
$\beta_{3,2011}$	0.0704403	0.07941698	<1
$\beta_{4,2011}$	0.1558145	0.15318808	>

Table 5 showed the standard errors of parameters model obtained from both system of SUR and individual models. Almost, the estimated standard errors for SUR-SAR were smaller than those for individual SUR. Thus, the results were considerably consistent with the claim that SUR system was more efficient than individual model. Therefore, the further researches are needed to explain why there are the values of standard errors for SUR system were still greater than the individual models.

CONCLUSION

Bayesian SUR-SAR can estimate the spatial and non-spatial contributions variables in spatial modeling which constructed from correlated (contemporaneous correlation) SAR models. To estimate parameters, MCMC simulation is conducted by Metropolis and Gibbs Sampler algorithms. Then, test the significance of the estimated parameters using CI, evaluate the existence of correlations among individual models in SUR system, and evaluate the efficiency of parameter estimation.

Based on the DHF data, we found the following implications: there was the similar pattern for the pattern of DHF spread in Bogor during 2009 – 2011, the nearby areas had a significant role to the incidence of DHF in an area in the city of Bogor, and the non-spatial contributions of DHF cases in Bogor during 2009 -2011 included in this model were dynamic. Gain efficiency of parameters estimation on modeling Bogor DHF with SAR model for each year during 2009-2011 could be obtained if we constructed with SUR system model.

REFERENCES

[1] Anggana HD. 2012. Penerapan Bayesian Spatial Autoregressive Model dalam Menaksir Kontribusi Faktor-faktor yang Mempengaruhi Kejadian DBD di Kabupaten Tasikmalaya [Skripsi]. Bandung(ID) : Universitas Padjadjaran.

[2] Anselin L. 1988. *Spatial Econometrics Methods and Models*. Dordrecht: Kluwer Academic Publishers.

[3] Fatmawati N. 2011. Aplikasi Model Autoregresif Spasial Regresi Campuran Dalam Menaksir Hubungan Antara ABJ dan Kepadatan Penduduk Dengan Penderita DBD di Kota Bogor. [Skripsi]. Bandung(ID) : Universitas Padjadjaran.

[4] Felmlee DH, Hargens LL. 1988. Estimation and Hypothesis Testing for Seemingly Unrelated Regressions : A Sociological Application. *Social Science Research* , vol.17, pp. 384-399.

[5] Griffiths W. 2001. *Bayesian Inference in the Seemingly Unrelated Regressions Model.* Melbourne: University of Melbourne.

[6] Kakamu K, Polasek W, Wago H. 2007. Spatial Agglomeration and Spill-over Analysis for Japanese Prefectures during 1991-2000.

[7] Kartika Y. 2007. Pola Penyebaran Spasial Demam Berdarah Dengue di Kota Bogor Tahun 2005. [Skripsi]. Bogor(ID) : Institut Pertanian Bogor.

[8] LeSage JP. 2005. *Bayesian Estimation of Spatial Regression Models*. Toledo : Department of Economics University of Toledo.

[9] Nadra. 2006. Pola Penyebaran Spasial dan Penerapan Model Regresi Auto-Gaussian pada Kasus Jumlah Penderita Demam Berdarah di Kota Bogor. [Skripsi]. Bogor(ID) : Institut Pertanian Bogor.

[10] Ntzoufras I. 2009. *Bayesian Modeling Using WinBUGS*. New York : John Wiley & Sons. Inc.

[11] Rahmat A. 2014. Analisis Spasial-Temporal untuk Mengkaji Faktor-faktor yang Mempengaruhi Sebaran Penyakit Demam Berdarah di Kota Bogor [Tesis]. Bogor(ID) : Institut Pertanian Bogor.

[12] Zellner A. 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias. *Journal of American Statistical Association,* vol. 57, pp. 348-368.