Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 20 No.2, Oktober 2015, p: 78-84

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

A SIMULATION STUDY OF LOGARITHMIC TRANSFORMATION MODEL IN SPATIAL E MPIRICAL BEST LINEAR UNBIASED PREDICTION (SEBLUP) METHOD OF SMALL AREA ESTIMATION

Hazan Azhari Zainuddin[1], Khairil Anwar Notodiputro[1], Kusman Sadik[1]

[1]Department of Statistics, Bogor Agricultural University, Indonesia
E-mail: hazanazhari78@gmail.com, khairilnotodiputro@gmail.com, kusmansadik@gmail.com

ABSTRACT

There have been many studies developed to improve the quality of estimates in small area estimation (SAE). The standard method known as EBLUP (Empirical Unbiased Best Linear Predictor) has been developed by incorporating spatial effects into the model. This modification of the method was known SEBLUP (Spatial EBLUP) since it incorporates the spatial correlations which exist among the small areas. The data obtained (variables of concern) usually have a large variance and tend to have a a nonsymmetric distribution and therefore tend to have nonlinear relationship pattern between concomitant variables and variables of concern. the results showed that the method SEBLUP using logarithmic transformation produces estimator more than the other methods.

Keywords : EBLUP, SAE, SEBLUP

INTRODUCTION

Various surveys are generally designed to estimate population parameters of national level. Problems will arise if the survey would like to obtain information for smaller areas, for example at the provincial level, district level, or sub-district level. The size of the sample at the level of the area is usually so small that the statistics obtained will have a large variance. To overcome this problem, developed a parameter estimation method called small area estimation methods (SAE).

Small area estimation has now become the world's attention statisticians very seriously. There have been many studies were developed both for the improvement of techniques and the development of methods and applications in a variety of cases and real problems faced. Fay and Herriot (1979) (referred to in Rao 2003) was the first researcher to develop a small area estimation based models. Models are developed and then become a reference in the development of small area estimation study further until today.

There are two basic assumptions in developing SAE models, namely, that diversity in the response variable subpopulations can be explained entirely by the diversity of relations corresponding to the additional information, called a fixed effect, then the assumption of specific subpopulations diversity can be explained by a random effect subpopulation. A combination of both of these assumptions form a model of the effect of a mixture (mixed model). One of the interesting properties of linear mixed models is the ability to guess a linear combination of the effect of fixed and random effects. Henderson (1963) (referred to in Rao 2003) developed a technique completion of a mixture of linear models, namely, Best Linear Prediction Unbiased Prediction (BLUP). This method is then studied further by Harville (1991) (referred to in Rao 2003) by first estimation variance components with the maximum likelihood method (maximum likelihood) and constrained maximum likelihood (restricted maxim um likelihood), so-called Empirical Best Linear Unbiased Prediction (EBLUP).

A few years later EBLUP method developed by incorporating spatial effect into the model. EBLUP estimators by observing the effect of spatially correlated random area known as the Spatial Unbiased Empirical best linear prediction (SEBLUP) method. SEBLUP method can improve the structural diversity of small area estimation models that have a spatial correlation between areas. The model used in the method based SEBLUP area, the reason for modeling spatial incorporated into the model SAE is a spatial data type modeling area. SEBLUP estimators have been used by Petrucci & Salvati (2004), Chandra, Salvati and Chambers (2007) to include spatial weighting matrix spatially nearest neighbors (nearest neighbors) into SEBLUP method.

The data obtained (variables of concern) usually have a large variance and tend to have a distribution pattern that is not symmetric and therefore caused no linear relationship pattern between concomitant variables and variables of concern. To resolve this problem, Kurnia (2009) using the logarithmic transformation variables of concern are then applied to the methods and the results EBLUP estimation better than usual EBLUP metho ds. In this research, the logarithmic transformation variables of concern would be used in the method SEBLUP, which is expected to be obtained estimators with better precision.

PRELIMINARY THEORY

Direct Estimation

Implementation of the survey conducted to estimate population parameters. The classical approach to estimate population parameters based on design of sampling (design-based), and the estimation resulting from the approach is called direct estimation. Data from this survey can be used to obtain a reliable estimate of the total and the average population of an area or domain with a large number of examples. However, when the direct estimators are used to a small area, it will cause a large standard error (Ghosh and Rao, 1994 referred to in Rao 2003).

Indirect Estimation

In small area estimation, there are two basic types of models used, the basic area level model and the basic unit level models (Rao 2003).

a) Basic area level models
A model based on the availability of supporting data that exists only for a particular area level, let $x_i = (x_{i1},\ x_{i2}, \dots, x_{ip})^T$ with parameters that will supposedly is θ_i are assumed to have a relationship with x_i. The supporting data used to build the model $\theta_i = {x_i}^T\beta + z_i v_i$, with i = 1,2,3, ..., m. And $v_i \sim N(0, \sigma_v^2)$, as a random effect. Conclusions regarding θ_i, can be determined by assuming that the direct estimation models y_i already available, namely, $y_i = \theta_i + e_i$ with i = 1,2,3, ... m. and sampling error $e_i \sim N(0, \sigma_i^2)$, with σ_i^2 unknown. Then the two models are combined to obtain a combined models: : $y_i = {x_i}^T\beta + z_i v_i + e_i$, with i = 1,2,3,...m. The model is a special form of linear mixed models.

b) Basic unit level model
Is a model in which the supporting data provided corresponding individually with response data, such$x_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})^T$ to obtain a regression model nested $y_{ij} = {x_{ij}}^T\beta + v_i + e_{ij}$ with i = 1,2,3, ... m. and j = 1,2,3, ... Ni, $v_i \sim N(0, \sigma_v^2)$ and $e_i \sim N(0, \sigma_i^2)$.

Spatial Empirical Best Linear Unbiased Prediction (SEBLUP)

Suppose that defined the vector $\hat{\theta} = \left(\hat{\theta}_1, \dots, \hat{\theta}_m\right)^T, v = (v_1, \dots, v_m)^T$ and $e = (e_1, \dots, e_m)^T$, and the matrix $\boldsymbol{X} = (x_1^T, \dots, x_m^T)^T$ and **Z**= diag $(z_1, \dots, z_m)$ Based on the definition of the vectors and matrices, the equation (1) in matrix notation is:

$$\hat{\theta} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \qquad (1)$$

The model in equation (1) assumes that there is random effect area, these effects are independent between areas. In fact, it is reasonable to say that there is a correlation between the adjacent area. Such correlations will decrease as the distance increases. This is in accordance with the first law of geography that put forward by Tobler (*Tobler's first low of geography)* in Schabenberger and Gotway (2005) which is a pillar study spatial data analysis, that "*everything is related to everything else, but near things are more related then distant things "*. SAE models to include the spatial correlation between areas was first introduced by Cressie (Cressie 1991 referred to in Rao 2003), assuming that the spatial dependence follow Conditional Autoregressive process (autoregressive conditional, CAR). SAE models are then developed further by several researchers, including Salvati (2004), Candra, Salvati, and Chambers (2007), Pratesi and Salvati (2008), assuming that the spatial dependency component that is inserted into the error of random factors follow the process

Simultaneous Autoregressive (Simultaneous otoregresif, SAR). Model SAR was first introduced by Anselin (Anselin 1992 referred to in Candra, Salvati, Chambers 2007) in which the random effect area vectors $\boldsymbol{v}$ satisfy area:

$$\boldsymbol{v}= \rho \boldsymbol{W}\boldsymbol{v}+ \boldsymbol{u} \qquad (2)$$

coefficient ρ in equation (2) is a spatial otoregresi coefficient indicates the strength of the relationship between the spatial random effect. The value of ρ ranging from -1 to 1. The value ρ> 0 indicates that an area with a high parameter values tend to be surrounded by other areas with high parameter values as well and an area with a low parameter value anyway. On the other hand, ρ <0 indicates that an area with a high parameter values are surrounded by other areas with a low parameter value, or vice versa (Savitz and Raudenbush 2009). **W** is a spatial weighting matrix, $\boldsymbol{v}$ is the random effect of the area and $\boldsymbol{u}$ is the error vector of random effects of an area with an average of zero and variance $\sigma_u^2 \boldsymbol{I}_m$ Equation (2) can be rewritten as follows:

$$\boldsymbol{v}= (\mathbf{I} - \boldsymbol{v}\mathbf{W})^{-1} \boldsymbol{u} \qquad (3)$$

I is the identity matrix with size $m \times m$. From equation (3) shows that the average $\boldsymbol{v}$ is 0 and the matrix koragam $\boldsymbol{v}$ **(G)** is as follows:

$$\mathbf{G} = \sigma_u^2 \,[(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}^{\mathrm{T}})]^{-1}$$

Equation (3) is inserted into the equation (1) yields:

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

Covarians matrix from $\hat{\boldsymbol{\theta}}$ with $\mathbf{R}$ = diag (σ_i^2) :

$$\mathbf{V} = \mathbf{R} + \mathbf{ZGZ}^{\boldsymbol{T}} = \text{diag}\,(\sigma_i^2) + \mathbf{Z}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}\,[(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}^T)]^{-1}\,\boldsymbol{Z}^T \qquad (4)$$

Spatial BLUP estimator for parameter θ_i with σ_u^2 , σ_i^2 and ρ known area

$$\tilde{\theta}_i^s(\sigma_u^2\,,\rho) = x_i\hat{\beta} + \mathbf{b}_i^{\mathrm{T}}\{\sigma_u^2\,(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}^{\mathrm{T}})^{-1}\}\mathbf{Z}^{\mathrm{T}} \times \{\text{diag}\,(\sigma_i^2) + \mathbf{Z}\sigma_{\boldsymbol{u}}^{\mathbf{2}}\,[(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}^T)]^{-1}\boldsymbol{Z}^T\}^{-1}(\hat{\theta} - \boldsymbol{X\hat{\beta}}) \qquad (5)$$

Where $\hat{\beta}$ = $(\boldsymbol{X}^T\,V^{-1}\,\boldsymbol{X})^{-1}\boldsymbol{X}^T V^{-1}\hat{\theta}$ and $\mathbf{b}_i^{\mathrm{T}}$ is a vector of size $1 \times n$ $(0, 0, \ldots 0,\ 1,\ 0, \ldots .0)$ with 1 pointing to the location of the i-th. Spatial BLUP estimator is obtained by inserting covarians matrix in equation (4) into the BLUP estimator. Spatial BLUP will be the same as if ρ = 0. It can be concluded that estimators of small area estimation models to include the spatial correlation will lead to ordinary small area estimation model (random effects are independent) if in fact there is no spatial correlation in the area of the observed area , However, if there is a spatial correlation and if we use the model of the usual small area estimation models were also less precise. So, using the usual small area estimation models in areas that have a spatial correlation will produce a variety of structures that do not fit, when the model chosen less precise it will generate greater error.

MSE calculation of Spatial BLUP may be obtained as in Rao (2003), namely:

$$MSE\left[\tilde{\theta}_i^S(\sigma_{\boldsymbol{u}}^{\mathbf{2}}, \rho)\right] = g_{1i}\,(\sigma_{\boldsymbol{u}}^{\mathbf{2}}, \rho) + g_{2i}\,(\sigma_{\boldsymbol{u}}^{\mathbf{2}}, \rho) \qquad (6)$$

As is the case with EBLUP estimator, SEBLUP estimator $\left(\hat{\theta}_i^S(\hat{\sigma}_u^2\,, \hat{\rho})\right)$ from Spatial BLUP obtained by replacing the value σ_u^2, ρ with the estimators. The assumption of normality of random effects is used to predict σ_u^2 dan ρ using both ML and REML procedure with the log-likelihood function has a global maximum and several local maximum (Patresi and Salvati 2005 referred to in Candra, Salvati, Chambers 2007). The probe can be obtained iteratively by using a scoring algorithm. Scoring algorithm requires a great starting point to get the maximum functionality. The estimation results are then used to probe against SEBLUP, with estimators EBLUP formula is:

$$\tilde{\theta}_i^s(\hat{\sigma}_u^2\,, \hat{\rho}) = x_i\hat{\beta} + \boldsymbol{b}_i^T\{\hat{\sigma}_u^2\,(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})(\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^T)^{-1}\}\boldsymbol{Z}^T \times \{\text{diag}\,(\sigma_i^2) + \mathbf{Z}\hat{\sigma}_u^2\,[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}^T)]^{-1}\boldsymbol{Z}^T\}^{-1}(\hat{\theta} - \boldsymbol{X\hat{\beta}}) \qquad (7)$$

$MSE\left[\tilde{\theta}_i^S, (\hat{\sigma}_u^2, \hat{\rho})\right]$ EBLUP spatial models with random effects are normally distributed, are:

$$MSE\left[\tilde{\theta}_i^S, (\hat{\sigma}_u^2, \hat{\rho})\right] = MSE\left[\tilde{\theta}_i^S, (\sigma_{\boldsymbol{u}}^{\mathbf{2}}, \rho)\right] + E\left[\tilde{\theta}_i^S, (\hat{\sigma}_u^2, \hat{\rho}) - \tilde{\theta}_i^S, (\sigma_{\boldsymbol{u}}^{\mathbf{2}},\ \rho)\right]^2 \qquad (8)$$

Form $E\left[\tilde{\theta}_i^S, (\hat{\sigma}_u^2, \hat{\rho}) - \tilde{\theta}_i^S, (\sigma_{\boldsymbol{u}}^{\mathbf{2}},\ \rho)\right]^2$ estimated by Taylor and denoted by $g_{3i}(\sigma_u^2, \rho)$ (Kackar and Harville in 1984 referred to in the Pratesi and Salvati 2008).

Estimators of $MSE\left[\tilde{\theta}_i^S, (\hat{\sigma}_u^2, \hat{\rho})\right]$ obtained by following the results of Harville and Jeske (Harville and Jeske 1992 in the Pratesi and Salvati 2008) and later developed into a model with generalized covariances by Zimmerman and Cressie (Zimmerman and Cressie 1992 in the Pratesi and Salvati 2008), namely

$$mse\left[\tilde{\theta}_i^S,(\hat{\sigma}_u^2,\hat{\rho})\right] \approx g_{1i}(\hat{\sigma}_u^2,\hat{\rho}) + g_{2i}(\hat{\sigma}_u^2,\hat{\rho}) + 2g_{3i}(\hat{\sigma}_u^2,\hat{\rho})$$

Where $\hat{\sigma}_v^2$ and $\hat{\rho}$ is the estimator obtained using REML method. If using $\hat{\sigma}_v^2$ and $\hat{\rho}$ ML estimators procedures, calculations $mse\left[\tilde{\theta}_i^S,(\hat{\sigma}_u^2,\hat{\rho})\right]$ as follows:

$$mse\left[\tilde{\theta}_i^S,(\hat{\sigma}_u^2,\hat{\rho})\right] \approx g_{1i}(\hat{\sigma}_u^2,\hat{\rho}) - b_{ML}^T(\hat{\sigma}_u^2,\hat{\rho})\nabla g_{1i}(\hat{\sigma}_u^2,\hat{\rho}) + g_{2i}(\hat{\sigma}_u^2,\hat{\rho}) + 2g_{3i}(\hat{\sigma}_u^2,\hat{\rho})$$

$b_{ML}^T(\hat{\sigma}_u^2,\hat{\rho})\nabla g_{1i}(\hat{\sigma}_u^2,\hat{\rho})$ form is an additional form of an additional bias of $g_{1i}(\hat{\sigma}_u^2,\hat{\rho})$.

Logarithmic Transformati

Defined a logarithmic transformation in linear mixed models (log-scale linear mixed model) as follows:

$$\log(y_{ij}) = \boldsymbol{x_i}^T\beta + v_i + \varepsilon_{ij} \qquad (9)$$

with ε_{ij} following the distribution iid $N(0,\sigma_i^2)$, random effect of area v_i following the distribution iid $N(0,\sigma_v^2)$ but if there is a spatial effect of the v_i following the distribution MVN (0, G). Kurnia (2009) explained that following EBLUP theory to model (9), namely EBLUP to the mean θ_i of $\log(y_{ij})$, then estimators for θ_i can be written as follows:

$$\hat{\theta}_i^{EBLUP} = \hat{\gamma}_i\hat{\theta}_i^D + (1-\hat{\gamma}_i)\,\boldsymbol{x_i}^T\hat{\beta} \quad (10)$$

with $\hat{\beta}$ obtained by weighted least squares method for parameter β regression of log-scale linear mixed model, where $\hat{\gamma}_i = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + n_i^{-1}\hat{\sigma}_i^2)$ and $\hat{\theta}_i^D = \frac{1}{n_i}\sum_{j\in s(i)}\log(y_{ij})$ is a direct estimator for θ_i based on data sample s (i) for the i-th area.

Because we want is an actual estimate for the median in each area to-i, then used a lognormal distribution properties to transform and forth from the model (10). Furthermore, it is assumed that the sampling distribution for $\hat{\theta}_i^{EBLUP}$ is $N\{\theta_i, Var(\hat{\theta}_i^{EBLUP})\}$. Thus, actual estimate value (raw-scale) for the i-th area is

$$\hat{\mu}_i = exp\left(\hat{\theta}_i^{EBLUP} + \frac{1}{2}\hat{v}_i^{EBLUP}\right) \quad (11)$$

With $\hat{v}_i^{EBLUP}$ is MSE estimators of $\hat{\theta}_i^{EBLUP}$. Then the MSE estimator for the mean estimator in equation (11) can be obtained as follows:

$$\hat{V}_i(\hat{\mu}_i) = e^{\hat{v}_i^{EBLUP}}\left(e^{\hat{v}_i^{EBLUP}} - 1\right)e^{2\hat{\theta}_i^{EBLUP}} \qquad (12)$$

In this study, will be applicable log-scale linear mixed models into SEBLUP method. by following SEBLUP theory to model (9), namely SEBLUP to the mean θ_i of $\log(y_{ij})$, then estimator for θ_i can be written as follows:

$$x_i\hat{\beta} + \boldsymbol{b}_i^T\{\hat{\sigma}_u^2(\boldsymbol{I}-\hat{\rho}\boldsymbol{W})(\boldsymbol{I}-\hat{\rho}\boldsymbol{W}^T)^{-1}\} \times\{\text{diag}(\hat{\sigma}_i^2) + \mathbf{Z}\hat{\sigma}_u^2[(\mathbf{I}-\hat{\rho}\mathbf{W})(\mathbf{I}-\hat{\rho}\mathbf{W}^T)]^{-1}\boldsymbol{Z}^T\}^{-1}\left(\boldsymbol{\hat{\theta}^D} - \boldsymbol{X\hat{\beta}}\right) \qquad (13)$$

With, $(\boldsymbol{\hat{\theta}^D})^T$: $(\hat{\theta}_1^D,\hat{\theta}_2^D,\hat{\theta}_3^D,\ldots\ldots\hat{\theta}_m^D)$ and $\hat{\theta}_i^D = \frac{1}{n_i}\sum_{j\in s(i)}\log(y_{ij})$.). as well as EBLUP, the method is desirable also SEBLUP actual estimator for the mean in each area to-i, thus obtained:

$$\hat{\mu}_i = exp\left(\hat{\theta}_i^{SEBLUP} + \frac{1}{2}\hat{v}_i^{SEBLUP}\right) \qquad (14)$$

With $\hat{v}_i^{SEBLUP}$ is MSE estimators of dari $\hat{\theta}_i^{SEBLUP}$. Then the MSE estimator for mean estimation in the equation (14) can be obtained as follows:

$$\hat{V}_i(\hat{\mu}_i) = e^{\hat{v}_i^{SEBLUP}}\left(e^{\hat{v}_i^{SEBLUP}} - 1\right)e^{2\hat{\theta}_i^{SEBLUP}} \qquad (15)$$

METHOD

Simulation Study

Simulations were performed to evaluate the good of the developed model. The simulation process is done by following these steps.

1. Make a map made in the form of a grid consisting of a small area m. Where m is to be tested are 25
2. Build a sample consisting of m small area
3. search for spatial contiguity weighting matrix Queen (W) based on a map that has been created.
4. This simulation uses the response variable Y(variable of concern) and an accompanying variable X. The model used to derive the value of the response variable log(y_{ij}) for small area i-th and j-th unit is as follows:

$$\log(y_{ij}) = \boldsymbol{x_{ij}}^T\beta + v_i + e_{ij} \quad , \quad i = 1,2,\ldots,25, \quad j = 1,2,\ldots,n\_i \qquad (16)$$

Where $\boldsymbol{x_{ij}}$is concomitant variables, v_i is random effect area, and e_{ij} is sampling error.

(a) x_{ij} value generated by a normal distribution $N(10,1.34)$. x_{ij} value obtained is used for the entire scenario in the simulation process.

(b) then $\boldsymbol{v} = (v_1,\ldots,v_m)^T$ raised by the spread Multivariat Normal

MVN ($\mathbf{0}$, $\mathbf{G}$), where $\mathbf{G}=\sigma_u^2[(\mathbf{I}-\rho\mathbf{W})(\mathbf{I}-\rho\mathbf{W}^T)]^{-1}$ is the variance-covariance matrix size 25×25. determined value $\sigma_u^2 = 2$ and $\rho = 0.77$

(c) Then, $\mathbf{e} = (e_{11}, e12\ ,...,eij,...,emNm)T$generated by a normal distribution $N(0, 0.34)$.

(d) Last value, set $\boldsymbol{\beta} = (12.7451\,, 0.02111)^T$ in order to obtain the following equation:
$\log(y_{ij}) = 12.7451 + 0.02111x_{ij} + v_i + e_{ij}\ \ , i = 1,2,\dots,25,\ j = 1,2,\dots,n_i.$ (17)

(e) Determine the value $y_{ij}$by entering the value x_{ij}, v_i and e_{ij} into the model or equation (17)

(f) Looking for the actual value y_{ij} with $y_{ij} = e^{\log(y_{ij})}$, so it can be said y_{ij} generated with a log-normal distributon.

5. Calculating the mean variable of concern samples in each small area as a direct estimation
$\hat{\bar{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, with $i = 1,2,\dots,25,\ ,\quad j = 1,2,\dots,n_i$
Then calculate the mean variable concomitant each sample in a small area
$\hat{\bar{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, with $i = 1,2,\dots,25,\ ,\quad j = 1,2,\dots,n_i$
Calculating the mean logarithmic variable of concern samples in each small area as a direct estimation
$\hat{\bar{y}}_i^{\,D} = \frac{1}{n_i}\sum_{j=1}^{n_i} \log(y_{ij})$, with $i = 1,2,\dots,25,\ ,\quad j = 1,2,\dots,n_i$
6. Find the value of $\hat{\bar{y}}_i^{\,EBLUP}$ using $\hat{\bar{y}}_i$
7. Find the value of $\hat{\bar{y}}_i^{\,EBLUP}$ using $\hat{\bar{y}}_i^{\,D}$. After that is done, behind the transformation. Stage 7 is called a log-scale method EBLUP
8. Find the value of $\hat{\bar{y}}_i^{\,SEBLUP}$ using $\hat{\bar{y}}_i^{\,D}$. After that is done, behind the transformation. Stage 8 is called a log-scale method SEBLUP
9. Perform steps (4) to step (8) of B = 1000
10. Calculating the value of μ_i which is the expected value of the estimates of the mean in each area.
   a. Find the value of $E[y_{ij}] = exp\left(E[\log(y_{ij})] + 12var\log(yij)\right)$
   b. Repeat the steps until B = 1000
   c. Find the value of $\mu_i = \frac{\sum_{l=1}^{1000} E[y_{ij}]_l}{1000}$
11. so it can be calculated the value of the Relative Bias (RB) and Relative Root Mean Squares Error (RRMSE) of parameter estimation results in each area as follows:

$$RB_{(i)} = \frac{1}{B}\sum_{l=1}^{B}\left(\frac{\hat{\theta}_{il}-\theta_i}{\theta_i}\right) \times 100\%$$

$$RRMSE_{(i)} = \frac{1}{\theta_i}\sqrt{\frac{1}{B}\sum_{l=1}^{B}\left(\hat{\theta}_{il}-\theta_i\right)^2} \times 100\%$$

12. Compare the value of RB and RRMSE between the direct estimation method, the method EBLUP, log-scale EBLUP methods, methods of log-scale SEBLUP.

RESULTS AND DISCUSSION

This simulation study conducted for four estimators, namely: (1) direct estimator for the original data (Direct), (2) EBLUP to the original data (EBLUP), (3) behind transformation EBLUP, equation 7. (log-scale EBLUP), and (4) behind SEBLUP transformation, equation 10. (log scale-SEBLUP).

From the simulation results, can be seen in Table 1 that when viewed from each area, RRMSE log-scale SEBLUP more smaller than the log-scale RRMSE EBLUP ,from 25 area, there are 16 RRMSE area is smaller. When compared with RRMSE EBLUP it can be seen that of the 25 areas, there are 25 RRMSE area is smaller. Then if RRMSE log-scale SEBLUP compared with direct estimators RRMSE it can be seen that of the 25 areas, there were 23 who RRMSE smaller.

Table 1. RB(*Relative Bias*) and RRMSE(*Relative Root Mean Square Error*)

Area	RB (relative bias)				RRMSE (relative root mean square error)			
	direct	EBLUP	log scale-EBLUP	log scale-SEBLUP	direct	EBLUP	log scale-EBLUP	log scale-SEBLUP
1	-0,73	-5,29	-11,49	-11,30	174,14	142,58	136,53	135,87
2	-7,92	-10,19	-15,30	-15,24	125,76	118,12	113,47	113,37
3	5,13	7,81	2,65	2,70	144,98	150,16	145,56	145,57
4	-5,74	-8,17	-11,84	-11,84	122,55	117,66	114,22	114,24
5	-0,15	0,27	-2,98	-2,97	153,23	152,11	150,47	150,62
6	18,82	17,05	12,95	12,90	213,02	207,76	204,69	204,20
7	0,99	-2,98	-9,93	-10,06	111,06	99,32	94,91	95,26
8	8,62	5,01	-0,77	-0,83	168,94	144,86	139,76	139,64
9	2,02	-0,30	-4,90	-4,98	196,65	190,93	185,56	185,05
10	8,79	2,87	-1,14	-1,16	155,68	153,81	148,98	148,91
11	11,17	7,28	0,81	0,79	131,89	117,47	111,99	112,04
12	-6,19	-4,87	-7,53	-7,55	118,85	122,87	120,15	120,07
13	2,71	0,83	-6,32	-6,32	129,00	120,66	114,49	114,36
14	6,51	-0,89	-6,22	-6,18	162,71	145,73	140,53	140,51
15	3,61	-1,44	-5,66	-5,61	145,22	129,88	125,19	125,28
16	12,21	9,58	4,93	4,86	149,25	139,46	137,25	137,10
17	0,83	0,72	-4,95	-5,04	168,40	164,10	159,24	158,94
18	-6,10	-5,61	-10,16	-10,20	117,82	112,66	107,79	107,83
19	4,66	-0,57	-4,79	-4,86	157,61	145,70	142,11	141,89
20	4,13	-1,06	-5,85	-5,87	149,81	144,60	140,40	140,11
21	-8,00	-7,93	-11,22	-11,20	94,95	93,50	90,94	91,00
22	0,67	-0,08	-7,25	-7,30	133,76	124,33	117,41	117,26
23	8,66	4,03	-2,46	-2,46	131,33	117,59	111,22	111,20
24	-12,59	-11,60	-14,97	-14,97	111,98	114,13	111,14	111,05
25	-3,66	-5,89	-9,01	-8,96	105,64	96,98	94,68	94,79

Table 2. Simulation Study Results Summary

	DIRECT	EBLUP	Log-Scale EBLUP	Log -Sale SEBLUP
ARRMSE(%)	142,9693	134,6786	130,3488472	130,2466936
ARB(%)	6,024427	4,893576	7,043548	7,046

From this simulation study shows that the proposed model well logscale SEBLUP gives better results especially for relative RMSE. Underestimate the nature of the MSE bias logscale SEBLUP has yet to be determined causes. However, indicated at least influenced by one things because it ignores the influence of σ_i^2 estimation.

CONCLUSIONS

Problems encountered in the small area estimation is the large variety of statistical estimators obtained directly so inefficient. To increase the effectiveness of the size of the sample, SAE add good information from the area itself, other areas and other surveys. SAE standard model as SEBLUP sometimes not able to explain the data well as strict linearity assumption of the model. Efforts to improve by modifying SEBLUP for models prior logarithmic transformation to increase the efficiency estimation.The proposed model, the log-sclae SEBLUP, gives better results as shown by the smallest RRMSE, although The relative bias is still quite large. However, in general there are indications that the proposed model can improve the efficiency estimation.

REFERENCES

Chandra H, Salvati N, Chambers R. 2007. Small Area Estiamtion for Spatially correlated populations a comparison of direct and indirect model-based methodes. *Statistics in transition* 8:887-906.

Kurnia A. 2009. *An Empirical Best Prediction Method for Logarithmic Transformation Model in Small Area Estimation with Particular Application to Susenas Data*. Master Student, Bogor Agricultular University.

Petrucci A, Salvati N. 2004a. Small area estimation using spatial information. The rathbun lake watershed case study. *Dipartimento di Statistica "G.Parenti" viale morgagani*, 59-50134.

Petrucci A, Salvati N. 2004a. Small area estimation considering spatially errors : the unit level random effects model. *Dipartimento di Statistica "G.Parenti" viale morgagani*, 59-50134.

Pratesi M, Salvati N. 2008. Small Area Estimation: the EBLUP estimator based on spatially correated random area effects. Statistical methodes and applications. Stat.Meth. & Appl. 17:113-141.

Rao JNK.2003. *Small Area Estiamtion*. London : Wiley.

Savitz NV, Raudenbush SW. 2009. Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology*, 39: 151.

Schabenberger O, Gotway CA. 2005. Statistical Methodes for Spatial Data Anlysis. Chapman & Hall /CRC.