IOSR Journal of Mathematics (IOSRJM)
ISSN: 2278-5728 Volume 2, Issue 4 (Sep.-Oct. 2012), PP 01-08
www.iosrjournals.org

Adaptation of Hierarchical Bayes SAE to Spatial Satscan

T. Siswantining[1], A. Saefuddin[2], K.A. Notodiputro[3], N. Nuryartono[4], I.W. Mangku[5]

[1]*Department of Mathematics , Faculty of Mathematics and Natural Science, University of Indonesia, Depok 16424,*

[2,3,4,5]*Department of Statistics, Faculty of Mathematics and Natural Science, Bogor Agricultural University, Bogor, Indonesia*

Abstract : *Scan statistic requires a large sample, whereas the real problem in the research typically uses small sample. By using Satscan software, this paper aims to replace the direct estimator (DE) with the estimator obtained from small areas (Hierarchical Bayes). Hierarchical Bayes Small Area Estimation (HB SAE) is more efficient than DE. Besides, it also can broaden the parameters prediction in a large area. In general, HB2 (using spatial nearest neighbor weighted) is better than the other HB, both in the simulation and the real data. In addition, analysis of a small data using HB2 SAE is resulting in better statistical properties, such as less biased and consistent.*

Keywords: *Direct Estimator (DE), Consistent, Efficient, Prediction, Spatial Hierarchical Bayes Small Area Estimation (HB SAE).*

I. INTRODUCTION

Spatial scan statistic requires a large sample because its hypothesis testing is based on the log of likelihood ratio (LLR). Spatial Satscan is software that can be used to do spatial scan statistic, i.e. to detect highly potential cluster. Unfortunately, many researches usually use small sample. One solution to overcome the small sample is to add or increase the sample size. However, in practice, this is rarely done due to the limitations of the situation, funds, time, and effort.

Small Area Estimation (SAE) statistical method can be used to analyze small samples. Because scan statistic requires a large sample, it is necessary to connect SAE to the spatial scan statistic in order to obtain a better estimator, compared with non-SAE estimator [1].

SAE has two kinds of estimators, i.e. direct and indirect estimator. Direct Estimator (DE), a simple SAE estimator, uses only domain-specific sample. However, the sample size of the domain is not large enough to support the reliability and accuracy [2,3,4]. The use of small sample in this estimator results in large standard error. Therefore, we need to find indirect estimators that can increase the effective sample size and lower the standard error[4].

SAE utilize information from outside (as an auxiliary variable), which is expected to improve DE. It is called indirect estimator. Indirect estimator can be grouped into Bayes method and non-Bayes method. Bayes method consists of Hierarchical Bayes (HB) and the empirical Bayes (EB). When compared with DE, the indirect estimator has a smaller MSE [4].

In this paper, we use HB approach in estimating parameter proportion. Its results then can be used in spatial scan statistic to replace the role of DE. This paper shows that, analytically, through simulation and real data application, the estimator obtained through HB SAE has better statistical properties, such as less biased and consistent, compared to DE.

II. POTENTIAL SOLVING FOR SMALL SAMPLE USING SAE IN SPATIAL SCAN STATISTIC

In theory, hypothesis testing in the scan statistic is based on the likelihood ratio,where the likelihood function is depend on the sample size. Likelihood ratio is a ratio between likelihood function based on the data compared with the likelihood function based on the null hypothesis is true. If the sample is in small size, the result of likelihood function is also small. It can make a difficulty to detect differences in the conclusions which is tending to be not significant. Therefore, the use of a small sample size analysis of the hypothesis testing will produce results as expected [5]. In addition, if the sample size is very small, the conclusion of the statistic test would be less obvious [6]. The maximum likelihood estimate has some optimum properties which distinguish it from all other large sample estimates [7].

Several studies have been conducted using the parameter estimation through SAE. They have shown that a variety of DE is greater than the range of indirect estimation. If the data is in the form of binary data, then appropriate method to use is EB SAE and HB SAE [8].

Because HB SAE result is better than DE, the role of DE then can be replaced with HB estimator. If the DE proportion in the likelihood function of the parameter space Ω is replaced with a proportion using the SAE results, it will be obtained that:

$$\hat{L}(\Omega) = \left(\hat{p}_{SAE}\right)^{n_Z}\left(1-\hat{p}_{SAE}\right)^{N(Z)-n_Z}\left(\hat{q}_{SAE}\right)^{n_G-n_Z}\left(1-\hat{q}_{SAE}\right)^{(N(G)-N(Z))-(n_G-n_Z)}\prod_{y_i\in G}\frac{N(y_i)!}{(N(y_i)-n_{y_i})!n_{y_i}!}$$

While the likelihood functions in the parameter space ω will be:

$$\hat{L}(\omega) = \left(\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{n_G}\left(1-\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{N(G)-n_G}\prod_{y_i\in G}\frac{N(y_i)!}{\left(N(y_i)-n_{y_i}\right)!n_{y_i}!}$$

So the likelihood ratio becomes:

$$\frac{\hat{L}(\Omega)}{\hat{L}(\omega)} = \frac{\left(\hat{p}_i^{SAE}\right)^{n_Z}\left(1-\hat{p}_i^{SAE}\right)^{N(Z)-n_Z}\left(1-\hat{q}_i^{SAE}\right)^{(N(G)-N(Z))-(n_G-n_Z)}}{\left(\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{n_G}\left(1-\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{N(G)-n_G}}.$$

This indicates that there is a potential settlement of the issue of small sample size in the scan statistic is solved using SAE estimation results, because the statistic used in scanning using the DE role replaced by statistical estimation results by using the SAE as suggestions on paper Siswantining *et al.* [1].

III. INTEGRATION SAE MODEL ON SCAN STATISTIC

3.1. Area Level Model in Hierarchical Bayes SAE

There are two basic models in the SAE model, i.e. area level and unit-level models. This paper only discusses the area level model. In the small area with the type of area-level models, the auxiliary variable $\mathbf{x_i} = (x_{i1},...,x_{ip})^T$ is only available to area level, where parameter(s), considered as θ_i, assumed to follow this model,

$$\theta_i = \mathbf{x}_i^T\boldsymbol{\beta} + v_i z_i\ , \quad i = 1,...,m \tag{1}$$

Where z_i is a positive constant, β is a vector of regression parameters, and v_i is a random effect are assumed to have a normal distribution with an identical and independent

$$E(v_i) = 0, \qquad V(v_i) = \sigma_v^2 \qquad . \tag{2}$$

To make a conclusion, it is assumed that the estimator directly available to the model

$$\hat{\theta}_i = \theta_i + e_i, \quad i = 1,...,m \tag{3}$$

which e_i is assumed tohave arange of sampling the normal distribution with mean $E\left(e_i \mid \theta_i\right) = 0$ and variance $V\left(e_i \mid \theta_i\right) = \psi_i$.

If the model in(1) and (3) is combined, the models obtained as follows:

$$\hat{\theta}_i = \mathbf{x}_i^T\boldsymbol{\beta} + v_i z_i + e_i, \quad i = 1,...,m \tag{4}$$

That model is a special case of the general linear mixed model (GLMM). Because the response variable used in this paper is a discrete variable, we used logit model to estimate the parameter [4]. Logit model, in the case of Hierarchy mixture, is more computational efficient than the other models [9].

Prior distribution for the β parameters in the model HB is flat priors, and the prior distribution for the σ_v^2 parameter is Inverse Gamma with parameters a and b [4]:

(i) $y_i \mid p_i \sim i.i.d$ Binomial (n_i, p_i)

(ii) $\xi_i = \text{logit}(p_i) = \mathbf{x}_i^T\boldsymbol{\beta} + v_i$; $v_i \sim i.i.d\ N(0,\sigma_v^2)$

(iii) β and σ_v^2 mutually independent ; $f(\beta) \propto 1$; $\sigma_v^2 \square\ IG(a,b),\ a \geq 0,\ b > 0$

From the algorithm Gibbs Sampling and Metropolis Hasting is generated Markov Chain $\left\{p_i^{(k)},.....,p_m^{(k)},\boldsymbol{\beta}^{(k)},\sigma_v^{2\,(k)}; k = d+1,.....,K = d+D\right\}$, used to obtain HB estimators for p_i and the posterior variance for p_i [4]:

$$\hat{p}_i^{HB} \approx \frac{1}{D}\sum_{k=d+1}^{d+D} p_i^{(k)} \text{ dan } V\left(p_i \mid \hat{p}\right) \approx \frac{1}{D-1}\sum_{k=d+1}^{d+D}\left(p_i^{(k)} - \hat{p}_i^{HB}\right)^2 \quad . \tag{5}$$

While DE for small data can be performed by using Microsoft Excel to calculate the proportion of the sample and the sample variance for each of these areas:

$$p_i = \frac{\sum_j y_{ij}}{n_i} = \frac{y_i}{n_i} \text{ and } \hat{V}(p_i) = \frac{pq}{n-1}, \quad q = 1-p$$

For large data, we can use Synthetic Estimator, namely SynE for large area, with the assumption that small area has the same characteristics as the large area [10].

The basic model, such as equation (4), assumes iid small area effect υ_i, but in fact or in applications it is often correlated among υ_i, such as geographical proximity. Therefore, this paper discusses the spatial and non spatial HB SAE, as has also been discussed before [11]. The values of these proportions will act as an input in Satscan.

3.2. Integration of Hierarchical Bayes SAE models in scan statistic

Scan Statistic is using basic LLR. LLR is the ratio of the two likelihood functions.

$$\nabla = \frac{\hat{L}(\Omega)}{\hat{L}(\omega)}$$

$$= \frac{\max_{Z \subseteq \mathbf{Z}} L(\Omega)}{\left(\frac{n_G}{N(G)}\right)^{n_G}\left(1-\frac{n_G}{N(G)}\right)^{N(G)-n_G}\prod_{y_i \in G}\frac{N(y_i)!}{\left(N(y_i)-n_{y_i}\right)!n_{y_i}!}}$$

$$= \frac{\left(\frac{n_Z}{N(Z)}\right)^{n_Z}\left(1-\frac{n_Z}{N(Z)}\right)^{N(Z)-n_Z}\left(1-\frac{n_G-n_Z}{N(G)-N(Z)}\right)^{(N(G)-N(Z))-(n_G-n_Z)}}{\left(\frac{n_Z}{N(Z)}\right)^{n_G}\left(1-\frac{n_G}{N(G)}\right)^{N(G)-n_G}}.$$

By replacing the role of DE to SAE, form LLR is obtained as follows:

$$\nabla = \frac{\left(\hat{p}_i^{SAE}\right)^{n_Z}\left(1-\hat{p}_i^{SAE}\right)^{N(Z)-n_Z}\left(1-\hat{q}_i^{SAE}\right)^{(N(G)-N(Z))-(n_G-n_Z)}}{\left(\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{n_G}\left(1-\frac{\sum_{i=1}^{m}\hat{p}_i^{SAE}}{m}\right)^{N(G)-n_G}}$$

IV. PROBLEMATIC INTEGRATION SAE MODEL INTO SPATIAL SCAN STATISTIC

Analytical and experimental ways can be used to prove the use of HB SAE in spatial scan statistic. To do hypothesis testing, we need to calculate test statistic to make inference statistical. However, we can not find the distribution of the test statistic in closed analytical form. To overcome this problem, we can use Markov Chain Monte Carlo (MCMC) simulation to test the hypothesis [12]. On the other side, experimental way can be done by the simulation and real data application.

4.1. Simulation approach

Simulation is conducted at SAE estimator and simulation on scan statistic, each of which is simulated for a small sample and large sample.

4.1.1. Simulation in small sample : 35 areas

In the simulation using random number generating, set some extreme values, i.e. the value of p = 0.2; 0.4; 0.8; 0.95. Because the simulation of the data distributed Binomial (n, p) the variance of p is pq/n. Each method SAE (from HB1 to HB4) performed simulations with repetition of 20 times to obtain the smallest MSE. To find simulation variance Bayes SAE, performed simulations using HB approach with 4 kinds of weighted,i.e. HB1 (using weighted spatial correlation), HB2 (using weighted spatial nearest neighbor), HB3 (weighted spatial distance), and HB4 (weighted non-spatial) [11]. To stimulate HB SAE Bayes method, use R.2.15.0 software. The results of the variance ratio can be seen in Fig. 1.

In general, it can be said that the HB method simulation is more efficient than the DE because its variance value ratio exceeds 1 in many areas. It can be said that the HB approach on 35 simulated areas has smaller variance (more efficient) than the DE.

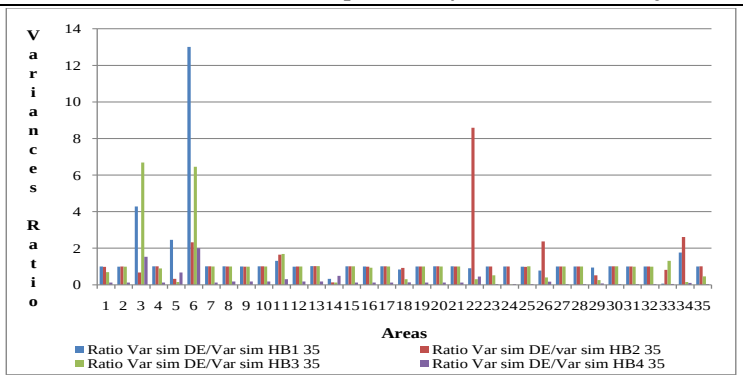

Figure 1. Variance ratio of the DE variance to HB SAE

4.1.2. Simulation scenario SAE for large data : 247 areas

In the 247 simulation areas, using random number generating, set some extreme values, i.e. the value of p = 0.05; 0.2; 0.4; 0.6; 0.8; and 0.95. The variance of p for Synthetic Estimator (SynE) in this problem is pq / n. Fig. 2 is a variance ratio of the simulated HB in 247 simulation areas which indicates that the value of the variance is ratio greater than 1.In fact, SynE variance is 25 times bigger than HB variance. It states that the HB method is more efficient than SynE. Based on these results, we can conclude that HB2 approach is better than the others.

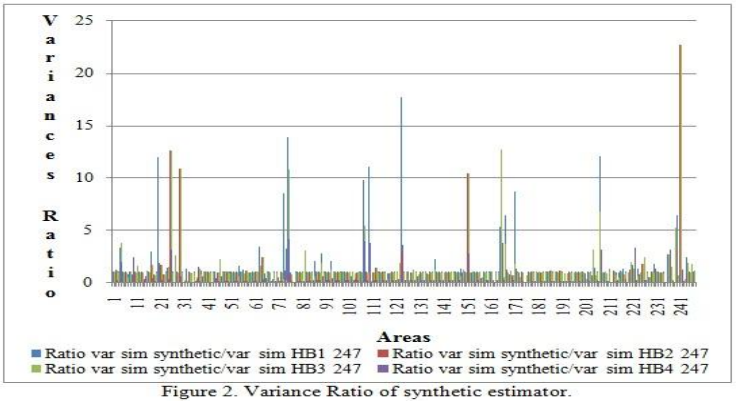

Figure 2. Variance Ratio of synthetic estimator.

4.2. Simulation of Scan Statistic

4.2.1. Simulation Scan Statistic for 35 areas

In his simulation, Ugarte divided the area into 5 zones: the 10th percentile, 25th percentile, 50th percentile, 75th percentile, and 90th percentile [13]. However, in this paper, for the small data of 35 areas, we divide the area into four zones, namely p = 0.2; 0.4; 0.8; and 0.95. Number of sample in every area is between 14 –16. The results of the scan statistic simulation show that the proportion of the area with other areas is very significant. Complete result of this simulation can be seen in Siswantining *et al*. [1].

By using any of the HB approach as shown in Table 1, we find that cluster 1 is the area 24. It could be argued that by using the scan statistic HB SAE consistent in the results obtained. Besides, the relative risk (RR) is higher when we use HB than DE, which means that the risk of individual to be classified in one cluster is bigger than DE.

Table1. Summary results of the HB scan statistic using simulation 35 areas.

	HB1	**HB2**	**HB3**	**HB4**
Ratio O/E	1.71	1.66	1.65	1.70
Relative Risk (RR)	1.75	1.69	1.68	1.73
Log Likehood Ratio (LLR)	6.213983	5.761638	5.717374	6.122081
p-value	0.057	0.073	0.073	0.057
Cluster 1	24	24	24	24
Cluster 2	-	26 (0.073)	34 (0.076)	34 (0.057)
Cluster 3	-	34 (0.073)	-	-

Notes : HB1 (using spatial correlation weighted), HB2 (using spatial nearest neighbor weighted), HB3 (spatial distance weighted), dan HB4 (non spatial weighted)

4.2.2. Simulation Scan Statistic for 247 areas.

Synthetic estimator simulation can be seen in Siswantining *et al.* [1], whereas simulation result of HB method can be seen in Table 2. The approach uses HB3 (weighted spatial distance), it is not statistically significant. The approach uses HB1 (weighted spatial correlation), HB2 (weighted spatial nearest neighbor) and HB4 (weighted non-spatial), all provide statistically significant results. The area that includes cluster 1 (MLC) for HB1, HB2 and HB4 also produces the same area. The approach uses HB1 produces the highest LLR compared with HB2. HB2 has value ratio O/E of the smallest, meaning that the value of observation is closer to the true value, so that HB2 method is better than other methods in the case of HB clusters lot number. From the LLR, for simulation of 35 areas (small data) and simulated 247 areas (large data) to the HB method, it appears that the data simulation LLR has a value 16 times greater in HB1 (weighted correlation), nearly 18 times greater for spatial weighted nearest neighbor, and 14 times greater for non-spatial HB weighted. In the large data it also has a value of ratio O/E is smaller than the small data, meaning that for large sample or large data turns HB2 produce an unbiased estimator (smaller bias) than the simulation with a small data.

Table 2. Summary of simulation scan statistic results with 247 areas.

	HB1	**HB2**	**HB4**
Ratio O/E	1.308	1.26	1.29
Relative Risk (RR)	1.544	1.50	1.51
Log Likehood Ratio (LLR)	96.644435	88.303149	87.643167
p-value	0.001	<0.00000000000000001	<0.00000000000000001
Cluster 1	224,223, 221, 222, 220, 169, 168, 213, 166, 225, 187, 165, 167, 186, 163, 212, 214, 164, 182, 176, 175, 183, 208, 211, 173, 243, 162, 184, 215, 244, 181, 172, 216, 209, 188, 246, 174, 239, 180, 242, 218, 161, 247, 185, 245, 177, 219, 238, 241, 217, 171, 179, 210, 158, 240, 178, 237, 61, 170, 160, 13, 235, 157, 154, 198, 232, 236, 234, 59, 204, 231, 205, 155, 199, 230, 62, 233, 60, 85, 194,228, 193	224, 223, 221, 222, 220, 169, 168, 213, 166, 225, 187, 165, 167, 186, 163, 212, 214, 164, 182, 176, 175,183, 208, 211, 173, 243, 162, 184, 215, 244, 181, 172, 216, 209, 188, 246, 174, 239, 180, 242, 218, 161, 247, 185, 245, 177, 219, 238, 241, 217, 171, 179, 210, 158, 240, 178, 237, 61, 170, 160, 13, 235, 157, 154, 198, 232, 236, 234, 59, 204, 231, 205, 155, 199, 230, 62, 233, 60, 85, 194, 228, 193, 197, 159, 206, 229, 156, 227, 63, 58, 151, 195, 53, 226, 153	224, 223, 221, 222, 220, 169, 168, 213, 166, 225, 187, 165, 167, 186, 163, 212, 214, 164, 182, 176, 175,183, 208, 211, 173, 243, 162, 184,215, 244, 181, 172, 216, 209, 188,246, 174, 239, 180, 242, 218, 161,247, 185, 245, 177, 219, 238, 241,217, 171, 179, 210, 158, 240, 178,237, 61, 170, 160, 13, 235, 157, 154, 198, 232, 236, 234, 59, 204, 231, 205,155, 199, 230, 62, 233, 60, 85, 194, 228, 193
Cluster 2	15 areas (0.021)		

For more details, the summary of the scan statistic results based on simulation data, both small and large sample, is shown at Table 3.

Use of HB estimator based on the amount of data suggests that large sample would increase the value of LLR between 15 to 18 times compared with the small sample. This is consistent with the theory that the likelihood function depends on the size of the sample. Based on the value of the significance of the area indicates rejection of the hypothesis that the large number of samples will reduce the value of up to 4 times significance for HB Bayes approach. From the RR indicates that the large samples will slightly lower the value of RR, i.e., the greater data then the sample will be more difficult to get to a particular cluster. From the ratio O/E indicates that if a large sample, the value of the ratio O/E decreased in HB Bayes approach. It can be said a large sample, then using HB methods of observation tends to reach the actual value (less biased). Especially for weighted spatial distance (HB3), apparently when a large sample, weighted spatial distance does not affect the scan statistic.This means that the greater the distance the more difficult spatiall to enter into MLC. When compared with the synthetic estimator, the HB estimator will produce consistent results for the large data.

Table3.Summary of differences in simulated small and large data on Satscan.

Esti mator	Small sample					Esti mator	Large sample				
	Ratio O/E	RR	LLR	p-value	Cluster		Ratio O/E	RR	LLR	p-value	Cluster
DE	1.44	1.53	16.56	0.0000 0034	1.2	SynE	1.53	2.53	413.5 4	$< 10^{-16}$	15 areas (0.021)
HB1	1.71	1.75	6.21	0.057	1.0	HB1	1.308	1.54	96.64	0.001	15 areas (0.021)
HB2	1.66	1.69	5.76	0.073	1.0	HB2	1.26	1.5	88.30	$< 10^{-16}$	
HB3	1.65	1.68	5.71	0.076	1.0	HB3			ns		
HB4	1.7	1.73	6.12	0.057	1.0	HB4	1.29	1.51	87.64	$< 10^{-16}$	

Note : DE (Direct Estimator), SynE (Synthetic Estimator), HB (Hierarchical Bayes), ns = not significant

4.3.Analysis of Real Data

Analysis of real data is done for poverty data from the Central Bureau of Statistics [14,15]. The data used to analyze the proportion of poor families are from Susenas (National Socioeconomic Survey) and PODES (Village Potential) 2008. Dependent variable is the calorie consumption per household per day (P1) which is 2100 kcal based on Indonesia's line poverty (LP). Every household below 2100 kcal is categorized as poor. Dependent variable obtained from the survey is used as a direct estimator (DE). Dependent variable is taken from the data Susenas 2008. While used as an auxiliary variable percentage of farm families (x_1), the number of families receiving ASKESKIN card in a year (x_2), the number of families electricity users (x_3), the number of schools (elementary, junior high, high school, university) (x_4), the number of families residing in slums (x_5), the number of Certificate of Ability (SKTM) last year (x_6), the number of educational institutions other skills (x_7), and the number of Indonesian Workers (x_8) data taken from PODES 2008 [11].

4.3.1.The results of parameter proportion estimation through HB1 HB4 SAE for calorie consumption per house hold per day (P1) for small data.

Fig. 3 shows the comparison of variance from DE and variance from HB. The results of the variance ratio between the variance HB and variance DE is exceeds 1. It is therefore HB SAE method is said more efficient compared with variance DE. In other words, HB SAE method can be used as a substitute DE as the input of the scan statistic in Satscan.

4.3.2. Scan statistic of small samples based on calorie consumption

Table 4 shows the hotspot detection by Satscan using DE and HB methods. Consumption of calories from village Garahan, Wringin Agung, Sumber Pinang, and Sumbersari was very less compared to other villages based on HB1 (weighted spatial correlation), and HB2 (weighted spatial nearest neighbor). On the other side, Karang Semanding village is the lowest in calorie consumption compared with other villages using HB3 (weighted spatial distance).Thus, HB1 and HB2 are less biased compared with DE and HB3.

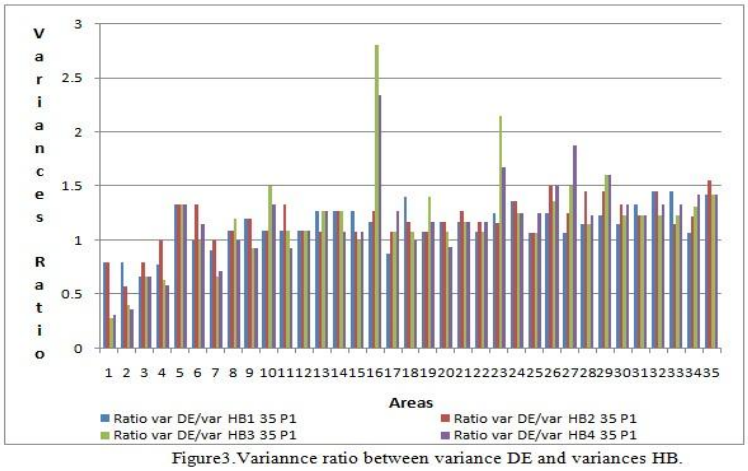

Figure3.Variannce ratio between variance DE and variances HB.

Table 4. Summary of 35 areas analyzed by scan statistic based on calories consumption

Estimator	Ratio O/E	RR	LLR	p-value	Cluster 1	Cluster 1	Cluster 1
DE	1.711	2.062	23.332	0.001	Wringin Agung, Sidodadi, Sumber sari, Ampel, Wringintelu, Sumber Pinang, Garahan	Suren, Pringgowiraw an (5.1979032)	Pace (1.49)
HB1	1.637	1.931	19.276	0.001	Wringin Agung, Sidodadi, Sumber sari, Ampel, Wringintelu, Sumber Pinang, Garahan		
HB2	1.637	1.931	19.27	0.001	Wringin Agung, Sidodadi, Sumber sari, Ampel, Wringintelu, Sumber Pinang, Garahan		
HB3	2.209	2.286	9.1602	0.003	Karang Semanding		

Note : Wringin Agung, etc is the name of area or village

4.3.3. Ratio between Variance synthetic estimator (SynE) and variance HB.

To predict the proportion of poverty in all areas (villages) of Jember, Indonesia, we use SynE. SynE is an estimator which is used to derive an indirect estimator for a small area on the assumption that small area has the same characteristics as the large area [4, 10]. The results of the variance ratio can be seen in Fig.4.It shows that the ratio is more than1, meaning that the variance of HB is more efficient than the SynE.

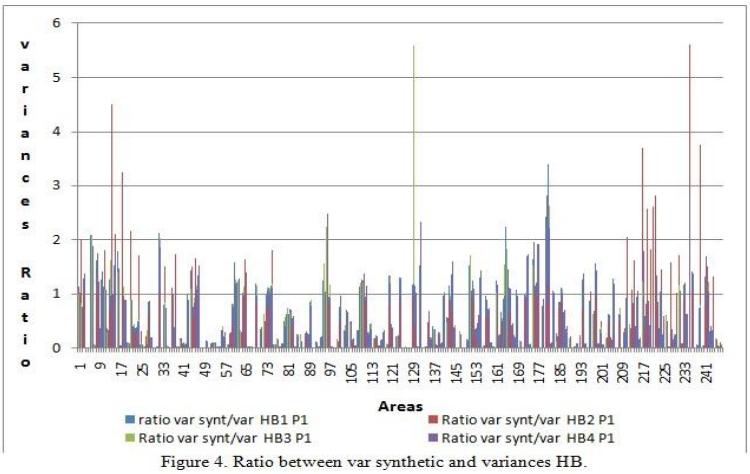

Figure 4. Ratio between var synthetic and variances HB.

4.3.4. The results of scan statistic for large data based on calorie consumption P1.

Table 5 shows a summary of scan statistic result for large sample or large data. It shows that LLR, grouping MLC and p-value of HB1 and HB2 has exactly the same value, but lower than the DE. However, ratio of O/E of HB2 and HB4 are smaller than HB1, which means that the HB2 and HB4 are less biased.

That table also indicates : the relative risk of HB1 is higher than the HB2 and HB4; LLR HB1 greater than HB2 and HB4; HB1 has a greater bias than HB2 and HB4; p-value of HB1 greater than the HB2 and HB4. This means that the p-value of HB1 is worse than the HB2 and HB4. In general, it can be said that, in large sample or large data, HB2 SAE is better than the other HB SAE.

Table 5.A scan statistic Satscan for large sample based on calorie consumption.

Estimator	RatioO/E	RR	LLR	p-value	Cluster 1	Cluster 1	Cluster 1
SynE	1.19	1.45	65.647	$< 10^{-17}$	120 areas		
HB1	1,52	1,66	93,572	$< 10^{-16}$	36 areas	14 areas	9 areas
HB2	1.29	1.51	87.670	$< 10^{-17}$			
HB4	1.29	1.51	87.670	$< 10^{-17}$	82 areas		

Note: Areas mean villages

V. CONCLUSION

In general, it can be said that the HB2 (weighted spatial nearest neighbor) SAE is better than the other HB to substitute the role of DE in determining the hotspots MLC. Analysis using HB2 is said well for small and large sample, both in simulated and real application of the data, where the statistical properties are better than the other HB. The statistical properties have less biased, higher RR, and consistent.

With the limitations that HB SAE that is not a closed form distribution, we suggest further studies using other approaches which have the form of closed form, for example, Bayes model i.e. the EB or for non-Bayes i.e. EBLUP SAE or Spatial EBLUP (SEBLUP).

REFERENCES

[1] T. Siswantining, A. Saefuddin, A.N. Khairil, N. Nunung and M. Wayan. Some Properties of Spatial Scan Statistic Bernoulli Model : Example Simulation for Small and Large Data Using Satscan. *IOSRJRM I, 1(6)*,2012, 21 – 26.

[2] S. Arima, G.S. Dattaand B.Liseoz. Objective Bayesian Analysis of aMeasurement Error Small Area Model.*Bayesian Analysis*,*7*, 2012, 363- 384.

[3] M. Ghosh, and J.N.K. Rao. Small Area Estimation: An Appraisal. *Statistical Science*.,*9*, 1994,255-93.

[4] J.N.K. Rao. *Small Area Estimation*. USA : Wiley-Interscience, 2003.

[5] G.P. Patil and W.L. Myers. *Digital Governance and hotspot Geoinformatics of Biodiversity Measurement, Comparison and Management in the Age of Indicators and Information Technology*. Center of Statistical Ecology & Environmental Statistics, Pennsylvania State University, 2009.

[6] J. Gehrung and Y. Scholz. The application of simulated NPP data improving the assessment of the spatial distribution of biomass in Europe. *Biomass and Bioenergy*, *33*, 2009,712 – 720.

[7] C.R. Rao. Efficient Estimates and Optimum Inference Procedures in Large Samples. *Journal of the Royal Statistical Society. Series B (Methodological), 24*, 1962, 46-72.

[8] A. Roy. *Empirical and Hierarchical Bayesian Methods With Application To Small Area Estimation.* Disertation. Graduate School of The University of Florida, 2007.

[9] R.B.Gramacyand N. G. Polson. Simulation-based Regularized Logistic Regression Bayesian Analysis. *Bayesian Analysis*, *7*, 2012,1-24.

[10] M.E. Gonzalez. Use and Evaluation of Synthetic Estimates. *Proceedings of The Social Statistics Sections.American Statistical Association*, 1973, 33 – 36.

[11] E. Sunandi. Model Spasial Bayes Dalam Pendugaan Area Kecil Dengan Peubah Respon Biner. Tesis S2. Sekolah Pascasarjana IPB, 2011.

[12] M.Kulldorff. A Spatial Scan Statistic. *Commun. Statist*.- Theory Meth., 1997, 26(6), 1481 – 1496.

[13] M.D.Ugarte,T. Goicoa, A.F. Militino. Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, 2009, 53, 2938 – 2949.

[14] Badan Pusat Statistik (BPS, Central Bureau of Statistic) Indonesia. *Data and information poverty. 2008*. BPS Pubs,Jakarta, Indonesia,2008.

[15] Badan Pusat Statistik(BPS, Central Bureau of Statistic) Indonesia. *Indonesian Statistic 2011*. BPS Pubs, Jakarta, Indonesia, 2012.