

Small Area Estimation of Expenditure Per-capita in Banyuwangi with Hierarchical Bayesian and Empirical Bayes Methods

Wirajaya Kusuma¹, Nur Iriawan¹, Irhamah¹

Abstract—One of the economic indicators that are widely used to measure the level of prosperity and welfare is per capita income. However, an accurate income data is difficult to be obtained. In Susenas this data is approached by using data on expenditures per capita. This study employ Hierarchical Bayes (HB) and Empirical Bayes (EB) methods to be applied to Small Area Estimation (SAE) to estimate the expenditure per-capita in Banyuwangi. The results showed indirect estimation using hierarchical Bayes and Empirical Bayes produce RMSE values smaller than the direct estimation. The HB method, on the other hand, produces smaller RMSE value than the EB method. Finally, this research suggests to use HB method to estimate the expenditure per-capita in Banyuwangi rather than direct estimation which is used nowadays.

Keywords—Hierarchical Bayes, Empirical Bayes, Expenditure Per-capita, Small Area Estimation, Root Means Square Error.

I. INTRODUCTION

The development of an area is said to succeed if the welfare level of an entire region includes the level of individual or household. One of the most widely used economic indicators for measuring prosperity and welfare is per capita income. However, accurate income data is difficult to obtain, so in Susenas activities this data is approached through household expenditure data. Household expenditure consisting of food and non-food expenditure can illustrate how the population allocates household needs.

Badan Pusat Statistik (BPS) usually conducts the Survei Sosial Ekonomi Nasional (Susenas) on per capita expenditures, this survey is designed to collect population social data on a relatively broad scope of district / city level. If the results of this survey are used to make predictions at smaller levels, such as sub-districts or villages, then it is likely to produce large biased and variance estimates caused by a less representative sample size to represent the population. Sources of data in a study are usually constrained on relatively small number of samples, one attempt is to increase the number of samples, but often the cost is quite expensive. Another effort that can be done is optimizing the data available with a Small Area Estimation (SAE).

Small Area Estimation (SAE) is a statistical technique to estimate subpopulation parameters whose sample size is small [1]. This estimation method utilizes data from large scale for estimating parameters on a smaller scale that is not sampled. A simple approximation of a small area based on a design-based application is called direct estimation. This direct estimation could not guarantee the accuracy when the

sample size in a small area is small or zero (does not sampled), so the statistics obtained will have a large variance. There is no rule to make a prediction, because it is not represented in the survey [2].

SAE has been implemented in several countries. Ndeng'e [3] from Kenya builds a poverty map in Kenya based on a combination of information from the Welfare Monitoring Survey (household survey) in 1997 with the 1999 Population Census. In Indonesia, Kurnia and Notodiputro [4] perform data simulations to evaluate some SAE standard techniques and apply SAE techniques with indirect methods on poverty data of West Java. Wardani [5] in the case study of per-capita expenditure estimation in Bogor City, the result of his research concluded that empirical Bayes estimation method with Jackknife approach resulted in Relative Root Mean Square Error (RRMSE) smaller than EBLUP method. Other studies that use SAE is Rumiati [6] researched the SAE with unequal probability sampling for binomial and multinomial responses using Empirical Bayes (EB).

Study about method of EB by Fausi [7] estimates per capita expenditure at the sub-district level in Sumenep by differentiating into mainland and island groups and a study with the same data was also performed by Darsyah [8] using the estimation method Kernel-Bootstrap approach. From two studies with different approaches, indirect estimates resulted in more precise estimates than direct estimates based on MSE values.

Various SAE methods have been developed especially regarding model-based methods as an alternative to direct estimation. The methods are Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB), and Hierarchical Bayes (HB). EBLUP method is an estimation of parameters that minimize Mean Square Error by substituting unknown variant components with variant estimators through sample data. In the EB method, the model parameters estimated from marginal distribution of

¹Wirajaya Kusuma, Nur Iriawan, Irhamah are with Department of Statistics, Faculty of Mathematics, Computing, and Data Science, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: wirajaya15@mhs.statistika.its.ac.id; nur_i@statistika.its.ac.id; irhamah@statistika.its.ac.id.

data, and then the inference is based on the posterior distribution estimated. In the HB method, parameter estimation is based on posterior distribution where Parameters are estimated with posterior averages, and precision is measured by its posterior variant [9].

The EB and HB methods are the more general method that can handle continuous data, binary and count. Therefore, this study compared two SAE models that are SAE model with HB method and SAE Model with EB method in expenditure per capita population per sub district in Banyuwangi.

II. LITERATURE REVIEW

A. Small Area Estimation

There are two types of models in SAE ie area level model and unit level model. The area level model is a model based on the availability of supporting data that exists only for a given area level. Let $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$ and the parameter to be expected is θ_i for the i-area assumed to be related to x_i . The supporting data is used to build the $\theta_i = x_i^T \beta + b_i v_i$ model with $i = 1, 2, \dots, m$ and $v_i \sim N(0, \sigma_v^2)$ as the random effect is assumed to be normally distributed. The general model of the area level also assumes that a direct survey estimate of the observed θ_i variable denoted as $\hat{\theta}_i$ is assumed that $\hat{\theta}_i = \theta_i + e_i$. Where the sampling error $e_i \sim N(0, \sigma_e^2)$ with σ_e^2 is known. The combination of the two models will form the equation (1) which is a model of mixed linear level area known as the Fay-Herriot model [10].

$$\hat{\theta}_i = X_i^T \beta + b_i v_i + e_i, \quad i = 1, 2, \dots, m \quad (1)$$

Where b_i is known to be a constant positive value.

B. Hierarchical Bayes

The HB approach to the area-level model in equation (1) is assumed that the prior distribution on the (β, σ_v^2) parameter model for cases with σ_v^2 is known and assumed to be 'flat' prior to β via $f(\beta) \propto 1$, and rewritten according to equation (1) for the HB model..

- i. $\hat{\theta}_i | \theta_i, \beta, \sigma_v^2 \sim N(\theta_i, \tau); \quad i = 1, 2, \dots, m$
- ii. $\theta_i | \beta, \sigma_v^2 \sim N(x_i^T \beta, b_i^2 \sigma_v^2); \quad i = 1, 2, \dots, m$
- iii. $f(\beta) \propto 1,$

For the case of unknown σ_v^2 , equation (2) becomes,

- i. $\hat{\theta}_i | \theta_i, \beta, \sigma_v^2 \sim N(\theta_i, \tau); \quad i = 1, 2, \dots, m$
- ii. $\theta_i | \beta, \sigma_v^2 \sim N(x_i^T \beta, b_i^2 \sigma_v^2); \quad i = 1, 2, \dots, m$
- iii. $f(\beta, \sigma_v^2) = f(\beta) f(\sigma_v^2) \propto f(\sigma_v^2),$

where $f(\sigma_v^2)$ is prior to σ_v^2 .

C. For σ_v^2 Unknown

In the case where σ_v^2 is unknown, use Gibbs sampling for the area-level model for (i) and (ii) of equation (2), assume the priorities β and σ_v^2 in equation (3) with the Gamma distribution with the shape parameter a and scale parameter b .

$$\sigma_v^{-2} \sim G(a, b), \quad a > 0, b > \quad (4)$$

σ_v^2 distributes gamma inverse $IG(a, b)$ with :

$$f(\sigma_v^2) \propto \exp\left(\frac{-b}{\sigma_v^2}\right) \left(\frac{1}{\sigma_v^2}\right)^{a+1} \quad (5)$$

The positive constants a and b are made very small. Gibbs conditional is proven through,

- i. $[\beta | \theta, \sigma_v^2, \hat{\theta}] \sim N_p \left[\beta^*, \sigma_v^2 (\sum_i \tilde{x}_i \tilde{x}_i^T)^{-1} \right]$
- ii. $[\theta_i | \beta, \sigma_v^2, \hat{\theta}] \sim N \left[\hat{\theta}_i^B(\beta, \sigma_v^2), \gamma_i \sigma_v^2 \right]; \quad i = 1, 2, \dots, m$
- iii. $[\sigma_v^{-2} | \beta, \theta, \hat{\theta}] \sim G \left[\frac{m}{2} + a, \frac{1}{2} \sum_i (\hat{\theta}_i - \tilde{x}_i^T \beta)^2 + b \right]$

Where,

$$\tilde{\theta}_i = \frac{\theta_i}{b_i}, \tilde{x}_i = \frac{x_i}{b_i}, \beta^* = \left(\sum_i \tilde{x}_i \tilde{x}_i^T \right)^{-1} \left(\sum_i \tilde{x}_i \tilde{\theta}_i \right) \quad (7)$$

All conditional Gibbs have a closed form so that MCMC samples can be generated directly from the conditional (i) - (iii).

The mean posterior of $(\theta_i | y)$ in the HB approach is used as an estimate of the positional point and variance of $V(\theta_i | y)$ as a measure of diversity. The Gibbs sampler method [11] with the Metropolis Hasting algorithm [12] can be used to find posterior mean and variance. Define MCMC sample as $\{(\beta^{(k)}, \theta^{(k)}, \sigma_v^{2(k)}), k = d + 1, \dots, d + D\}$ with posterior mean and variance.

$$\hat{\theta}_i^{HB} = \frac{1}{D} \sum_{k=d+1}^{d+D} \hat{\theta}_i^B(\beta^{(k)}, \sigma_v^{2(k)}) = \hat{\theta}_i^B(\dots) \quad (8)$$

and

$$\hat{V}(\theta_i | \hat{\theta}) = \frac{1}{D} \sum_{k=d+1}^{d+D} g_{1i}(\sigma_v^{2(k)}) + \frac{1}{D-1} \sum_{k=d+1}^{d+D} \left[\hat{\theta}_i^B(\beta^{(k)}, \sigma_v^{2(k)}) - \hat{\theta}_i^B(\dots) \right]^2 \quad (9)$$

For more efficient estimators can be obtained from the results of closed-form exploration of equation (3) for σ_v^2 is known.

$$\hat{\theta}_i^{HB} = \frac{1}{D} \sum_{k=d+1}^{d+D} \hat{\theta}_i^H(\sigma_v^{2(k)}) = \hat{\theta}_i^H(\dots) \quad (10)$$

and

$$\hat{V}(\theta_i | \hat{\theta}) = \frac{1}{D} \sum_{k=d+1}^{d+D} \left[g_{1i}(\sigma_v^{2(k)}) + g_{2i}(\sigma_v^{2(k)}) \right] + \frac{1}{D-1} \sum_{k=d+1}^{d+D} \left[\hat{\theta}_i^H(\sigma_v^{2(k)}) - \hat{\theta}_i^H(\dots) \right]^2 \quad (11)$$

D. Empirical Bayes

The Empirical Bayes Method (EB) is one of the approaches that can be used on SAE based on the bayess method. The first step taken on the bayess method is to obtain a posterior distribution for the observed parameter denoted $f(\theta_i | y_i, \beta, \sigma_v^2)$, assuming β and σ_v^2 are known. However, in the EB method, the inferences obtained are based on the estimate of the posterior distribution of θ_i by including the estimation values of β and σ_v^2 that are $f(\theta_i | y_i, \beta, \sigma_v^2)$.

Data from variable support (auxiliary variables) are included within the model. Supporting data available only

up to the area level is $\mathbf{X}_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$, then the model for the Empirical Bayes approach using the model in equation (1) is also known as the fay-Herriot model where $v_i \sim N(0, \sigma_v^2)$ and $e_i \sim N(0, \psi_i)$, v_i and e_i are independent. β and σ_v^2 are unknown while ψ_i is assumed to be known.

Let σ_v^2 and ψ_i be symbolized by A and D_i , then the bayes estimator for θ_i by following the bayes model :

- i. $y_i | \theta_i \sim N(\theta_i, D_i)$.
- ii. $\theta_i \sim N(\mathbf{X}_i^T \beta, A)$ is the prior distribution for θ_i , $i = 1, 2, \dots, m$.

Bayes model's explanation is given as follows :

$$f(y_i | \theta_i) = \frac{1}{\sqrt{2\pi D_i}} \exp \left[-\frac{1}{2D_i} (y_i - \theta_i)^2 \right] \quad (12)$$

and

$$\pi(\theta_i) = \frac{1}{\sqrt{2\pi A}} \exp \left[-\frac{1}{2A} (\theta_i - x_i^T \beta)^2 \right] \quad (13)$$

So that

$$f(y_i, \theta_i | \beta, A) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi D_i}} \exp \left[-\frac{1}{2D_i} (y_i - \theta_i)^2 \right] \frac{1}{\sqrt{2\pi A}} \exp \left[-\frac{1}{2A} (\theta_i - x_i^T \beta)^2 \right] \quad (14)$$

for $x_i = (x_{1i}, x_{2i}, \dots, x_{mi})^T$ and $\theta_i = (\theta_{1i}, \theta_{2i}, \dots, \theta_{mi})^T$, Consider two exponential functions regardless of factor $(-1/2)$ on $f(y_i, \theta_i | \beta, A)$,

$$\begin{aligned} & \frac{1}{D_i} (y_i - \theta_i)^2 + \frac{1}{A} (\theta_i - x_i^T \beta)^2 \quad (15) \\ & = \left(\frac{1}{D_i} + \frac{1}{A} \right) \left[\theta_i - \frac{\left(\frac{y_i}{D_i} + \frac{x_i^T \beta}{A} \right)}{\left(\frac{1}{D_i} + \frac{1}{A} \right)} \right]^2 + a_i^* \end{aligned}$$

with a_i^* is constant and does not contain θ_i so that,

$$\begin{aligned} (\theta_i | y_i, \beta, A) & \sim N \left[\frac{\left(\frac{y_i}{D_i} + \frac{x_i^T \beta}{A} \right) \left(\frac{1}{D_i} + \frac{1}{A} \right)^{-1}}{\left(\frac{1}{D_i} + \frac{1}{A} \right)^{-1}} \right] \\ (\theta_i | y_i, \beta, A) & \sim N \left[x_i^T \beta + \frac{A}{A+D_i} (y_i - x_i^T \beta), \left(\frac{AD_i}{A+D_i} \right) \right] \quad (16) \end{aligned}$$

Based on that formula, we get a bayes estimator for θ_i :

$$\hat{\theta}_i^B = E(\theta_i | y_i, \beta, A) = \mathbf{X}_i^T \beta + (1 - B_i)(y_i - \mathbf{X}_i^T \beta) \quad (17)$$

with $B_i = D_i / (A + D_i)$

$$MSE(\hat{\theta}_i^B) = Var(\theta_i | y_i, \beta, A) = AD_i / (A + D_i) \quad (18)$$

When the parameter A is known, the β in the above formula can be estimated by the Maximum Likelihood method. But in reality, A is not known, to estimate the parameter β also uses the Maximum Likelihood Estimation (MLE) or Restricted/Residual Maximum Likelihood (REML) method. Estimator Uses REML consistently despite the violation of the normality assumption [13]. Because A and β are estimated then obtained an estimate of Empirical Bayes as follows:

$$\hat{\theta}_i^{EB} = \mathbf{X}_i^T \hat{\beta} + (1 - \hat{B}_i)(y_i - \mathbf{X}_i^T \hat{\beta}) \quad (19)$$

with $\hat{B}_i = D_i / (\hat{A} + D_i)$

Based on the Bayes method, obtained:

$$MSE(\hat{\theta}_i^{EB}) = \hat{A} D_i / (\hat{A} + D_i) \quad (20)$$

The MSE estimator is underestimated because of the estimation of A and β values. This can be corrected using the jackknife approach. The jackknife's method is one of

the most frequently used methods in surveying because of its simple concept [14].

E. Mean Square Error (MSE) Jackknife

The procedure for applying the jackknife approach in estimate MSE is as follows :

1. Calculate the estimator of \hat{M}_{1i} as :

$$\hat{M}_{1i} = g_{1i}(\hat{\sigma}_v^2) - \left(\frac{m-1}{m} \right) \sum_{l=1}^m [g_{1i}(\hat{\sigma}_{v(l)}^2) - g_{1i}(\hat{\sigma}_v^2)] \quad (21)$$

where $g_{1i}(\hat{\sigma}_{v(l)}^2)$ obtained by deleting the l th observation from the full data set $g_{1i}(\hat{\sigma}_v^2)$.

2. Calculate the estimator of \hat{M}_{2i} as :

$$\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^m \left((\hat{\theta}_{i(l)}) - (\hat{\theta}_i) \right)^2 \quad (22)$$

where $(\hat{\theta}_{i(l)})$ obtained by deleting the l th observation from the full data set $(\hat{\theta}_i)$.

3. Calculate the jackknife estimator of MSE as :

$$MSE(\hat{\theta}_i^E) = \hat{M}_{1i}(\sigma_v^2) + \hat{M}_{2i}(\sigma_v^2) \quad (23)$$

Jackknife methods developed by Jiang, Lahiri, and Wan can be used for all models for SAE, including mismatched models and for cases that are not normally distributed [1].

F. Per-Capita Expenditure

Average expenditure per capita monthly indicates the totals of expenditure of each household member within a month, while the definition of a household is a group of people who inhabit some or all physical buildings and usually live together and eat from one kitchen (BPS 2003). In one household may consist of one, two, or more heads of households. The expenditure per capita is formulated as follows :

$$y = \frac{p}{q} \quad (24)$$

Where y is expenditure per capita, p is the amount of household expenditure a month and q is number of household members.

III. METHOD

A. Exploration of Per-Capita Expenditure Data

The proposed methods are applied to the per-capita expenditure data in Banyuwangi District, which is available in Susenas 2015. There are 23 sub-districts with only one sub district, i.e. Siliragung sub district, is not in sample.

TABLE 1.
DESCRIPTIVE STATISTICS OF PER CAPITA EXPENDITURE.

Statistics	Per-Capita Expenditure (Rp) (x100.000)
N	23
Minimum	4,6992
Maksimum	12,5531
Mean	7,7463
Variance	3,346
Std. deviation	1,8292

Table 1 show that the average expenditure per capita in Banyuwangi is Rp 774.630 where the biggest expenditure is in Banyuwangi sub district with total expenditure per capita equal to Rp 1.255.300,- and the lowest amount of expenditure is Rp 469.900,- that belongs to Kalipuro sub district.

In figure 1 show the distribution pattern of expenditure per-capita in Banyuwangi Regency forms a normal distribution pattern. By using EasyFit v5.5 the normality test result using Anderson-Darling method obtained the value of AD of 0.4389 greater than 2.5018 with ($\alpha = 5\%$) which means that the failure to reject H_0 , which means that variable expenditure per capita is normal distribution.

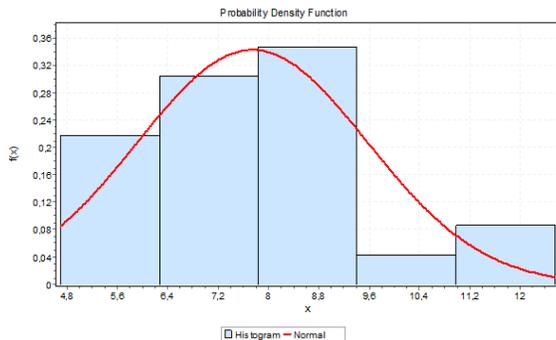


Figure 1. Per-Capita Expenditure Histogram

B. Exploration of independent variable data

Estimation of expenditure per capita is done with the help of five independent variables. Descriptive of the independent variables is presented in Table 2.

TABLE 2.
DESCRIPTIVE STATISTICS OF INDEPENDENT VARIABLES

Variable	Mean	Minimum	Maximum	Std. Deviation
X_1	665	46	3594	735
X_2	16,06	7,38	30,27	7,06
X_3	12194	3395	25040	5468
X_4	18980	6519	36646	7463
X_5	3,02	2	3,75	0,34

Table 2 show that, the average of population density (X_1) in Banyuwangi is 665. This means that the average area of one km² is inhabited by 665 residents. The most densely populated sub district is Banyuwangi Sub-district of 3594 people / km², and the smallest in Kecamatan Tegaldlimo is 46 people / km². For the percentage of poor people (X_2), the average value is 16.06%, which means that average of 100 populations, there are 16 poor people, where the largest percentage of poor people in Kecamatan Licin is 30.27%, and the smallest in Kecamatan Gambiran is 7.38 %. For educational problems, the average number of residents who are schools (X_3) in Banyuwangi is 12194 people, While for welfare problems in terms of the presence or absence of electricity services from PLN (X_4), Banyuwangi residents have subscribed to electricity PLN with an average of 18980 customers in each sub district. The average number of household members (X_5) equal 3.02. This shows that there are 3 household members per household.

To determine whether there is a linear relationship between each independent variable to expenditure per capita in Banyuwangi, then the correlation test is done.

$H_0 : \rho_{xy} = 0$

$H_1 : \rho_{xy} \neq 0$

With a significance of 5% ($\alpha = 0.5$), the results from this correlation test are presented in Table 3.

TABLE 3.
CORRELATION BETWEEN VARIABLE X AND VARIABLE Y

Variable	Pearson correlation	P-Value
X_1	0,561	0,005
X_2	-0,145	0,510
X_3	0,138	0,529
X_4	0,120	0,585
X_5	0,114	0,606

Table 3 show that p-value less than ($\alpha = 0,05$) is population density (X_1) which means that population density has significant linear relation to expenditure per capita in Banyuwangi.

C. SAE Model on Expenditure Per-Capita Using HB Method

The Small Area estimation method with the HB approach is used to estimate expenditure per capita at the sub district level in Banyuwangi District. The estimation is done using the help of Win BUGS software. In estimating the $\hat{\theta}_i$ we first estimate β and σ_v^2 through MCMC method with Gibbs sampling algorithm.

In this study, Markov chain convergence was obtained after burn-in as much as 50 out of 20.000 iterations performed, with number of thin is 10.

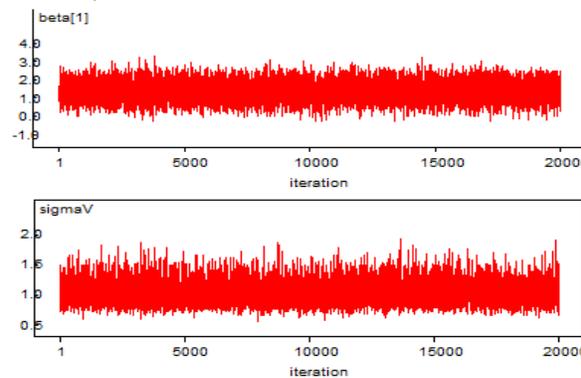


Figure 2. Some Examples of Trace Plot Parameters β and σ_v^2

The result of the trace plot indicates that the Markov chain has converged because the parameter estimation value not formed the up and down pattern.

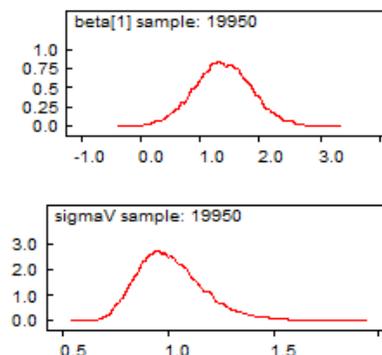


Figure 3. Some Examples of Density Plot Parameters β and σ_v^2

The density plot for parameter β , it shows that the priority density form is relatively normal distribution in accordance with its full conditional function, as well as for parameter σ_v^2 showing a smooth density plot.

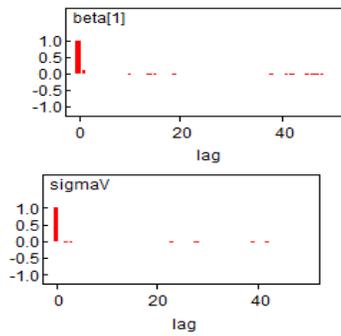


Figure 4. Some Examples of Autocorrelation Plot Parameters β and σ_v^2

The autocorrelation plot has shown a cut off since lag 0 indicating the MCMC sample is independent. From the iteration result on MCMC process, we can get parameter estimation for β and σ_v^2 parameter shown in Table 4.

TABLE 4.
PARAMETERS ESTIMATION OF HB METHOD TO PER-CAPITA EXPENDITURE IN BANYUWANGI

Parameters	Mean	Standard Deviation	Interval	
			2,50%	97,50%
$\hat{\beta}_0$	7,5530	0,5283	6,5230	8,5880
$\hat{\beta}_1$	1,3850	0,4780	0,4413	2,3240
$\hat{\beta}_2$	-0,4099	0,4509	-1,2980	0,4692
$\hat{\beta}_3$	-0,5055	0,2632	-1,0230	0,0100
$\hat{\beta}_4$	-0,3234	0,3000	-0,9124	0,2584
$\hat{\beta}_5$	0,1009	0,3907	-0,6541	0,8680
SigmaV	1,0170	0,1621	0,7580	1,3910

Table 4 show that parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ have a significant effect to the per-capita expenditure. This is shown by the 95% credible interval generated for $\hat{\beta}_0$ and $\hat{\beta}_1$ do not contain a zero value.

TABLE 5.
STATISTIC VALUES OF PER CAPITA EXPENDITURE USING HB METHOD

Statistics	Per-Capita Expenditure (Rp) (x100.000)
N	23
Minimum	5,981
Maximum	11,540
Mean	7,527
Variance	1,365
Standard deviation	1,168

Table 5 show that the average expenditure per capita in Banyuwangi with Hirarcichal Bayes methods of Rp 752.700. where the biggest expenditure is in Banyuwangi sub district with total expenditure per capita equal to Rp 1.154.000,-, and the lowest amount of expenditure is Rp 598.100 that belongs to Kalipuro sub district.

D. SAE Model on Expenditure Per-Capita Using EB Method

For doing the Empirical Bayes model, we first estimate the variance of the random effect factor (A) using the Restricted Estimation Maximum Likelihood (REML) method and is obtained the value of 1,6975. The next is to estimate the value of $\hat{\beta}$ using Maximum Likelihood Estimation.

TABLE 6.
PARAMETER ESTIMATION VALUE OF B

X_i	$\hat{\beta}_i$
X_0	7,7352
X_1	1,0044

The Small Area Estimation model using Empirical Bayes method based on (7) is as follows:

$$\hat{\theta}_i = 7,7352 + 1,0044X_1 + (1 - B_i)(y_i - (7,7352 + 1,0044X_1)) \quad (24)$$

Where $B_i = D_i / (1,6975 + D_i)$.

D_i is the value of sampling error variance that is assumed to be known. The D_i value is estimated from the s_i^2/n_i value which is the ratio between the variance of expenditure per capita value and the number of samples in each sub district.

After the model is obtained, the next step is to estimate the expenditure per capita from the surveyed sub-districts. The following is a general overview of per-capita expenditure estimates using the Empirical Bayes method.

TABLE 7.
STATISTIC VALUES OF PER CAPITA EXPENDITURE USING EB METHOD

Statistics	Per-Capita Expenditure (Rp) (x100.000)
N	23
Minimum	4,8980
Maximum	11,9616
Mean	7,5110
Variance	1,8390
Standard Deviation	1,3560

Table 7 show that the average expenditure per capita in Banyuwangi with Empirical Bayes methods of Rp 751.100, where the biggest expenditure is in Banyuwangi sub district with total expenditure per capita equal to Rp 1.196.160,-, and the lowest amount of expenditure is Rp 489.800 that belongs to Kalipuro sub district.

E. Comparison of Estimation Results Between HB Method and EB Method of Per-Capita Expenditure

After estimating per-capita expenditure using both direct and indirect estimates (HB method and EB method), the next step is to estimate the MSE value from the second result of the estimate. In this, study applied jackknife method to correct the bias of the estimator. Figure 5 shows the MSE value of direct estimation and indirect estimation (HB method and EB method).

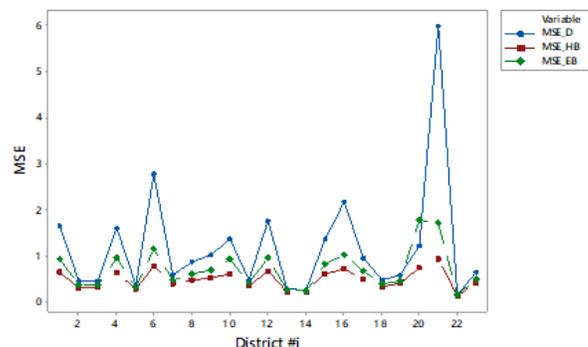


Figure 5. Comparison of MSE Value Between Direct Estimation with Indirect Estimation (HB Method and EB Method)

Based on Figure 5, the MSE of direct estimates tends to be higher than the MSE of indirect estimation. MSE values indirect estimation methods of HB (MSE_HB) and EB (MSE_EB) methods are more precise than direct estimation methods (MSE_D). It can also be seen from the box plot comparison of MSE values from direct estimates and indirect estimates (HB method and EB method) in Figure 6.

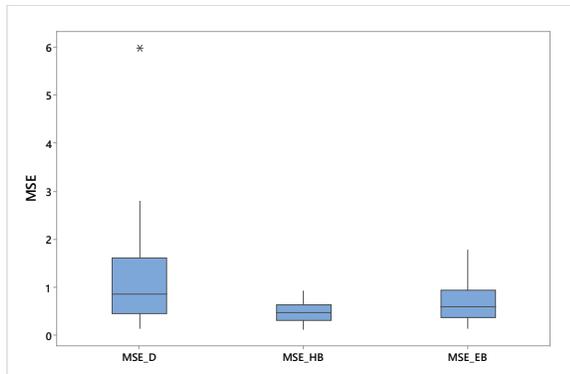


Figure 6. Boxplot of MSE Value Between Direct Estimation with Indirect Estimation (HB Method and EB Method)

Based on Figure 6, the MSE value of indirect estimation is generally smaller than the MSE value of the direct estimation. It can be seen as the MSE value of direct estimation, there is a large value. MSE value that be an outlier is the MSE of the estimated expenditure per capita in Giri Subdistrict while indirect estimation MSE there are no outliers. This indicates that the value of indirect estimation MSE is more precise than the direct estimation.

Evaluation of the three approach of estimation, (direct and indirect approach) can be determined by comparing the value of their RMSE after employing the jackknife's method. A smaller RMSE value indicates that the approach has good accuracy.

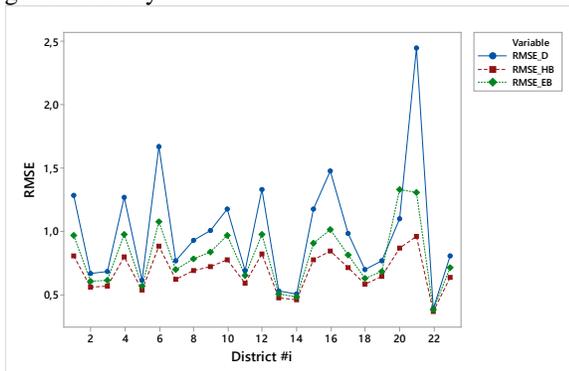


Figure 7. Comparison of RMSE Value Between Direct Estimation with Indirect Estimation (HB Method and EB Method)

Figure 7 shows that the RMSE value of the indirect estimator smaller than the direct estimator, and the RMSE points to the HB method show a smaller value than the EB method in all the sampled sub districts.

In Susenas data of 2015, Siliragung subdistrict is not sampled so it will be estimated using the best model that is HB model. According to Rao [1], the concept of synthetic estimation can be used to estimate the expenditure per capita on the District that is not surveyed, with the

assumption that the behavior between the sub district in Banyuwangi is same (the same value). The expected value of the Small Area Estimation model is $x_i^T \beta$, So expenditure per capita is calculated by the formula :

$$\hat{\theta}_i = x_i^T \hat{\beta} \quad (25)$$

The estimation of expenditure per capita in Siliragung sub-district equal to Rp 805.675,7.

IV. CONCLUSION

Small area estimation of per-capita expenditure using Jackknife's HB approach and Jackknife's EB approach had more accurate results than direct estimation. These two indirect estimation can be employed than the traditional direct one. Thus improving the RMSE value is very significant than the direct estimator, although data has a sampling error variance is not homogeneous and diversity of big sub districts. Between two indirect methods, HB estimation method with Jackknife approach produces smaller RMSE values compared with EB estimation method with Jackknife approach in predicting expenditure per capita per sub district in Banyuwangi.

REFERENCES

- [1] J. N. K. Rao, *Small Area Estimation*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003.
- [2] N. G. N. Prasad and J. N. K. Rao, "The Estimation of the Mean Squared Error of Small-Area Estimators," *J. Am. Stat. Assoc.*, vol. 85, no. 409, p. 163, Mar. 1990.
- [3] G. Ndeng'e, "Geographic Dimensions of Well-Being in Kenya: Where are the Poor?," Nairobi, Kenya, 2003.
- [4] A. Kurnia and K. A. Notodiputro, "Penerapan Metode Jackknife dalam Pendugaan Area Kecil," *Forum Stat. dan Komputasi*, vol. 11, no. 1, pp. 12–16, 2006.
- [5] A. D. Wardani, "Perbandingan Metode Empirical Bayes (EB) dan Empirical Best Linear Unbiased Prediction (EBLUP) pada Pendugaan Area Kecil (Studi Kasus Pendugaan Pengeluaran Per Kapita di Kota Bogor)," Institut Pertanian Bogor, 2008.
- [6] A. T. Rumiati, "Model Bayes untuk Pendugaan Area Kecil dengan Penarikan Contoh berpeluang Tidak Sama Pada Kasus Respon Binomial dan Multinomial," Institut Pertanian Bogor, 2012.
- [7] H. Fausi, "Small Area Estimation Terhadap Pengeluaran per Kapita di Kabupaten Sumenep dengan Metode Empirical Bayes," Institut Teknologi Sepuluh Nopember, Surabaya, 2011.
- [8] M. Y. Darsyah, "Small Area Estimation Terhadap Pengeluaran Per Kapita Pada Level Kecamatan di Kabupaten Sumenep dengan Pendekatan Kernel-Bootstrap," Institut Teknologi Sepuluh Nopember, Surabaya, 2013.
- [9] M. Ghosh and J. N. K. Rao, "Small Area Estimation: An Appraisal," *Stat. Sci.*, vol. 9, no. 1, pp. 55–76, Feb. 1994.
- [10] R. E. Fay and R. A. Herriot, "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *J. Am. Stat. Assoc.*, vol. 74, no. 366, p. 269, Jun. 1979.
- [11] A. E. Gelfand and A. F. M. Smith, "Sampling-Based Approaches to Calculating Marginal Densities," *J. Am. Stat. Assoc.*, vol. 85, no. 410, p. 398, Jun. 1990.
- [12] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *Am. Stat.*, vol. 49, no. 4, p. 327, Nov. 1995.
- [13] J. Jiang, "Linear and Generalized Linear Mixed Models and Their Applications," in *Linear and Generalized Linear Mixed Models and Their Applications*, New York, NY: Springer New York, 2007, pp. 1–50.
- [14] J. Jiang, P. Lahiri, and S.-M. Wan, "A Unified Jackknife Theory for Empirical Best Prediction with M-estimation," *Ann. Stat.*, vol. 30, no. 6, pp. 1782–1810, Dec. 2002.