Spatial Fay-Herriot Model for Estimating Expenditure in Bangka Belitung Province, Indonesia

Ahmad Risal[1]; Siti Muchlisoh[2]

[1] BPS - Statistics Indonesia, Jakarta – ahmadrisal50@gmail.com
[2] Polytechnic of Statistics STIS, Jakarta – sitim@stis.ac.id

Abstract:

In Indonesia, the local government has the right to self-regulate the government affairs known as regional autonomy system. One indicator that assist government in carrying out development of the region is expenditure per capita. STIS, a campus for future employments in Statistics Indonesia has done survey named Praktik Kerja Lapangan (PKL), to estimated expenditure up to village level in Bangka Belitung Province with Small Area Estimation (SAE), Fay-Herriot Model. Fay-Herriot hasn't included spatial influences into model. Meanwhile, there's possibility of spatial influence on expenditure. Hence, this research will estimate SAE method by including spatial influence or Spatial Fay-Herriot (SFH) Model. Data used were raw data PKL, Potential Village and shapefile of Bangka Belitung. Based on Mean Square Error (MSE), MSE value of estimated SFH is smaller than direct estimation. It means, estimation with the SFH provide more precise estimation results than direct estimation. So, SFH is better for further investigation.

Keywords:

Small Area Estimation; Fay-Herriot Model; Spatial Fay-Herriot Model; Expenditure

1. Introduction:

Since January 1st, 2001, the Indonesian government system went through a transformation from centralized to decentralized known as regional autonomy. In this format, local governments have the right, authority, and obligation to regulate and manage their own government affairs and the interests of the local community. Therefore, program development by local governments began to be directed at the level of a smaller area, ranging county, district, or even village. In order for the development program to be carried out more effectively and efficiently, it is necessary to have the availability of data up to the micro scale.

During a present, providing data through surveys conducted by Statistics Indonesia – Badan Pusat Statistik (BPS) as the official provider of statistics is not entirely capable of directly estimating small areas to districts and villages yet. If an estimation is made directly to the level of a small area, it will produce a low precision value because the number of samples is insufficient. A solution to this problem is to provide a budget to increase the number of samples, so that the survey design carried out is able to provide a statistical output of small areas.

The next problem arises because of budget constraints, so information about an indicator is not available evenly in a small area. By paying attention to the information needs of small areas and seeing the conditions of limited resources, a statistical method that is able to meet the availability of information needs to be applied even though the resources owned are limited. This method is known as the Small Area Estimation.

Small Area Estimation (SAE) is a technique that is often used and is expected to produce better precision to estimate smaller areas. Estimates in SAE are based on the model, so additional information is needed from variables that have a relationship with the variables being observed which are called covariates. The additional information is in the form of previous census data or administrative data from the area concerned. All this additional information must be related to the parameters observed (Rao, 2003).

Poverty is a problem faced by every region in Indonesia. In addressing these problems made an effort to make the estimation of the poorer regions. This estimation is expected to help the government to reduce poverty more accurately. The main indicator used by the BPS in determining someone is said to be poor or not is the average per capita expenditure.

So far, BPS has obtained per capita expenditure data per region from a survey named National Socio-Economic Survey (SUSENAS). However, published results have limitations in the number of samples for estimation in small areas. The number of samples used can only estimate up to the district level or even the province. This situation does not meet the needs of the regional government, which requires local statistical data to the lower area level.

Politeknik Statistika STIS as a campus for future employments in BPS, already done a survey named Field Work Practice – Praktik Kerja Lapangan (PKL) in 2017. The aim of the field work was conducted on the study of poverty and income distribution in the Bangka Belitung Province using Fay-Herriot model. The use of Fay-Herriot model to estimate parameters has not incorporated the spatial influence into the model. On the other hand, it does not rule out the possibility of spatial influence on expenditure per capita in a region. Therefore, Fay-Herriot model was developed by incorporating spatial influences into the model. The Fay-Herriot model estimator takes into account the random influence of the spatial correlated area known as the Spatial Fay-Herriot model. Spatial Fay-Herriot model can improve the variety structure of the estimation model for small areas that have spatial correlations between areas. The model used in the area-based on Spatial Fay-Herriot model is because the spatial modelling included in the SAE model is modelling the spatial data type area. The estimator of Spatial Fay-Herriot model has been used by Pratesi & Salvati (2007) by entering a weighting spatial matrix, spatially nearest neighbours into the Spatial Fay-Herriot model.

Based on that, this study examines the estimated expenditure per capita in Bangka Belitung Province at 2017 by using direct estimation and Spatial Fay-Herriot model. After that based on the results of the estimation, carried out a comparison from that two method of estimation method to determine the most appropriate estimation method. so that we hope to produce estimates of expenditure per capita to the level of small areas with good precision can be achieved.

Several studies have been conducted relating to the Spatial Fay-Herriot Model. Matualage (2012) estimated the per capita expenditure in Jember Regency, East Java Province using the Spatial Fay-Herriot model. The results showed that the per capita expenditure of each village in Jember Regency obtained with estimators of Fay-Herriot model was more diverse when compared to village per capita expenditures produced with the Spatial Fay-Herriot model with an average per capita expenditure for each village with the Fay-Herriot model higher than the average expenditure per capita village with the Spatial Fay-Herriot model. In addition, the Spatial Fay-Herriot model is better used to estimate the average village/district expenditure per capita in Districts Jember with the participation variables compared to the direct estimator method or the the Fay-Herriot model. Whereas when compared to the direct estimator, the Fay-Herriot model estimator RRMSE value is not much different.

2. Methodology:

The data used in this study are secondary data from various sources. The following are details of the data used:

- Data used as direct estimation, namely the average per capita expenditure of each sub-district and village in the Bangka Belitung Islands Province with 7053 samples. This data is estimated from per capita expenditure for each region selected as a sample in Field Work.
- Village Potential (PODES) 2014 as covariates in Small Area Estimation. The variables chosen are as follows:
  - The main source of income for the majority of the population
    - District: agriculture (X_1); mining and quarrying (X_2); processing industry (X_3); wholesale/retail trade and restaurants (X_4); transportation, warehousing and communication (X_5); and services (X_6).
    - Village: mining and quarrying (d_1); wholesale/retail trade and restaurants (d_2); transportation, warehousing and communication (d_3); and services (d_4)
  - Number of hospital facilities (X_7)

- Number of village unit cooperatives – Koperasi Unit Desa (KUD) that are still active/operating (X_8)

➢ Queen type of spatial weighting matrix originating from the shapefile of Bangka Belitung Islands Province.

3. Small Area Estimation:

Small Area Estimation (SAE) is an indirect estimation method that combines survey data with other supporting data such as from the previous census. The supporting data must contain variables with the same characteristics as the survey data. This is done so that smaller areas can be estimated and provide a better level of precision.

Based on data availability, SAE is grouped into two types, namely area-level and unit-level. The study in this paper used area-level of SAE. This is because the supporting data (covariates) available only reaches the area level. The area level model connects the small area direct estimator with supporting data from other domains for each area, that is $\mathbf{x_i^T} = \left(x_{1i}, \dots, x_{P_i}\right)$. The parameter of small area that will be estimated is θ_i. The linear model that explains these relationships is:

$$\theta_i = \mathbf{x_i^T}\boldsymbol{\beta} + z_i v_i \qquad i = 1,2, \dots, m \tag{1}$$

Where:

$\beta = \left(\beta_0, \dots, \beta_p\right)^T$ is SAE regression coefficient estimates

z_i = positive constant

v_i = random effect area, assumed $v_i \sim$ iid $N(0, \sigma_v^2)$

m = number of observations (area)

In making conclusions about the population under equation 1, it is assumed that the direct estimation value is known $\hat{\theta}_i$ and can be written as follows:

$$\theta i = \theta i + ei \qquad i = 1, 2 \dots, m \tag{2}$$

where e_i is a sampling error assumed to be $e_i \sim N(0, \Psi_i)$.

The SAE model for area level consists of two levels of the model component, namely the indirect estimation model component showed by equation (1) and the direct component showed by equation (2). If the models in equations (1) and (2) are combined they will form the following equation:

$$\hat{\theta}_i = \mathbf{x_i^T}\boldsymbol{\beta} + z_i v_i + e_i \qquad i = 1,2, \dots, m \tag{3}$$

The above model has not included the effect of spatial correlation in it. The SAE model which included the spatial correlation between areas was first introduced by Cressie (Cressie 1991 referred to in Rao 2015), assuming spatial dependence following the model of Conditional Autoregressive (CAR) process. The SAE model was later developed by other researchers, including Pratesi and Salvati (2007) by assuming that the spatial dependence included in the error component of random effect area follows the model of Simultaneous Autoregressive (SAR) process. The SAR model (Spatial Fay-Herriot Model) itself was first introduced by Anselin (Anselin 1992) where area v random influence vectors is as follows:

$$\mathbf{v} = \rho \mathbf{W} \mathbf{v} + \mathbf{u} \tag{4}$$

Where:

ρ = spatial Autoregressive coefficient

W = spatial weighted matrix

v = random effect area

u = error vector of random effect area

4. Result:

4.1. Direct Estimation of Per Capita Expenditure

Direct estimates of the per capita expenditure can only be done in regions that have at least a sample of the Field Work Practices (PKL STIS). Bangka Belitung Islands Province has a total of 43

sub-districts and 387 villages. PKL STIS just have sample on 42 sub-districts and 135 villages with number of samples are 7053 samples. So, the direct estimation is done only on that sample.

Table 1. Descriptive Statistics on Per Capita Expenditure Based on Direct Estimation

Statistic	Subdistrict	Village
Number of observations	42	135
Mean	IDR.1,238,894	IDR.1,208,024
Std. Deviation	IDR.232,380	IDR.285,505
Minimum	IDR.880,850	IDR.512,963
Median	IDR.1,168,565	IDR.1,184,435
Maximum	IDR.1,788,768	IDR.1,877,401

Based on the results of direct estimation, per capita expenditures for sub-districts and villages were obtained as shown in table 1. At the sub-district level, sub-districts that have the lowest average per capita expenditure is the districts in South Bangka District, Tukak Sadai and sub-districts that have the highest average per capita expenditure is Rangkui in Pangkalpinang City. At the village level, village which has the lowest per capita expenditure average is Peradong village, West Bangka District and the village which has the highest average per capita expenditure is Masjid Jamik Village, Pangkalpinang City.

4.2. Spatial Fay-Herriot of Per Capita Expenditure

In estimating the Spatial Fay-Herriot model, a spatial weighted matrix is needed. In this study, the spatial weighted matrix used is the Queen-type which has been standardized in the row. Spatial weighted type queen contiguity takes into account the proximity of a region to another region. If the area is right around the observation area, then the area is given code one (1) whereas if it is not right around the observation area then the area is given code zero (0). The Illustration is on Figure 1.

Figure 1. Illustration of Queen Contiguity Matrix

Queen		
1	1	1
1	i	1
1	1	1

After obtaining a spatial weighting matrix, a spatial autocorrelation test will be conducted using the Moran's I. Test. A summary of the results of the Moran's I test is presented in Table 2. The hypotheses used are as follows:

H0: I = 0 (no spatial autocorrelation)

H1: I ≠ 0 (there is spatial autocorrelation)

Based on the test results of the Autocorrelation in Table 2, it is shown that using the queen type spatial weighting matrix shows there is spatial autocorrelation in the random effect area at the village level.

Table 2. Spatial Autocorrelation Test Results on Average per Capita Expenditure

Statistics	Sub-District	Village
Moran's I	-0.1404	-0.0074
Z count	-1.0151	1.5197
P-value	0.8450	0.0643*

*sign at level significant of 10%

The next step is applying the Spatial Fay-Herriot model with the REML procedure. The covaites selection method that will be used to estimate the per capita expenditure is done by backward elimination. The result of using backward elimination at the sub-district level shows that from the 8 covariates used, only the remaining 3 variables are left, namely variables X_1, X_5 and X_6. Whereas for

the village level, from all of the covariates used, only remaining five variables, they are variables d_1, d_2, d_3, d_4 and x_7. A summary of the estimation from covariates is presented in Table 3.

Table 3. Estimated coefficient of covariates in the Spatial Fay-Herriot Model

Sub-district			Village		
Variable	$\hat{\beta}$	p-*value*	Variable	$\hat{\beta}$	p-*value*
(Intercept)	1295961	0.0000	(Intercept)	1061677	0.0000
x_1	-23180	0.0000	d_1	186360	0.0001
x_5	85008.2	0.0000	d_2	610388	0.0000
x_6	92033.5	0.0000	d_3	494445	0.0000
			d_4	378201	0.0023
			x_7	235344	0.0359

The estimation of random effect variance (σ_v^2) is carried out by the maximum likelihood (MLE) method. The random effect variance calculation using the R program produces a random effect variance value of 3918068286 at the sub-district level. At the village level, the results of the random effect variance are 34761096201. Besides estimating σ_v^2, the Spatial Fay-Herriot model also estimates the coefficient of spatial autoregression (ρ). At the sub-district level, ρ value is -0.981 and at the village level the value is 0.1961.

Statistical values from the estimation of sub-district and village level per capita expenditure is presented in Table 4.

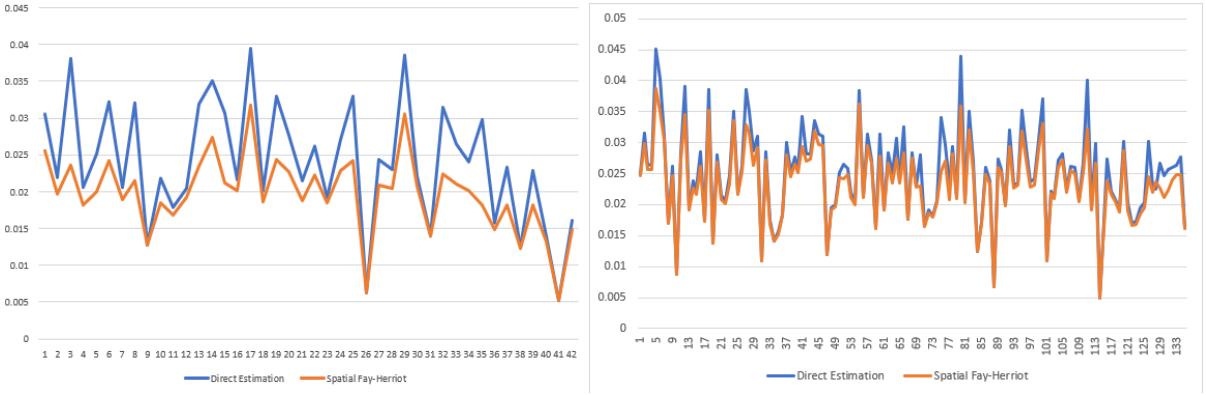
Table 4. Descriptive Statistics on Per Capita Expenditure Based on Spatial Fay-Herriot Model

Statistics	Sub-District	Village
Number of observations	42	135
Mean	IDR.1,240,177	IDR.1,192,000
Std. Deviation	IDR.212,025	IDR.266,385
Minimum	IDR.854,922	IDR.537,006
Median	IDR.1,175,266	IDR.1,158,716
Maximum	IDR.1,723,181	IDR.1,834,688

4.3. Comparison of MSE and RRMSE

To see the estimation method that gives the best estimation results, a comparison of the estimation results between the direct estimation method and the Spatial Fay Herriot Model will be compared. Comparisons will be made through Mean Square Error (MSE). After that, MSE value used to count Relative Root Mean Square Error (RRMSE) value. In this section a comparison is made for each sub-district and village.

Figure 2. Comparison of RRMSE of expenditure per capita sub-district (a) and village (b)



(a) (b)

Based on Figure 2, it can be seen at the sub-district and village levels, the RRMSE value of the Spatial Fay-Herriot model estimation method is better than the direct estimate. Therefore, it can be said that the Spatial Fay-Herriot model is the best method of estimating the value of expenditure per capita at the sub-district and village level.

5. Conclusion:

Based on the results discussed earlier, it can be seen that at the sub-district level, the largest value of per capita expenditure is in Rangkui sub-district that is IDR.1,788,768 and the lowest is in Tukak Sadai sub-district. While at the village level, the largest value of per capita expenditure is in Masjid Jamik village that is IDR.1,877,401.and the lowest value is in peradong village. After that, based on the RRMSE value we can see that the RRMSE value of the estimated Spatial Fay-Herriot model is smaller than the direct estimation. This shows that estimation with the Spatial Fay-Herriot model provide more precise estimation results than direct estimation. So we can say that the Spatial Fay-Herriot model (Small Area Estimation) is better for further investigation.

References:

Anselin L. (1992). *Spatial econometrics: method and models*. Boston: Kluwer.

Matualage D. (2012). *Metode prediksi tak bias linier terbaik empiris spasial pada area kecil untuk pendugaan pengeluaran per kapita: studi kasus Kabupaten Jember Provinsi Jawa Timur* [tesis]. Bogor: Institut Pertanian Bogor.

Rao JNK. (2003). *Small Area Estimation*. New York: John Wiley and Sons.

Rao JNK, Molina I. (2015). *Small Area Estimation, second edition*. New York: John Wiley and Sons.

Pratesi M, Salvati N. (2008). *Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Statistical methods and applications*, Stat. Meth. & Appl.