APPLICATION OF SMALL AREA ESTIMATION OF MULTIVARIATE FAY-HERRIOT MODEL (MFH) FOR THE AVERAGE OF PER CAPITA EXPENDITURE IN VILLAGE LEVEL

Hasbi Nuryadin, Budi Susetyo, Kusman Sadik

Abstract— The National Socio-Economic Survey samples are designed to produce estimates of paramaters of planned domains (provinces and districts). The estimation of unplanned domains (sub-districts and villages) has its limitation to obtain reliable direct estimates. One of possible solutions to overcome this problem is employing small area estimation technique. Univariate Fay Herriot (UFH) model become a popular method to produce reliable characteristic of interest in small area. However, for the cases of some related multiple characteristics of interest can be used by Multivariate Fay-Herriot (MFH) model. This research focused on comparing three estimations methods (direct estimation, UFH model and MFH model) based on the average expenditure per capita on type of food and non food at villages level in Bogor District, West Java. Using data taken from National Sosio-Economic Survey, during period of March 2016. The best method chosen by smallest root mean square errors (RMSE).As a result we managed to identify that the best estimate was given by MFH model which had the smallest RMSE.

Index Terms— small area estimation, multivariate, expenditure

1 INTRODUCTION

The Central Bureau of Statistics (BPS) is a non-ministerial government agency directly responsible to the President. Its existence and main task and function is regulated in Law No. 16 of 1997 on Statistics. Under the law, one of the roles that must be run by BPS is to provide data for the government and the public. This data is obtained from self-censuses or surveys as well as from other department or government agencies as secondary data. Census and routine surveys are conducted by BPS every year and periodically. Censuses conducted by BPS are Population Census (SP), Agricultural Census (ST), Economic Census (SE) and Village Potential (PODES), while surveys conducted include National Socio-Economic Survey (SUSENAS), National Labor Force Survey (SAKERNAS), Indonesia Demographic and Health Survey (SDKI), Intercensal Population Survey (SUPAS) and many other types of surveys. All these things, both census and surveys aimed to assist the governments in the development planning, monitoring and evaluation / assessment of development or other important policies.

The survey conducted by BPS is generally designed to provide statistical information for the national, provincial and district/city level. For example, Susenas is a survey activity aimed at collecting information/data on population, health, education, family planning (KB), housing and consumption and expenditure that can only provide statistics to the district/city level. Whereas for statistical information at smaller levels such as sub district or villages have not been able to be presented by this survey because of the small sample size and big variety of values. Moreover, not all villages are the examples in Susenas, so the estimation cannot be conducted in the non-selected villages as an example.

The need for statistical information to the sub-district and village levels is very important for the current policy stakeholders, since the policy programs undertaken have directly touched to the sub district or village. For example, the average expenditure of per capita household consumption per village used to calculate village-level poverty indicators is a statistic whose availability needed for policy programs in order to reduce the number of village poverty.

According to Giusti et.al possible solutions for the problem of providing statistical information at a smaller level such as sub-districts or villages is by increasing the number of samples so that direct estimation become more reliable or by using small area estimation techniques [1]. Increasing the number of instances will cause the cost to be expensive and time consuming. The application of small area estimation techniques has the potential to overcome these problems on the basis of existing surveys.

Small area estimation is a model-based indirect estimation technique that is conducted by utilizing additional information from the surrounding area. Such additional information is in the form of prior census data or administrative data of the concerned area. All of this additional information should be related to the parameters observed [2]

There are several methods of small areas estimation that have been used, including Best Linear Unbiased Prediction (BLUP), Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB), and Hierarchical Bayes (HB). The BLUP

- *Hasbi Nuryadin is currently pursuing masters degree program in Department of Statistics in Bogor Agricultural University, Indonesia, PH +62-85248998990. E-mail: hasbi.nuryadin@gmail.com*
- *Budi Susetyo is a Lecturer, Department of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: erfiani_ipb@yahoo.com*
- *Kusman Sadik is a Lecturer, Department of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: bagusco@gmail.com*

and EBLUP methods are used to be limited to continuous response variables, whereas EB and HB can be used more widely for both continuous respond variable or binary and count data. Based on the availability of data of participant variable, the model of small area estimation is divided into the area level model and unit level model, the area level model is often called the Fay-Herriot model.

Several studies of small area estimation related to the average of per capita household expenditure have been widely used, such as Ningtyas using EBLUP method to estimate per capita expenditure in Brebes Regency [3]. Fausi used the Emphirical Bayes (EB) method to estimate per capita expenditure in Sumenep [4] and Wardani districts compared EB and EBLUP methods to generate per capita expenditure estimation in Bogor City [5]. The research concluded that EB and EBLUP methods resulted in better estimators than direct estimators.

The three studies are still using one response variable (univariate) which is average of per capita expenditure. According to Datta et al. the estimation efficiency resulting from the calculation using multiple multivariate responses correlated better than using a single variable [6]. As it is known, the average of per capita expenditure itself is the sum of the per capita expenditure of food and non-food expenditures, which these two variables are likely to correlate each other [7]. So this study uses both variables to estimate the average of per capita expenditure of food and non food per village of Bogor Regency in 2016 using Multivariate Model Fay Herriot (MFH).

The purpose of this research is to compare the estimation results with direct estimation method, Univariate Fay Herriot (UFH) model and Multivariate Fay Herriot (MFH) model in the case of estimation of average household expenditure per capita of food and non-food at village level in Bogor Regency 2016.

2 Research Methode

2.1 Data

The data used in this study comes from BPS. The response variable in this research is the average of per capita expenditure of food and the average of non-food per capita expenditure is derived from the SUSENAS data in Bogor Regency, March 2016. As the participant variable in each village is obtained from the data of PODES of Bogor Regency 2014 where the variables selection of the available variables refers to the results of variables that have a significant contribution to the per capita consumption expenditure model of food and non-food. Table 1 is the detail of the variables used in the research.

TABLE 1
RESEARCH VARIABLES

Variables	Data source
per capita expenditure on food (Y1)	Susenas 2016
per capita expenditure on non food (Y2)	Susenas 2016
Percentage of number of families of electricty users (X1)	Podes 2014
Percentage of agriculturel families (X2)	Podes 2014

2.2 Data Analysis Procedures

The stages of analysis in this research are as follows:

1. Conducting data exploration, i.e. checking the data distribution and correlation on the average data of per capita expenditure of food and non-food per village.
2. Estimating the average of per capita consumption expenditure on food and non-food per capita consumption expenditure by direct estimation method.
3. Estimating per capita consumption expenditure per month of food and non-food and Root Means Square Error (RMSE) with EBLUP method using UFH Model.
4. Estimating per capita consumption expenditure per month of food and non-food and Root Means Square Error (RMSE) with EBLUP method using MFH Model.
5. Selecting the best model by comparing the RMSE value on each estimation model.

3 Result and Discussion

3.1 Data Exploration

The population of Bogor Regency based on the population projection of 2016 as much as 5,715,009 people which consists of 2,920,288 male and 2,794,721 female population. Compared to the population of 2015, Bogor Regency experienced a population growth of 4.68 percent. The population density reached 2.146 people / km2 with the highest population density in Bojonggede sub district with the density of 11,389 people / km2 and the lowest density in Tanjungsari sub district was 399 persons / km2.

The average of per capita expenditure of Bogor Regency in 2016 is Rp 1,090,973. This value is higher than in 2015 which reached Rp 906.682. The percentage comparison of per capita expenditure average on food and non-food in 2016 shows that the average of per capita expenditure on food is greater than the average of per capita expenditure on non food of 52.93 percent versus 47.07 percent, indicating that the pattern of public expenditure in Bogor Regency has already lead to the fulfillment of secondary and tertiary needs. The following description of the average of per capita expenditure in Bogor District in 2015-2016:

TABLE 2
AVERAGE OF PER CAPITA EXPENDITURE

average of per capita expenditure	2015	2016
Food	424,909 (46,86%)	513,506 (47,07%)
Non Food	482,773 (53,14%)	577,467 (52,93%)
Total	906,682 (100%)	1,090,973 (100%)

Susenas in March 2016 for Bogor Regency includes 1,166 sample households spread in 111 villages from 434 villages in 39 sub-districts from a total of 40 districts. The average number of samples per village / sub district amounts to 10 households. The variables of concern in this research were the average of per capita expenditure on food (Y1) and the average of non-food per capita expenditure (Y2) per village in Bogor Re-

gency. Furthermore, the correlation value between the two variables is 0.818 with p-value 0.000.

3.2 Estimation Results of Per Capita Expenditure Average on Food and Non-food

The results of the regression coefficients and the various random components of each variable are presented in Table 3.

TABLE 3
ESTIMATES REGRESSION COEFFICIENTS AND VARIOUS RANDOM COMPONENTS OF UFH AND MFH MODELS

Variabel	Coefficient	UFH	MFH
Y1	$\hat{\beta}_0$	13.110	13.123
	$\hat{\beta}_2$	-0.007	-0.008
	$\hat{\sigma}^2_{v1}$	0.075	0.033
Y1	$\hat{\beta}_0$	9.493	11.411
	β_1	0.035	0.016
	$\hat{\beta}_2$	-0.011	-0.013
	$\hat{\sigma}^2_{v2}$	0.142	0.087

Table 4 presents the application results of direct estimation methods, UFH and MFH methods on the average data of per capita expenditure on food and non-food per village in Bogor Regency. Based on Table 4, the statistical values of per capita expenditure on both food and non-food with direct estimation methods are greater than the UFH and MFH models, except for the minimum values. The standard deviation and range values on direct estimation are also greater than UFH and MFH, indicating that direct estimation results are more widespread than UFH and MFH estimation.

TABLE 4
STATISTICAL VALUES OF PER CAPITA EXPENDITURE ON FOOD AND NON FOOD (000 RP)

Statistic	Methods		
	Direct	UFH	MFH
	Per Capita Expediture on Food		
Average	485.635	449.410	444.421
Standard Deviation	196.381	96.908	79.196
Minimum	179.297	224.829	237.266
Median	452.082	445.875	447.966
Maximum	1,208.962	932.970	766.798
	Per Capita Expediture on Non Food		
Average	467,782	407,123	400,817
Standard Deviation	426,408	173,385	169,454
Minimum	101,246	110,669	141,823
Median	378,455	394,058	374,279
Maximum	3.673,281	1,200,588	1,288,336

3.3 Method Comparison

The selection of the best model to estimate the average of per capita expenditure on food and non-food per village is conducted based on the resulted RMSE. The best model is the model with the smallest RMSE. The comparison of direct estimation of RMSE, UFH and MFH from the average of per capita expenditure of food per village can be seen in Table 5.

Based on Table 5 and Figure 1, it is seen that the average RMSE value of food per capita expenditure on food is smaller in the MFH small area modeling method compared to the RMSE of UFH small area modeling and direct estimations. On average, the RMSE on the direct estimation method of 0.136 decreased to 0.052 in the UFH small area modeling method then decreased again to 0.023 on the MFH method. Furthermore, the RMSE value of non-food per capita expenditure is also the same as the average RMSE value of per capita expenditure on food which shows the RMSE method of MFH is smaller than the UFH method and direct estimation. The average RMSE value on the direct estimation of 0.130 decreased to 0.071 on the UFH method and decreased again to 0.024 on the MFH method.

In zone establishment step was also done an estimation of log likelihood ratio value. This value will later be the reference in determining hotspots from the existing hotspot candidates. Log likelihood ratio value is sorted from the biggest to the smallest.

TABLE 5
RMSE OF PER CAPITA EXPENDITURE ON FOOD AND NON FOOD

Statistic	Methods		
	Direct	UFH	MFH
	Per Capita Expediture on Food		
Minimum	0.003	0.003	0.003
Q1	0.080	0.041	0.022
Average	0.136	0.052	0.023
Median	0.206	0.050	0.023
Q3	0.235	0.062	0.026
Maximum	2.068	0.086	0.036
	Per Capita Expediture on Non Food		
Minimum	0.004	0.004	0.003
Q1	0.073	0.050	0.021
Average	0.130	0.071	0.024
Median	0.206	0.073	0.023
Q3	0.221	0.094	0.025
Maximum	2.240	0.149	0.038

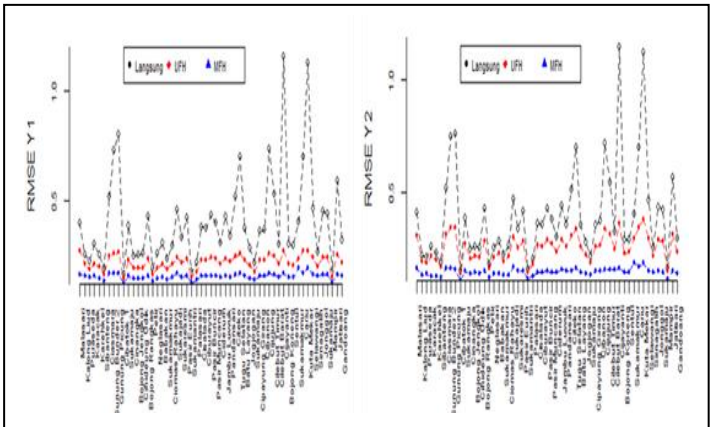

Fig. 1. RMSE Plot of per capita expenditure on food and non food

The decrease in RMSE values on the UFH method indicates that the random effects of the area and the participant variable work to improve the direct estimation based solely on survey data, and the impairment of RMSE values on the MFH method improves the estimation result of the UFH method based solely on survey data using a single response regardless of correlation with other response data. It can be said that the overall method of small areas estimation of MFH is more reliable than direct prediction and small areas estimation of UFH.

The evaluation of estimated quality is conducted by looking at the Relative Root Mean Square Error (RRMSE) generated by all three methods. Figure 2 presents the RRMSE distribution of per capita expenditure estimation on food and non-food generated by all three methods. In general the MFH method is seen to have a median and smaller range. From the results in Table 6, it is also seen that the quality of the MFH small area estimation generates a smaller RRMSE average value of per capita income on food and non-food per village than the two other estimation methods. This indicates that the small areas estimation through the MFH method approach in this case has better quality than the direct estimation and UFH.

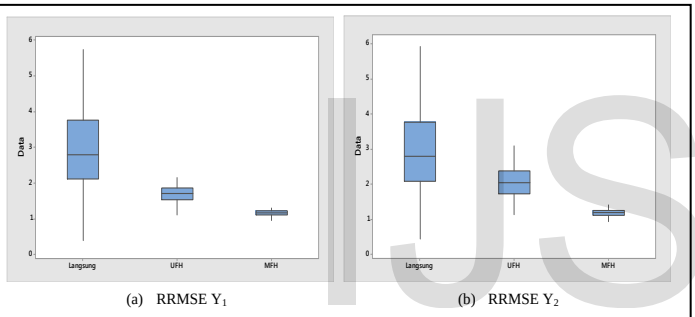

Fig. 2. Boxplot RRMSE of . of per capita expenditure on food (a) and non food (b)

TABLE 6
RRMSE OF PER CAPITA EXPENDITURE ON FOOD AND NON FOOD

Statistic	Methods		
	Direct	UFH	MFH
	Per Capita Expediture on Food		
Average	3.107	1.699	1.161
Standard Deviation	1.569	0.291	0.140
Minimum	0.392	0.387	0.377
Median	2.794	1.744	1.179
Maximum	10.956	2.315	1.539
	Per Capita Expediture on Non Food		
Average	3.135	2.055	1.186
Standard Deviation	1.576	0.452	0.138
Minimum	0.447	0.442	0.430
Median	2.795	2.054	1.193
Maximum	11.337	3.103	1.592

4 Conclusion

The conclusion of this research is the estimation of the average per capita expenditure on food and non-food per village in Bogor regency by UFH Model and MFH Model better than direct estimation because it can increase the efficiency of the estimated value of average of per capita expenditure on food and non food per village. It can be seen from the generated RMSE, the UFH Model and the MFH Model generating a smaller RMSE than the direct estimation without having to add sample sizes. In addition, the small-area estimation model produces a better estimation when using correlated variables calculated together in one model (MFH Model) rather than a model with only one variable (UFH Model).

The model used in this study is the area level model (Fay-Herriot) and the concerned variables are the continuous data of normal distribution. For further research, it can be conducted by using unit level model and for discrete data variable and its distribution is not normal.

Acknowledgment

The authors wish to thank BPS-Central Bureau of Statistic for providing data in this research. This work was supported in part by a grant from BPS.

References

[1] C.Giusti, S. Marchetti, M. Praseti, N. Salvati, "*Robust small area estimation and oversampling in estimation of poverty indicators*". *Survey Research Methods*. pp 155–163, 2012

[2] JNK Rao. 2003. Small Area Estimation. New Jersey (US): John Wiley & Sons, Inc.

[3] R. Ningytas, "*The Application of EBLUP of Small Area Estimation Model on per capita expenditure in Brebes District*," *Jurnal Gaussian* vol.4, no.4,pp. 977-986, 2015.

[4] H. Fausi, "Small Area Estimation of Prediction of Sumenep District's Per Capita Expenditure by using EB Method," Master Thesis, Dept. of Statistics, Tenth of November Institute of Technology, Surabaya, 2011.

[5] AD. Wardani, "Comparison of EB and EBLUP Method on Small Area Estimation (Case Study on Prediction of Per Capita Expenditure of Bogor)," Undergraduate Thesis, Dept. of Statistics, Bogor Agricultural Univ., Bogor, 2008.

[6] Datta GS, Fay RE, GhoshM, "*Hierarchical and empirical Bayes multivariate analysis in small area estimation," In: Proceedings of Bureau of the Census 1991 Annual Research Conference, US Bureau of the Census, Washington, DC,* pp. 63–79. 1991.

[7] Central Bureau of Statistic, *Consumption Expenditure of Population of Indonesia by Province 2016*. Jakarta, pp. 15, 2016.