Qonnita Makrufa (BPS Statistics Indonesia), Ayesha Tantriana (BPS Statistics Indonesia)

Analysis of Poverty in a Village Level of Indonesia with Small Area Estimation: Case in Bangkalan District

Goals formulated in Sustainable Development Goals (SDGs) and national social policies indicate the need for analyzing poverty at local levels. Poverty measurement with sample surveys can't directly produce poverty indicators at lower level (e.g. district, sub-district, or village). The difficulty confronted is that detailed household surveys are rarely representative at lower levels and they don't cover sufficient number of households to yield statistically reliable poverty estimation when disaggregated (Elbers, et al., 2000). At the same time, census data have no variable to measure poverty. Large cost will be needed if a country wants to get variables to measure poverty at local levels. Due to the data limitation, one of the solutions is using Small Area Estimation (SAE) method. Small area estimation is a statistical technique to estimate parameters for small sub-population. The estimation technique combines two data sources which are household sample surveys and the comprehensive coverage of a census. Simple estimation on a small area needs to be used because direct estimation is not able to provide sufficient accuracy when the sample size in the area is small so that the statistic results will have a large variant or even the estimation can't be acquired because the results are biased (Rao, 2003).

In measuring poverty, Statistics Indonesia uses the basic needs approach. In this approach, poverty is considered as the inability to fulfill basic needs (food and non-food) which is measured by the expenditure. The data sources are National Socio-economic Survey (SUSENAS) in 2016 and Population Census in 2010. The response variables in this study is poverty level at the level of sub-districts in Bangkalan District in East Java, Indonesia. To gain the poverty rate in sub-district or village level, this paper implements two approaches, Elbers, Elbers, Lanjou (EEL) method and Empirical Bayes (EB) method. These methods are compared to see which method is better in estimating the small area using SUSENAS and census data. The statistic programs used in this research are software R, Stata, SPSS and QGIS for poverty mapping.

There are nine steps to gain the poverty rate. First step is selecting concomitant X variables that affect or describing poverty level. There are nine variables available in survey and census data. Concomitant dichotomy variables predicted to affect and describe poverty rate are school participation (X1), education level (X2), employment status (X3), job sector (X4), floor type (X5), toilet facility (X6), source of drinking water (X7), internet access (X8), and household size (X9). Second step is analyzing the poor household descriptively. The result shows that 19 of 100 household in Bangkalan district living below the poverty line.

Next step is the simulation of the variables in regression model and assumption test. The result in this step shows that the best model is constructed from five variables: school participation

(X1), education level (X2), job sector (X4), floor type (X5), and household size (X9). There are three assumptions that have been tested, namely normality, multicollinearity, and homoscedasticity. Normality and multicollinearity assumptions are fulfilled, but not for the homoscedasticity. Heteroscedasticity causes estimators of the regression coefficients no longer efficient in both small and large samples (asymptotically). In addition to that, the variance generated will either underestimate or overestimate. In this paper, the way to overcome the violation of homoscedasticity assumptions is by estimating the robust standard error in each coefficient.

Next is calculating the percentage of poor households in population census data. The estimated value of per capita expenditure for households in the SP2010 data will be used to calculate the percentage of poor households in each village. The estimated value of per capita household expenditure is obtained by multiplying the estimated regression coefficients on SUSENAS data with variables on census. By using bootstrap resampling, the percentage of poverty at the district level is 18.4 percent. The fifth step is simulating the variable of the nested error linear regression model and assumption test. In EB, expenditure models are formed using the best models obtained in the ELL method.

Next is estimating the percentage of poor household using EB method. The parameter from SUSENAS data distribution which has been estimated before, is used to estimate the expenditure of household in census data. Simulation with monte carlo method is implemented to count the expenditure of household in census data. The simulation has been done until get convergence of the value of estimation. After calculating the estimated percentage of household poverty in the EB and ELL methods in previous steps, a comparison of the results can be obtained. Compared results from both methods are estimated values of presentations of poor households at the village level, aggregation of the percentage of poor households at the district level, MSE and RRMSE for village level estimates, and village level poverty mapping. In the EB method, the villages that have the highest percentage of poor households are in Kanegarah village (70%) and the lowest is Mlajah village (9%). However, in the ELL method, village with the highest percentage of poor households is Kokop village (43.96%) and the lowest is Tanjung Jati village (1.42%). It can be seen that there are obvious differences in estimation results between the two methods, but not far enough. The Pearson correlation coefficient indicates that villages with high poverty rate in the EB estimation also have high poverty rate in the ELL estimation results. The result also shows that both EB and ELL method have a small bootstrap MSE value. This shows that the estimation results obtained through both methods are good and reliable. But to map the estimated poverty rate, this paper uses EB estimation as it has less MSE bootstrap and RRMSE score than EEL. Lastly, this paper maps poverty rate at sub-district level using the EB estimation. Poverty analysis on a map makes it easy to detect some villages that have high percentage of poor household.