ABSTRACT

Title of Thesis: SMALL AREA ESTIMATION OF SUBDISTRICT POVERTY RATES ON JAVA ISLAND, INDONESIA

Fnu Yuniarti
Master of Science, 2022

Thesis Directed by: Professor Partha Lahiri
Department of Mathematics
and Joint Program In Survey Methodology

Poverty representing the deprivation of well-being is a global concern. The provision of poverty statistics for small area levels becomes crucial and mandatory to accelerate the poverty eradication program. This research undertakes poverty rate estimation for subdistrict levels on Java Island using small area estimation techniques. The results would give opportunities to better understand the poverty situation in small areas levels on Java Island. The main data source to calculate the poverty estimates is the Social and Economic Survey of March 2020. The direct estimation method, logistic fixed effects, and the logistics mixed effect models are employed to generate the poverty estimates. The model-based techniques use datasets from Population Census 2010 and Village Potential Enumeration Survey 2014 to derive the covariates. The direct technique does not show plausible estimates due to the sample size insufficiency. The logistic model improves the direct estimates but the variation between subdistricts is quite small which is unreasonable. The logistic mixed-effects model does not work perfectly but it gives good

estimates for several subdistricts.

SMALL AREA ESTIMATION OF SUBDISTRICT POVERTY RATES ON JAVA ISLAND, INDONESIA

by

Fnu Yuniarti

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2022

Advisory Committee:
Professor Partha Lahiri, Chair/Advisor
Professor Paul J. Smith
Professor Takumi Saegusa

Acknowledgments

First and foremost, I intend to convey sincere gratitude to my advisor, Professor Partha Lahiri for giving me an opportunity to work in his area of expertise. I started learning the topic of small area estimation in 2018. I found his name almost everywhere, in books and papers. After I got accepted into the Department of Mathematics, University of Maryland, I realized that he is one of its faculty members. The first words that came to mind were "I should learn something from him before I leave this country!". My wish came true when he gave his approval to serve as my thesis advisor. It is truly my great pleasure to learn about small area estimation from a famous expert. Thank you for your invaluable advice and guidance so that I could complete this project.

I would also like to thank my colleagues at the BPS-Statistics Indonesia for providing various micro datasets. The availability of these data has opened an opportunity to explore the Indonesian poverty cases at the small area levels. After further improvement, I believe this project will be helpful in the provision of official small area statistics in the future.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BAPPENAS	Badan Pembangunan dan Perencanaan Nasional (Ministry of National Planning)
BPS	Badan Pusat Statistik (Indonesia National Statistics Office)
GDP	Gross Domestic Product
GLMM	Generalized Linear Mixed Model
GLM	Generalized Linear Model
HCI	Head Count Index
MPI	Multidimensional Poverty Index
MSE	Mean Square Error
NSO	National Statistics Office
OECD	Organization for Economic Co-operation and Development
OPHI	Oxford Poverty and Human Development Initiative
PODES2014	Potensi Desa 2014 (Villages Potential Enumeration Survey)
RMSE	Root Mean Square Error
RSE	Relative Standard Error
SAE	Small Area Estimation
SDG	Sustainable Development Goal
SE	Standard Error
SP2010	Sensus Pensusuk 2010 (Population Census 2010)
SUSENAS	Survey Sosial Ekonomi National (National Social-Economic Survey)
TNP2K	Tim Nasional Percepatan Penanggulangan Kemiskinan
UNECE	United Nations Economic Commission for Europe
UNDESA	United Nations Department of Economic and Social Affairs
UNDP	United Nations Development Programme

Chapter 1: Introduction

Deprivation of well-being has occurred everywhere in the form of poverty. Those who lack access to productive resources or cannot ensure sustainable living have a high chance to suffer from this phenomenon. Haughton and Khandker (2009) assert that poverty is a multidimensional issue as it is related to many aspects, such as inadequate income or consumption, low access to education and health care, or even the absence of political freedom. Hence, poverty eradication requires a comprehensive formulation of efforts.

Poverty has become a global concern since it happens all over the world. The increase in morbidity, mortality, malnutrition cases, the unsafe environment, and the inconducive political situation are some dreadful impacts of poverty incidence (UNDESA, 2010). Poverty also could lead to humiliation and even inhuman treatment by a higher power (Ng et al., 2013). Since this issue is crucial, United Nations set poverty eradication as goal 1 among the 17 Sustainable Development Goals (SDGs). Meanwhile, the Indonesian government considers it an urgent national problem and then defines the poverty alleviation acceleration program through Presidential Decree No. 15 2010 (TNP2K, 2021). Moreover, the Indonesian government also has incorporated this SDG goal into National Development Planning System (BAPPENAS, 2020) to ensure the poverty eradication development program would meet the global target, reducing the proportion of people living in poverty at least by half in 2030 (United Nations General Assembly, 2015).

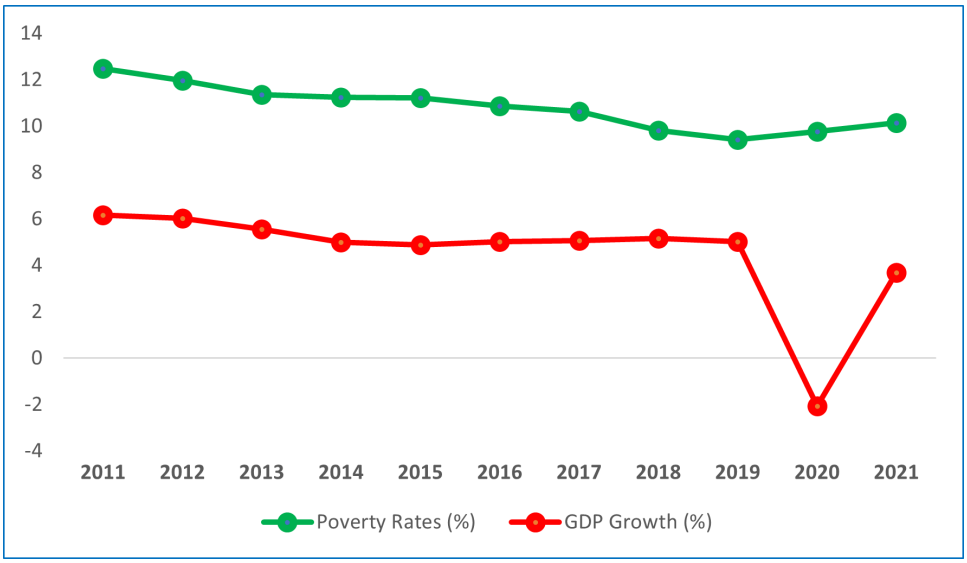

Figure 1.1: National Gross Domestic Product (GDP) Growth and Poverty Rate (March), 2011-2021.

History proves that economic growth is the best way to improve the life quality of people including those at the very bottom (Dani, 2007; Son and Kakwani, 2004). Dollar and Kray (2002) found that the growth of income of the poor has a one-to-one relationship with overall economic growth. This means that higher economic growth would benefit the poor proportionately. Roemer and Gugerty (1997) also provide evidence of the strong relationship between economic growth and poverty reduction but the difference in growth-oriented policies would give a different rate of poverty reduction. Figure 1.1 shows approval of the above statement that the poverty rates would decrease given the growing economy.

DKI Jakarta is the capital city of Indonesia. It is considered the center of governmental activities and economic growth at the same time. Banten, situated on the west end of Java Island bordering DKI Jakarta, is also popular as the center of economic development since many mega businesses and manufacturers are built there. Banten has attracted many people in other regions,

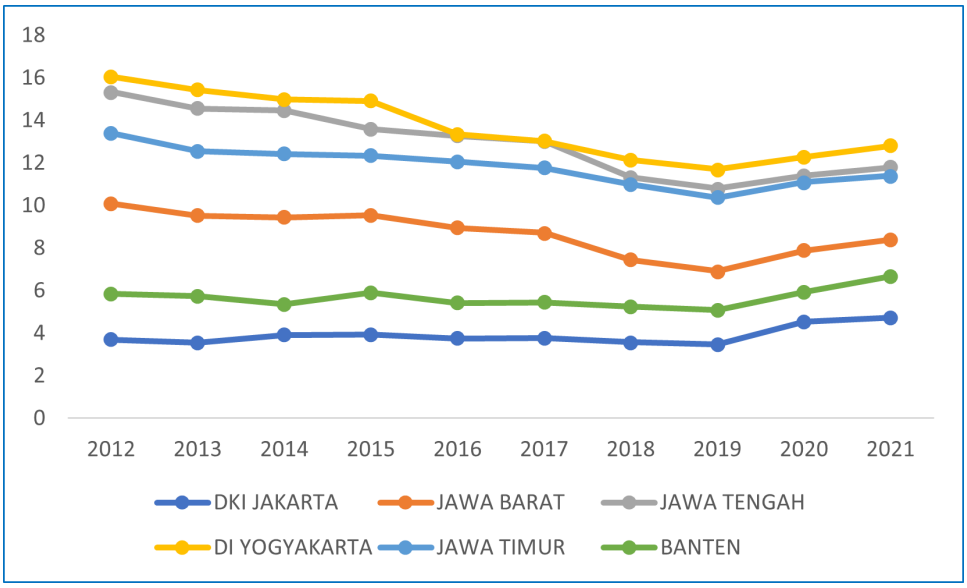

Figure 1.2: Poverty rates six provinces on Java Island, 2012-2021.

even from other islands, to come and find decent jobs. From Figure 1.2, it can be seen that the poverty rates in the center of economic growth are lower than in other regions. The six provinces highlighted in Figure 1.2 are all located on Java Island where public infrastructures are information technologies are fast-growing compared to other islands.

These facts raise a prejudice that the poverty eradication programs being developed give more focus on the areas referred to as the center of economic growth. The possible reasons behind these shallow policies are that the area where economies are well developed tend to have abundant resources. Thus, poverty elimination projects can be implemented straightforwardly in these areas. The other possible reason is the unavailability of poverty statistics in small area levels, such as villages or subdistricts. Since the number of those living under the poverty line is unidentified in these area domains, it is difficult for policy makers to focus on the most vulnerable areas. Hence, the provision of small area poverty statistics is extremely important in this situation.

Local governments are fundamental actors in implementing policies for the vulnerable people as they understand the real situation as it is (Pratesi and Salvati, 2016). One of the tasks of local governments is ensuring the effectiveness of service delivery (Mudalige, 2020). Therefore, if small area statistics are made available, it could encourage their serious involvement in the poverty eradication programs, including impact monitoring and the effectiveness assessment.

Indonesia NSO regularly releases poverty statistics for national, province, and district levels. The national and provincial levels are released twice a year while the district level is provided annually. This research intends to estimate poverty statistics at the lower area level which is not served by Indonesia NSO. The results of this research would open a chance poverty incidence can be monitored at the small area levels, particularly by regional governments.

Chapter 2: Literature Review

2.1 Poverty Measures

UNECE (2017) highlighted the evolution of poverty measures from the very modest, called the physical subsistence approach, to the basic needs, relative deprivation, wealth deprivation, time poverty and multidimensional poverty criterion. The first technique has gotten many critics since it treats people as consumers of physical goods only and ignores their social needs (Lister, 1990) such as access to health care and educational opportunities (Jones, 1994).

Walker (2015) argues that the perspective of poverty is not just about the absence of money or money-like resources. He asserts that the non-monetary dimension of poverty could be the main reason to prolong poverty and its perpetuation. Therefore, the last approach to examining poverty from a multidimensional view is the most recommended by many organizations, especially those concerned with poverty eradication in all forms such as United Nations and World Bank. UNDP and Oxford Poverty and Human Development Initiative then proposed a multidimensional poverty index (MPI) to measure this kind of poverty issue. MPI considers 10 indicators representing three crucial dimensions: health, education, and standard of living (OPHI, 2018).

MPI is extremely useful for capturing well-being deprivation beyond monetary measures (Akinyetun et al., 2021) since rural people are less active in monetary-oriented transactions compared to those living in urban areas (Tefera et al., 2016). Yet, the implementation of this

approach for monitoring the change in welfare over time is arduous. Burchi et al. (2011) reveal that MPI overlooks the distribution of deprivation among the poor. Moreover, MPI tends to produce higher poverty rates compared to the monetary-based techniques (Rippin, 2011). Thus, for countries with (severe) budget constraints, it is difficult to set priority for the most impoverished.

Indonesia chose the basic need needs to approach in terms of expenditures to measure people's welfare levels. Indonesia considers poverty as an economic inability to meet the basic needs of food and non-food as quantified from the expenditure side (BPS, 2021). Wong (2012) claims that this method is convenient and easy to operationalize for regular poverty monitoring and ensures the policies reach the neediest.

The most fundamental element for measuring poverty is the poverty line. It is a minimum acceptable standard to classify a household as poor or not poor (Ravallion, 1998). Households having average monthly basic needs expenditures below the standard poverty line are considered poor families. People living in poor families are viewed as the poor.

Assessing basic needs expenditures requires a basket of minimally necessary commodities (goods and services) that are different across countries or regions (Watson, 2014). OECD (2013) recommends the goods and services included in the calculation are those obtained through the purchase in the market, as in-kind income from employers, self-employment, property, or investment, from the household's own production, and as in-kind transfers from other households and businesses. Indonesia has allowed 52 kinds of food commodities into the basket and then adds 51 non-food commodities for urban areas and 47 non-food commodities for rural areas since 1998 (BPS, 2021).

Foster et al. (1984) suggests a statistic called Head Count Index or HCI-P0 to estimate

poverty rates defined by

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left[\frac{z - y_i}{z}\right]^\alpha, \tag{2.1}$$

where z is the poverty line, y_i is the monthly average per capita expenditures for those living under the poverty line, q is the number of people living under the poverty line, and n is the total population. HCI-P0 is simply obtained by taking $\alpha = 0$. Substituting α with 1 and 2 will result in a poverty gap index and poverty severity index respectively.

2.2 Small Area Estimation Models

"Small area estimation" (SAE) is a popular term that refers to a technique to estimate statistics for a small area domain. The term "small area" generally represents any geographic domain for which statistics cannot be provided with adequate precision using direct estimation methods (Rao, 2003; Rao and Molina, 2015). Jiang and Lahiri (2006b) reveal the reasons behind the deficiency of reliable direct estimates. One of them is due to the sampling design of a particular data source (survey) which aims to produce reliable statistics for larger areas but is not dedicated to the smaller areas. Datta (2009) also affirms that reliable precision cannot be achieved if the sample for the area of interest is not large enough. In other words, the insufficiency of sample size is the common excuse to implement a small area estimation technique.

According to Rao and Molina (2015), the principle of the SAE technique is to "borrow strength" from related or similar areas to increase the effective sample size. The values from related or similar areas are incorporated into the estimation process through explicit or implicit linking models and statistical models which connect the variable of interest with supplementary

information derived from other data sources. Pfeffermann (2013) encourages the use of supplementary or auxiliary information with good predictive power to produce reliable small area estimates. The most common source of supplementary information are the census and administrative records (Tarozzi and Deaton, 2009) since they do not contain sampling errors (Lohr and Prasad, 2003).

Under the model-based approach, there are two estimation models popularly used based on the availability of response variables (Jiang and Lahiri, 2006a). The first is area-level models which can be employed when area-specific variables of concern are available. Meanwhile, if the available variables denote unit-specific information then unit level models are good to consider. The type of variable of interest in the small areas is another crucial issue in the small area modeling. When the variables take the form of binary data such as poverty status (poor or not poor) and employment state (employed or unemployed), then logistic models are known to have good properties for estimation (Hwang and Shin, 2011).

2.2.1 Logistic Model

Denote Y as a binary response variable with two possible outcomes, 1 (success) and 0 (failure). This kind of variable has probabilities $P(Y = 1) = \pi$ for the successful event and $P(Y = 0) = (1 - \pi)$ for the failure event. The number of successes within n independent observations is then viewed to have a binomial distribution. Let $\{y_i, i = 1, ..., n\}$ are observations under a binary response variable Y. Agresti (2015) defines the logistic regression model as

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^{p} \beta_j x_{ij}, \tag{2.2}$$

where x_{ij} are the independent variables in the number of p, β_j is the rate of increase or decrease of the π_i curve with respect to x_{ij} which also acts as a fixed effect parameter, and π_i is the binomial parameter. The above equation is often called the logit model since it makes use of the logit link function to relate π_i with the linear predictors.

In the context of small area estimation, i may represent a small area of interest or unit of concern. For simplicity, let i denote a household unit since this study will apply the above model at the household level. In relation to poverty analysis, success and failure can be replaced with poverty status, 1 for poor and 0 for not poor. Thus, π_i in the above model indicates the probability of a household being poor.

The likelihood equations of the model 2.8 are given by

$$\sum_{i=1}^{n} n_i(y_i - \pi_i)x_{ij} = 0, j = 1, ..., p \tag{2.3}$$

To solve equation (2.3), one of the techniques suggested by Agresti (2015) is the Newton–Raphson iterative method. Let

$$u_j^{(t)} = \frac{\partial L(\beta)}{\partial \beta_j}\bigg|_{\beta^{(t)}} = \sum_i (s_i - n_i\pi_i^{(t)})x_{ij}, \tag{2.4}$$

and

$$h_{ab}^{(t)} = \frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b}\bigg|_{\beta^{(t)}} = -\sum_i x_{ia}x_{ib}n_i\pi_i^{(t)}(1 - \pi_i^{(t)}), \tag{2.5}$$

where $s_i = n_i y_i$. For the case of ungrouped data, n_i is equal to 1. The approximation of $\hat{\pi}_i$ at iteration t is obtained by the following formula

$$\pi_i^{(t)} = \frac{\exp(\sum_{j=1}^{p} \beta_j^{(t)} x_{ij})}{1 + \exp(\sum_{j=1}^{p} \beta_j^{(t)} x_{ij})}. \tag{2.6}$$

Then, the estimation value of β is given by

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1} u^{(t)} \tag{2.7}$$

where H is the Hessian matrix with entries of h_{ab}.

2.2.2 Logistic Mixed Model

Jiang and Nguyen (2021) recommend the use of mixed models or mixed effect models when there are correlations among the data. The correlations are normally specified via random effects. Jiang and Lahiri (2006) mention that this type of model is appropriate for small area estimation since the random effects could explain the between area variations which are difficult to define through fixed effects. Rao (2003) calls these variations random area-level effects.

Many studies have proven that the incorporation of an area random effect gives better estimates than a model containing fixed effect only. For estimation based on logistic models, Lehtonen and Veijanen (2016) and Farrel et al. (1997) discover that the logistic mixed model produces better results than the logistic model. These facts are some reasons why mixed effect models are widely used in many fields.

Let $\{y_{ij}\}$ denote j-th observation in i-th area. Rao and Molina (2015) suggest the logistic mixed model for unit level estimation as below

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x_{ij}^T\beta + \upsilon_i, \tag{2.8}$$

where π_{ij} is the Bernoulli parameter, x_{ij} are unit-level covariates, β is the fixed effect parameter, and υ_i is the random effect parameter having an independent $N(0, \sigma_v^2)$ distribution.

Jiang and Lahiri (2001) interpret the above equation as a three-stage model as follows

(i) Conditional on the parameter π_{ij}, $\{y_{ij}\}$ are independent Bernoulli distributed variables with $P(y_{ij} = 1|\pi_{ij}) = \pi_{ij}, i = 1, \ldots, d; j = 1, \ldots, n_i$.

(ii) Conditional on the random effect υ_i, logit $(\pi_{ij}) = log\,(\pi_{ij}/(1-\pi_{ij})) = x_{ij}^T\beta + \upsilon_i, i = 1, \ldots, d; j = 1, \ldots, n_i$.

(iii) $\{\upsilon_i, i = 1, \ldots, d\}$ are independent $N(0, \sigma_v^2)$ distributed where σ_v^2 is an unknown variance.

Agresti (2015) notices that the logistic mixed model is a special case of the generalized linear mixed model (GLMM) under two stages of assumptions. The first stage assumes that conditional on υ_i, y_{ij} are described by the general linear model (GLM) where the expected values are linked to a linear predictor, logit $[E(y_{ij}|\upsilon_i)] = x_{ij}^T\beta + \upsilon_i$. At this stage, υ_i is a known offset. The second stage admits υ_i to follow $N(0, \sigma_v^2)$ distribution. The likelihood function for a GLM is called marginal likelihood which refers to the marginal distribution of y_{ij} after integrating out the random effect.

The conditional probability of y_{ij} given υ_i is defined as

$$P(y_{ij} = 1|\upsilon_i) = \frac{\exp(x_{ij}^T\beta + \upsilon_i)}{1 + \exp(x_{ij}^T\beta + \upsilon_i)}, \tag{2.9}$$

$$P(y_{ij} = 0|\upsilon_i) = \frac{1}{1 + \exp(x_{ij}^T\beta + \upsilon_i)}, \tag{2.10}$$

Thus, the conditional probability of $y_i = (y_{i1}, ..., y_{in_i})^T$ is given by

$$g(y_i|\upsilon_i; \beta) = \prod_{j=1}^{n_i} P(y_{ij} = 1|\upsilon_i)^{y_{ij}} P(y_{ij} = 0|\upsilon_i)^{1-y_{ij}} \tag{2.11}$$

and the marginal likelihood function of the i-th area is

$$L_i(\beta, \sigma_v^2) = \int g(y_i|\upsilon_i; \beta) f(\upsilon_i, \sigma_v^2) d\upsilon_i \tag{2.12}$$

where $f(\upsilon_i, \sigma_v^2)$ is the normal probability density function for υ_i. Hence, the full marginal likelihood function for all areas is given by

$$L = \prod_{i=1}^{d} L_i(\beta, \sigma_v^2) \tag{2.13}$$

Fitting the marginal likelihood function of GLMM is tricky and can be computationally intensive. Therefore, Agresti (2015) and Jiang and Nguyen (2021) recommend using the Gauss–Hermite Quadrature method. Kim et.al (2013) illustrate the detail of this approach as follows.

Denote the variance-covariance matrix of the υ_i as $\Sigma(\sigma_v^2)$ and reparametrize $\upsilon_i = \Gamma u_i$ where $\Gamma\Gamma^T = \Sigma(\sigma_v^2)$ and u_i has a standard normal distribution with density $\phi(u_i)$. Thus, equation 2.12 can be written as

$$L_i(\beta, \sigma_v^2) = \int g(y_i|\upsilon_i; \beta, \Gamma) \phi(u_i) du_i \tag{2.14}$$

The Gauss–Hermite Quadrature approximated marginal likelihood is given by

$$L_i(\beta, \sigma_v^2) \approx \sum_q g(y_i|b_q; \beta, \Gamma) w(b_q) \tag{2.15}$$

where b_q is a vector of quadrature points and $w(b_q)$ is the related weights. The dimension of b_q is the same as that of u_i.

2.3 Parametric Bootstrap

In statistical analysis, the bootstrap can be implemented in two ways, parametric and nonparametric based. Nonparametric bootstrap repeatedly resamples the original data with replacement to approximate the population distribution or make statistical inferences such as standard error and confidence interval. There is no assumption employed in this approach. Meanwhile, parametric bootstrap utilizes the extra information about the population distribution to make inferences. Both are computer-intensive methods useful for assessing statistical accuracy or estimating accuracy measures (Shao Tu, 1995).

Parametric bootstrap assumes that the distribution of the data is known. This technique is performed by developing a good model under the original data and then using the estimated parameters to simulate the samples. Hence, the performance of parametric bootstrap immensely relies on the pertinence of the assumptions. The estimation using the parametric bootstrap is generally more reliable than that of the nonparametric version if the proposed model is closed to the population distribution (Lee, 1994).

Efron (1979, 1982) have shown many applications of bootstrap for estimating standard deviation, bias, and other probabilistic properties under linear and non-linear functions and even

smoothed distribution functions. This technique is proven to be effective even for more complex parametric settings. Bootstrap is considered simple and computationally fast, so it become prevalent in statistical analysis. Efron and Tibshirani (1993) claim that 50 bootstrap replication is often enough for estimating standard error and seldom does one need more than 200.

Lahiri (2003) calls bootstrap the supplest method of analysis since it can solve challenging statistical problems efficiently, such as imputation and small area estimation. In small area analysis, this approach is very popular to derive the mean square error (MSE) as studied by Jiang and Lahiri (2001), Butar and Lahiri (2003), and Hall and Maiti (2006).

Suppose θ is the target parameter to estimate. If the cumulative density function (CDF) F of the population is known, then θ can be written as a function of F as shown below:

$$\theta = t(F) \tag{2.16}$$

If F is unknown, an empirical CDF $\hat{F}_n$ can be computed to produce the plug-in estimator

$$\hat{\theta} = t(\hat{F}_n). \tag{2.17}$$

Thus, the MSE of $\hat{\theta}$ is given by

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = E((\hat{\theta} - E(\hat{\theta}))^2) + (E(\hat{\theta}) - \theta)^2 \tag{2.18}$$

From the above formula, we know that

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2 \tag{2.19}$$

Hence, the bootstrap for $MSE(\hat{\theta})$, particularly under logistic mixed models, can be conducted as follows.

(i) Fit the unit-level model to get estimated parameters $\hat{\beta}$ and σ_v^2 using the Gauss–Hermite Quadrature method.

(ii) Generate area random effect υ_i from $N(0, \sigma_v^2)$ distribution, $i = 1, \ldots, d$.

(iii) Generate response variables from the model as

$$\pi_{ij}^* = \frac{\exp(x_{ij}^T\hat{\beta} + \upsilon_i)}{1 + \exp(x_{ij}^T\hat{\beta} + \upsilon_i)} \tag{2.20}$$

(iv) Calculate $\hat{\theta}_i^*$, the target parameter for each area, $i = 1, \ldots, d$.

(v) Repeat steps (i) to (iv) for B times to get $\{\hat{\theta}_{i1}^*, \hat{\theta}_{i2}^*, ..., \hat{\theta}_{iB}^*\}$ or each area, $i = 1, \ldots, d$.

(vi) Calculate the $Var(\hat{\theta})$ and $Bias(\hat{\theta})$ for each area as follows

$$Var(\hat{\theta}) = \frac{\sum_{b=1}^{B}(\hat{\theta}_{ib}^* - \overline{\hat{\theta}}_i^*)^2}{B - 1} \tag{2.21}$$

and

$$Bias(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{ib}^* - \hat{\theta}_i \tag{2.22}$$

where

$$\overline{\hat{\theta}}_i^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{ib}^* \tag{2.23}$$

and $\hat{\theta}_i$ is a statistic derived directly from the samples in area i.

(vii) Calculate the $MSE(\hat{\theta})$ for each area as

$$MSE(\hat{\theta}_i) = Var(\hat{\theta}_i) + (Bias(\hat{\theta}_i))^2 \tag{2.24}$$

Chapter 3: Methodology

3.1 Data Sources And Variables

This research makes use of the Social and Economic National Survey (SUSENAS) March 2020 for estimating poverty rates at the subdistrict level. The unit of observation of SUSENAS March 2020 is household. This survey covers around 345,000 households or 34,500 census blocks and is dedicated to producing statistics at the district level. SUSENAS is conducted under two-stage sampling with the following scheme (BPS, 2019).

Stage 1:

(i) Select 40 percent of the census blocks (n') from the population of census blocks (N) recorded through Population Census 2010 (SP2010) using probability proportional to size (PPS) sampling of the number of households. The selection is performed independently between urban and rural strata in every district. At this stage, the probability of the census block being selected is M_{ghi}/M_{gh}, where M_{ghi} is the number of households in the i-th census block of the j-th stratum M_{ghi} of the g-th district. Meanwhile, M_{gh} represents the number of households j-th strata of gth district.

(ii) Select n census blocks from n' using systematic sampling. The value of n is determined by the allocation rule and is considered sufficient for district level estimation.

Stage 2: Select 10 households from the total number of households in the sampled census blocks of size n using systematic sampling. Before making a selection, all households are stratified based on the head of household's education.

In addition to the SUSENAS 2020, this research also uses Population Census 2010 (SP2010) and Village Potential Enumeration Survey 2014 (PODES2014) as the sources of the auxiliary variables. Both datasets are available at the village level. The auxiliary variables derived from SP2010 are

1. The proportion of the population with high school education or higher (HScGr).
2. The proportion of the population working in the non-agricultural sector (NoAgr).
3. Dependency ratio, (DP).
4. The proportion of the working population (Work).

The dependency ratio specifies the number of those likely to depend on the working-age people for support. The term working-age refers to an age group between 15 to 65 considering that most people in this group are economically active. The dependency ratio is then calculated as the number of those aged below 15 and above 64 divided by those aged 15 to 64. The auxiliary variables obtained from the PODES2014 are specified below:

1. The proportion of villages with transportation infrastructure improvement in the last three years (Trans).
2. The proportion of villages with education infrastructure improvement in the last three years (Edu).

3. The proportion of villages with health infrastructure improvement in the last three years (Health).

4. The proportion of villages with economic infrastructure improvement in the last three years (Eco).

5. The proportion of villages with access to agricultural business loans in the last three years (LoAg).

6. The proportion of villages with access to non-agricultural business loans in the last three years (LoNAg).

7. The proportion of villages with access to agricultural business grants in the last three years (GrBus).

8. The proportion of villages receiving production skills improvement programs in the last three years (ProIm).

9. The proportion of villages receiving marketing skills improvement programs in the last three years (MarIm).

3.2 Poverty Line

District level poverty lines based on SUSENAS March 2020 released by Indonesian National Statistics Office (NSO) are employed to define one's poverty status at the subdistrict level in this research. One living in a household with average monthly basic needs expenditures below the poverty line is considered poor. The poverty line is composed of food and non-food poverty lines.

The food poverty line is using 52 food commodities, while the non-food poverty line is derived from 51 non-food commodities from urban areas and 47 non-food commodities from rural areas.

3.3 Software

Poverty rates and the corresponding standard error (SE) and relative standard error (RSE) are generated using R software under the *survey* package. This package provides *svydesign* to declare the survey design and *svyby* to compute the statistics of interest based on the predefined survey design. Upon the generation of poverty rate direct estimates, the logistic models are fitted using *glm* function under *stats* package. Meanwhile, logistic mixed effect models are simulated through *glmer* functions under *lme4* package. The simulation of mean square error (MSE) employing the bootstrap method also rely on this software.

3.4 Analysis Procedures

The analysis conducted for this research is summarized through Figure 3.1.

The unit of observation of SUSENAS March 2020 is household. Every household has records on the number of household members and individual weight Individual weight means the number of people can be represented by an individual. Thus, the poverty rate direct estimate for every subdistrict is calculated by modifying the formula of Head-Count Index (HCI-P0) in Chapter 2 as presented below:

$$P_{0i} = \frac{\sum_{j=1}^{n_i} [nHH]_{ij} w_{ij} x_{ij}}{\sum_{j=1}^{n_i} [nHH]_{ij} w_{ij}} \tag{3.1}$$

where $[nHH]_{ij}$ is the number of people living in the i-th subdistrict, j-th household, w_{ij} is the

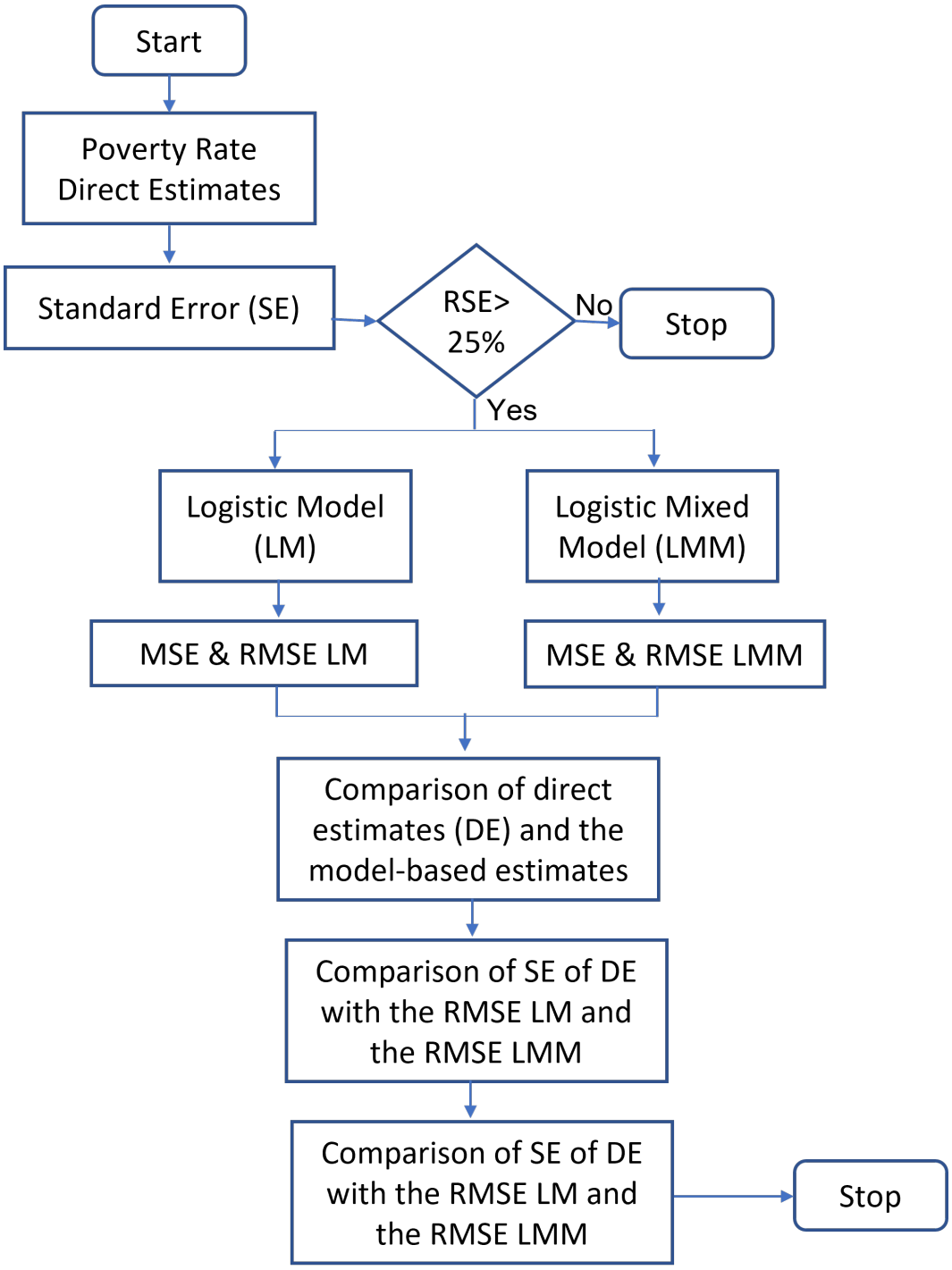

Figure 3.1: Analysis Flowchart.

individual weight, x_{ij} is the household's poverty status, and n_i are the number of poor households and the total households in the i-th subdistrict, respectively.

The SE and RSE of the poverty rate estimate in each subdistrict are then calculated by the following formula (Deming, 1950).

$$SE_i = \left[\frac{\sum_{j=1}^{n_i} w_{ij}[nHH]_{ij}(x_{ij} - \overline{\overline{x}}_i)^2}{\sum_{j=1}^{n_i} w_{ij}[nHH]_{ij}} \right]^{0.5} \tag{3.2}$$

where

$$\overline{\overline{x}}_i = \frac{\sum_{j=1}^{n_i} w_{ij}[nHH]_{ij}x_{ij}}{\sum_{j=1}^{n_i} w_{ij}[nHH]_{ij}} \tag{3.3}$$

and x_{ij} is the poverty status of a household, $[nHH]_{ij}$ is the size of the household, and w_{ij} is the corresponding individual weight. Thus, the RSE is given by

$$RSE_i = \frac{SE_i}{P_{0i}} \tag{3.4}$$

Indonesia NSO adheres to a principle that poverty estimates deserve official statistics if the RSE of estimation is less than 25 percent. If this condition is violated, then model-based statistical estimation is performed to reduce the RSE. This study attempts to apply the logistic model and logistic mixed model to handle the above issue.

A logistic model is performed by setting the household's poverty status as the response variable and the 13 covariates as independent variables. This process is followed by a backward elimination method to obtain a suitable model containing useful fixed effects only. The behavior of these effects is observed to determine whether they contribute to the poverty estimation in reasonable directions. If some variables exhibit improper directions, such as the more economic

relief delivered to the villages causes the higher possibility of households in a particular subdistrict being poor, the involvement of these variables in the model fitting is overlooked. Thus, the final model considered the most appropriate one is the model containing all significant fixed effects and the effects behave in the expected direction.

Running a mixed-effects logistic model using 13 fixed effects and one random effect is tiresome, particularly when the number of observations is also large enough. It takes a long time to converge and often encounters hardships related to the very large eigenvalues. Therefore, this research will use the final logistic model as a benchmark for performing a mixed-effects logistic model to estimate subdistrict poverty rates. Covariates included in the final logistic model are adopted in the mixed-effects logistic model fitting process then add the area level variation as the random effect. A backward elimination method is also applied in this effort to meet the most appropriate model.

The poverty rate estimates under the logistic and mixed-effects logistic models are given by

$$\hat{\pi}_i = \frac{\sum_{j=1}^{n_i} [nHH]_{ij} w_{ij} \hat{\pi}_{ij}}{\sum_{j=1}^{n_i} [nHH]_{ij} w_{ij}} \tag{3.5}$$

where $[nHH]_{ij}$ is the number of people living in the i-th subdistrict, j-th household, w_{ij} is the individual weight, $\hat{\pi}_{ij}$ is the probability of a household being poor resulted from logistic or mixed effects logistic models, and n_i are the number of poor households and the total households in the i-th subdistrict, respectively.

After obtaining poverty rate estimates using both models, the mean square errors (MSE) are estimated using the bootstrap technique. The component of MSE is variance and the bias of

the estimates. In Chapter 2, it has been discussed that 50 bootstrap replication is often enough for estimating standard error, and seldom does one need more than 200 (Efron and Tibshirani, 1993). Hence, this research will apply 100 bootstrap replication to derive this statistic.

The root mean square error (RMSE), the square root form of MSE, is a common measure to pick the best models. The lower RMSE is the better. In a small area estimation study, this statistic is usually compared with the standard error (SE) of the direct estimates to highlight the performance of the model-based estimation over the direst estimation technique.

In addition to the RMSE criteria for selecting the models, this research also uses the poverty rates at the district level released by Indonesia NSO to assess the reliability of the model-based estimates. The poverty rates at the subdistrict level produced by logistic and mixed-effects logistic models are aggregated to the district level and then are contrasted with those official statistics.

Chapter 4: Results Analysis

4.1 The Sample Distribution

The Law of Badan Pusat Statistic No. 15 2021 about The Codes and Names of Statistics Work Area mentions that there are 2,146 subdistricts on Java Island as of Semester 1 2021. SUSENAS March 2020, the main dataset in this research, covers 2,124 subdistricts as the samples. This means, there are 22 subdistricts unsampled by SUSENAS March 2021.

SP2010 and PODES2014 are two other datasets in this research used as the source of auxiliary variables. Since they are full enumeration surveys, the information of all subdistricts is recorded. SP2010 and PODES 2014 use the statistics work area frames of 2010 and 2014, respectively.

From the different statistics work area frames of the three datasets, it can be inferred that some subdistricts enumerated through SP2010 and PODES2014 were split up and merged with new names. Hence, those new subdistricts covered in SUSENAS March 2020 could not be found in SP2010 and PODES2014. Therefore, some subdistricts do not have auxiliary information to borrow strength from. These subdistricts are then excluded from this research. The final total number of subdistricts for small area analysis is 2,104 as summarized in Table 4.1.

The average population in each district on Java Island is about 70 thousand people. The smallest district is inhabited by 10 thousand people. From Figure 4.1, it is clear that many

Table 4.1: Number of subdistricts on Java Island.

Province	Number of Districts	Number of Subdistricts		
		Frame[1]	SUSENAS[2]	SAE[3]
DKI Jakarta	6	44	44	44
Jawa Barat	27	627	614	602
Jawa Tengah	35	576	573	570
Yogyakarta	5	78	77	77
Jawa Timur	38	666	661	657
Banten	8	155	155	155
Total	**119**	**2,146**	**2,124**	**2,104**

[1]The number of subdistricts from the statistics work area frames semester 1 2021
[2] The number of subdistricts in SUSENAS March 2021
[3] The number of subdistricts used for the SAE analysis

subdistricts have very small sample sizes covered by SUSENAS March 2021. A subdistrict named Cibinong has a population of more than 427 thousand but the sample size in this area is 146 people only. Meanwhile, Pakualaman, which is inhabited by 10 thousand people only, covers 53 samples. SUSENAS March 2020 is designed to produce district-level estimates. Hence, when it is used to generate lower-level statistics, then the sample representativeness is questioned. This becomes a strong reason why small area estimation models are required to produce better estimates.

4.2 Poverty Direct Estimates

Figure 4.2 represents the poverty rate direct estimates of 119 districts on Java Island. These figures are calculated based on SUSENAS March 2020 and have been released by Indonesia NSO as official statistics. The estimated range of poverty rates is from 2.29 to 22.78 percent with a mean of 9.93 percent.

When this survey is used to compute poverty statistics for lower-level areas, such as subdistricts,

Figure 4.1: The Sample Distribution of Each District Grouped by Province.

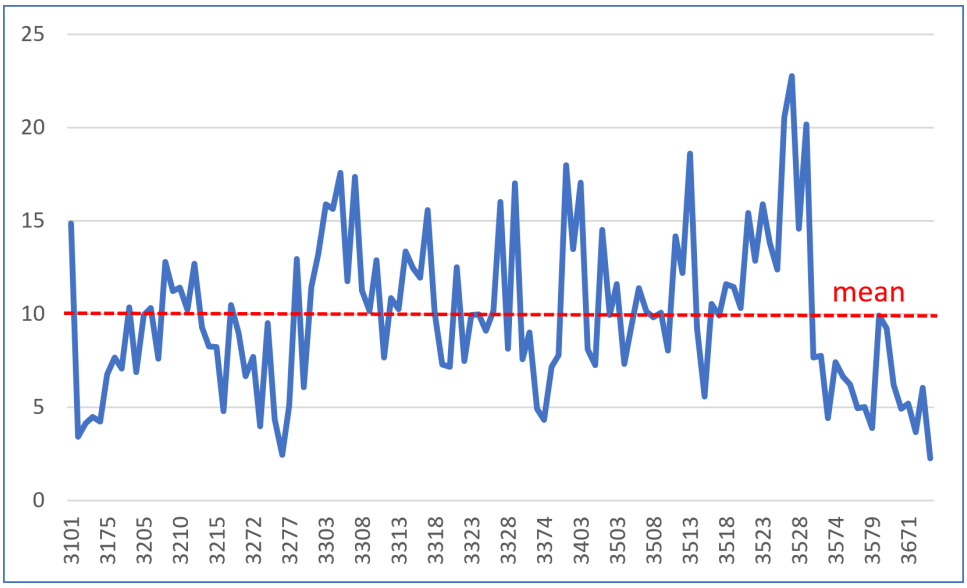

Figure 4.2: Poverty Rate of 119 Districts on Java Island, March 2020.

the results could be very different from the district-level estimates. Yet, the district level estimates are important as benchmarks to check the reliability of the subdistrict level estimates. When the direct technique produces subdistrict level estimates significantly far from those district level, the sum of poor people in each subdistrict should be ensured to be the same as the number of people in the corresponding district.

The direct poverty estimates using SUSENAS March 2020 for subdistrict level are presented in figure 4.3 It reveals the implausible estimates since numerous subdistricts are experiencing 0 percent of poverty rates which is impossible. Out of 2,104 subdistricts, 645 of those have 0 percent poverty rates. Many subdistricts also show extreme poverty estimates, more than 50 percent, and one subdistrict even experiences almost 90 percent of poverty incidence. The direct estimation technique performs poorly in this case.

The RSEs given in Figure 4.4 confirm the above findings that the direct estimates are not

Figure 4.3: Poverty Rates in 2,104 Subdistricts on Java Island, March 2020.

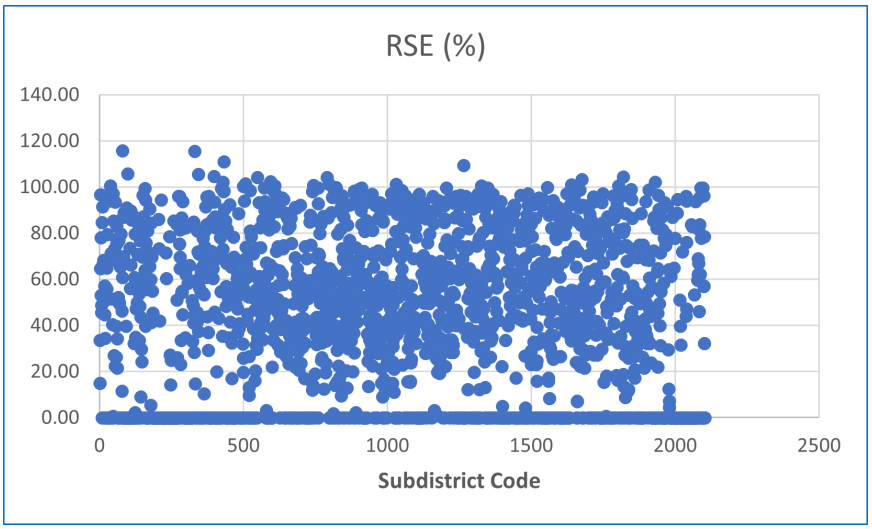

Figure 4.4: RSEs of Poverty Rate Direct Estimates 2104 Subdistricts on Java Island, March 2020.

satisfactory. The RSEs produced under direct estimation techniques are extremely high above 25 percent. It was mentioned in Chapter 3 that Indonesia NSO could release figures as official statistics only when the RSEs do not exceed 25 percent. Hence, small area estimation methods need to be applied to improve the estimates.

4.3 Small Area Poverty Models

Since the direct estimation could not produce reliable results for the subdistrict level, logistic and logistic mixed models are employed to improve the precision of poverty rate estimation. The models are applied at the household (unit) level of SUSENAS March 2020. The household's poverty status acts as the response variable. The 13 predictors (covariates) are derived from SP2010 and PODES2014 as detailed in Chapter 3.

Table 4.2 reveals that there is a strong positive correlation between the proportion of people

with a high school education or higher (HscGr) and the proportion of those working in the non-agricultural sector (NoAgr). It indicates that the non-agricultural labor market becomes a targeted sector to find decent jobs for those having a high school education or higher. The number of people holding a high school education or higher (HscGr) also correlated moderately with the dependency ratio (DP) in the opposite direction. This type of correlation is reasonable since the dependency ratio measures the dependency level of the non-woking age over the working age.

Table 4.2: Pearson coefficients of correlation between covariates.

	Trans	Edu	Health	Eco	LoAg	LoNAg	GrBus
Trans	1	0.33	0.16	0.2	0.2	0.23	0.15
Edu	0.33	1	0.25	0.33	0.34	0.28	0.27
Health	0.16	0.25	1	0.24	0.1	0.26	0.34
Eco	0.2	0.33	0.24	1	0.4	0.3	0.37
LoAg	0.2	0.34	0.1	0.4	1	0.4	0.41
LoNAg	0.23	0.28	0.26	0.3	0.4	1	0.35
GrBus	0.15	0.27	0.34	0.37	0.41	0.35	1
ProIm	0.12	0.13	0.39	0.24	0.25	0.31	0.49
MarIm	0.02	0.09	0.33	0.21	0.15	0.23	0.41
HscGr	-0.21	-0.32	0.25	-0.17	-0.37	-0.14	0.05
NoAgr	-0.16	-0.27	0.31	-0.16	-0.36	-0.07	0.03
DR	0.22	0.26	-0.07	0.21	0.3	0.16	0.05
Work	-0.04	0	-0.04	0.04	0.08	-0.05	0.07

	ProIm	MarIm	HscGr	NoAgr	DP	Work
Trans	0.12	0.02	-0.21	-0.16	0.22	-0.04
Edu	0.13	0.09	-0.32	-0.27	0.26	0
Health	0.39	0.33	0.25	0.31	-0.07	-0.04
Eco	0.24	0.21	-0.17	-0.16	0.21	0.04
LoAg	0.25	0.15	-0.37	-0.36	0.3	0.08
LoNAg	0.31	0.23	-0.14	-0.07	0.16	-0.05
GrBus	0.49	0.41	0.05	0.03	0.05	0.07
ProIm	1	0.65	0.19	0.16	-0.04	0.08
MarIm	0.65	1	0.28	0.22	-0.12	0.04
HscGr	0.19	0.28	1	0.82	-0.69	0.07
NoAgr	0.16	0.22	0.82	1	-0.48	-0.2
DP	-0.04	-0.12	-0.69	-0.48	1	-0.31
Work	0.08	0.04	0.07	-0.2	-0.31	1

Table 4.2 also explains that the more villages receiving production skills improvement programs in the last three years (ProIm), the more possibility for the villages to get marketing skills improvement programs (MarIm) as well. It gives a hint that both skills improvement programs are delivered as one package of policy to complement each other.

A logistic model involving full covariates has been performed to compute the probability of 101,291 households being poor. Following the backward selection methods, the interim model which covers all significant fixed effects is achieved and summarized in Table 4.3. This model comprises 11 significant fixed effects including the intercept term. This means that 10 covariates are meaningful to the model-based estimation. However, 4 variables exhibit odd behavior namely the proportion of the working population (Work), the proportion of villages with access to agricultural business loans (LoAg), the proportion of villages that received production skills improvement programs (ProIm), and the proportion of villages received marketing skills improvement programs in the last 3 years (MarIm). They have a positive contribution to the model which means that the more people in the subdistrict are working, the higher probability of a household being poor. Moreover, households could potentially be poorer when more villages in the particular subdistrict received agricultural business loans and also production and marketing skills improvement.

The above findings do not make sense since loans and skill improvement are delivered as part of economic relief. The goal of these programs is to improve well-being so that people can step outside poverty. The more people economically active in the labor market should contribute to better family income so that they could bear basic needs expenses. Yet, those four covariates give effects in unaccepted directions.

Since in-depth research is required to find the true direction of those 4 variables in affecting

Table 4.3: The output of the interim logistic regression model.

Coefficients:	Estimate	Std. Error	z value	Pr($>\|z\|$)	
(Intercept)	-4.3274396	0.3957121	-10.936	$<$ 2.00E-16	***
HscGr	-0.0124082	0.0020486	-6.057	1.39e-09	***
NoAgr	-0.0068088	0.0010159	-6.702	2.05e-11	***
DR	0.0111829	0.0026925	4.153	3.28e-05	***
Work	0.0203703	0.0034861	5.843	5.12e-09	***
Health	-0.0013221	0.0005776	-2.289	0.022074	*
LoAg	0.0019883	0.0005177	3.841	0.000123	***
LoNAg	-0.0019633	0.0005502	-3.569	0.000359	***
GrBus	-0.0012560	0.0005750	-2.185	0.028924	*
ProIm	0.0017004	0.0005982	2.843	0.004476	**
MarIm	0.001978	0.0007880	2.511	0.012036	*
Residual deviance:	45973				
Degrees of freedom:	101280				
AIC:	45998				

Significance code:***(0.1%), **(1%), *(5%)

poverty estimates, the roles of these variables in the poverty modeling are revoked from the analysis to avoid misleading or misinterpretation. A logistic model omitting these covariates has been refitted, followed by a backward elimination process to get a decent model. The summary of this final model is given in Table 4.4.

Table 4.4: The output of the final logistic regression model.

Coefficients:	Estimate	Std. Error	z value	Pr($>\|z\|$)	
(Intercept)	-2.2240622	0.1373694	-16.19	$<$ 2.00E-16	***
HscGr	-0.0078691	0.0019427	-4.051	5.11E-05	***
NoAgr	-0.0106343	0.0008928	-11.912	$<$ 2.00E-16	***
DR	0.0069199	0.0025183	2.748	6.00E-03	**
Residual deviance:	46074				
Degrees of freedom:	101287				
AIC:	46082				

Significance code:***(0.1%), **(1%), *(5%)

A logistic mixed model has also been enforced to improve the precision of poverty estimates derived under the direct method. The source of fixed effects for this model are covariates comprised in the final logistic model. Whereas, the random effect comes from the subdistrict variations. The summary of the model fitting is presented in Table 4.5. This table discloses the proportion of the population with a high school education or higher (HScGr) and the dependency ratio (DP) do not exhibit significant effects in the model development. Hence, this model is then refitted by applying a backward selection technique to derive an appropriate mixed logistic model as shown in Table 4.6.

Table 4.5: The output of the interim logistic mixed effects regression model.

Random effects:					
Groups	Name	Variance	Std.Dev.		
Subdist_Code	(Intercept)	1.063	1.031		
Number of obs:	101291				
Fixed effects:					
	Estimate	Std. Error	z value	Pr($>\|z\|$)	
(Intercept)	-2.709717	0.294819	-9.191	$<$ 2.00E-16	***
HscGr	-0.003081	0.004146	-0.743	0.457	
NoAgr	-0.011669	0.001924	-6.065	1.32e-09	***
DR	0.006900	0.005393	1.279	0.201	
Residual deviance:	43440.7				
Degrees of freedom:	101286				
AIC:	43450.7				

Significance code:***(0.1%), **(1%), *(5%)

From the summary of the final logistic and the final logistic mixed effect model expressed in Tables 4.4 and 4.6, it can be deduced that covariates from PODES2014 are not meaningful in the estimation process. In other words, the contribution of these covariates to both models is considered statistically insignificant. SP2010 is the only beneficial data source for estimating poverty rates in this research.

Table 4.6: The output of the final logistic mixed effect regression model.

Random effects:					
Groups	Name	Variance	Std.Dev.		
Subdist_Code	(Intercept)	1.066	1.032		
Number of obs:	101291				
Fixed effects:					
	Estimate	Std. Error	z value	Pr($>\|z\|$)	
(Intercept)	-2.813557	0.258769	-10.873	<2e-16	***
NoAgr	-0.012772	0.001214	-10.524	<2e-16	***
DP	0.009109	0.004490	2.029	0.0425	*
Residual deviance:	43441.2				
Degrees of freedom:	101287				
AIC:	43449.2				

Significance code:$^{***}(0.1\%)$, $^{**}(1\%)$, $^{*}(5\%)$

Figure 4.5 tries to contrast the direct estimates with the poverty statistics produced by final logistic and mixed logistic models. The mixed logistic model gives more diverse estimates than the logistic model. The poverty estimates under the logistic model are smooth enough which means that the poverty rates for all subdistricts have only a small variation. Subdistricts in DKI Jakarta are a clear example of this case. The poverty estimates for each subdistrict are nearly the same except in the first and second subdistricts. These two areas are located on a small island separated from the main Java Island which belongs to the rural region. It makes sense if the poverty rates in these two areas are higher than in other subdistricts in DKI Jakarta. Yet in general, the logistic model generates overly smooth poverty rates for subdistricts in this province.

When statistics are disaggregated to lower levels of area, the estimation results would normally be more heterogeneous than that of the higher levels. In this research, the mixed logistic model looks better than the logistic model since the dynamic of the subdistrict-level poverty estimates are visible. Unfortunately, this model does not work properly for many subdistricts due

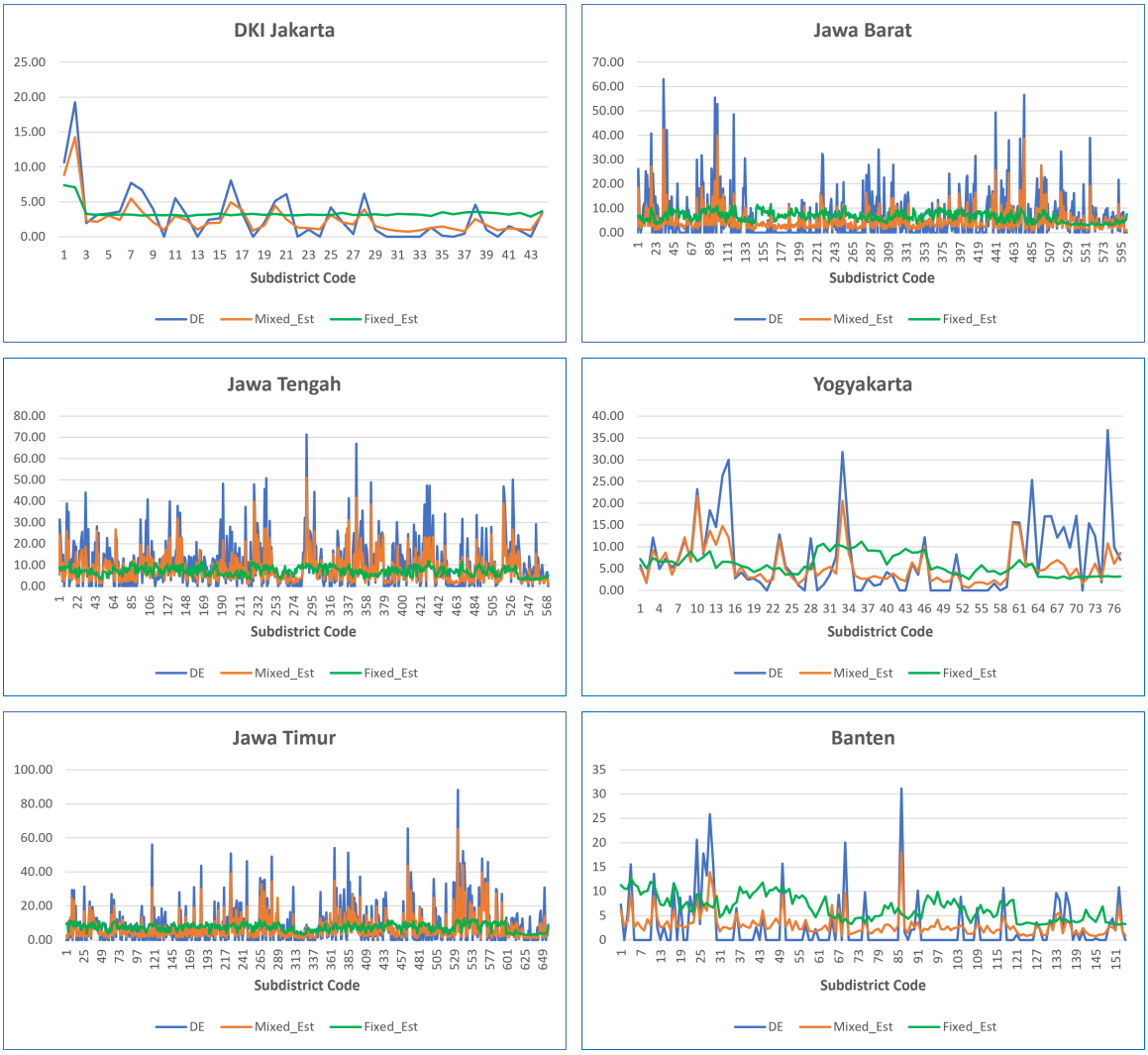

Figure 4.5: Comparison of direct estimates (DE) with final logistic model estimates (Fixed_Est) and final mixed logistic model estimates (Mixed_Est).

to the very high poverty estimates.

In terms of root mean square error (RMSE), the logistic model seems to perform better than the mixed logistic model as shown in Figure 4.6. In many subdistricts, the RMSEs under the logistic model are significantly lower than those in the mixed effect model. Normally, a researcher relies on this accuracy measure for selecting the best model. Yet in this research, the lower RMSEs do not imply better estimates.

After aggregating the subdistrict estimates under both models to the district level, the poverty estimates using the logistic model tend to be lower than the official statistics. Numerous aggregation estimates using the mixed effect model are significantly distant from the official statistics. However, some of them approach the official statistics for several districts. Hence, Figure 4.7 confirms the previous finding that the lower RMSEs do not always imply better estimates.

Figure 4.6: Comparison of the standard error of direct estimates (SE_DE) with the RMSE of the final logistic model estimates (RMSE_Log) and final mixed logistic model estimates (RMSE_Mixed).

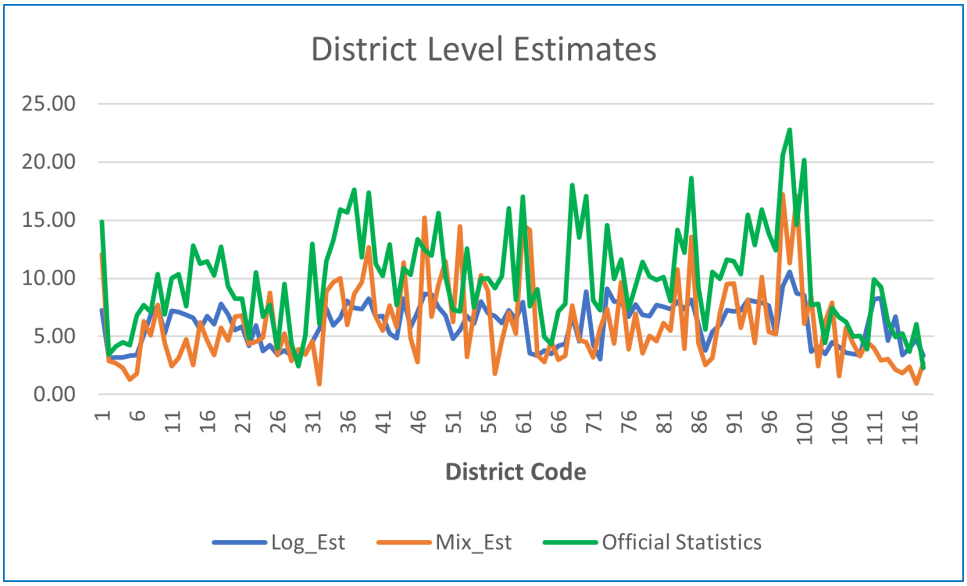

Figure 4.7: Comparison of the official statistics poverty rates with logistic and mixed logistic model estimates at the district level.

Chapter 5: Conclusion and Future Research

This research put effort to estimate poverty rates at the subdistrict level on Java Island which was never done before by the Indonesia NSO. The data source for the poverty calculation is Social and Economic Survey (SUSENAS) March 2020 under district estimation sampling design. After running three different statistical techniques to estimate poverty rates at the subdistrict level, several findings need to be highlighted as follows.

1. None of the covariates from the Village Potential Enumeration Survey (PODES2014) is helpful for the model-based estimation. Hence, the small area poverty estimation borrows strength from data produced by Population Census 2010 (SP2010) only.

2. In general, the logistic model performs poorly for subdistrict level estimation even though its RMSEs are better than those of the mixed logistic model. It produces poverty statistics with small variations between subdistricts.

3. The mixed effects model also performs weakly for many subdistricts. Yet after comparing the aggregation estimates for the district level, some estimations are close enough to the official statistics. In this case, the mixed effect model works slightly better than the fixed-effects model.

4. The expected result from this research is the aggregation of subdistrict estimates to the

district level are close enough to the official statistics. The possible reasons why the model-based estimation does not meet the goal of the research are detailed below.

a. There are 649 subdistrics on Java Island Sampled non-poor households only under SUSENAS March 2020. Hence, the direct poverty estimates for these subdistricts are zero. This asserts that the applied sampling design is not suitable for subdistrict level estimation.

b. There is a big time lag between data sources. It accounts for the missing covariates for 20 subdistricts covered by SUSENAS March 2020 since some subdistricts were split and merged. As the result, subdistricts with missing covariates are excluded from the modeling.

c. The weights assigned to the unit observations are possibly not reliable anymore for lower-level estimation.

d. There are 22 subdistricts not covered by SUSENAS March 2020 are also excluded from the analysis.

The possible future research to improve the above issues are

1. Replace the covariates with variables derived from the population census 2020 (SP2020). Since SP2020 is the latest complete enumeration survey, this will ensure that all subdistricts covered by SUSENAS March 2020 will have covariates.

2. Since 30 percent of subdistricts in SUSENAS March 2020 have zero poverty incidence, considering a zero-inflated model probably would be helpful. If this model is applied, then

the response variable should be transformed from the poverty rate to the number of poor people.

3. Apply an imputation technique to the 20 subdistricts uncovered by SUSENAS March 2020. The Indonesia NSO normally uses sister-city criteria for area imputation. The missing information for a particular subdistrict is imputed by the information from another subdistrict sharing social and economic characteristics similarities. This effort will be laborious as it requires substantial additional research for finding a suitable sister city for every subdistrict. Otherwise, the mean value imputation method can be considered instead of the prementioned technique.

4. Replicate the study conducted by Buil-Gil et al. (2020) entitled "Nonparametric Bootstrap and Small Area Estimation to Mitigate Bias in Crowdsourced Data". Their study is conducted in the following steps.

    a. Define the optimal sample size for the area of interest. Since the estimation level for the research is subdistrict level, the optimal sample size (n_{opt}) for every subdistrict needs to be determined.

    b. Calculate pseudo-sampling weights for every area based on the original sample size (n) and the optimal sample size (n_{opt}). These weights possibly work better than those assigned in the survey.

    c. Sample n_{opt} from n then calculate the statistics of interest. In this research, the statistic of interest is the poverty rate.

    d. Repeat step c to produce the bootstrap estimates of the statistics of interest. If this

technique is applied to this research, then the bootstrap poverty estimate for every subdistrict is obtained.

e. Perform model-based small area estimation method with statistics produced in step d as the response variable.

Bibliography

[1] Akinyetun, T., Alausa, J., Odeyemi, D., Ahoton, A. (2021). Assessment of the Prevalence of Multidimensional Poverty in Nigeria: Evidence From Oto/Ijanikin, Lagos State. Journal of Social Change. 13. 10.5590/JOSC.2021.3.2.03.

[2] Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. John Wiley Sons, Inc.

[3] BAPPENAS. (2020). Pedoman Teknis Penyusunan Rencana Aksi Tujuan Pembangunan Berkelanjutan (Tpb) / Sustainable Development Goals (Sdgs). Jakarta: Kedeputian Bidang Kemaritiman dan Sumber Daya Alam.

[4] BPS. (2019). Panduan Praktis Desain Survei. Direktorat Pengembangan Metodologi Sensus dan Survei.

[5] BPS. (2021). Penghitungan dan Analisis Kemiskinan Makro Indonesia 2021. Fungsi Statistik Kerawanan Sosial.

[6] Buil-Gil, D., Solymosi, R., Moretti, A. (2020). Nonparametric Bootstrap and Small Area Estimation to Mitigate Bias In Crowdsourced Data: Simulation Study And Application to Perceived Safety. In C. Hill, P. Biemer, T. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, L. Lyberg (Eds.), Big data meets survey science (pp. 487-517). (Big Data Meets Survey Science). John Wiley Sons Ltd. https://doi.org/10.1002/9781118976357.ch16.

[7] Burchi, F., Fanzo, J., Frison, E. (2011). The role of food and nutrition system approaches in tackling hidden hunger. International journal of environmental research and public health, 8(2), 358–373. https://doi.org/10.3390/ijerph8020358

[8] Butar Butar, F., & Lahiri, P. (2003). On measures of Uncertainty Of Empirical Bayes Small-Area Estimators. Journal of Statistical Planning and Inference - J STATIST PLAN INFER. 112. 63-76. 10.1016/S0378-3758(02)00323-3.

[9] Dani, R. (2007). One Economics, Many Recipes: Globalization, Institutions, and Economic Growth. Princeton University Press.

[10] Datta, G. (2009). Model-Based Approach to Small Area Estimation. Handbook of Statistics. 29. 251-288. 10.1016/S0169-7161(09)00232-6.

[11] Deming, W.E. (1950). Some Theory of Sampling. John Wiley and Sons.

[12] Dollar, D. & Kraay, A. (2002). Growth Is Good for the Poor. Journal of Economic Growth, 7(3), 195–225. http://www.jstor.org/stable/40216063.

[13] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics, Vol. 7, No. 1. (Jan., 1979), pp. 1-26.

[14] Efron, B. (1982) The Jackknife, The Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia. http://dx.doi.org/10.1137/1.9781611970319.

[15] Efron, B. & Tibshirani, R.J. (1993). An Introduction to The Boostrap. Chapman Hall.

[16] Farrell, P.J., MacGibbon, B., Tomberlin, T.J. (1997). Empirical Bayes Small-Area Estimation Using Logistic Regression Models and Summary Statistics, Journal of Business Economic Statistics, 15:1, 101-108, DOI: 10.1080/07350015.1997.10524692.

[17] Foster, J., Greer, J., Thorbecke, E. (1984) A Class of Decomposable Poverty Measures, Econometrica, 52, no. 3, 761-6.

[18] Hall, P., Maiti, T. (2006). On Parametric Bootstrap Methods for Small Area Prediction. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 68(2), 221–238. http://www.jstor.org/stable/3647567.

[19] Haughton, J. & Khandker, S.R. (2009). Handbook on Poverty and Inequality.World Bank Publications.

[20] Hwang, H. & Shin, K. (2011). Logistic Regression Type Small Area Estimations Based on Relative Error. The Korean Journal of Applied Statistics, 24(3), 445(453). DOI: 10.5351/KJAS.2011.24.3.445.

[21] Jiang, J. & Lahiri, P. (2001). Empirical Best Prediction for Small Area Inference with Binary Data. Annals of the Institute of Statistical Mathematics. 53. 217-243. 10.1023/A:1012410420337.

[22] Jiang, J. & Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. Test, 15, 1-96. https://doi.org/10.1007/BF02595419.

[23] Jiang, J. & Lahiri, P. (2006). Estimation of Finite Population Domain Means - A Model Assisted Empirical Best Prediction Approach. Journal of The American Statistical Association. In press.

[24] Jiang, J. & Nguyen, T. (2021). Linear and Generalized Linear Mixed Models and Their Applications. 2nd edition. Springer.

[25] Jones, L. J. (1994). The Social Context of Health and Health Work. Basingstoke: Macmillan Press Ltd.

[26] Kim, Y., Choi, Y., Emery, S. (2013) Logistic Regression With Multiple Random Effects: A Simulation Study of Estimation Methods and Statistical Packages, The American Statistician, 67:3, 171-182, DOI: 10.1080/00031305.2013.817357.

[27] Lahiri, P. (2003). On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation. Statistical Science, 18(2), 199–210. http://www.jstor.org/stable/3182850.

[28] Lee, S. (1994). Optimal Choice Between Parametric and Non-Parametric Bootstrap Estimates. Mathematical Proceedings of the Cambridge Philosophical Society. 115. 335 - 363. 10.1017/S0305004100072121.

[29] Lehtonen, R. & Veijanen, A. (2016). Model-Assisted Methods for Small Area Estimation of Poverty Indicators. Analysis of Poverty Data by Small Area Estimation. Wiley.

[30] Lister, R (1990). The Exclusive Society: Citizenship and the Poor. London: Child Poverty Action Group.

[31] Lohr, S.L., & Prasad, N.G. (2003). Small Area Estimation with Auxiliary Survey Data. Canadian Journal of Statistics, 31.

[32] Mudalige, P. W. (2020). Performance of Decentralized Local Service Delivery in Developed Countries: Case Studyof Japan. European Scientific Journal, ESJ ,16 (35), 40. https://doi.org/10.19044/esj.2020.v16n35p40.

[33] Ng, A., Farinda, A., Kan, F., Lim, A., Ting, T. (2013). Poverty: Its Causes and Solutions. World Academy of Science, Engineering and Technology, Open Science Index 80, International Journal of Humanities and Social Sciences, 7(8), 2471 - 2479.

[34] OECD. (2013). OECD Framework for Statistics on The Distribution of Household Income, Consumption And Wealth. Paris: OECD Publishing. Available from http://www.oecd.org/statistics/302013041e.pdf.

[35] OPHI. (2018). Global Multidimensional Poverty Index 2018: The Most Detailed Picture to Date of the World's Poorest People. Oxford.

[36] Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. Statistical Science, 28(1), 40–68. http://www.jstor.org/stable/43288411.

[37] Pratesi, M. & Salvati, N. (2016). Introduction on Measuring Poverty at Local Level Using Small Area Estimation Methods. Chapter1, Analysis of Poverty Data by Small Area Estimation. John Wiley Sons.

[38] Rao, J. N. K. (2003) Small Area Estimation. New York: Wiley.

[39] Rao, J.N.K. & Molina, I. (2015). Small Area Estimation. 2nd Edition. New Jersey: Wiley.

[40] Ravallion, M. 1998. Poverty Lines in Theory and Practice: Living Standards Measurement Study. World Bank: Working Paper No. 13.

[41] Rippin, N. (2011). A Response to the Weaknesses of the Multidimensional Poverty Index (MPI): The Correlation Sensitive Poverty Index (CSPI). Briefing paper. German Development Institute.

[42] Roemer, M. & Gugerty ,M. (1997). Does Economic Growth Reduce Poverty?, Consulting Assistance on Economic Reform (CAER) Discussion Paper No. 4, Cambridge, MA: Harvard Institute for International Development.

[43] Shao, J. & Tu, D. (1995) The Jackknife and Bootstrap. Springer-Verlag, Inc., New York. http://dx.doi.org/10.1007/978-1-4612-0795-5.

[44] Son, H. & Kakwani, N. (2004). Economic Growth and Poverty Reduction: Initial Conditions Matter. UNDP.

[45] Tarozzi, A. & Deaton, A. (2009). Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas. The Review of Economic and Statistics, 91 (4), p. 773.

[46] Tefera, T., Balana, B., Woldeamanuel, T., Tarefasa, S., Black, H. Measuring Multidimensional Rural Poverty Using Combination of Methods - A Case Study from Southern Ethiopia, Journal of Poverty, Investment and Development, vol.29, 2016.

[47] TNP2K. (2021). Pedoman Produksi Produk Pengetahuan Sekretariat TNP2K. Jakarta: Unit Pengelola Pengetahuan (Knowledge Management) Sekretariat TNP2K.

[48] UNDESA. (2010). The World's Women 2010: Trends and Statistics.

[49] UNECE. (2017). Guide on Poverty Measurement. ECE/CES/STAT/2017/4.

[50] United Nations General Assembly. (2015). Transforming our World: The 2030 Agenda for Sustainable Development.

[51] Walker, R. (2015). Multidimensional Poverty. GSDRC Professional Development Reading Pack no. 22. Birmingham, UK: University of Birmingham.

[52] Watson D.D. (2014). Poverty and Basic Needs. In: Thompson P., Kaplan D. (eds) Encyclopedia of Food and Agricultural Ethics. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6167-4_442-1.

[53] Wong, S. (2012). Understanding Poverty: Comparing Basic Needs Approach and Capability Approach. SSRN Electronic Journal. 10.2139/ssrn.2066179.