Name of author: Yuniarti  
**Organization:** Badan Pusat Statistik  
**Contact address:** Jl. Dr. Sutomo 6-8 Jakarta, 10710  
**Contact phone:** +62 85647559086  
**Email:** yuna@bps.go.id  

**Title of Paper:**  
Small area estimation for monitoring SDGs at the sub-national level  

**Abstract**

Badan Pusat Statistik (BPS) or Statistics Indonesia is responsible for serving data and information for the sustainable development goals (SDGs) through its regular or collaborative surveys. To meet this commitment, BPS established initiatives for SDGs indicators development, such as integrating new question items in the current available survey instruments, integrating new surveys in the current available surveys and exploring big data as potential data source. Indonesia also committed to implement SDGs up to sub-national level including employment and child labor which are necessary to monitor the implementation of the SDGs under Goal 8. Budget cutting in 2016 resulted to a 75 percent samples size reduction in the National Labor Force Survey (SAKERNAS). In addition, normal SAKERNAS samples are not designed to present child labor information. To fill in the gap, BPS employed small area estimation (SAE) method to estimate unemployment rate in 2016 and the proportion of children aged 5-17 years engaged in child labor in 2017. The paper presents results of using the SAE method for district level estimates for both indicators in Java Island and are further disaggregated by sex and geographical location (urban/rural).  

**Keywords:** labor, SAE, SDGs, data integration, subnational
I. Contents

<table>
<thead>
<tr>
<th>I. Contents</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>II. Introduction</td>
<td>3</td>
</tr>
<tr>
<td>III. Implementing SDGs in Sub-National Level</td>
<td>3</td>
</tr>
<tr>
<td>A. Commitment on Vertical Integration</td>
<td>3</td>
</tr>
<tr>
<td>B. BPS’ Roles on Monitoring and Evaluation</td>
<td>4</td>
</tr>
<tr>
<td>IV. SAE for Producing SDGs Indicators in Sub-National Level</td>
<td>6</td>
</tr>
<tr>
<td>A. Emerging Issues</td>
<td>6</td>
</tr>
<tr>
<td>B. Small Area Estimation (SAE) Method</td>
<td>7</td>
</tr>
<tr>
<td>C. Variables and Data Sources</td>
<td>8</td>
</tr>
<tr>
<td>D. Estimation of Unemployment Rate 2016</td>
<td>9</td>
</tr>
<tr>
<td>E. Estimation of Percentage of Working Children 2017</td>
<td>12</td>
</tr>
<tr>
<td>V. Conclusion</td>
<td>14</td>
</tr>
<tr>
<td>VI. References</td>
<td>16</td>
</tr>
</tbody>
</table>
II. Introduction

Commitment for implementing SDGs until sub-national level rises a consequence to work hard, collaborate, and cooperate with all country’s elements. National government level should create strategic plan for SDGs implementation then bring it down to local level. The way of SDGs localization is different among countries. In Indonesia, SDGs is not only incorporated into national development plan, but also in sub-national development plan. It means, SDGs is implemented in local level under local legal framework.

A substantial challenge in the SDGs enforcement is preparing effective tools for progress monitoring and evaluation. Local government should invite statistical office in order to provide indicators of achievements. BPS has taken part in this effort. Moreover, has conducted some endeavors to meet the monitoring and evaluation requirements, such as integrating new question items in the current available survey instrument, integrating new survey in the current available survey and exploring big data as potential data source. Special preparation for district level, BPS has explored the utilization of small area estimation methods to constrict the data or information availability gaps.

In this research, small area estimation method is implemented to estimate unemployment rate 2016 and proportion of child labor 2017. Both indicators are set up for every district in Java Island and disaggregated by geographic region and sex. The auxiliary variables as sources of estimation strength are acquired from population census 2010 (SP2010) and Villages Potential Enumeration 2014 (PODES2014). The result would be revealed in the next discussion.

III. Implementing SDGs in Sub-National Level

A. Commitment on Vertical Integration

Sub-national level, specifically local governments, hold crucial role in every segment of country’s development. The aggregate of local development process would determine the success of national growth. Vadaveloo and Singaravelloo (2013) argued that local development offers a practice that is a part of a process of social change based on the sharing of integrity, skills, knowledge and experience. The goal is to build assets that increase the capacity of residents to improve their quality of life (Green and Haines, 2012). Both also stated that development controlled by local governments provides a better match between the assets and the need of the communities, because local governments are closer to the action spot (Oduro-Ofori, 2011) and have access to more resources (Morgan, 2009).

Indonesia has declared a commitment to implement SDGs in sub-national level through Presidential Decree No. 59/2017 regarding Achieving the Sustainable Development Goals. This regulation manages clearly the institutional arrangement, implementation strategy for SDGs goals and targets, monitoring, evaluation and reporting mechanism, financial policy, and coordination between national and sub-national governments. Then the next work is incorporating SDGs into National Development Planning System (Figure 1).
Currently, Indonesia is embracing decentralization governance system called autonomy. Decentralization brings decision making closer to citizens either through administrative reforms or devolution to lower level of government (Čapková, 2005). But local governments tend to have various local issues that different among regions. Therefore, Indonesia prepared a five year national action plan, 6 months after Presidential Decree was ratified, then forwarded to local government as guidance. There are 15 Provinces have constructed regional action plan on SDGs as of 2018.

B. BPS’ Roles on Monitoring and Evaluation

Regional action plan on SDGs is Indonesian strategy for localizing the SDGs through legal framework. It means that responsibility is disaggregated across government levels to achieve SDGs targets (Patole). The assessment on local progress will be conducted on each goal. Monitoring and evaluation SDGs progress would be challenging as every region has specific characteristics on social, economic, geographic, and governmental issues (ICLEI, 2015). Therefore, stakeholders should merge with statistics office to develop indicators required.
BPS handles a responsibility for proving data and information concerning SDG indicators through regular and collaborative surveys. The result of capacity assessment to compile indicators for SDGs monitoring mentioned that BPS would provide 96 indicators through its regular surveys (Figure 2). Wherein, 31 of them are global indicators and the remains are proxy of global indicators. In addition, BPS has constructed some initiatives to fill SDGs data gaps as follows:

1. Inserting new question items in the current available survey instrument, such as
   - Inserting 8 questions about household’s food insecurity experience adopted from FAO in National Social Economic Survey (SUSENAS) 2017 to produce data for Food Insecurity Experience Scale (FIES).
   - Inserting questions related to disability status in SUSENAS.

2. Integrating new survey in the current available survey
   - Conducting Water Quality Survey 2015 (pilot survey) integrated with SUSENAS to produce proxy indicator for access to safe drinking water.

3. Conducting new surveys
   - Conducting Women Life Experience Survey 2016 in collaboration with Ministry of Women Empowerment and Child Protection to produce indicator prevalence of violence against women.

4. Adopting and adapting standard method for calculating global SDG indicators
   - Adopting Prevalence of Undernourishment (PoU) calculation method suggested by FAO, applied to SUSENAS and supporting data from Ministry of health.

5. Exploring the possibility of using big data as data sources
   - Utilizing a roaming data to improve the foreign tourism indicator
IV. SAE for Producing SDGs Indicators in Sub-National Level

A. Emerging Issues

Mainstreaming SDGs to local level emerged a consequence on regular tracking progress. SDGs implementation progress can be measured sensibly when a country has effective indicators’ design. Effective indicators must be measurable, relevant, reliable and comprehensible (ICLEI, 2015). But most countries, including Indonesia, often face a problem on the data generation for SDGs. An issue regard to data availability on required disaggregation has forced countries to compromise with their readiness on producing SDGs statistics.

Most of BPS’ household surveys are designed to produce estimations in 3 level areas, national, provincial and district. One of BPS’ major household surveys is National Labor Force Survey (SAKERNAS) organized twice a year in February and August. This survey is designed to generate employment indicators, such as unemployment rate, proportion of formal and informal employment, and average earnings per hour. SAKERNAS-August is prepared for district level estimation.

SAKERNAS-August usually considers 300,000 households as its sample respondents. Due to budget limitation in 2016, this survey should work with 25% of usual sample size. It means, only 75,000 households were involved as survey’s respondents. This sample size is adequate for national and provincial level, but not district level. As the sample size was back to normal in 2017, BPS suffers from employment indicator gaps for every district in 2016. Absolutely, the absence of employment indicators has caused a disappointment for district governments as they could not rely on SAKERNAS for employment policy withdrawal.

Another issue is children around the world are routinely engaged in many forms of works, whether paid or unpaid, not harmful to harmful. A lot of reasons lie behind this incidence. Statistics should be able to capture their involvement in the labor force. Evidence on their number or prevalence would contribute to nurturing children’s rights.

Indonesian statistics office has not ready to provide child labor indicators. SAKERNAS with current sampling design is not powerful to produce this indicator. With 300,000 sample households, some districts suffer from zero respondents. The incidence of child labor could not be well captured through the respective sample households. Therefore, this study would explore a substantial statistical approach to close the information gaps for both employment and child labor indicators. Both indicators are important to support SDGs under goal 8.
B. Small Area Estimation (SAE) Method

The two emerging issues as mentioned above have built an initiative to use small area estimation (SAE) methods for producing SDGs indicators. Rao and Molina (2015) explained that small area estimation deals with the problem of producing reliable estimates of parameters of interest and the associated measures of uncertainty for subpopulations of a finite population for which samples of inadequate sizes or no samples are available. Subpopulation refers to both “small geographic area” (districts, sub districts or villages) and “small domain”, such as age, sex, race group of people within a large geographical area (Ghosh and Rao, 1994). Having only a small sample (and possibly an empty sample) in a given area, the only possible solution to the estimation problem is to borrow strength (information) from other related data set to increase effective sample size (Rao, 2003; Pfeffermann, 2002).

Generally, SAE could be executed based on design or model (Rao and Molina, 2015). Design based approach fully relies on sampling design. Meanwhile, model based approach relies on auxiliary information incorporated in to estimation model. Regard to the availability of auxiliary variables, SAE could estimate the interest information for both area and unit level. In order to fulfill SDGs need, BPS will employ model based approach for area (district) level under Fay-Herriot model.

Fay-Herriot model assumes that parameter estimator, $\hat{\theta}_i = g(\bar{Y}_i)$, is related to area-specific auxiliary data $x_i = (x_{1i}, x_{2i}, ..., x_{pi})^T$ through linear model as follow:

$$\theta_i = x_i^T \beta + v_i, \quad i = 1, 2, ..., m \text{ (area)} \quad (1)$$

where $\beta = (\beta_1, \beta_2, ..., \beta_i)^T$ is the $px1$ vector of regression coefficients and $v_i$’s are area-specific random effects assumed to be independent and identical distributed $v_i \sim (0, \sigma_v^2)$.

Inference on the small area means $\bar{Y}_i$ is made under assumption that direct estimators $\bar{Y}_i$ are available. Again it is assumed that

$$\bar{Y}_i = g(\bar{Y}_i) = \theta_i + e_i, \quad i = 1, 2, ..., m \quad (2)$$

where $e_i$’s are known independent sampling error $e_i|\theta_i \sim (0, \psi_i)$.

Combining equation 1 and 2 then we obtained what is called Fay-Herriot model

$$\hat{\theta}_i = x_i^T \beta + v_i + e_i$$

where $v_i$ and $e_i$ are independent.
The next step is applying empirical best linear unbiased prediction (EBLUP) to estimate small area statistics under Fay-Herriot model. The linear and unbiased (BLUP) estimator $\tilde{\theta}_i = x_i^T \tilde{\beta} + \tilde{v}_i$ which minimize $\text{MSE}(\tilde{\theta}_i) = E(\tilde{\theta}_i - \theta_i)^2$ is

$$\tilde{\theta}_i^{BLUP} = x_i^T \tilde{\beta} + \tilde{v}_i$$

where

$$\begin{align*}
\tilde{\beta} &= \left( \sum_{i=1}^m y_i x_i x_i^T \right)^{-1} \sum_{i=1}^m y_i x_i y_i \\
\tilde{v}_i &= y_i (y_i - x_i^T \tilde{\beta}) \\
y_i &= \frac{\sigma^2_v}{\sigma^2_v + \psi_i}
\end{align*}$$

In practice, $\sigma^2_v$ is unknown then $\tilde{\theta}_i^{BLUP}$ depends on $\sigma^2_v$ through $\tilde{\beta}$ and $y_i$:

$$\tilde{\beta} = \tilde{\beta}(\sigma^2_v), \quad \tilde{\theta}_i^{BLUP}(\sigma^2_v)$$

The empirical BLUP (EBLUP) of $\theta_i$ is done by replacing $\sigma^2_v$ in the BLUP by an estimator $\hat{\sigma}^2_v$. The $\sigma^2_v$ can be estimated using maximum likelihood (ML) or restricted maximum likelihood (REML) method.

$$\hat{\theta}_i^{EBLUP} = \hat{\theta}_i^{BLUP}(\hat{\sigma}^2_v), \quad i = 1, 2, ..., m$$

$$\hat{\theta}_i^{EVLUP} = \hat{y}_i y_i + (1 - \hat{y}_i) x_i^T \tilde{\beta}$$

BPS entrusts R software to perform EBLUP estimation.

C. Variables and Data Sources

This study intends to estimate two SDGs indicators under goal 8, unemployment rate 2016 and proportion of children aged 5-17 years engaged in child labor 2017 for district level. Unemployment rate is useful measure of the underutilization of the labor supply. Meanwhile, proportion of child labor is matter for eliminating of the worst forms of child labor. Both indicators are calculated using SAKERNAS-August and disaggregated by geographical location (urban/rural) and sex.

As Indonesia is an archipelago country, every major island tends to have different social, economic, and geographical characteristics. This study will focus on districts’ estimation in Java Island with assumption that they share characteristic similarities. Java Island consists of 119 districts which represent 6 provinces, i.e. DKI Jakarta, Jawa Timur, Jawa Tengah, DI Yogyakarta, Jawa Timur, and Banten.
Small area estimation with model based approach requires auxiliary variables from other related sources. The estimation of unemployment rate 2016 and proportion of child labor 2017 will borrow information from several data sets with detailed variables as follows:

1. Population Census 2010 (SP2010)
   - X1: proportion of male population, 2015 (SP2010 projection)
   - X2: dependency ratio, 2015 (SP2010 projection)
   - X3: proportion of population high school graduates and higher, 2010
   - X4: proportion of population work in agricultural sector, 2010
   - X5: proportion of literate people, 2010
   - X6: proportion of population capable to speak Bahasa, 2010

2. Villages Potential Enumeration 2014 (PODES2014)
   - X7: proportion of villages with access to public transportation, 2014
   - X8: proportion of villages with access to agricultural business loan, last 3 years
   - X9: proportion of villages with access to non-agricultural business loan, last 3 years
   - X10: proportion of villages received grant for productive business improvement, last 3 years
   - X11: proportion of villages with production skills improvement program, last 3 years
   - X12: proportion of villages with marketing skills improvement program, last 3 years
   - X13: proportion of villages with community strengthening program, last 3 years
   - X14: proportion of villages with public infrastructures improvement, last 3 years

3. Additional variable
   - X15: proportion of 4 major sectors (manufacturing, agriculture, construction, trade) to total Gross Regional Domestic Products (GRDP), 2015

D. Estimation of the Unemployment Rate 2016

✓ Unemployment Rate 2016

Estimation model:

\[
\hat{\theta}^{UR}_{i} = \hat{\gamma}_{i}y_{i} + (1 - \hat{\gamma}_{i})(-0.447 + 1.124x_{3i} - 0.088x_{3i} - 0.075x_{4i} - 0.056x_{11i} + 0.049x_{12i})
\]
Figure 3. Unemployment Rate in Java Island by District, 2016

✓ Urban Unemployment Rate 2016

Estimation model:
\[
\hat{\theta}_i^{URBAN} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)(-0.536 + 0.748x_{1i} + 0.101x_{2i} + 0.140x_{6i} + 0.028x_{7i} - 0.055x_{11i} + 0.049x_{12i} + 0.053x_{15i})
\]

Figure 4. Urban’s Unemployment Rate in Java Island by District, 2016

✓ Rural Unemployment Rate 2016

Estimation model:
\[
\hat{\theta}_i^{RURAL} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)(-0.353 + 0.900x_{1i} + 0.207x_{2i} - 0.120x_{4i} + 0.142x_{5i} - 0.249x_{6i})
\]
✓ **Male’s Unemployment Rate 2016**

Estimation model:

\[
\tilde{\theta}_i^{MALE} = \hat{y}_i y_i + (1 - \hat{y}_i)(-0.392 + 0.819x_{3i} + 0.165x_{2i} - 0.069x_{4i} - 0.040x_{11i})
\]

✓ **Female’s Unemployment Rate 2016**

Estimation model:

\[
\tilde{\theta}_i^{FEMALE} = \hat{y}_i y_i + (1 - \hat{y}_i)(-0.695 + 1.453x_{8i} + 0.071x_{9i} - 0.053x_{11i} + 0.088x_{12i} - 0.054x_{13i})
\]

The higher unemployment rate 2016 tends to occur in western part of Java Island, either the total unemployment rate (Figure 3) or its disaggregation by geographic region (Figure 4 and 5) and sex (Figure 6 and 7). Most districts located in the concerned area are industrial center, including all districts in DKI Jakarta. Since manufacturing sector still confidently drives the country’s economy, it should be realized that industrial center will also actively pull people from diverse region. People tend to lean on this sector to find decent works. The consequence from over moving people to industrial center is the increasing of unemployment rate.
E. Estimation of the Proportion of Working Children 2017

Originally, one of the purposes of this study is estimating the proportion of child labor 2017 by districts in Java Island. UNICEF explained that standard indicator definition for child labor for SGDs metadata of indicator 8.7.1 are

1. Age 5-11 years: at least 1 hour of economic work or 28 hours of unpaid household services per week.
2. Age 12-14 years: at least 14 hour of economic work or 28 hours of unpaid household services per week.
3. Age 15-17 years: at least 43 hour of economic or unpaid household services per week.

This definition is seizing profound attention of BPS and stakeholders. Discussion on this matter is still conducted intensively. As this definition is excluded, this study would produce estimation of the proportion of working children as proxy indicator for the proportion of child labor. The targeted respondent of SAKERNAS is people aged 10 years and above. Thus, the estimation of working children will cover those of aged 10-17 years.

✓ Working Children 2017

Estimation model:
\[ \hat{\theta}_{i}^{WC} = \hat{y}_i y_i + (1 - \hat{y}_i)(0.144 - 0.061x_{3i} + 0.035x_{10i} - 0.113x_{14i} + 0.029x_{15i}) \]

✓ Urban Working Children 2017

Estimation model:
\[ \hat{\theta}_{i}^{URBAN} = \hat{y}_i y_i + (1 - \hat{y}_i)(0.229 - 0.111x_{2i} - 0.180x_{3i} - 0.089x_{4i} + 0.032x_{10i} - 0.079x_{14i}) \]

![Figure 8. Proportion of Working Children in Java Island by District, 2017](image)
Rural Working Children 2017

Estimation model:

\[
\hat{\theta}_{i}^\text{RURAL} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)(-0.022 + 0.042x_{4i} + 0.060x_{13i} + 0.070x_{15i})
\]

Male Working Children 2017

Estimation model:

\[
\hat{\theta}_{i}^\text{MALE} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i)(-0.154 + 0.648x_{4i} - 0.194x_{2i} - 0.161x_{3i} + 0.036x_{11i})
\]
Female Working Children 2017

Estimation model:

$$\hat{\theta}_i^{\text{FEMALE}} = \hat{Y}_i + (1 - \hat{Y}_i)(0.213 - 0.066x_{3i} - 0.034x_{4i} + 0.038x_{10i} - 0.186x_{14i} + 0.031x_{15i})$$

Figure 12. Female’s Proportion of Working Children in Java Island by District, 2017

The proportion of working children in Java Island is closely randomly distributed among districts (Figure 8). But comparing the number between geographical regions, rural areas deposit higher prevalence of working children than that of urban areas (Figure 9 and 10). Statistics claimed that urban areas relatively have less percentage number of poor people compared to rural areas. It rises a though if there is strong correlation between poverty incidence and prevalence of working children. In addition, higher prevalence of working children in rural area is may also affected by sturdy culture or rural wisdom. People, including children, have high willingness to help each other in any kind of works. Contrasting the statistics by sex, it is known generally that the proportion of female worker is less than that of male (Figure 11 and 12).

V. Conclusion and Way Forward

Santoso (2017) stated that monitoring and evaluation to the SDGs achievements should be able to answer the following question:

1. Strategy and direction: Have we done correctly?
2. Management and governance: Have we implemented the plan effectively?
3. Output: have we gotten expected outputs?
4. Absorption: Are people able to access and share welfare?
5. Outcome: What changes have been made?
6. Context: How changes in political, economic, social and organizational aspects would influence the plan and expected result?
All those points are not designated for central government only, but also committed local entity, especially district government. Local governments prone to data availability lack. Small area estimation method offers more comprehensive matrix of SDGs monitoring data and information in order to answer those questions.

Small area statistics delivered in this study have not reached their perfect shape yet. For instance, the unemployment rate 2016 estimation model considered proportion of population high school graduates and higher 2010 (X3) in negative direction. It seems impossible as most formal employment tend to take well educated labors. Employers often put high school graduate as labors’ minimum education requirements. But a trend in society is undeniable that bachelor graduates and higher demand reasonable works which are aligned with the worth of their education background. This tendency often contributes to high unemployment rate. Hence, people who hold high school and bachelor (or higher) education should be detached. It may improve the estimation model.

Generating reliable small area statistics is intricate and time consuming. Experts often confront these following issues during their research:

1. Time lag between estimated indicator and its auxiliary variables are too long. This study can be a respective example. Proportion of working children is estimated for 2017 but some auxiliary variable are taken from population census 2010. It brings a notion to employ more recent data set as source of auxiliary variables to gain more reliable results. For the case of Indonesia, the possible attached data set is recent survey. For upcoming study, SAKERNAS 2017 used to estimate proportion of working children will be engaged with National Social Economic Survey (SUSENAS) 2016 as source of auxiliary variables. As the auxiliary variables contain sampling error, the small area estimation method applied will be modified by involving error measurement.

2. Implement single estimation method EBLUP for all types of estimated variables. EBLUP could be implemented when the response variable is continuous under normal distribution assumption. When response variable is discrete, Bayesian approach can be adopted. We should be careful with variables in the percentage form. Applying EBLUP to percentage without special consideration could result out of range estimation.

3. Rare characteristics, usually less than 10% of the population, are tough to model (McEwin and Elazar, 2006). Originally, this study has also explored the estimation of unemployment rate 2016 for district level disaggregated by age groups. Writer defined age groups of a 5 year and found that the number of unemployed people in the age groups of above 35 is extremely rare. Consequently, the estimation system is difficult to produce convergent results. For future study, unemployment rate for small area will be estimated through different stream to avoid that problem. The number of unemployed people (as numerator) and the number of people aged 15 years and above (as denominator) will be estimated separately.

4. Determining the quality gate indicators for output approval is often troublesome. If small area estimation aims to capture current condition, then conducting ground check will be helpful to confirm the estimation results. But when estimation is for past, such as unemployment rate 2016, benchmarking the estimation output by its history would
be better. Otherwise, in-depth discussion with subject matters is necessary for estimation output acknowledgment. The respective subject matters could be related ministries or other agencies.

Small area estimation result is useful for closing data gaps in sub national level. With all its imperfections, small area statistics are highly important as bases for implementing action to help the vulnerable people. Direction for local development is getting obvious in order to achieve the SDGs. No more blind decisions established as various small area estimation methods can be hired.

VI. References


