TECHNICAL NOTES ON THE GENERATION OF 2015 SMALL AREA ESTIMATES OF POVERTY

I. Background

The need for local level poverty statistics, particularly for planning and targeting has been recognized even before the Sustainable Development Goals (SDGs). However, available official poverty statistics directly estimated from the Family Income and Expenditure Survey (FIES) are only available, at most, at the provincial level. This is in consideration of the sampling designs of household surveys like the 2015 FIES, which has regions as domain and can therefore only provide reliable estimates for regions and some provinces.

One solution to address this concern is the use of small area estimation (SAE) technique. There are numerous SAE techniques that can actually be used to generate statistics at the local area. One of these techniques is a methodology developed by Elbers, Lanjouw and Lanjouw (ELL) of the World Bank (WB). Such methodology requires the use of census and survey data sets conducted on the same year. For the case of the Philippines, this started in the year 2000. Consequently, a Poverty Mapping Project implemented by the former National Statistical Coordination Board (NSCB) in 2005 with funding assistance from the World Bank, used the ELL method to generate the Municipal and City level poverty statistics for 2000. The project made use of the FIES, LFS and CPH data sets that were all gathered in the same year, 2000, as required in the methodology. More so, the methodology in the project made use of a single regression model[1] for the whole country to predict the family income per capita in logarithmic form.

An update of the 2000 Municipal and City level poverty estimates was again generated by the former NSCB with funding assistance from the WB. This time, however, is different from the previous project since municipal and city level poverty estimates were generated for 2003, a year when there was no census and only a nationwide survey was conducted, which is the usual source of poverty statistics. Thus, in updating the small area poverty estimates from the census year 2000 to the intercensal year 2003, a

[1] Regression is a statistical tool used to predict one variable using other variables/information. For example, one can predict a salesperson's total yearly sales using information on age, education and years of experience of the sales person.

slightly different approach was used. Variables were limited to time invariant variables. And the same methodology was adopted for 2006, 2009 and 2012 city and municipal level poverty estimates. But for 2015, since the survey and the census were conducted in the same year, the same methodology used in the 2000 was applied except that regional models were developed instead of a national model, which was used in the 2000 exercise recognizing varying characteristics of regions.

For reference years 2000, 2003, 2006, 2009, 2012 and 2015, different SAE of poverty regression models were developed. However, it should be noted that the same ELL methodology was implemented across the reference years to generate city and municipal level poverty estimates. The models may be different but the methodology utilized is the same. Using the standard errors generated, one can compare the estimates. However, it must be noted that in 2011, refinements in the computation of the poverty threshold was adopted and back estimates were only generated and utilized in the project starting with the 2006 sae on poverty. Hence, the poverty data series for municipalities and cities that are considered comparable are for 2006, 2009, 2012 and 2015.

Further, the resulting estimates are still treated as a result of study/exercise and are not yet part of the official poverty statistics, which Philippine Statistics Authority (PSA) regularly generates. The methodology for the generation of city and municipal level has yet to be further reviewed and studied for approval and adoption by the PSA Executive Board.

II. Methodology

Similar to previous exercises conducted by PSA in the generation of city and municipal level poverty estimates, the small area estimation technique called the Elbers, Lanjouw and Lanjouw (ELL) methodology of World Bank was employed in the generation of the 2015 city and municipal level poverty estimates. The general idea of the ELL is to "borrow strength" from multiple data sources to generate local level poverty estimates. In particular, for this case, it combines information from a household survey, which contains data on income or expenditure, but has limited sample size given the complexity of information collected, and the census, which has complete coverage and therefore can provide reliable estimates at the local level, but does not have income or expenditure information, which are used for poverty estimation.

A. Data Source

In the generation of the 2015 municipal and city level poverty estimates in the Philippines, the following data sets were used:

a. 2015 Official Provincial Poverty Thresholds

These are the provincial poverty thresholds (disaggregated by urban and rural areas), which are part of the official poverty statistics released by the PSA on 27 October 2016 based on PSA Board Resolution No. 1, Series of 2017-171 and 171 Annex. It may be noted further that these are not consistent with the updated 2015 first semester poverty thresholds released with 2018 on 10 April 2019, which already considered the updating of the market basket for the collection of prices for the Consumer Price Index (CPI).

b. 2015 Family Income and Expenditure Survey (FIES)

The FIES is a nationwide survey conducted by the PSA every three years. It uses a 70-page questionnaire to collect information on household income and expenditure, as well as, some socio-demographic characteristics of the family. It is the main source of income and expenditure data, which are used for the estimation of the official poverty statistics in the country. The 2015 FIES is a regular module of the Integrated Survey of Households (ISH), with sampling design still based on the 2013 Master Sample with regions as its domain. It was conducted among 40,056 sample households, distributed across the 17 regions in the country.

c. January 2016 Labor Force Survey (LFS)

The LFS is another regular module of the ISH conducted by the PSA every quarter of the year. It collects data on the demographic and socio-economic characteristics of household members 15 years old and over and the major source of official employment data in the country. The FIES is a rider to the LFS. Hence, most, if not all, of sample households in the 2015 FIES are also available in the January 2016 LFS. This is particularly useful for the small area estimation as the FIES and LFS can be combined to get family and individual level information.

d. 2015 Population Census (POPCEN)

In general, the Census of Population, which is a complete enumeration of the population in the country is conducted at least every ten years. It is a vital source of information on the composition of the population and characteristics of their housing units. It covers all areas under the jurisdiction of the Philippines as defined by the 1987 Constitution. In between the decennial census, a mid-decade census is usually conducted mainly to update the population in the country. Unlike the decennial census, this has more limited information on demographic and housing characteristics. Nonetheless, it still contains a wealth of information that can be useful for the small area poverty estimation.

In summary, the data sources for the 2015 municipality and city level poverty estimates are as follows:

Table A1. Data Sources

Data	*Reference Period*	*Source Institution*	*Frequency of Conduct*	*Part of the System of Designated Statistics (SDS)*
Census of Population	2015	PSA	Usually every 5 years	Included in the SDS
Family Income and Expenditure Survey	2015	PSA	Every 3 years	Included in the SDS
Labor Force Survey	January 2016	PSA	Every quarter	Included in the SDS
Official Provincial Per Capita Poverty Thresholds (PT)	2015	PSA	Every 3 years	Included in the SDS

Note: The SDS is a mechanism that identifies and generates the most critical and essential statistics required for social and economic planning/analysis based on approved criteria.

B. Concepts and Definitions

1. **Poor** – Based on Republic Act 8425, otherwise known as Social Reform and Poverty Alleviation Act, dated 11 December 1997, the poor refers to individuals and families whose income fall below the poverty threshold as defined by the government and/or those that cannot afford in a sustained manner to provide their basic needs of food, health, education, housing and other amenities of life. It may be estimated in terms of percentages (poverty incidence) and total number of poor families (magnitude of poor families).

2. **Poverty Threshold** - the minimum income required for a family/individual to meet the basic food and non-food requirements. Basic food requirements are currently based on 100% adequacy for the Recommended Energy and Nutrient Intake (RENI) for protein and energy, and 80% adequacy for other nutrients. On the other hand, basic non-food requirements is indirectly estimated by obtaining the average ratio of food to total basic expenditure of a reference group of families around the subsistence incidence in the 2000, 2003, 2006 and 2009 FIES. Total basic expenditure covers: 1) clothing and footwear; 2) housing; 3) fuel, light, water; 4) maintenance and minor repairs; 5) rental of occupied dwelling units; 6) medical care; 7) education; 8) transportation and communication; 9) non-durable furnishings; 10) household operations; 11) personal care & effects; and 12) food. Mathematically, poverty threshold is computed as:

$$Poverty\ Threshold = \frac{Food\ Threshold}{Food\ Expenditure/Total\ Basic\ Expenditure}.$$

Food Expenditure (FE) is actual food expenditure of families in the FIES that are within the +/- ten percentile of the food threshold while *Total Basic Expenditure (TBE)* is total basic expenditures of families in the FIES that are within the +/- ten percentile of the food threshold. The average FE/TBE generated from the 2000, 2003, 2006 and 2009 FIES is currently being used. This is equivalent to 0.6983.

3. **Poverty Incidence** - the proportion of families/individuals with per capita income less than the per capita poverty threshold to the total number of families/individuals. This is usually expressed in percent.

4. **Standard Error (SE)** – a measure of the precision/accuracy of the estimate or gives the average amount the estimate differs from the actual value in the population. SE is inversely proportional to the sample size; the larger the sample size, the smaller the standard error of the estimate.[2]

5. **Coefficient of Variation (CV)** – a measure of the reliability of an estimate. CV (usually expressed in percent) is estimated as the ratio of the standard deviation/error relative to the estimate. Relatively small CV of an estimate reflects

[2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2013). An introduction to statistical learning: with applications in R. New York: Springer, p.66.

smaller variability between repeated measures and thus, indicates a more reliable or consistent estimate.

6. **Confidence Interval** – refers to the upper and lower limit of values that the true value of the estimate will lie given a certain level of confidence. The wider confidence interval for a particular estimate at a certain level of probability, will mean the large variety of estimates the true value can be equal to, hence, caution is required when using the estimate.

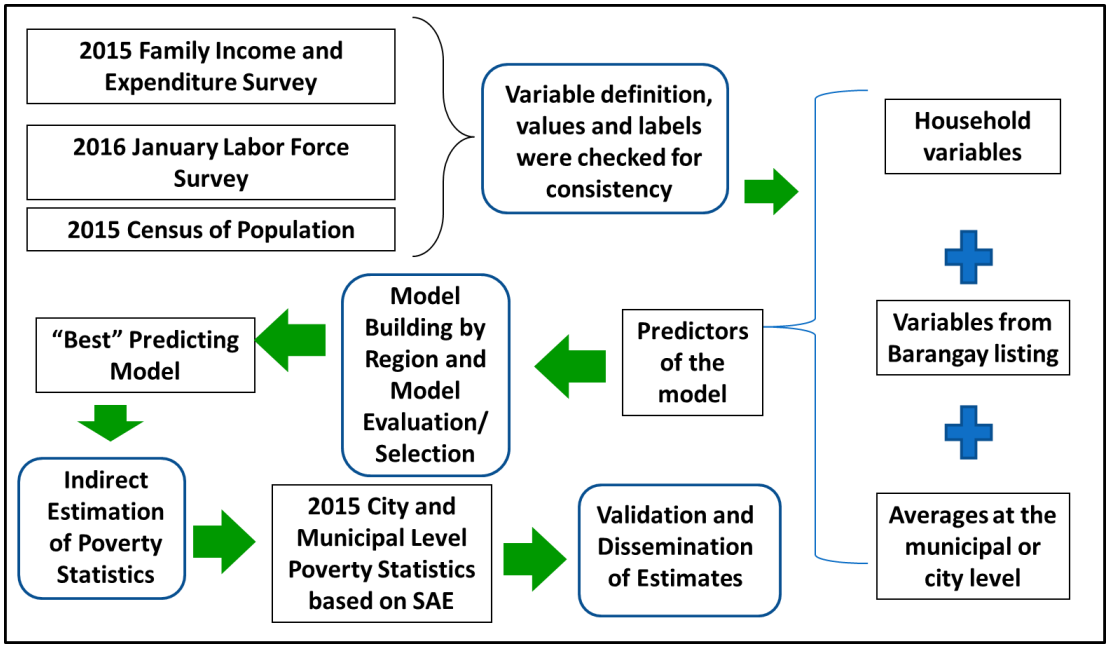
C. Implementation of Methodology

This section presents a general perspective on small area estimation following the ELL method implemented in the Philippines for the censal year 2015. As the methodology used in this updating is similar to the previous poverty mapping project up to a certain extent, most part of this section are lifted directly from the previous *2012 Municipal and City Level Poverty Estimates* report.

Basically, the procedure is as follows:

- Use the household survey data, in this case a merged datafile of the 2015 FIES and the January 2016 round of the LFS, to estimate a model of per capita income (Y) as a function of variables that are common to both the household survey and the census (X's). (*see Statistical Modeling*)
- Use the resulting estimated equation/model to predict per capita income for each household in the census.
- Compare the predicted per capita income with the corresponding poverty threshold and tag as Poor=1 if per capita income is below per capita poverty threshold and Poor=0 otherwise.
- The number of poor are then aggregated for small areas, such as cities and municipalities to estimate corresponding poverty incidence and other poverty measures.

To illustrate, the procedure can be characterized as follows:



a. Selection of Explanatory Variables

Similar to the earlier poverty mapping project, X can be classified into two types: the survey variables, at the household or individual level (e.g., educational attainment of household head, etc.); and the census-derivable location variables, which correspond to barangay or municipal means (e.g., existence of a market in the barangay). It is important that X used in modelling should be (a) available both in the survey and census; (b) comparable and/or consistent with both the survey and census (i.e., X follows the same definition in both survey and census) and (c) have survey and census statistics (mean value) that match.

It may be noted that the overall objective is to compute city and municipal level poverty statistics, with reliable and/or acceptable levels of precision. This can be done by modelling income using various X's and fitting the resulting model using its census counterpart. Once this has been done, predicted (per capita) income for all family units in the population will be derived. Effectively, strength is borrowed from the census which has a larger coverage than the survey. Note that such procedure requires that the variables constituting X should also be available from the census. In addition to availability, comparability is also an essential component in order to make the substitution of X with its census counterpart valid in the computation of predicted (per capita) income.

Selection of survey explanatory data can be done by examining the survey and census questionnaires to identify which questions elicit equivalent information. In several cases, equivalence may be achieved by collapsing some categories of answers. When common variables have been identified, the appropriate summary statistics are compared for the survey and the census data. For variables to be considered as consistent, summary statistics for the census data should be within the confidence interval of the survey. Comparability assessment is not required for the case of location-effect variables as these are mainly sourced from the census, which were merged with the survey; and as long as the geographic configurations between survey and census are the same.

b. Statistical Modelling

This section provides a brief discussion of the regression modelling for per capita income (Note: Please refer to the Estimation of Local Poverty in the Philippines released in 2005 for discussion of statistical concepts such as multicollinearity, heteroscedasticity, modelling, and bootstrapping.).

Recall that the dependent variable Y is expressed at the household level. To capture a significant amount of variability of Y, it is operationally useful to consider many variables correlated to poverty including interactions among the variables. Interactions of explanatory variables with urbanity were also considered. These approaches created more household-level auxiliary data.

Separate models were fitted for each geographic region. The objective is to tailor the model to account for the differences of geographic regions in the country, such as spatial peculiarities. The set of geographic *barangays* comprise the clusters. Per geographic region, computing through PovMap begins in the estimation of the income function,

$$\ln Y_{ij} = E[\ln Y_{ij} \mid \underline{\mathbf{X}}_{ij}] + u_{ij} \qquad (1)$$

where Y_{ij} is the per capita income of j^{th} household in i^{th} cluster, $\underline{\mathbf{X}}$ is the explanatory variable and u is the error component. This error component u_{ij} can be attributed into two components: variability among the clusters and variability among households. Thus, we can represent u_{ij} as,

$$u_{ij} = h_i + e_{ij} \qquad (2)$$

where h_i is the cluster component and e_{ij} is the household component. For each region, a number of candidate models were estimated. As mentioned earlier, estimation of these models was implemented using PovMap.

c. Development and Selection of Final Model

After model estimation and fitting of parameter estimates to census, it is necessary to undo the log transform used for Y, also implemented through PovMap. The set of official provincial poverty thresholds for the year 2015 was used to compute poverty estimates. Bootstrap estimates were summarized by their mean and standard deviation giving a point estimate and standard error for the desired level of disaggregation. Bootstrapping is used to provide accurate estimates of the standard errors. As imputed income depends non-linearly on the stochastic variables involved (the estimated model parameters, the correlated error terms), computing the standard errors analytically will be very demanding.

Assessment of candidate models for each region involved comparison of similarity of (subset of) parameter estimates, consistency of the relationship of the variable based on previous studies on poverty correlates and similarity of small area estimates, in addition to basic statistical criterion such as adjusted R squares, among others. This approach of assessment is also useful in identifying over-fitted models, aberrant fluctuations as well as robustly significant variables. Further, the resulting model-based poverty estimates at the regional levels were also compared to direct survey estimates, to ensure that resulting regional and provincial estimates from the SAE will not deviate much to officially released poverty estimates. The generated CV's were also another consideration such that small CV's among city and municipal level estimates are desired.