Philippine Institute for Development Studies
Surian sa mga Pag-aaral Pangkaunlaran ng Pilipinas

Small Area Estimation of Poverty Statistics

Zita Villa Juan-Albacea

DISCUSSION PAPER SERIES NO. 2009-16

July 2009

For comments, suggestions or further inquiries please contact:
The Research Information Staff, Philippine Institute for Development Studies
5th Floor, NEDA sa Makati Building, 106 Amorsolo Street, Legaspi Village, Makati City, Philippines
Tel Nos: (63-2) 8942584 and 8935705; Fax No: (63-2) 8939589; E-mail: publications@pids.gov.ph
Or visit our website at http://www.pids.gov.ph

Small Area Estimation of Poverty Statistics[1]

Zita Villa Juan-Albacea, PhD[2]

Abstract

In response to high demands for lower level poverty estimates, the National Statistical Coordination Board releases provincial estimates, in addition to the national and regional, starting with the 1997 FIES. However, estimates of the coefficients of variation (CV) of several provincial estimates indicate that the resulting poverty measures are not reliable. Making a decision based on unreliable poverty statistics is very risky especially if the decision to be made relates to the welfare of poor families. Such unreliable poverty statistics may also lead to incorrect targeting of the right beneficiaries of the poverty alleviation program. Hence, this paper provides alternative ways of coming up with sub-national statistics (i.e. provincial and municipal/city-level data) that yield lower CVs than those of the official ones. This refers to the small area estimation (SAE) technique, a model-based approach to produce provincial or even municipal-level data. With a good predicting model, the SAE technique has a lot of potential in providing reliable sub-national estimates for poverty reduction efforts.

Keywords: small area estimates (SAE), poverty statistics, synthetic estimation, composite estimation, regression-synthetic, empirical best linear unbiased prediction estimator (EBLUP), Elbers, Lanjouw and Lanjouw (ELL) estimation procedure

---

[1] This paper is part of a bigger report for the UNDP-funded project titled "Comprehensive Documentation and Analysis of the Official Poverty Estimation Methodology of the Philippines" implemented by PIDS in cooperation with the National Economic and Development Authority (NEDA) and the Technical Committee on Poverty Statistics (TCPOVSTAT).

[2] Dean and Professor, College of Arts and Sciences, UPLB

Small Area Estimation of Poverty Statistics[1]

Zita Villa Juan-Albacea, PhD[2]

I. Introduction

Poverty is a major concern not only of the Philippines but the whole world as well. The government of the different countries around the world has their own programs and projects to target the people in poverty with the aim of alleviating them. Not to be left out is the government of the Republic of the Philippines. It has exerts its efforts to eradicate poverty or at least to lessen it.

However, many think that the Philippines has not been successful in the implementation of its poverty alleviation programs. In an Asian Development Bank (ADB) Report in 2005, it was reported that the Philippines is not successful in alleviating the conditions of those in poverty because of the following reasons:

- weak macroeconomic management;
- employment issues;
- high population growth rates;
- an underperforming agricultural sector and an unfinished land reform agenda;
- governance issues including corruption and a weak state;
- conflict and security issues, particularly in Mindanao; and
- disability

One of the key factors in the formulation of appropriate policies on poverty as well as on the proper implementation of poverty alleviation programs is correct and timely statistics. Timely and reliable statistics are very much needed to formulate policies in alleviating people from poverty. Likewise, statistics are needed to target or pinpoint the true beneficiaries of the program to have a successful implementation.

In the Philippines, the government monitors the poverty situation using official poverty statistics. Per Executive Order No. 352, *Designation of Statistical Activities that will Generate Critical Data for Decision-Making of the Government and Private Sector*, the computation of the estimates is the responsibility of the

---

[1] This paper is part of a bigger report for the UNDP-funded project titled "Comprehensive Documentation and Analysis of the Official Poverty Estimation Methodology of the Philippines" implemented by PIDS in cooperation with the National Economic and Development Authority (NEDA) and the Technical Committee on Poverty Statistics (TCPOVSTAT).

National Statistical Coordination Board (NSCB) using the official poverty estimation methodology approved by the NSCB Executive Board.

The development of the official poverty estimation methodology was assigned to the Technical Committee on Poverty Statistics (TC PovStat), the then Technical Working Group on Income and Poverty Statistics. The TC PovStat has a multi-sectoral representation consisting of noted experts in the area of poverty statistics coming from the academe, producers and users of poverty statistics from both government and non-government organization.

Official poverty statistics are computed based on the final results of a nationwide survey conducted by the National Statistics Office (NSO). Specifically, the Family Income and Expenditure Survey (FIES) is the official source of income and expenditure data of families in the country. FIES is conducted every three years, the latest of which was conducted in 2006.

The poverty statistics is an income-based statistics which also make use of low-cost, nutritionally adequate menus. These menus are used in the estimation of the food thresholds and are determined by Food and Nutrition Research Institute (FNRI) of the Department of Science and Technology (DOST). To cost the menus, price data on agricultural commodities are obtained from the Bureau of Agricultural Statistics (BAS) while the price data on non-agricultural commodities are generated by the NSO.

Latest statistics released by the NSCB last March of 2008 indicates that approximately 27 out of 100 Filipino families did not earn enough in 2006 to satisfy their basic food and non-food requirements. This is based on an annual per capita poverty threshold equal to 15,057 pesos. When translated in population count, this means that 33 out of 100 Filipinos in 2006 had income short of the minimum cost of satisfying the basic requirements. The 26.9% national poverty incidence estimate has a coefficient of variation equal to 1.3%, which indicates that the national estimate is reliable.

At the regional level, the estimates of the poverty incidence among families and the corresponding coefficients of variation of these estimates are shown in Table

---

[2] Dean and Professor, College of Arts and Sciences, UPLB

1. In the year 2006, the poorest region or the region with the highest poverty incidence is the Autonomous Region of Muslim Mindanao (ARMM) while the richest region is National Capital Region or NCR. All of the regional estimates are said to be reliable. The most reliable estimate is that of the Bicol region while the estimate with the highest coefficient of variation is the one for the National Capital Region.

Table 1. Regional poverty incidence among families estimated using 2006 FIES.

REGION	POVERTY INCIDENCE (%)	COEFFICIENT OF VARIATION (%)
National Capital Region (NCR)	7.1	6.6
Region I – Ilocos	26.2	5.4
Region II – Cagayan	20.5	6.9
Region III – Central Luzon	16.8	5.1
Region IV-A – CALABARZON	16.7	4.3
Region IV-B – MIMAROPA	43.7	5.0
Region V – Bicol	41.8	3.5
Region VI – Western Visayas	31.1	4.3
Region VII – Central Visayas	30.3	4.2
Region VIII – Eastern Visayas	40.7	5.3
Region IX – Zamboanga Peninsula	40.2	5.5
Region X	36.1	5.0
Region XI – Davao	30.6	5.0
Region XII	33.8	4.8
Cordillera Autonomous Region (CAR)	28.8	5.7
Autonomous Region of Muslim Mindanao (ARMM)	55.3	5.0
CARAGA Region	45.5	3.8

Source: National Statistical Coordination Board (2008) 2006 Official Poverty Statistics of the Philippines

The poorest province in 2006 in terms of poverty incidence of families is Tawi-Tawi (78.9%), followed by Zamboanga del Norte (63%) and Maguindanao (62%). The ten poorest provinces were comprised of six provinces in Mindanao, three in Luzon, and only one in the Visayas. Table 2 shows the estimates of the provincial poverty incidence among families of the top ten poorest provinces and the corresponding standard errors and coefficients of variation of the estimates. Three out of the ten estimates have estimates with coefficients of variation greater than 10% but less than 20%. Hence, the estimates of the top ten poorest provinces have acceptable measure of reliability.

Table 2. Top ten poorest provinces based on their estimates of poverty incidence among families and the standard error and coefficient of variation of the estimates obtained from the 2006 FIES.

PROVINCE	REGION	POVERTY INCIDENCE (%)	STANDARD ERROR (%)	COEFFICIENT OF VARIATION (%)
Tawi-Tawi	ARMM	78.9	7.1	9.0
Zamboanga del Norte	Region IX	63.0	5.1	8.1
Maguindanao	ARMM	62.0	3.8	6.1
Apayao	CAR	57.5	9.5	16.6
Surigao del Norte	CARAGA	53.2	3.6	6.8
Lanao del Sur	ARMM	52.5	6.6	12.6
Northern Samar	Region VIII	52.2	8.4	16.0
Masbate	Region V	51.0	3.3	6.5
Abra	CAR	50.1	4.0	8.0
Misamis Occidental	Region X	48.8	4.7	9.6

Source: National Statistical Coordination Board (2008) 2006 Official Poverty Statistics of the Philippines

Looking more closely at the reliability of the provincial estimates, Table 3 shows the distribution of the coefficients of variation of provincial poverty incidence estimates for families in year 2006. If a coefficient of variation less than 10% implies reliable estimate, then 44% of the estimates are said to be very reliable. This is almost similar to the percentage of the estimates with acceptable measure of reliability, that is, coefficients of variation that are less than 20% but greater than 10%. Only 11 estimates are unreliable. However, since there is no sampled poor household in the province of Batanes, there is no measure of reliability for that estimate. Because of this, the zero estimate of poverty incidence of Batanes is very unreliable. The provincial poverty incidence estimate for families with the highest coefficient of variation is that of Aurora. The Aurora Province in Central Luzon Region has a poverty incidence estimate equal to 31.6% and this estimate has a coefficient of variation equal to 47.1%.

Table 3. Distribution of the coefficients of variation of provincial estimates of poverty incidence estimates among families.

COEFFICIENT OF VARIATION (%)	FREQUENCY	RELATIVE FREQUENCY (%)
< = 10	37	44
10.01 - 20.00	33	43
20.01 - 30.00	7	8
30.01 - 40.00	3	4
40.01 – 50.00	1	1

Source: National Statistical Coordination Board (2008) 2006 Official Poverty Statistics of the Philippines

The eleven provinces with the highest coefficients of variation are shown in Table 4. The estimates for these provinces are considered statistically unreliable but have to be used by the government in absence of indicator to base their decisions. Making a decision based on unreliable poverty statistics is very risky especially if the decision to be made relates to the welfare of poor families. Such unreliable poverty statistics may also lead to incorrect targeting of the right beneficiaries of the poverty alleviation program.

Table 4. Eleven provinces with unreliable poverty incidence estimates.

PROVINCE	REGION	POVERTY INCIDENCE (%)	STANDARD ERROR (%)	COEFFICIENT OF VARIATION (%)
Aurora	Region III	31.6	14.9	47.1
Bataan	Region III	6.8	2.2	33.0
Basilan	ARMM	31.6	9.7	30.7
Quirino	Region II	15.9	4.8	30.3
Ilocos Norte	Region I	17.1	4.8	27.8
Siquijor	Region VII	22.3	5.6	24.9
Cotabato City	Region XII	38.0	9.3	24.5
Nueva Viscaya	Region II	12.7	3.1	24.3
Biliran	Region VIII	31.4	7.4	23.7
Catanduanes	Region V	37.3	8.2	22.1
Isabela City	Region IX	43.0	9.5	22.1

Source: National Statistical Coordination Board (2008) 2006 Official Poverty Statistics of the Philippines

Also, the Province of Batanes in Cagayan Valley Region has no reported poor sampled household in 2006 and hence, no reliability measure was reported for its poverty incidence estimate equal to 0%. As mentioned before, this estimate is very unreliable and misleading. In 2000 revised estimate for the province of Batanes, its poverty incidence is 10.4% with a coefficient of variation equal to 36.5%, the highest among the provincial estimates reported in that year. In 2003, the estimate for the Province of Batanes is 6.3% based on a single sampled poor household. Thus, the reported decrease in the estimates from 10.4% to 0% in six years may lead to risky decisions on the welfare of the people residing in the Province of Batanes since the decision is based on unreliable statistics.

With the release of these estimates, certain issues on the reliability of provincial poverty estimates were raised. Likewise, the integrity of the official methodology in generating sub-national official statistics was questioned.

II. Ways to Generate Sub-national Poverty Statistics

As observed, smaller disaggregation like provincial level results to a number of unreliable estimates. Further disaggregation like municipality level will surely lead to higher standard error and consequently, high coefficients of variation. The main reason for this is the insufficient number of respondents or no respondent at all in a small domain. FIES is one such nationwide survey which when used to compute statistics of small domains result to estimates with large standard errors, in other words, unreliable statistics.

Increasing the sample size or redesigning the survey is an option. Currently, the FIES is conducted annually with around 50,000 sample households and a budget of almost 75M pesos. To change the design of the survey and have the provinces as domain of estimation, it was found that there is a need to quadruple the total number of households, like taking around 200,000 households and this would mean increasing also the budget to around 300M pesos. This cost makes the conduct of survey impossible since it cannot be supported financially by the government.

Another alternative is to collect data at the small domain. Local government can collect the required data needed to monitor poverty situation in their localities, that is, the local government of the different provinces can finance and undertake their own data collection. In fact, some provinces have taken this option. However, the local government realized that this option is time consuming and also very costly. Thus, this is not an efficient method of getting the needed official sub-national poverty statistics.

Emerging small area estimation techniques can provide alternative ways to generate poverty statistics at the small area level with acceptable standard errors. These are generally referred to as indirect methods of generating sub-national poverty statistics. The first set of techniques is commonly referred to as the traditional indirect techniques. This set includes the demographic, synthetic and composite methods of estimation.

Another set of techniques, which is becoming more popular recently is the model-based approach of estimating poverty statistics using data coming from different sources like nationwide surveys, most recent census and administrative records. This set of techniques includes the regression-synthetic, empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB) and the hierarchical Bayes (HB) techniques. Another model-based approach technique that is gaining popularity recently is the one developed by World Bank through Elbers, and the couple named Lanjouw. This method became popularly known as the *Elbers, Lanjouw and Lanjouw (ELL) method of estimation*.

Some of these SAE techniques have been studied and were found to be more efficient to use in estimating official poverty statistics like poverty incidence. The results of these studies are discussed in the next section.

III. SAE Techniques Used to Estimate Sub-national Poverty Statistics

Several SAE techniques are being studied and proposed to improve the methodology of estimating official poverty statistics in the Philippines. These set of techniques include synthetic, composite, empirical best linear unbiased predictor (EBLUP) and Elbers, Lanjouw and Lanjouw (ELL) method.

A. Synthetic Estimation of Poverty Incidence

Purcell and Kish (1979) described the *synthetic estimation* procedure as a technique, which uses sample data to estimate at some higher level of aggregation the variable of interest for different subclasses of the population; then it scales these estimates in proportion to the subclass of incidence within the small domains of interest. The estimates obtained are referred to as synthetic estimates since these estimates are not directly obtained from survey results. The procedure obtained unbiased estimates using survey results for a large area and these unbiased estimates were used to derive estimates for sub-areas with the assumption that the small areas have the same characteristics as the large area. Hence, the method borrows information from similar small domains in order to increase the accuracy of the resulting estimates.

Albacea (1999) applied synthetic estimation procedure in estimating poverty incidence among households using the 1994 FIES and the 1990 Census of Population and Housing (CPH) data sets. The total number of poor households and total number of households at the provincial or city level are estimated using this procedure. The synthetic estimates of these totals are used to obtain an estimate of poverty incidence, which is referred to as synthetic estimator of poverty incidence.

The large domain considered in this estimation procedure is the subgroup of households grouped according to the gender of the household head. In FIES, the households are classified according to the gender of the household head. The estimated number of poor households headed by a male or that by a female as well as the number of households headed by a male or that by a female are obtained. The estimates are computed for all provinces and cities. These estimates are aggregated at the regional level, that is, total number of households and that of poor households in provinces and cities comprising a region are accumulated. Hence, the counts of the households per subgroup are made at the regional level. This aggregation is expected to produce estimates with smaller variances since the provinces within a region are more homogeneous compared to all provinces nation-wide. Such aggregation takes into account the use of regional poverty thresholds instead of the national poverty threshold in classifying poor households.

Similarly, in the 1990 CPH, there are available information on some associated variables at the provincial or city (small area) level within the subgroup of households (large domain) grouped according to the gender of the household head. The associated variable used in this study is the indicator of whether the household's main source of income is agriculture or not. Data from 1990 CPH are used to obtain the number of households in a province with agriculture as their main source of income. These counts are further classified according to the gender of the household head.

With the number of households in a province with agriculture as their main source of income as the auxiliary variable, denoted X_{ij}, where i stands for a province or city and j for the subgroup (that is, whether the household is headed by a male or female), the synthetic estimator of poverty incidence was formulated as a ratio of two synthetic estimates of total counts, that is,

$$\hat{p}_i^S = \frac{\hat{NP}_i^S}{\hat{N}_i^S}$$

where $\hat{NP}_i^S$ is the synthetic estimator of the total number of poor households in the i^{th} province or city computed as $\hat{NP}_i^S = \sum_j \hat{NP}_{ij} = \sum_j \frac{X_{ij}}{X_{.j}} \hat{NP}_{.j}$ and $\hat{N}_i^S$ is the synthetic estimator of the total number of households in the i^{th} province or city obtained as $\hat{N}_i^S = \sum_j \hat{N}_{ij} = \sum_j \frac{X_{ij}}{X_{.j}} \hat{N}_{.j}$.

Such estimators are biased because there will be departures from the underlying assumptions of homogeneity of rates and the weights $X_{ij} / X_{.j}$ are based on data taken four years earlier than the survey and the structure of the population may have changed through time.

Table 5 shows the statistical properties of the design-based and synthetic estimates of poverty incidence. It was observed that the generated synthetic estimates have the same average value of the design-based estimates but the range of estimates is smaller among the synthetic estimates. Also, negative mean square error (MSE) was observed in synthetic estimation because of the instability of the MSE estimator for the synthetic procedure. Nevertheless, the procedure generated estimates with small coefficient of variation resulting to a more reliable set of estimates for the poverty incidence.

Table 5. Statistical properties of the estimates of provincial or city level poverty incidence among households using design-based and synthetic estimation procedures as applied to the 1994 FIES and 1990 CPH data sets.

PROPERTY	DESIGN-BASED ESTIMATES	SYNTHETIC ESTIMATES
Average poverty incidence estimate	0.38	0.38
Range of poverty incidence estimates	0.05-0.85	0.08-0.60
Average MSE or variance	0.0036	-0.000025
Range of MSE or variance	0.0001-0.0257	-0.001-0.0004
Average coefficient of variation (%)	19.15	1.93
Range of coefficient of variation (%)	4.8-127.08	0.13-4.51
Percentage of CV values at most 10%	27%	100%

Source: Albacea, ZVJ and AIN Gironella, (2001) Estimation of Provincial Poverty Incidence of the Philippines

B. Composite Estimation of Poverty Incidence

To correct the potential bias of a synthetic estimator against the instability of a design-based or direct estimator, these two estimators were combined into one. The resulting estimator, which is commonly referred to as composite estimator, is generally expressed as:

$$\hat{Y}_i^C = w_i \hat{Y}_{1i} + (1 - w_i)\hat{Y}_{2i}$$

where $\hat{Y}_{1i}$ is a design-based or direct estimator, $\hat{Y}_{2i}$ is an indirect estimator and w_i is a chosen weight with a value between 0 and 1, inclusive. Usually, the design-unbiased or direct estimator $\hat{Y}_i$ is considered as $\hat{Y}_{1i}$ while the synthetic estimator $\hat{Y}_i^S$ as $\hat{Y}_{2i}$.

In the study of Albacea (1999), a composite estimator of the poverty incidence among households was formulated which is a weighted average of a design-based $\hat{p}_i$ and a synthetic estimator $\hat{p}_i^S$, that is,

$$\hat{p}_i^C = w_i \hat{p}_i + (1 - w_i)\hat{p}_i^S$$

where w_i is a weight, which has a value between 0 and 1, based on the variability of the observations in the data set. The optimal weights for combining two estimators generally depend on the mean square errors of the estimators and their covariance. These quantities would generally be unknown but can be estimated from the data.

The results of the study indicate an improvement of the reliability of the estimates. Compared to the design-based or direct estimation technique, the composite estimation generated more reliable provincial estimates of poverty incidence as shown in Table 6. Higher percentage of estimates with at most 10% was obtained in composite estimation compared to design-based estimates. Also, the range of values of the estimates are narrower compared to design-based but wider compared to synthetic estimates. Unlike in synthetic estimation, composite estimation did not result to negative estimates of mean square error.

Table 6. Statistical properties of the estimates of provincial or city level poverty incidence among households using design-based and composite estimation procedures as applied to the 1994 FIES and 1990 CPH data sets.

PROPERTY	DESIGN-BASED ESTIMATES	COMPOSITE ESTIMATES
Average poverty incidence estimate	0.38	0.38
Range of poverty incidence estimates	0.05-0.85	0.06-0.79
Average MSE or variance	0.0036	0.0026
Range of MSE or variance	0.0001-0.0257	0.0001-0.0183
Average coefficient of variation (%)	19.15	14.97
Range of coefficient of variation (%)	4.8-127.08	4.16-82.76
Percentage of CV values at most 10%	27%	37%

Source: Albacea, ZVJ and AIN Gironella, (2001) Estimation of Provincial Poverty Incidence of the Philippines

C. Regression-Synthetic Estimation of Poverty Incidence

This regression-synthetic estimation is a model-based approach, which utilizes the linear regression model in predicting the poverty incidence. There are two types of models that are used in this approach and these are:

- Aggregate or area level models that relate the small area means to area-specific auxiliary variables. Such models are essential if unit (element) level data are not available.

- Unit level models that relate the unit values of the variable of interest to unit-specific auxiliary variables.

Between these two types the first one is more commonly used since unit level data are seldom available. In using the area level modeling approach, the model to be fitted has two types of errors, namely; the model error v_i, and the sampling error e_i, The model estimation is a two-stage procedure where the variability due to the modeling process is first estimated using ordinary least squares procedure. Once this is estimated, together with sampling variance, the weights are formulated and weighted least squares procedure is applied. The predicted values using the weighted least squares regression estimates serve as the regression-synthetic estimates.

This procedure was used in the study of Perez (2007) where he estimated poverty incidence among population using the 2000 FIES and CPH data sets. Table 7 shows the distribution of the coefficients of variation of the design-based and regression-synthetic estimates (referred to as model-based in his study) of the poverty incidence among population. A substantial increase in the percentage of estimates with coefficients of variation less than 10% was attained when the model-based approach was used. However, a number of regression-synthetic estimates have higher coefficients of variation compared to the design-based estimates.

Table 7. Distribution of the coefficients of variation of the design-based and regression-synthetic estimates of the provincial level poverty incidence among population using 2000 FIES and CPH data sets.

COEFFICIENT OF VARIATION (%)	DESIGN-BASED (%)	REGRESSION-SYNTHETIC (%)
< = 10	34.1	86.6
10.01 - 20.00	63.4	7.4
20.01 - 30.00	2.4	1.2
30.01 - 40.00	0	2.4
> 40	0	2.4

Source: Perez, RG (2007) Design-based and Model-based Estimation of Poverty Measures at the Provincial Level in the Philippines

D. EBLUP Estimation of Poverty Incidence

The empirical best linear unbiased prediction estimator is similar to a composite estimator but this time it combines the direct or design-based unbiased estimator with the regression-synthetic estimator. It is a model-based approach that is gaining popularity. Rao (2003) said that this approach is becoming popular because of the following:

- Model-based methods make specific allowance for local variation through complex error structures in the models that link the small areas.
- Efficient indirect estimators can be obtained under the assumed models.
- Models can be validated from the sample data.
- Methods can handle complex cases such as cross-sectional and time-series data.
- Stable area specific measures of variability associated with the estimates can be obtained, unlike the overall measures for synthetic and composite estimators.

The EBLUP had been used in several studies to model not only poverty incidence but also other poverty statistics like subsistence incidence or food poverty incidence, poverty gap and severity of poverty. It was even used to estimate these parameters at the lower level of disaggregation like municipalities and village or barangay level. Some of these studies are successful in improving the reliability of estimates.

Albacea and Pacificador (2003) used the EBLUP method in estimating poverty incidence among households using the 2000 FIES and CPH data sets in their study entitled “Targeting the Poor in the Philippines”. This study is commissioned by the National Statistics Office (NSO) with funding from the New Zealand Government through the Asian Development Bank (ADB) Technical Assistance Project to NSO. Table 8 shows the distribution of the coefficients of variation of the design-based and EBLUP estimates of the poverty incidence among households. A substantial increase in the percentage of estimates with coefficients of variation less than 10% was attained when the model-based approach was used. More so, the method was able to lessen the number of estimates with coefficient of variation greater than 20%.

Table 8. Distribution of the coefficients of variation of the design-based and EBLUP estimates of the provincial level poverty incidence among households using 2000 FIES and CPH data sets.

COEFFICIENT OF VARIATION (%)	DESIGN-BASED (%)	EBLUP (%)
< = 10	50.0	61.0
10.01 - 20.00	48.0	38.0
> 20	2.0	1.0

Source: Albacea, ZVJ and AYPacificador (2003) Targeting the Poor in the Philippines,

One of the major recommendations of the external reviewers of the study is the use of other sources of auxiliary variables like administrative data in addition to the CPH data set. This is in anticipation of the estimation of poverty statistics during the non-censal years, that is, no census is done at the same time the nationwide survey is being conducted. Hence, a follow-up study was conducted by Albacea (2003) using the same estimation procedure and data sets but this time additional auxiliary variables were used coming from administrative data sources.

Table 9 shows the distribution of the coefficients of variation of the design-based and EBLUP estimates of the poverty incidence among households obtained from the follow-up study. A higher percentage of estimates with coefficients less than 10% were attained from this study but higher percentage of estimates was also observed for coefficients greater than 20%.

Table 9. Distribution of the coefficients of variation of the design-based and EBLUP estimates of the provincial level poverty incidence among households using 2000 FIES, CPH and administrative data sets.

COEFFICIENT OF VARIATION (%)	DESIGN-BASED (%)	EBLUP (%)
< = 10	50.0	65.0
10.01 - 20.00	48.0	30.0
> 20	2.0	5.0

Source: Albacea, ZVJ (2003) Estimating Philippine Provincial Poverty Incidence Using Administrative Data

The study of Perez in 2007 also makes use of the EBLUP estimation procedure in estimating poverty incidence among population. Table 10 shows that again an improvement was attained on the quality of estimates when EBLUP was used instead of the design-based or direct estimation procedure. More reliable estimates were generated using the EBLUP procedure.

Table 10. Distribution of the coefficients of variation of the design-based and EBLUP estimates of the provincial level poverty incidence among population using 2000 FIES and CPH data sets.

COEFFICIENT OF VARIATION (%)	DESIGN-BASED (%)	EBLUP (%)
< = 10	34.1	47.6
10.01 - 20.00	63.4	51.2
> 20	2.4	1.2

Source: Perez, RG (2007) Design-based and Model-based Estimation of Poverty Measures at the Provincial Level in the Philippines

Applying the EBLUP estimation procedure using the 2003 FIES, 2000 CPH and administrative data sets to estimate poverty incidence among families resulted also to a better set of estimates compared to the reported official poverty measure. Table 11 shows the distribution of the coefficients of variation of the EBLUP estimates compared to the distribution of the coefficients of the official poverty incidence statistics. More so, the province of Batanes has a more reliable estimate as compared to the official statistic. The poverty incidence of Batanes is estimated using the predicting equation obtained in the regression-synthetic procedure, a component of the EBLUP estimation procedure.

Table 11. Distribution of the coefficients of variation of the official provincial level poverty incidence among families and EBLUP estimates using 2003 FIES, 2000 CPH and administrative data sets.

COEFFICIENT OF VARIATION (%)	OFFICIAL METHODOLOGY (%)	EBLUP (%)
< = 10	33.3	39.3
10.01 - 20.00	54.7	53.6
> 20	12.0	7.1

D. ELL Estimation of Poverty Incidence

The Elbers, Lanjouw and Lanjouw (ELL) estimation procedure was designed specifically for the small-area estimation of poverty measures based on per capita household expenditure. When it was applied to the Philippines case, the target variable is the log-transformed of the income, since poverty measures in the country is income-based. It is a procedure that requires the existence of both the nationwide survey and the census data sets on the same year.

This procedure was first applied in the Philippines by the Poverty Team of the NSCB with funding assistance from the World Bank ASEM Trust Fund. The study in 2005 produced small area estimates of poverty in the country at the provincial, municipal and city level by combining the 2000 FIES and Labor Force Survey (LFS) data sets with the 2000 CPH data set. Estimates of poverty based on expenditure and income were generated. Table 12 shows the statistical properties of the estimates produced using this procedure as compared to direct estimates. The results of this study show that the reliability of the official poverty statistics can be improved using an SAE technique.

Table 12. Statistical properties of the estimates of provincial poverty incidence among population using direct and ELL estimation procedures based on the 2000 FIES, LFS and 2000 CPH data sets.

PROPERTY	OFFICIAL ESTIMATES	ELL-BASED ESTIMATES
Average poverty incidence estimate	0.41	0.40
Range of poverty incidence estimates	0.06-0.71	0.07-0.68
Average standard error	0.0425	0.0174
Range of standard error	0.0098-0.1041	0.0072-0.0323
Average coefficient of variation (%)	11.74	4.85
Range of coefficient of variation (%)	4.50-39.88	2.39-13.81

Source: NSCB-WB (2005) Estimation of Local Poverty in the Philippines.

The data requirement of the ELL Technique is its limitation since it makes the procedure less useful during the non-censal year. However, recent development in the procedure provided a solution to this limitation. A study conducted by the NSCB in 2006 entitled “intercensal Updating of Small Area Poverty Estimates” through the World Bank Trust Fund for Statistical Capacity Building generated 2003 provincial, municipal and city level poverty estimates using a modified ELL Technique. In 2003 no census data were collected but FIES and LFS were conducted in that year. The study used time-invariant auxiliary variables in model-building process. The time-invariant variables were constructed using the 2000 census data set to serve as predictors of per capita income in year 2003 which is a non-censal year. This approach made it possible to use the ELL Technique even there is no census conducted during the year that the FIES and LFS were conducted as in the year 2003.

This modified ELL Technique was able to generate poverty statistics for provinces, municipalities and cities of the Philippines. Table 13 shows the statistical properties of the 79 provincial estimates of poverty incidence among population produced using the modified ELL Technique as compared to official estimates. The distribution of generated estimates is almost the same to that of the official estimates but the estimates have smaller mean square errors and coefficients of variation. Hence, it can be said that the estimates are more reliable, precise and accurate.

Table 13. Statistical properties of the estimates of provincial poverty incidence among population using direct and modified ELL estimation procedures based on the 2003 FIES, LFS and 2000 CPH data sets.

PROPERTY	OFFICIAL ESTIMATES	Modified ELL-BASED ESTIMATES
Average poverty incidence estimate	0.38	0.34
Range of poverty incidence estimates	0.049-0.685	0.048-0.68
Average standard error	0.043	0.018
Range of standard error	0.09-0.184	0.006-0.041
Average coefficient of variation (%)	12.0	5.9
Range of coefficient of variation (%)	5.0-47.1	2.0-18.4

Source: NSCB-WB (2009) Intercensal Updating of Small Area Poverty Estimates.

In terms of reliability, the estimates generated using the modified ELL technique are at least with an acceptable coefficient of variation. Unlike the official estimates, none of the estimates generated using the modified ELL technique have coefficients greater than 20% as shown in Table 14.

Table 14. Distribution of the coefficients of variation of the official provincial level poverty incidence among population and modified ELL estimates using 2003 FIES and LFS, and 2000 CPH data sets.

COEFFICIENT OF VARIATION (%)	OFFICIAL METHODOLOGY (%)	Modified ELL Technique (%)
< = 10	44	94
10.01 - 20.00	50	6
> 20	6	0

Source: NSCB-WB (2009) Intercensal Updating of Small Area Poverty Estimates.

More so, the procedure was able to generate estimates for 1622 municipal and cities in the country for the year 2003. Although only 85% of the 1622 estimates are with at least acceptable coefficients of variation (as seen in Table 15), the modified ELL technique was able to provide reliable estimates for 1,380 municipalities and cities that the direct estimation technique cannot provide. These 1,380 reliable estimates can be used by local planners in the implementation of poverty alleviation program.

Table 15. Distribution of the coefficients of variation of the modified ELL estimates at the municipal and city level poverty incidence among population based on the 2003 FIES and LFS, and 2000 CPH data sets.

COEFFICIENT OF VARIATION (%)	Frequency	Percentage
< = 10	623	38
10.01 - 20.00	757	47
> 20	242	15

The Research Team who conducted the study believed that the estimates can be better improved with better predicting models to use.

IV. Lessons Learned

The small area estimation procedures offered alternative ways to obtain reliable provincial or city or even municipal or city level poverty statistics with reported coefficients of variation smaller than that of the usual reported official poverty statistics. However, notice that in most cases, the application of small area estimation technique in the Philippines to produce provincial or even municipal or city level poverty statistics proved to be successful because of the availability of census and survey data with the same reference period. Most of the model-based techniques require the presence of both the census and survey data sets.

The census is conducted every ten years but policy-makers demand frequent estimates and therefore, there is a need to generate these estimates more frequency than every ten years. Time-invariant variables provide also an option to remedy the situation. However, construction of such variables is also a challenge to handle, in addition to the fact that only few such variables can be found to exists.

An alternative is to use administrative data that is collected regularly. However, if administrative data are used, it should be measured consistently across all areas covered to obtain consistent resulting estimates.

Still another alternative is the use of time-invariant variables as done in the modified ELL methodology. Such procedures proved to be useful especially in the generation of poverty statistics at the municipal and city level.

Similarly, for consistency, definitions and concepts used in compiling data that will be used in the model should be applied consistently across the whole data set. The definition of a ‘household’ in census must be the same in FIES. Otherwise, the association between the household income observed in FIES and the household characteristics observed in the census will not be determined correctly.

The absence of sampled observations in an area prevents the generation of direct estimates for that area. In this case, an indirect estimate in the form of predicted values using the regression model can serve as an estimate for that area.

Validation procedures are needed to investigate the reliability of the estimates generated from the SAE methodology. It is not enough to look at the point estimates and compared it to the direct estimates. One must investigate further on the statistical properties of the generated estimates. Specifically, the distributional properties of the estimates have to be studied.

V. Recommendations

Although the small area estimation procedures resulted to reliable provincial or city or even municipal level poverty statistics, the estimates can still be improved by getting better predicting model. As presented most of the studies are in poverty incidence, the goodness of the procedure as applied to other poverty statistics can also be looked at. Likewise, other validation procedures have to be formulated. External validation techniques can be used to assess the SAE estimates.

VI. References

Albacea, ZVJ. (1999) *Small Area Estimation of Poverty Incidence in the Philippines*, Unpublished PhD Dissertation, University of the Philippines Los Banos.

____________and AIN Gironella (2001) *Estimation of Provincial Poverty Incidence of the Philippines*, Paper presented at the 53rd Session of the International Statistical Institute, 22-29 August 2001 in Seoul, South Korea

____________ and AY Pacificador Jr.(2003) *Targeting the Poor in the Philippines* National Statistics Office Monograph on Studies on Small Area Estimation, ADBTA 3566: Improving Poverty Monitoring Surveys

____________ (2003) *Estimating Philippine Provincial Poverty Incidence Using Administrative Data* National Statistics Office Monograph on Studies on Small Area Estimation, ADBTA 3566: Improving Poverty Monitoring Surveys

____________ (2006) *Small Area Estimation of Poverty Statistics* Development Fund/Diamond Jubilee Professorial Chair, University of the Philippines Los Baños

Asian Development Bank (2005) Poverty in the Philippines: Income, Assets, and Access

GHOSH, M. and J.N.K. RAO (1994) *Small Area Estimation: An Appraisal*, Statistical Science, 9:1:55-93.

NATIONAL STATSTICAL COORDINATION BOARD (2008) *2006 Official Poverty Statistics of the Philippines*

____ – WORLD BANK (2005) *Estimation of Local Poverty in the Philippines*

____ – WORLD BANK (2009) *Intercensal Updating of SAE Poverty Statistics*

Perez, RG (2007) *Design-based and Model-based Estimation of Poverty Measures at the Provincial Level in the Philippines*, Unpublished MS Thesis, University of the Philippines Los Baños

Purcell, N.J. and L. Kish (1979) *Estimation for Small Domains*, Biometrics, 35 48:3-18

Rao, JNK (2003) *Small Area Estimation*, Wiley Series in Survey Methodology