RESEARCH ARTICLE

Small Area Estimates of Poverty in Region III

Caesar B. Cororaton
Virginia Polytechnic Institute and State University
ccororaton@gmail.com

Arlene B. Inocencio, Alellie B. Sobreviñas
De La Salle University, Manila Philippines

Emmanuel DC. Barnedo
University of the Philippines

Official estimates of the Philippine Statistical Authority (PSA) poverty incidence prior to 2018 are reliable up to the regional level as these are consistent with the sampling domain of the data used for estimation. Below this level, the PSA poverty estimates are unreliable because of very large sampling errors. The high level of unreliability makes the estimates less useful for poverty policy targeting. This paper addresses this concern by combining the Family Income and Expenditure Survey (FIES) and the Census of Population and Housing (CPH) data and using an increasingly accepted technique for small area estimation (SAE). We estimate the poverty incidence for the Central Luzon provinces and show that our coefficients of variation are considerably smaller than those of PSA indicating higher reliability.

Keywords: small area estimates, poverty estimation, Central Luzon, Philippines

JEL Classifications: C18, C83, I32

The key source of information on poverty in the Philippines is the Family Income and Expenditure Survey (FIES). The statistical authority which started collecting data for the FIES since 1985 used Philippine regions as the sampling domain. It was only during the latest round in 2018 when the data collecting authority started using Philippine provinces as the sampling domain[1]. Regions as the sampling domain for household survey in the county may provide inadequate information for poverty policy analysis because poverty incidence varies considerably across provinces in each region. The provincial poverty estimates which the statistical authority provides are unreliable for poverty policy targeting because of very large sampling errors. To overcome this data gap, the paper will combine FIES data with the data from the Census of Population and Housing (CPH) using techniques available in the literature on small area estimates (SAE) to calculate poverty in seven provinces in Region III.

There are several papers in the literature that employ SAE to calculate poverty at a more disaggregated level. There are basically two steps involved in the estimation. First, household survey data is used to estimate the relationship between income or expenditure and household characteristics such as household size, education, age, gender, employment and marital status of household heads, housing types, household asset ownership, etc. Second, the disaggregated census data on the same household characteristics are inserted into the estimated relationship in the first step to generate estimates of poverty for smaller geographic areas. The papers of Minot (1998, 2000, 2002) applied this 2-step approach to analyze district-level poverty in Vietnam. Henstchel et al (2002) also developed a similar method of combing household survey and census data to calculate poverty incidence in Ecuador. The World Bank (2005) also conducted local estimation of poverty for 2000 in the Philippines using SAE that combined the data from the 2000 FIES[2] and the 2000 CPH. The household welfare indicator in the study used both household expenditure-based and income-based measures. The poverty measures derived in the study at the regional, provincial and municipal levels have on the whole acceptably small standard errors.

Region III is one of the regions in the Philippines the least poverty. In 2015, the overall poverty incidence in the Philippines was 21.6 percent. In Region III, the poverty incidence was 11.2 percent. Table 1 indicates that poverty varies considerably across the seven provinces in Region III. Bataan has the smallest incidence of 2 percent, while Aurora was the highest at 26.3 percent. However, these provincial estimates are unreliable because of high coefficient of variations.

Small Area Estimation of Poverty

Small area estimation (SAE) requires data from the Family Income and Expenditure Survey (FIES) and the Census of Population and Housing (CPH). The FIES data is used to estimate econometrically the parameters of an income-generating equation, while the CPH data is used to predict the small area estimates of poverty using the estimated income-generating equation. That is, the CPH data is substituted into the explanatory variables of the income-generating equation to predict the values of the dependent variable. It is therefore important that the explanatory variables in the income-generating equation exist in both the FIES and the CPH database.

The income-generating equation takes the following log-linear form

Table 1. *Official Poverty Estimates in Region III*

	Poverty Incidence, %	Coefficient Variation
Region III	11.2	12.20
Aurora /a/	26.3	-
Bataan /b/	2.0	74.32
Bulacan /b/	4.5	30.23
Nueva Ecija	22.6	13.05
Pampanga /b/	4.9	26.67
Tarlac	18.1	16.99
Zambales /b/	16.8	20.66

/a/ Caution in utilizing the estimate for these provinces due to its very small sample size.

/b/ Coefficient of variation of 2015 poverty incidence among population is greater than 20%.

(4.1) $\ln(y_i) = X_i\beta + \varepsilon_i$

where y_i is the per capita income of household i, X_i, is a *(1 x k)* vector of household characteristics of i, k the number of household characteristics, b is a $(k\ x\ 1)$ vector of coefficients, and ε_i is a random disturbance term distributed as N(0,σ), i.e., ε_i normally distributed with mean zero and standard deviation of σ. \

Following Hentschel *et al.* (2000), the expected probability that household i with characteristics X_i is poor is expressed as

(4.2) $\mathrm{E}[P_i|X_i, \beta, \sigma^2] = \Phi\left[\frac{lnZ - X_i\beta}{\sigma}\right]$

where E is an expected operator, P_i is 1 if the household i is poor and 0 otherwise, Z per capita poverty income threshold, and Φ cumulative standard normal function.

Given the estimated regression coefficients in (4.1), the household characteristics data of household i in CPH, X_i^C, can be inserted into (4.2) to predict the expected probability that household being poor. That is,

(4.3) $\mathrm{E}[P_i|X_i, \beta, \sigma^2] = \Phi\left[\frac{lnZ - X_i^C\beta}{\sigma}\right]$

For a given geographic area such as region or province, following Hentschel *et al.* (2000), the proportion of the population living in households that are below the poverty threshold is estimated as the mean of the probabilities that individual households are poor, that is,

(4.4) $\mathrm{E}[P_i|X^C, \beta, \sigma^2] = \sum_{i=1}^{N} \frac{m_i}{M} \Phi\left[\frac{lnZ - X_i^C\beta}{\sigma}\right]$

where m_i is the size of household i, M is the total population of the area in question, N is the number of households, X^C is an (N x k) matrix of household characteristics whose data come from the CPH.

Hentschel *et al.* (2000) have noted that a simple headcount poverty incidence, which is the usual indicator of poverty, is a biased estimator of poverty if intra-household inequality is present. Intra-household inequality refers to the inequality in the allocation of resources within household, which Haddad and Kanbur (1990) have demonstrated to exist using Philippine data. The mean of the probabilities in (4.4), however would yield an unbiased estimate of poverty even if intra-household inequality is present because of the random component in (4.1), ε_i. Because of this random element no household has a zero probability of being poor or nonpoor given its characteristics.

The variance of (4.4) is calculated as follows

(4.5) $var(P^*) = \left(\frac{\partial P^*}{\partial \hat{\beta}}\right)' var(\hat{\beta}) \frac{\partial P^*}{\partial \hat{\beta}} + \left(\frac{\partial P^*}{\partial \hat{\sigma}^2}\right)^2$

$\frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^{N} \frac{m_i^2 P_i^*(1-P_i^*)}{M^2}$

where n is the sample size in the regression in (4.1) and k the number of household characteristics in the regression. Thus, n, k and σ^2 are from the estimated regression equation, while m_i, M and N are from the CPH. In (4.5) the partial derivatives of P^* with respect to the estimated parameters are estimated using the following formula

(4.6) $\frac{\partial P^*}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{m_i}{M} \left(\frac{-x_{ij}}{\hat{\sigma}}\right) \Phi\left[\frac{lnZ - X_i\hat{\beta}}{\hat{\sigma}}\right]$

(4.7) $\frac{\partial P^*}{\partial \hat{\sigma}^2} = -\frac{1}{2} \sum_{i=1}^{N} \frac{m_i}{M} \left(\frac{lnZ - X_i'\hat{\beta}}{\hat{\sigma}^3}\right) \Phi\left[\frac{lnZ - X_i\hat{\beta}}{\hat{\sigma}}\right]$

Minot (2002) has noted that (4.5) has two key components: (i) the "regression error", which is captured in the first two terms of (4.5), and (ii) the "idiosyncratic error", in the third term. The regression error is due to the uncertainty regarding the true value of β and σ in the regression in (4.1). This error is estimated by the covariance matrix of β and the estimated variance σ^2, as well as the effect of this variation on P^*. Even if the regression parameters, β and σ, are estimated correctly, there can still be errors in the predicted per capita income because of household-specific factors. This is the "idiosyncratic error".

Data

SAE of poverty for Region III in the paper combines the 2015 merged FIES-Labor Force Survey (LFS) and the 2015 CPH. Since both the FIES and the LFS are part of the Integrated Survey of Households (ISH), they can be merged. The merged FIES-LFS provides a richer set of variables for matching with the data in the CPH than the FIES only.

Income-Generating Equations

Table 2 presents the income-generating equation estimated using the 2015 CPH. Two sets of regression results are presented, ordinary least squares (OLS) estimates and estimates with correction for possible heteroskedasticity. The estimated coefficients in both regressions are the same. Only the standard error, the t-stat and the F-stat have changed as a result of the correction. In both regressions, the number of observations is 3,237 households.

The types of materials used in houses are also statistically significant predictors of per capita income. Roof and walls which are made of strong materials can indicate high per capita income. However, single unit houses built using light materials is negatively related with per capita income.

The dependent variable is natural logarithm of per capita income of households. The regression results indicate that households with large sizes are strongly associated with low per capita income. The negative sign of the coefficient on household size (*totmem*) implies that, all other things constant, each additional household member is associated with a 20 percent reduction in per capita income. The coefficient of *totmem* suggests that the effects of household size on per capita income is non-linear.

Table 2: 2015 Income-Generating Equation

	Ordinary least squares				**Corrected for possible heteroskedasticity**			
lnpcy1	**Coefficient**	**Std. Err.**	**t-stat**	**P>|t|**	**Coefficient**	**Robust Std. Err.**	**t-stat**	**P>|t|**
totmem	-0.205***	(0.012)	-16.742	0.000	-0.205***	(0.014)	-14.830	0.000
totmem2	0.009***	(0.001)	11.042	0.000	0.009***	(0.001)	8.985	0.000
dm_roof1	0.113***	(0.037)	3.101	0.002	0.113***	(0.033)	3.385	0.001
dm_wall1	0.158***	(0.036)	4.406	0.000	0.158***	(0.033)	4.752	0.000
sing_wall2	-0.124***	(0.044)	-2.836	0.005	-0.124***	(0.037)	-3.294	0.001
head_age	0.009***	(0.001)	12.581	0.000	0.009***	(0.001)	12.270	0.000
dm_HHedu1	-0.151***	(0.020)	-7.428	0.000	-0.151***	(0.022)	-6.882	0.000
dm_HHedu3	0.184***	(0.027)	6.818	0.000	0.184***	(0.033)	5.598	0.000
no_spouse	-0.100***	(0.023)	-4.423	0.000	-0.100***	(0.024)	-4.165	0.000
dm_spedu3	0.132***	(0.026)	5.023	0.000	0.132***	(0.032)	4.172	0.000
kids_female_coed	1.288***	(0.209)	6.176	0.000	1.288***	(0.241)	5.354	0.000
per_nofw	0.692***	(0.117)	5.910	0.000	0.692***	(0.116)	5.947	0.000
per_2560hhall	0.313***	(0.054)	5.798	0.000	0.313***	(0.052)	5.998	0.000
mem2560_noed	-0.891**	(0.388)	-2.295	0.022	-0.891***	(0.256)	-3.478	0.001
mem2560_coed	0.652***	(0.103)	6.315	0.000	0.652***	(0.112)	5.822	0.000
per_domhelper	3.093***	(0.405)	7.634	0.000	3.093***	(0.555)	5.575	0.000
dm_urbrur	0.066***	(0.019)	3.539	0.000	0.066***	(0.020)	3.243	0.001
dum_bat	0.229***	(0.049)	4.661	0.000	0.229***	(0.052)	4.407	0.000
dum_bul	0.126***	(0.039)	3.216	0.001	0.126***	(0.037)	3.365	0.001
dum_pam	0.106***	(0.041)	2.620	0.009	0.106***	(0.039)	2.717	0.007
dum_nec	-0.152***	(0.041)	-3.725	0.000	-0.152***	(0.039)	-3.865	0.000
dum_tar	-0.121***	(0.043)	-2.791	0.005	-0.121***	(0.044)	-2.772	0.006
dum_aur	-0.190***	(0.071)	-2.689	0.007	-0.190	(0.129)	-1.475	0.140
Constant	10.630***	(0.078)	136.776	0.000	10.630***	(0.077)	137.269	0.000
Number of obs .	3,237	F(23, 3213)	133.5		Number of obs.	3,237	F(23, 3213)	118.03
R-squared	0.489	Prob > F	0.000		R-squared	0.489	Prob > F	0.000
Adj R-squared	0.485	Root MSE	0.480		Root MSE	0.480		

Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 3: Definition of Variables

Variables	Description
lnpcy1	ln of per capita income (*dependent variable*)
totmem	total members in the household
totmem2	total members in the household squared
dm_roof1	dummy roof1=1 if roof is made of strong material (galvanized, iron, al, tile), 0 otherwise
dm_wall1	dummy wall1=1 if wall is made of strong materials, 0 otherwise
sing_wall2	interaction: (single house dummy) x (dummy if wall is made of light materials)
head_age	age of household head
dm_HHedu1	dummy_HHedu1 =1 if household head has elementary education, 0 otherwise
dm_HHedu3	dummy_HHedu3 =1 if household head has college education, 0 otherwise
no_spouse	household head has no spouse
dm_spedu3	dummy_spedu3 =1 if spouse has college education, 0 otherwise
per_2560hhall	proportion of members between 25 and 60 years old
kids_female_coed	interaction: (proportion of sons/daughters) x (dummy for spouse with college education)
per_domhelper	proportion working as domestic helpers within household
mem2560_noed	interaction: (members ages between 25 & 60) x (dummy with no education)
per_nofw	proportion of members who are overseas Filipino workers
mem2560_coed	interaction: (members ages between 25 & 60) x (dummy with education)
dm_urbrur	dummy urbrur=1 if urban, 0 if rural
dum_bat	dummy bat=1 if Bataan, 0 otherwise
dum_bul	dummy bult=1 if Bulacan, 0 otherwise
dum_pam	dummy pam=1 if Pampanga, 0 otherwise
dum_nec	dummy nec=1 if Nueva Ecija, 0 otherwise
dum_tar	dummy tar=1 if Tarlac, 0 otherwise
dum_aur	dummy baur=1 if Aurora, 0 otherwise

The age of household head is positive and statistically correlated with per capita income. However, the marginal effects of age on income is small: one additional year in age leads to 0.9 percent increase in per capita income, all other things constant.

However, the education attainment of the household head is significant predictor of per capita income. Household head with college education increases per capita income by 18 percent, all other things constant, while household head with elementary education reduces household per capita income by 15 percent. Both of these effects are statistically significant.

Household head without a spouse tends to have lower per capita income. Spouse is a major factor affecting per capita income, especially if the spouse has a college education. Spouse with college education increases per capita income by 13 percent.

The variable *kids_female_coed* is an interaction between the proportion of sons/daughters in the households and the education of spouse. This factor is not only statistically significant, but the partial effect as reflected in the magnitude of the coefficient (1.288) is also relatively significant. These regression results imply the importance of education of spouse[4].

Overseas Filipino Workers (OFWs) are major contributors to household income, as reflected in the coefficient of *per_nofw*, which is the proportion of the number of OFWs in the household. Each additional percentage increase in the proportion OFW members in the household overseas increases per capita income by 69 percent.

Other working members in the household also contribute to per capita income, as reflected in the coefficient of *per_2560hhall*, which is the proportion of household members between 25-60 years old. The coefficient is positive and statistically significant. The educational attainment of these members is also critical as reflected in the interaction two variables, *mem2560_noed* (no education) and *mem2560_coed* (with college education). Members with no education, captured in the former interaction, reduces per capita

Table 4. Comparison between SAE and Official Poverty Estimates in Region III, 2015

		Official Estimates			Small Area Estimates			Official-SAE % Difference Population Under Poverty
	Population	Poverty Incidence, %	Coefficient Variation	Population Under Poverty	Poverty Incidence, %	Coefficient Variation	Population Under Poverty	
Region III	11,184,068	11.2	12.20	1,251,602	11.7	0.21	1,313,759	5.0
Aurora /a/	213,683	26.3	-	56,144	18.6	1.69	39,785	-29.1
Bataan /b/	755,296	2.0	74.32	15,320	7.3	0.58	54,765	257.5
Bulacan /b/	3,283,215	4.5	30.23	146,541	5.9	0.30	195,130	33.2
Nueva Ecija	2,147,656	22.6	13.05	484,304	20.5	0.28	439,226	-9.3
Pampanga /b/	2,602,279	4.9	26.67	128,435	7.5	0.34	194,270	51.3
Tarlac	1,361,763	18.1	16.99	246,967	17.9	0.40	243,964	-1.2
Zambales /b/	820,176	16.8	20.66	137,983	17.9	0.47	146,618	6.3

income, while those with college education, captured in the latter, contribute significantly.

The number of domestic helpers in the household is also a predictor of per capital income. Households with domestic helpers have relatively higher per capita income, than those without domestic helpers.

The coefficients of the dummy variables for the different provinces in Region III indicate that households in located in Bataan, Bulacan and Pampanga have relatively and statistically higher per capita income those than in Aurora, Nueva Ecija, Tarlac and Zambales[5].

Small Area Estimates of Provincial Poverty in Region III

Table 4 presents and compares the official and the SAE of poverty in Region III and the seven provinces. The official poverty incidence in the region in 2015 11.2 percent, with a CV of 12.2 percent. The SAE estimate is very close at 11.7 percent, but with very low CV of 0.21 percent. The SAE estimate results in higher number of people below the poverty threshold compared to the official estimate in the region by 5 percent.

The SAE and official provincial poverty estimates differ notably in several provinces in the region both in terms of incidence, CV of the estimates, and the number of population under the threshold. In Aurora, the official poverty incidence of 26.3 percent is higher than the SAE estimate of 18.6 percent. This difference translates to lower number of population below poverty by 29.1 using SAE compared to the official estimates. The largest gap in the estimates is in Bataan: official poverty incidence is 2.0 percent while SAE is 7.3 percent. As a result, the number of people under poverty is 257.5 percent higher using SAE compared to the official estimates. There are also significant differences in poverty estimates in Bulacan, the largest province, and Pampanga, the regional capital. Relatively smaller difference in poverty estimates is observed in Tarlac, Zambales, and Nueva Ecija.

Conclusion and Insights

Poverty targeting requires accurate information on the number of people under the poverty threshold, the location of households, and the individual household characteristics. In the Philippines, the main source of information on households is the FIES. Since 1985, the statistical authority in the country conducts survey using regions as the sampling domain. It was only in 2018 when the data collecting authority started conducting survey of households using provincial sampling domain. Poverty estimates which are representative at the regional level are too broad for an efficient implementation of poverty targeting policies because poverty incidence varies significantly across provinces in a region, and across municipalities in a province. Although the statistical authority publishes provincial poverty estimates, the estimates have to be considered with great caution because of very small sample sizes. Thus, the official provincial poverty estimates have large coefficient of variations.

To address this data gap on poverty, the literature suggests small area estimates (SAE) of poverty by combining data from household survey with the very rich source of household information in the census of housing and population. In general, SAE adopts a 2-step approach, where in the first step an income-generating function is estimated using household survey data, and in the second stage disaggregated census data on the same household characteristics are inserted into the estimated relationship in the first step to generate estimates of poverty for smaller geographic areas. The paper adopts the 2-step approach to estimate provincial poverty incidence in Region III.

The SAE of poverty in Region III is very close to the official estimate. However, the coefficient of variation of the former is considerably smaller than the latter. There are notable differences between the official provincial poverty incidence (which have large coefficient of variations) and the provincial SAE (which very small coefficient of variations.) As a result, the number of population below the threshold differ considerably between the two sets of poverty estimates across the seven provinces in Region III.

There are two areas which the paper recommends for further research: (i) replicate the SAE exercise to cover all 17 regions in the Philippines; and (ii) once the 2018 provincial-based FIES is released, conduct SAE by estimating provincial income-generating equations and by inserting the 2015 census information on the same household characteristics into the equation to generate estimates of municipal poverty incidence. This set of information is critical for an efficiency implementation of poverty alleviation policies.

Notes

[1] As of this writing, the Philippine Statistical Authority (PSA) has not released the 2018 FIES. Currently, there are 17 regions in the Philippines. Each of the region is composed of several provinces. The country has 88 provinces at the moment.

[2] Merged Labor Force Survey (LFS) and FIES was used.

[3] Improved health and nutrition are additional benefits from higher educational attainment of spouse is established in the literature, but are not accounted for in the present specification of the regression equation.

[4] Zambales is the reference province in the provincial dummy variables, thus omitted in the regression.

References

Cororaton, C, A. Inocencio and A. Sobreviñas, A. Baulita and E. Barnedo. 2018. "The Study on Key Drivers of Poverty Reduction in Central Luzon (Region III)" Final Draft Report Submitted to the National Development and Economic Authority.

Haddad, L., and R. Kanbur. 1990. 'How serious is the neglect of intra household inequality', The Economic Journal, 100, 866-81.

Hentschel, J., J. Lanjouw, P. Lanjouw, and J. Poggi. 2000. "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador". *The World Bank Economic Review*. Vol 14. No 1

Minot, N. 1998. "Generating Disaggregated Poverty Maps: An Application to Viet Nam". Markets and Structural Studies Division, Discussion Paper No. 25. International Food Policy Research Institute, Washington. D.C."

Minot, N. 2000. "Generating Disaggregated Poverty Maps: An Application to Viet Nam, 2000". *World Development*, Vol. 28 No. 2: 319-331.

Minot, N. and B. Baulch. 2002. "Poverty Mapping with Aggregate Census Data: What is the Loss in Precision?" Presented at the Conference "Understanding Poverty and Growth in Sub-Saharan Africa", Centre for Studies of African Economies, Oxford University, 18-20 March 2002.

Minot, N., B. Baulch and M. Epprecht. 2003. "Poverty and inequality in Vietnam: Spatial patterns and geographical determinants", International Food Policy Research Institute, Washington, D.C. and Institute of Developments Studies, University of Sussex.

World Bank. 2005. "Local Estimation of Poverty in the Philippines". Manuscript