10th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 1-2, 2007

Responding to the Needs for Decentralized Planning: Small Area Poverty Estimates

by

Joseph M. Addawe, Arturo M. Martinez, Jr., and Raymond S. Perez

For additional information, please contact:

Author's name	:	Joseph M. Addawe
Designation	:	Statistical Coordination Officer IV
Affiliation	:	National Statistical Coordination Board
Address	:	403 Sen. Gil Puyat Avenue, Makati City
Tel. no.	:	(0632) 896-5390
E-mail	:	jm.addawe@nscb.gov.ph

Co-author's name	:	Arturo M. Martinez, Jr. / Raymond S. Perez
Designation	:	Statistical Coordination Officer II / Statistical Coordination Officer I
Affiliation	:	National Statistical Coordination Board
Address	:	403 Sen. Gil Puyat Avenue, Makati City
Tel. no.	:	(0632) 896-5390
E-mail	:	am.martinez@nscb.gov.ph /rs.perez@nscb.gov.ph

Responding to the Needs for Decentralized Planning: Small Area Poverty Estimates[1]

by

Joseph M. Addawe, Arturo M. Martinez Jr., and Raymond S. Perez[2]

ABSTRACT

Micro-level estimation of poverty has been increasingly becoming prominent among statistical and economic tools geared towards assessment of welfare indicators. The strength of this methodology comes from its capability to address the limitation imposed by detailed household surveys that include reasonable measures of income or expenditures but are rarely representative at low levels of disaggregation to yield statistically reliable poverty estimates. In particular, information from census (or other large sample) data, which are of sufficient size to allow disaggregation at local level but do not have data on welfare indicators, is combined with detailed welfare information available from household surveys. This is implemented by constructing a model of (log per capita) income / expenditure using the survey data set, restricting regressors to those that can be linked to households in both data sets. The expected level of poverty indicator is then estimated given the census-based observable characteristics of the population of interest using the "synthetic" parameter estimates derived from the model.

In the Philippines, official estimates of poverty incidence are available at the national, regional and provincial levels. In response to the need of stakeholders, NSCB implemented a project, *Estimation of Local Poverty in the Philippines*, through funding assistance of World Bank ASEM Trust Fund. The project's outputs include poverty statistics for the 1,623 municipalities, with the year 2000 (being both a census and survey year) as the reference period.

At present, the estimates are being updated for the year 2003, a survey but a non-censal year in the country. However, the small area estimation methodology gets a little complicated when the reference years of the two data sources are not the same. (Hence, the linking of explanatory variables from survey to census is not straightforward). When census is conducted less frequently than household surveys, small area poverty estimation for intercensal years requires the availability of either panel data or time-invariant regressors that can be extracted from the data sets.

This paper discusses the methodology of constructing welfare model(s) for censal and intercensal years. Further, the paper enumerates the empirical usefulness of small area poverty estimates to policy-makers, stakeholders and antipoverty organizations in the country, based on the experience gained in the dissemination and utilization of the outputs of the project on the *Estimation of Local Poverty in the Philippines* project.

I. Introduction

The National Statistical Coordination Board (NSCB) has regularly produced official poverty estimates for the Philippines up to regional level. In 2003, NSCB has come up with official poverty statistics at the provincial level for 1997 and 2000. But recently, the National Statistics Office (NSO) implemented some changes in the design of the Family Income and

[1] Paper presented during the 10th National Convention of Statistics (NCS)

[2] Statistical Coordination Officer IV, Statistical Coordination Officer II, and Statistical Coordination Officer I respectively, of the National Statistical Coordination Board. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the NSCB.

Expenditure Survey (FIES), the main source of income and expenditure data used for poverty estimation. With these changes, official poverty estimation at the provincial level is usually accompanied by strong caveats on the statistical reliability of some estimates. On the contrary, there has been an increasing demand from policy makers and planners for a more disaggregated poverty statistics for a more effective aid programs targeting. The KALAHI project of the National Anti-Poverty Commission (NAPC) was developed to strategize reducing poverty in the country. But resources allocated to the project need to be targeted to most vulnerable barangays or communities to ensure that poverty reduction efforts provide maximum impact and benefit those who need them most.

In response to the need of stakeholders, the National Statistical Coordination Board (NSCB) implemented a poverty-mapping project in 2005 using small area estimation technique for the year 2000 and came-up with regional, provincial and municipal poverty estimates for the whole Philippines. This undertaking was with funding assistance from the World Bank ASEM Trust Fund. At present, NSCB is currently undertaking an intercensal update of the previous small area project using 2003 survey data. The main purpose of these projects were to generate 2003 regional, provincial and municipal level estimates of poverty incidence, gap, and severity; using the income-based poverty lines with an improved precision of the poverty estimates.

This paper presents the methodology, results and some practical policy uses of combining the data from the 2000 Family Income and Expenditure Survey (FIES) and Labor Force Survey (LFS) with 2000 Census of Population and Housing (CPH) using small area estimation technique to come up with regional, provincial and municipal poverty rates. It also discusses some preliminary results and steps taken to update the small area poverty estimates using 2003 FIES and LFS with 2000 CPH.

II. What is Small Area Poverty Estimation?

The Small Area Estimation (SAE) (Rao, 1999; Gosh and Rao, 1994) is a statistical technique designed for improving sample survey estimates through the utility of auxiliary information from the census. It combines the information in surveys and censuses to generate more robust poverty estimates with the same level of disaggregation typically specified by surveys. Small area estimation is also able to generate reliable estimates at a lower level of disaggregation (i.e., provincial, city/municipal).

The application of small area estimation to poverty is relatively new especially to developing countries. But recently, SAE technique and its variants have been applied to a number of other welfare indicators aside from poverty incidence. Several countries in Latin America (Ecuador, Nicaragua, Guatemala, Brazil), in Africa (Kenya, Uganda, South Africa, Morocco, Mozambique) and in Asia (Bangladesh, Viet Nam, Cambodia, Thailand, Lao, Azerbaijan) have used small area estimation to measure more disaggregated poverty or other welfare indicators. Small area estimation have been widely used to measure poverty at local level within countries but a number of studies have already used SAE to measure not only poverty but also some related indicators like malnutrition (Fujii 2004 and Bangladesh Bureau of Statistics 2004), education (Fujii 2003), child mortality (Balk et. al 2004) and even disabled groups of the population (Hoogeven 2003).

III. Methodology

The NSCB small area estimation project as well as this paper used a variant of the small area estimation techniques, a technique developed by a research team at the World Bank (Chris Elbers, Jean O. Lanjouw, and Peter Lanjouw), and is called the ELL methodology. Typically, small area estimation approach begins with a household survey, to acquire a reliable estimate of household income (y) and calculate more specific poverty measures linked to a poverty line. A common set of explanatory variables (x) at the household unit level in both the survey and census is then used to estimate the statistical relationship between y and x in the survey. Once a robust model has been identified for the survey, apply the final model to the census data at household-unit level to predict per capita household income (including error estimate). This household unit data set can then be aggregated to small areas such as municipalities or even barangays, to obtain more robust estimates of the percentage of households living below poverty line. Theoretical discussions of the small area estimation technique are discussed in detail in the papers by Elbers, C, Lanjouw, J. and Lanjouw, P.(2002, 2003).

Box 1. Basic Steps of small area estimation for poverty mapping [3]

“Small area estimation for poverty mapping involves detailed analysis of two main source of data: a household survey; and a population census. In the first stage of the analysis the two data sources are subjected to very close scrutiny with an eye towards identifying a set of common variables. In the second stage the survey is used to develop a series of statistical models which relate to per capita income or consumption to the set of common variables identified in the first stage. In the final stage of the analysis the parameter estimates from the second stage are applied to the population census and used to predict income or consumption for each household in the population census. Once such a predicted income or consumption measure is available for each household in the census, summary measures of poverty (and/or inequality) can be estimated for a set of households in the census. Statistical tests can be performed to assess the reliability of the poverty estimates that have been produced.

First stage: This involves a painstaking comparison of common variables across the household survey and the population census. The idea here is to identify variables at the household level which are defined in the same way in both the household survey and census. One way to check is to look at means and percentiles of the variables in both data sources. A concurrent exercise that can be carried out is the compilation of a database at a level of aggregation higher than the household, which can be inserted into the household level survey and census databases. These variables serve as proxy indicators to unobserved factors that might be affecting household income or consumption. These sub-district level data may comprise a wide range of variables (i.e. construction of schools, public spending figures, infrastructure availability etc.) *In this paper municipal means and barangay means were generated.*

Second stage: This stage involves the econometric estimation of models of income or consumption on the set of household level survey data. A large number of diagnostic tests will need to be carried out, and a large number of different specifications are likely to be experimented with.

Third stage: This stage is associated with the imputation of income or consumption into the census data at the household level, and the estimation of poverty measures at a variety of levels of spatial disaggregation. Software has been developed which allows this phase to be carried out mechanically.”

IV. Implementation of small area poverty estimation in the Philippines for 2000 [4]

“

Selection of auxiliary data

The auxiliary data *X* used to predict the target variable *Y* can be classified into two types: the survey variables, obtainable or derivable from the survey at household or individual level, and the location variables applying to particular larger geographic units. The latter include averages of census variables at a municipal level.

It is important to note that any auxiliary variables used in modeling and predicting should be comparable in the estimation (survey) data set and the prediction (census) data set. In the case of survey variables, we begin by examining the survey and census questionnaires to find out which questions in each elicit equivalent information. In some cases equivalence may be achieved by collapsing some categories of answers. For example the categories recording educational attainment are different in the census and survey data,

[3] Lanjouw, Peter. (2003) “Estimating Geographically Disaggregated Welfare Levels and Changes”. *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Technique and Tools*. New York: Oxford University Press.

[4] This section is lifted from Chapter 4 of “Estimating Local Poverty in the Philippines” published by NSCB in cooperation with World Bank (2005). This chapter is jointly written by the project’s foreign consultants, Dr. Stephen Haslett and Dr. Geoffrey Jones.

but by focusing on broader categories of no education, elementary education, high school and college we were able to produce education variables which were comparable. When common variables have been identified the appropriate statistics are compared for the survey and census data. In the case of categorical data we compare proportions in each category: for numerical data, such as household proportion of children, we compare the means and standard deviations. For this purpose confidence intervals can be calculated for the relevant statistics in the survey data set, taking account of the stratification and clustering in the sample design. The equivalent statistic for the census data should be within the confidence interval for the survey. In some cases variables were dropped at this stage. For example, tenure status (own, rent, rent free with consent or rent free without consent) was found not to match sufficiently well.

The inclusion of location effect variables should be straightforward since they can be merged with the survey and census data using indicators for the geographical unit to which each household or individual belongs. This can be problematic in practice however, because of changing boundaries and the creation of new provinces, municipalities and barangays. The FIES survey and the 2000 census used different barangay classification so that it was not possible to merge with both survey and census in a comparable way at barangay level. Furthermore there was no barangay information in the census long-form. As an alternative, municipal-level census means were calculated for both short- and long-form census variables and these merged successfully with the survey data. Even at municipal level there were difficulties, particularly in Mindanao region with the creation of provinces 82 (Compostela Valley) and 98 (Marawi City and Cotabato City). We used the 2000 census coding in all cases, recoding the survey data to make it compatible.

Once all usable auxiliary data have been assembled, it may be necessary to delete some case or variables where there are missing values or outliers. In our case the educational attainment of the spouse was missing in a large number of households where a spouse was present, so this variable, although possibly useful in the regression modeling, had to be dropped.

First stage regressions

The selection of an appropriate model is a difficult problem. There are a large number of possible predictor variables, with inevitably a good deal of multicollinearity. Some of these are numerical (e.g. famsize denoting family size), some represent different values of a categorical variable (e.g. hms_sing, hms_mar, etc denoting marital status of head of household) and some are ordered categories (e.g. fa_xs to fa_xxxl denoting floor area). If we also include two-way interactions, there are well over a thousand variables to choose from. (A "two-way interaction" is the product of two basic or "main-effect" variables). Squares or other transformations of numerical variables could also be considered. With this, we must be careful not to over-fit, so the number of predictors included in the model should be small compared to the number of observations in the survey, but there is also the problem of selecting a few variables from the large number available which appear to be useful, only to find (or even worse, not find) that an apparently strong statistical relationship in the survey data does not hold for the population as a whole.

The search for significant relationships over such a large collection of variables must inevitably be automated to a certain extent, but we have chosen not to rely entirely on automatic variable selection methods such as stepwise or best-subsets regression. We have, in general, instead adopted the principle of hierarchical modeling, in which higher-order terms such as two-way interactions are included in the model only if their corresponding main-effects are also included. Thus we begin with main-effects only, and add interaction and nonlinear terms carefully and judiciously. We look not just for statistical significance but for a plausible relationship. For example, the effect of household size on log expenditure was expected to be nonlinear, with both small and large households tending to have larger per capita expenditure. The square of household size, centered on the mean, was added and found to be significant.

Some implementations of ELL methodology have fitted separate models for each stratum defined by the survey design. This has the advantage of tailoring the model to account for the different characteristics of each stratum, but it can increase the problem of over-fitting if some strata are small. First the country was divided into 31 domains, each domain comprising the urban or rural barangays of one region. (There were 16 regions but one, NCR, has no rural barangays). An initial model was fitted to the whole country, using the combined FIES/LFS data and selecting variables based on plausibility of the estimated relationship as well as statistical significance. Census means were not used at this stage, but we were still able to achieve an R^2 over 60 percent. The purpose of this stage was to identify a reduced subset of useful variables and hence diminish the risk of including spurious relationships through automatic selection from a large pool of candidates. We then tried a "domain-based" approach, fitting separate models for each domain but chosen from the reduced variable set, and a "global" approach, expanding out our initial model to include separate intercepts in each domain. In both approaches census means were added to reduce the cluster-level residual variation, but their use was kept to a minimum, as they were only available at municipal level and therefore likely to lead to spurious relationships, because the number of candidate census means is comparatively large relative to the number of sampled municipalities. Although our final model was a single model, it was in fact a compromise between the global and domain-based approaches, with a strong emphasis on the global but with a few coefficients in the global model being allowed to vary through the use of interaction terms.

In our own model, there are required interaction terms so that in this sense our model is not the extreme single model. There are domain specific constants, urban/rural effects and also the corresponding interaction terms. The domain-specific constants, or intercepts, in our models can be seen to be quite similar, although the differences were statistically significant overall. The intercepts for the rural areas were significantly lower than the corresponding urban intercepts in each region, indicating as expected a generally lower average per capita income and expenditure in the rural barangays. For the income model, the impact of the variables coed and fa_xxxl (which relate to college level education and the largest floor area classification respectively) was found to be different in the urban and rural areas, and the impact of the education variables hsed and coed was reduced in ARMM.

We departed from the usual ELL implementation in our use of a single-stage, robust regression procedure for estimating model, rather than the two-stage procedure of ordinary least squares followed by estimation of a variance matrix for generalized least squares. This gives the advantages of properly accounting for the survey design and obtaining consistent estimates of the covariance matrices in a single step (Skinner et al., 1989; Chambers and Skinner, 2003). These covariance matrices were saved, along with the parameter estimates and both household- and cluster-level residuals, for implementation of the prediction step.

Heteroscedasticity modeling

We amended the regression model for the household-level variance to prevent very small residuals from becoming too influential. We used a slightly different amendment:

$$L_{ij} \equiv \ln\left(\frac{\hat{e}_{ij}^2 + d}{A - \hat{e}_{ij}^2}\right) = Z_{ij}a + r_{ij}$$

where d is a small positive constant and A is chosen to be just larger than the largest $\hat{e}_{ij}^2$ (e.g. d = 0.0001, $A = 1.05 \times \max \hat{e}_{ij}^2$). These choices can be justified empirically by graphical examination of the L_{ij}, which should show neither abrupt truncation nor extreme outliers. The predicted value of the household-specific variance, using the delta method, then becomes:

$$s^2_{e,ij} = \left[\frac{AB_{ij} - d}{1 + B_{ij}}\right] + \frac{1}{2}\hat{s}^2_r \left[\frac{(A+d)B_{ij}(1 - B_{ij})}{(1 + B_{ij})^3}\right]$$

where $B = e^{Z?}$. There was actually very little heteroscedasticity and this step could arguably have been omitted. However it was noticeable that in the domain-based models there were some variables which were consistently being selected for their significance, so it was thought better to include this aspect of the model in all models tested, even though the effects are slight.

Simulation of predicted values

Simulated values for the model parameters a and *ß* were obtained by parametric bootstrap, i.e. drawn from their respective sampling distributions as estimated by the survey regressions. Simulation of the cluster-level and standardized household-level effects h_i and e^*_{ij} presents several possible choices. A parametric bootstrap could be used by fitting suitable distributions (e.g. Normal, *t*) to the residuals and drawing randomly from these. We chose here a non-parametric bootstrap in which we sample with replacement from the residuals, i.e. from the empirical distributions. A total of 100 bootstrap predicted values Y^b_{ij} were produced for each unit in the census and for each target variable.

Production of final estimates

Since a log transform was applied in modelling income and expenditure, we first undo this transformation by exponentiating, e.g. predicted expenditure $E^b_{ij} = e^{Y^b_{ij}}$. The predicted values can then be accumulated at the appropriate geographic level. We used primarily municipal and provincial levels, but in addition produced separate estimates for urban and rural poverty estimates at provincial level. Regional level estimates were also produced for comparison with the FIES-based estimates.
For the income and expenditure information the census units are households and the target variables per capita average values, so the accumulation needs to be weighted by

household size. Thus for example the formula for P_R^b the *b*th bootstrap estimate of poverty incidence in area *R* is amended to:

$$P_R^b = \sum_{ij \in R} n_{ij} \cdot I(E_{ij}^b < z) \Big/ \sum_{ij \in R} n_{ij}$$

where n_{ij} is the size of household *ij* in *R*.

The 100 bootstrap estimates for each region, e.g. $P_R^1 \ldots P_R^{100}$ were summarized by their mean and standard deviation, giving a point estimate and a standard error for each area. A detailed theoretical and practical discussion on the use of small area estimation for poverty mapping in the Philippines in 2000 is presented in the publication *Local Estimation of Poverty in the Philippines* by NSCB and the World Bank 2005.
"

V. Updating small area poverty estimates in the Philippines for the intercensal year 2003

With the intent to build up on the results of the 2000 Poverty Mapping Project, the National Statistical Coordination Board (NSCB) is currently undertaking a project to update the small area poverty estimates for 2003 also being funded by the World Bank.

The project entitled Intercensal Updating of Small Area Poverty Estimates commenced in May 2006. With guidance from Dr. Peter Lanjouw, Dr. Roy Van der Weide and Dr. Zita Albacea, NSCB used a slightly different approach as compared to the 2000 exercise. For the 2003 intercensal small area poverty estimation, NSCB is using PovMap, a software developed by Development Research Group of the World Bank for small area poverty estimation, although data preparation (stage 0) was still done using Stata software and later imported to PovMap for the modeling and generation of estimates. (In the 2000 SAE exercise, Stata was used in the data preparation, modeling and generation of estimates). It was also agreed that for the 2003 intercensal small area poverty estimation exercise, domain based models using regional disaggregation would be used instead of a single model for the whole country. This is to capture the heterogeneity of the regions. Due to this approach that later on in the project, 2000 SAE would have to be generated also to be able to determine whether the changes in the estimates are driven by the model or actual changes from 2000 to 2003. This would also serve as guidance for 2003 users.

Selection of auxiliary data

Small area poverty estimation requires a close correspondence between the survey and the census variables, as this is a pre-requisite to having reliable estimates. Considerable time and attention is therefore devoted to identifying common variables. While census is conducted less frequently than household surveys, small area poverty estimation for intercensal years requires the availability of either panel data or use of time-invariant regressors that can be extracted from the data sets. Time invariant variables were identified and generated during the data preparation and during initial model building phase. Time invariant variables include barangay / municipal characteristics (cluster variables) and some household-level education variables. Ideally, these variables must be defined such that the characteristic of interest is not expected to change from 2000 to 2003 and hence, time invariant. Examples of barangay/ municipal characteristics that are used as candidate regressors include an indicator whether a barangay has access to main roads, an indicator whether a barangay has a newspaper circulation, among others. At the household level, education variables are defined in this fashion: 1- if household head is at least high school graduate, 0 – otherwise, etc. It must be reiterated that these variables were constructed such that the characteristics of interest under consideration are not expected to change, at least between 2000 and 2003. For a complete list of candidate regressors, readers may refer to Appendix 3. Having mentioned this, it can be noted that there is a very short list of time invariant variables that can be defined at the household level. Consequently, resulting models may only capture a limited proportion of variation among household units. A possible way around this is to generate interaction between the cluster and household-level variables that were defined to be time invariant. Reassessment of the data quality and initial construction of poverty indicator models are continuously being done. This is to assure the robustness of resulting models and estimates.

Regression modeling and simulation of estimates

Survey and census data have been prepared in Stata and are imported to PovMap software for the regression modeling and estimates generation. PovMap is a software that allows the user to select the variables to include in the model manually or automated. Once the income or consumption model is generated, the cluster and household effect, the idiosyncratic model, the simulation and simulation results can be readily generated in the software. In the use of the PovMap software Initial exercises for 6 selected regions(Region

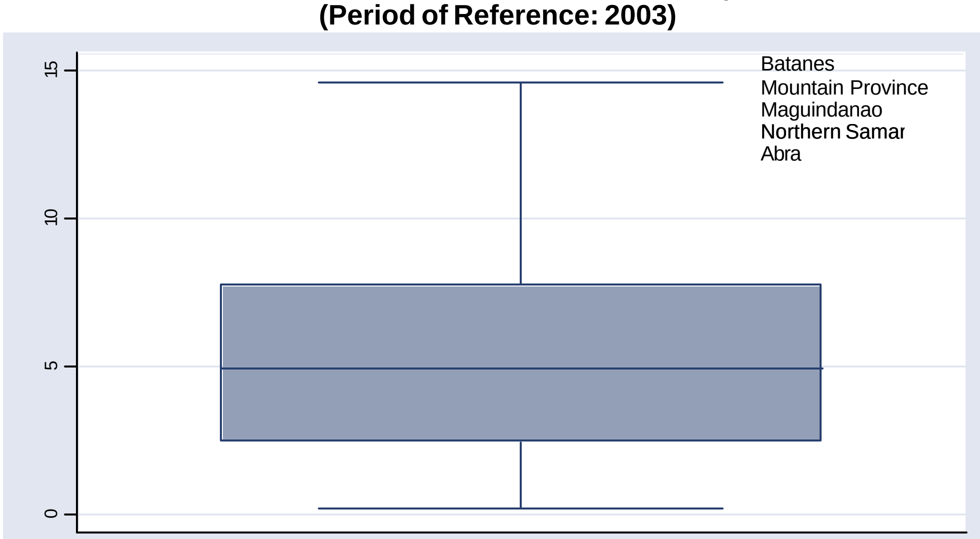
1, 5, 7, 10, CAR and Caraga) for 2003 were continuously revised until such time an estimate is generated with considerable reliable estimate. Steps taken in the refinement of model construction to come up with improved estimates are the following:

a. Construction of different types of models using different numbers of regressors. In the NSCB exercise, vn, vn/2 and vn/4 number of regressors were used to model weighted and unweighted survey data.

b. The development of significant relationships and regressors over large number of variables was initially automated to a certain extent, using some variable selection methods such as stepwise regression. Later on, variables were added one at a time when we have determined reasonable number of consistent and significant relationships on each domain. Carefully, the signs of the parameter estimates, its magnitude and its statistical significance from 0 are monitored and evaluated.

c. Note that PovMap provides an option to compute poverty statistics at different level of disaggregation. For this intercensal small area estimation exercise, regional, provincial and municipal estimates are generated using the software. During the development of the models and generation of initial estimates, measures were done to study and look for possible reasons behind deviations between SAE estimates are not consistent with official estimates at the regional and provincial levels.

Absolute Differences Between Official Provincial Poverty Incidence with SAE (Period of Reference: 2003)



Noting the data limitations involved in computing official provincial poverty statistics, it can be expected that there may be some considerable deviations between this set and the set of model-based small area estimates. For instance, significant differences between official and preliminary SAE for the following provinces: Batanes, Mountain Province, Maguindanao, Northern Samar and Abra, have been noted. The sample size and composition for these provinces, in the survey may be one of the potential causes of disparity (e.g., In the survey, Batanes has small sample size while no observations from urban areas in Abra were sampled in 2003).

Two of the major key assumptions for this methodology are:

(a) Independent variables in the model are time-invariant. Unlike the 2000 poverty mapping project in which the reference period of the two data sets coincide (i.e, both a survey and a census were conducted in 2000), only a household survey was conducted in 2003. To be able to generate small area statistics of poverty, we need to map the model constructed in 2003 survey to 2000 census. Since ideally, census is a complete coverage of the entire population, this mapping will likely provide statistical reliable estimates at the lower level of disaggregation.

(b) Migration (at least among municipalities) between 2000 and 2003 is not statistically significant. Otherwise, if we can not assume that the characteristics of the population living in a particular municipality, in 2003 are the same with that of 2000's population of interest, then the validity of mapping the model constructed using 2003 survey (plus auxiliary data) to 2000 census may be jeopardized. In particular, if this assumption does not stand on solid grounds, then poverty statistics generated through this intercensal updating procedure for a particular municipality may be biased.

VI. Evaluation of small area estimates for 2000 Poverty Mapping project

The following table shows the comparison of regional level head count poverty incidence using the SAE domain based model and single model as against the official estimates in 2000.

Comparison of regional level per capita poverty incidence estimates for 2000

Region	Survey-based (FIES)[5]		SAE (domain-based models)		SAE (global model)	
	Poverty Incidence	SE	Poverty Incidence	SE	Poverty Incidence	SE
NCR	0.0771	0.0061	0.0693	0.0051	0.0829	0.006
CAR	0.3817	0.0184	0.361	0.012	0.3825	0.0159
Region I	0.3558	0.0195	0.3427	0.0123	0.3397	0.0124
Region II	0.3168	0.0253	0.3354	0.0127	0.3553	0.0144
Region III	0.2089	0.0123	0.224	0.008	0.2347	0.0075
Region IV	0.2619	0.012	0.2812	0.0059	0.2923	0.0066
Region V	0.5281	0.0208	0.499	0.0126	0.4948	0.0133
Region VI	0.4572	0.0159	0.416	0.0097	0.434	0.0092
Region VII	0.3831	0.0199	0.3708	0.0119	0.3965	0.0118
Region VIII	0.4529	0.0237	0.4376	0.0114	0.4449	0.0163
Region IX	0.4555	0.0264	0.4512	0.0144	0.4718	0.0185
Region X	0.3993	0.022	0.4007	0.0172	0.3822	0.0155
Region XI	0.3691	0.0219	0.3633	0.0101	0.3656	0.0144
Region XII	0.5361	0.0213	0.5332	0.0118	0.5157	0.0131
ARMM	0.6446	0.025	0.612	0.016	0.6337	0.0171
Caraga	0.502	0.023	0.5287	0.0119	0.5064	0.0153

*using old regional configuration estimate

VII. Dissemination and Validation of the 2000 SAE

Provincial validation exercises were conducted in order to assess the acceptability and consistency of the SAE estimates generated with the available indicators at the municipal level as well as with the expert opinion and assessment of key informants. These activities were conducted in two provinces (Batanes and Palawan) of Luzon, one each for Visayas (Bohol) and Mindanao (Surigao del Norte). Participants were from the local government units, the academe and non-government organizations. A National dissemination forum was also conducted to: a) present the project report, providing details on the methodology, variables used, diagnostic tests undertaken, and the results; b) serve as a venue for the exchange of ideas and discussion of the provincial and municipal level poverty estimates produced through the project; and c) develop awareness among national government agencies, the academe, non-government organizations, local government units, and other institutions/organizations of the importance of the poverty mapping methodology and the results generated. Comments and suggestions were compiled and incorporated to

[5] This is different from the latest official poverty statistics for 2000. The estimates provided in the table above used the old set of poverty lines. This set of estimates were provided to facilitate comparability with small area estimates depicted in the table because the latter also used the old set of poverty thresholds. The latest official head count poverty rates stand at 33 and 30 percent in 2000 and 2003, respectively.

the future generation of estimates. Poverty estimates generated were compared with the official/survey estimates. It showed that the estimates were more or less close to each other, not only at the national level but also at the regional level with urban/rural disaggregation. There could be regions that did not match well, but they were within the bounds of acceptable statistical uncertainty. Also, for the ranking of the poorest forty provinces, there were few provinces that were included in the poorest forty based on the SAE but were not in the official statistics. But in general, the agreements of the rankings were good.

VIII. Empirical uses of the 2000 small area poverty estimates

The small area estimates generated for 2000 was disseminated to the local government units, academe, national government agencies and non-government organizations to provide policy-makers, stakeholders and anti-poverty organizations in the country a basis or to provide support in there planning activities in targeting localities for poverty intervention and alleviation. Following are some examples wherein SAE estimates were directly or indirectly used by the government agencies.

a. Small Area Poverty estimates for 2000 were recently used as one of the criterion in identifying target municipalities for Phase II of the Cordillera Highland Agricultural Resources Management (CHARM) project by the Department of Agriculture. CHARM is a DA project which seeks to reduce poverty and improve the quality of life of rural communities. It was first implemented in 16 municipalities in the provinces of Abra, Benguet and Mt. Province. In the proposed CHARM Phase II, it will cover 37 poor municipalities in 6 provinces in the Cordillera Administrative Region (CAR).

b. In region 1, most of the municipal governments used the 2000 SAE poverty estimates as part of their local MDG monitoring and action plans.

c. The Department of Social Welfare and Development (DSWD) used the 2000 small area poverty estimates as one of the indicators in determining pilot areas to be covered by its upcoming plan to adopt the conditional cash transfer strategy for poverty alleviation and social assistance. The Conditional Cash Transfer (CCT) is a scheme which provides money to poor families, on the condition that they make investments in human capital like sending their children to schools or bringing them to health centers regularly. This strategy has been proven successful in a number of Latin American countries in alleviating poverty.

d. The National Nutrition Council in Region VII used the SAE as input in assessing the nutritional situation of different municipalities in the region.

e. The estimates were also used by PhilHealth as one of their basis in identifying local government units / beneficiaries of PhilHealth assistance.

f. Non-government organizations used the estimates in determining priority municipalities in Leyte for sponsorship program for schooling of indigent children, and for micro-enterprise development projects.

The diversity of fields on which the results of the 2000 poverty mapping project were used as inputs, can be noted. This may affirm the operational usefulness of having poverty statistics at lower levels of disaggregation, in policy making and decentralized planning of different institutions in the society.

IX. Conclusion

By combining survey with census data, small area poverty estimates were derived up to municipal level. For 2000, a single model supplemented by different urban and rural effects within each region was developed to predict log per capita household income. The estimates are in general consistent with official estimates. Small area estimates also show that they are to some extent more precise than official estimates as shown in their low standard errors. The series of validation exercises and forums conducted provided some inputs to the improvement of the estimates as well as some explanations to the estimates generated. It has also shown that small area estimates are generally consistent with participants' assessment as well as general acceptability to the pubic.

The 2000 small area poverty estimates have been extensively disseminated to the public that it has been very useful for targeting interventions in poverty alleviation.
For 2003, the methodology shows that with the use of time invariant variables, small area poverty estimates can be updated on a non-census year, using more assumptions. These assumptions are necessary so that the mapping between 2003 survey to 2000 CPH can be considered, up to a certain extent, valid.

It is interesting to note that the usefulness of 2000 small area poverty statistics found its way to help different sectors in the society in planning appropriate intervention programs to alleviate economic hardship in the country. Hence, the commitment and political will to advocate and generate better statistics may facilitate a more efficient forms of decentralized planning geared towards significant poverty reduction.

References

Bangladesh Bureau of Statistics and the United Nations World Food Program. 2004. "*Local Estimation of Poverty and Malnutrition in Bangladesh*."

Elbers, C., Lanjouw, J. and Lanjouw, P. 2002. "*Micro-Level Estimation of Welfare*." World Bank.

Fujii, Tomoki. 2003. "*Commune-level Estimation of Poverty Measures and its Application in Cambodia*." U.C. Berkeley.

Fujii, Tomoki. 2004. "*Micro-level Estimation of Child Nutrition Indicators in Cambodia*." U.C. Berkeley.

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

Hoogeven, J., Emwanu, T., and Okwi, P. 2003. "*Updating Small Area Welfare Indicators in the Absence of a New Census*." World Bank.

Hoogeven, Johannes G., Thomas Emwanu, Paul Okiira Okwi. 2003. "*Updating Small Area Welfare Indicators in the Absence of a New Census*." World Bank.

Lanjouw, Peter. 2003. "Estimating Geographically Disaggregated Welfare Levels and Changes". *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Technique and Tools*." New York: Oxford University Press.

Lanjouw, Peter. 2004. "*The Geography of Poverty in Morocco: Micro-level Estimates of Poverty and Inequality from Combined Census and Survey Data*." World Bank.

National Statistical Coordination Board (NSCB) and World Bank. 2005. "*Estimation of Local Poverty in the Philippines*."

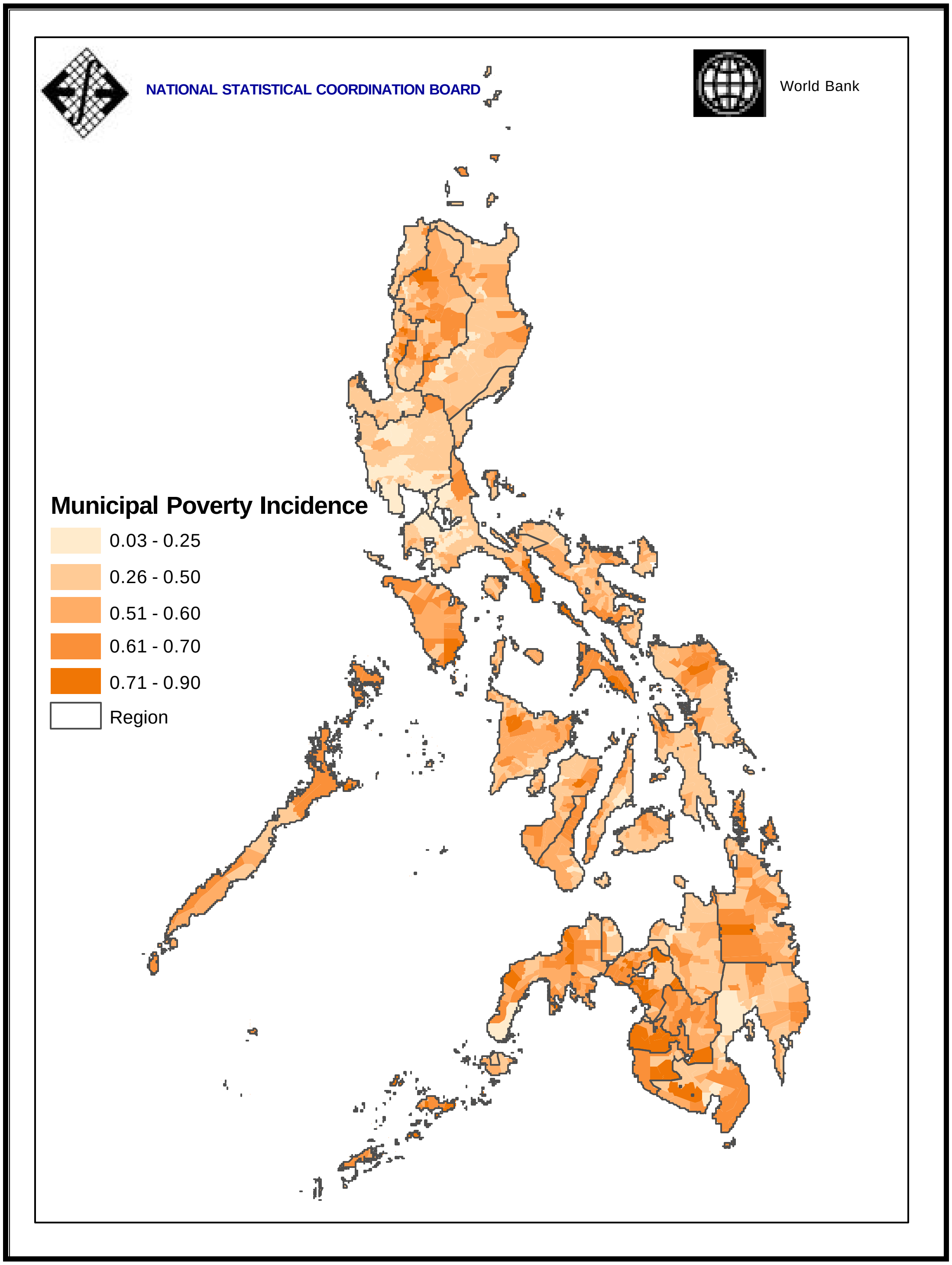
Appendix

Appendix 1, Municipal Poverty Map of Philippines for 2000

Appendix 2, List of candidate regressors for Intercensal Updating of Poverty Statistics

Appendix 3, Municipal Poverty Map of Region III for 2003 (using preliminary intercensal small area estimates)

Appendix 1: SAE 2000 results (poverty map)[6]



6 This figure is lifted from Estimating Local Poverty in the Philippines publication of NSCB in cooperation with World Bank (2005).

Appendix 2: List of candidate regressors for Intercensal Updating of Poverty Statistics[7]

Cluster level variables (municipal or barangay – level):

VARIABLE		DEFINITION
1	hea_rel_mus	proportion of household heads in the barangay whose religion is Islam
2	hea_rel_oth	proportion of household heads in the barangay whose religion is not Islam but is not unknown
3	per_disa	proportion of household members in the barangay who have disability
4	head_nohere	proportion of household heads in the barangay who did not live in the same city/municipality five years ago
5	head_abroad	proportion of household heads in the barangay who lived in a foreign country, five years ago
6	per_indig	proportion of household members in the barangay who are considered indigenous people
7	hou_repair	proportion of houses in the barangay that need major repair
8	hou_dilap	proportion of houses in the barangay that are condemned/dilapidated
9	hou_reno	proportion of houses in the barangay that require/ are under renovation
10	hou_const	proportion of houses in the barangay that are under construction
11	hou_unfconst	proportion of houses in the barangay which can be considered as unfinished construction
12	hou_nrprtd	proportion of houses in the barangay whose state of repair was not reported
13	hou_9600	proportion of houses/building in the barangay which were constructed in 1996 or later
14	bgy_towncity	1 if barangay is a part of the town/city proper or former poblacion of the municipality, or poblacion/city district
15	bgy_streets	1 if the barangay has a street pattern, i.e. networks of streets of at least three (3) streets or roads
16	bgy_highway	1 if the barangay is accessible to the national highway
17	bgy_provcap	1 if the barangay has a tow n/city hall or provincial capitol
18	bgy_church	1 if the barangay has a church, chapel or mosque with religious service at least once a month
19	bgy_park	1 if the barangay has a public plaza or park for recreation
20	bgy_cemetery	1 if the barangay has a cemetery
21	bgy_market	1 if the barangay has a market place or building were trading activities are carried on at least once a wwek
22	bgy_elschool	1 if the barangay has an elementary school
23	bgy_hischool	1 if the barangay has a highschool
24	bgy_college	1 if the barangay has a college/university
25	bgy_library	1 if the barangay has a public library
26	bgy_hosp	1 if the barangay has a hospital
27	bgy_health	1 if the barangay hasa puericulture center/barangay health center
28	bgy_hall	1 if the barangay has a barangay hall
29	bgy_housprj	1 if the barangay has housing projects (government or private)
30	bgy_news	1 if the barangay has a newspaper circulation
31	bgy_telep	1 if the barangay has telephone
32	bgy_teleg	1 if the barangay has telegraph
33	bgy_elep	1 if the barangay has electric power
34	bgy_post	1 if the barangay has postal service

[7] This list is only limited to single effect variables. As mentioned in the text, interactions among these variables were also considered as potential independent variables during the model building procedure.

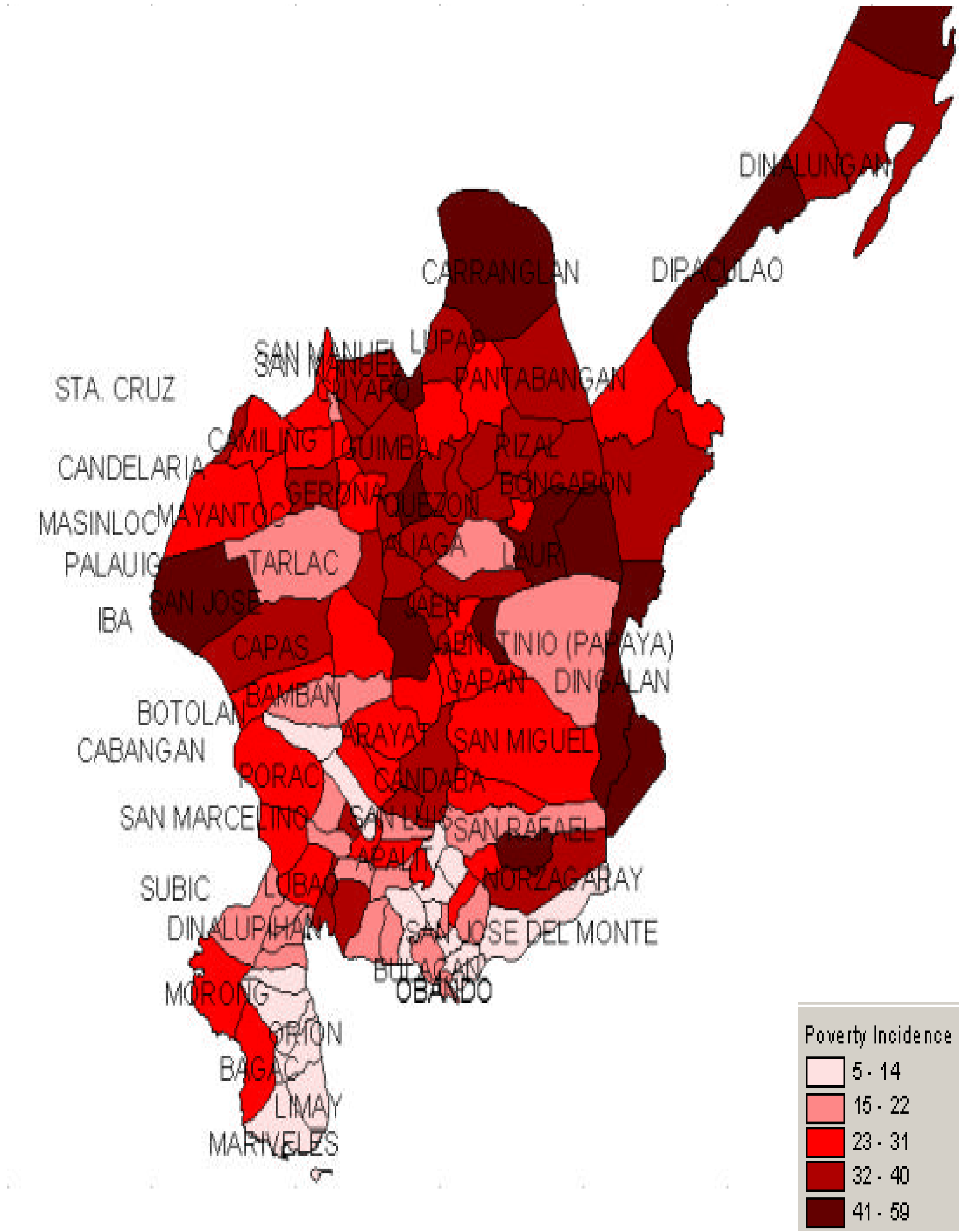
	VARIABLE	DEFINITION
35	bgy_comwork	1 if the barangay has community works system
36	bgy_nstore	average number of (wholesale store, department store, bazaar, hardware store, drugstore, sari-sari store and other store with current merchandise worth P600 or more; gasoine station) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
37	bgy_nfactory	average number of (manufacturing establishments like rice or corn mill, tailor or dress shop or shoe factory, furniture factory, blacksmith shop) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
38	bgy_nrepair	average number of (auto repair shop, vulcanizing shop and other repair shops) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
39	bgy_ncafe	average number of (restaurants, cafeteria, or refreshment parlor excluding temporary restaurants, cafeteria, or refreshment parlor; beauty parlor; barber shop; industry shop; funeral parlor; and other personal services establ;ishments) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
40	bgy_nhotel	average number of (hotel dormitory, and other lodging places) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
41	bgy_nplay	average number of (recreational establishments like theater or movie house, night club, cabaret, bar, beer garden, billiard hall, bowling alley, pool room, etc.) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
42	bgy_nbank	average number of (banking institution, pawnshop, financing/investment or insurance company or agency, etc.) in the barangay (where the value of the variable in the census is 10 if there are more than 10 establishment of this type)
43	Per_ind_1t5	% of persons employed in agriculture, hunting and forest
44	Per_ind_45	% of persons employed in construction
45	Per_ind_52	% of persons employed in retail trade
46	Per_ind_60	% of persons employed in land transport
47	Per_wor_abr	
48	Per_wor_prh	% who worked for private household
49	Per_wor_pre	% who worked for private establishment
50	Per_wor_gov	% who worked for private government
51	Per_nonphi	% of non-Philippine citizens
52	Per_lit	% of persons 5 and older who can read in some language
53	Per_taga	% of persons 5 and older who speak Filipino/Tagalog
54	Per_eng	% of persons 5 and older who speak English
55	Per_school	% of persons ages 5 to 18 who attended school from June 99-March 2000
56	Per_sch_cit	% of persons ages 5 to 18 who attended school in same city/municipality
57	Per_sch_abr	% of persons ages 5 to 18 who attended school in foreign country
58	Hou_li_ele	% of households that use electricity for lighting
59	Hou_ren	% of houses that are rented
60	Hou_renf1	% of houses that are rent-free with consent of owner
61	Hou_renf2	% of houses that are rent-free without consent of owner
62	Hou_acq_2	% of houses constructed by owner
63	Hou_gar_tru	% of households with pick-up by truck
64	Hou_own_rad	% of households who have radio
65	Hou_own_tv	% of households who have TV
66	Hou_own_ref	% of households who have refrigerator

VARIABLE		DEFINITION
67	Hou_own_vcr	% of households who have VCR
68	Hou_own_tel	% of households who have telephone
69	Hou_own_was	% of households who have washing machine
70	Hou_own_veh	% of households who have motorized vehicle
71	Hou_lan_res	% of households that own other residential lands
72	Hou_lan_ag1	% of households that own agricultural lands
73	Hou_lan_ag2	% of households that own agricultural lands acquired through CARP
74	Hou_lan_oth	% of households that own other agricultural lands
75	Hou_coelpg	% of households that use electricity or lpg for cooking
76	Hou_waduns	% of households that use an unsanitary water source for drinking
77	Hou_notoi	% of households with no toilet
78	Hou_untoi	% of households with unsanitory (open pit) toilet

Unit level, time invariant variables (household – level):

VARIABLE		DEFINITION
1	extended_fam	1 if household is extended
2	hea_ategrad	1 if household head has at least finished grade 6
3	hea_athsgrad	1 if household head has at least finished high school
4	hea_atleasthh	1 if household head has at least finished 4th year highschool
5	hea_atlowed	1 if household head has at least finished grade 5
6	hea_lowed	1 if household head has at least completed pre-school and at most finished grade 5
7	hh_kids	1 if household has at least a member who is son/daughter of the household head
8	single_fam	1 if household does not have "extended family members"

Appendix 3: Municipal Poverty Map of Region III for 2003[8]



8 This figure may be subject to some errors. The estimates that were used to generate this figure are results from preliminary computations and are subject to further scrutiny by the project staff and consultants. Hence, this may not be the exact figure to be published once the Intercensal Updating of Small Area Poverty Statistics project concludes. The purpose of providing this figure in the appendix is to illustrate the possible output(s) that will be generated from the project.