10-2008

How well can we target aid with rapidly collected data? Empirical results for poverty mapping from Cambodia

Tomoki FUJII
Singapore Management University, tfujii@smu.edu.sg

Published in World Development, October 2008, Volume 36, Issue 10, Pages 1830-1842
http://doi.org/10.1016/j.worlddev.2007.10.013

How Well Can We Target Aid with Rapidly Collected Data? Empirical Results for Poverty Mapping from Cambodia

TOMOKI FUJII *
Singapore Management University, Singapore

Summary. — We compare commune-level poverty rankings in Cambodia based on three different methods: small-area estimation, principal component analysis using aggregate data, and interviews with local leaders. While they provide reasonably consistent rankings, the choice of the ranking method matters. In order to assess the potential losses from moving away from census-based poverty mapping, we used the concentration curve. Our calculation shows that about three-quarters of the potential gains from geographic targeting may be lost by using aggregate data. The usefulness of aggregate data in general would depend on the cost of data collection.

Key words — Cambodia, Asia, poverty, small-area estimation, principal component, targeting

1. INTRODUCTION

Poverty reduction is a top priority for international organizations, governments, and nongovernmental organizations. The aid resources available for poverty reduction are, however, severely constrained in many countries. Minimizing the leakage of aid resources to the non-poor is a key to maximize poverty reduction with the limited amount of resources available.

One way to minimize such leakage is to target resources geographically. That is, policymakers can move resources to the poorest parts of the country. Geographic targeting can be quite effective when poverty is unevenly distributed across the country, and this proves to be the case in many countries.

Geographic targeting has two notable advantages. First, it is conceptually straightforward and relatively easy to administer. We only use the information of the location of residence for geographic targeting. Thus, there is virtually no additional cost for data collection to check the eligibility of aid programs. Second, it can be combined with other forms of targeting such as self-targeting. Food-for-work targeted to poor areas is a well-known example of a combination of geographic targeting and self-targeting. Coaby, Grosh, and Hoddinott (2004) found that geographic targeting—as well as proxy-means testing and self-targeting—is associated with an increased share of benefits going to poor people in their study of 122 targeted antipoverty policies in 48 countries.

Designing an effective policy for geographic targeting requires detailed information on the location of the poor. However, such information is hardly readily available. Household socio-economic surveys help us analyze the situation of poverty, but they often fail to provide us with poverty estimates at a spatially aggregated level. Recent development of the poverty mapping methodology has enabled us to estimate poverty indicators at a level of small areas and analyze the local-level relationship between poverty and other geographic factors. Some of the evidence of increased use of poverty maps can be found in Henninger and Snel (2002) and CIESIN (2006).

There are a number of commonly used poverty mapping methodologies. Davis (2003) provides an excellent overview of various poverty mapping methodologies. He identifies six major

* Final revision accepted: October 4, 2007.

methodologies, some of which have several variants, and discusses their strengths and weaknesses. The six methodologies are (i) small-area estimation (SAE), (ii) multivariate weighted basic-need index (MWBI), (iii) combination of qualitative information and secondary data, (iv) extrapolation of participatory approaches, (v) direct measurement of household-survey data, and (vi) direct measurement of census data.

If all the methodologies lead to an identical map, the choice of the methodology would not be an issue. We could simply choose the one that requires the least cost to produce. In practice, it is unlikely that different methodology leads to the same results. Yet, empirical evidence on how much the choice of the methodology matters is scarce.

This paper has two purposes. First, we compare the results obtained by SAE and MWBI, two of the most commonly used poverty methodologies. SAE typically combines a census dataset with a socio-economic survey dataset. It is built on a rigorous statistical basis, and provides both the point estimates and standard errors of poverty indicators for small geographic areas. However, it is computationally very intensive and the data requirement is relatively high. Further, since the maps are made for the year of census, which is typically conducted only once in a decade, the standard SAE methodology does not allow us to produce maps frequently.

An MWBI is constructed from more than one indicators of interest that are related to the human basic needs. Such indicators include the literacy rate and proportion of dwellings without electricity. An MWBI requires weights for each of the basic-needs indicator. The weights can be determined by a statistical technique or given subjectively by the researcher. The MWBI methodology is relatively simple, and the data requirement is not as demanding as the small-area estimation. However, the results are not always easy to interpret because the weights do not have strong theoretical foundations.

We also check the two different results against the commune classification database (CCDB), which contains the subjective ranking of poverty of the communes in the same district. This additional piece of information gives us an idea about which map reflects the situation of poverty more accurately in various parts of the country. It also helps us to explore why there may be differences between the two results.

This approach is somewhat similar to Suharyo *et al.* (2005). In order to verify the quality of the poverty maps for Indonesia in the field, they compared the SAE poverty rankings with the ranking obtained through focus group discussions. They found that the two rankings are consistent at sub-district level.

The second purpose of this paper is to evaluate how much we could fail to capture the efficiency gains if we create a poverty map using rapidly collected data. For this purpose, we take the SAE poverty map as a benchmark case. That is, we assume that the SAE poverty map allows us to target resources optimally at a given level of aggregation. This assumption is necessary for calculating the changes in poverty measures before and after geographic targeting, because only the SAE methodology provides us with poverty measures; the MWBI and CCDB only provide us with the ranking. Under this assumption, we can also evaluate how much potential gain there is from geographic targeting, and how much we lose from moving away from the census-based (SAE) poverty mapping,

This paper is structured as follows: In Section 2, we summarize the SAE poverty mapping in Cambodia. Section 3 discusses the methodology and dataset we used to create an MWBI poverty map in Cambodia. In Section 4, we describe the CCDB. In Section 5, we compare the SAE poverty map and the MWBI poverty map as well as the CCDB in Cambodia. In Section 6, we consider the implications for geographic targeting, and Section 7 concludes.

2. SAE POVERTY MAPPING

The basic idea of the SAE poverty mapping is straightforward. An indicator of interest, such as consumption or income, is regressed on other variables using the survey dataset. Using the regression coefficients, we can impute the indicator to the census dataset. Elbers, Lanjouw, and Lanjouw (2002, 2003) analyze statistical properties of the estimators of poverty and inequality indicators at a level of small geographic areas. They also developed a simulation technique to calculate the point estimates as well as their associated standard errors using a unit-record census dataset and a household socio-economic survey dataset. Their methodology was first applied to Ecuador and has been subsequently applied to a number of countries. Some of the past SAE poverty-mapping exercises are summarized in Henninger and Snel (2002). Besides creating poverty maps, the SAE methodology has a number of applications. It has been applied to the analysis of inequality (Elbers, Lanjouw, Mistiaen, Özler, & Simler, 2003), child malnutrition (Fujii, 2005), and geographic targeting (Elbers, Fujii, Lanjouw, Özler, & Yin, 2007).

Formally, the SAE methodology works as follows: Let y_{ch} be the *per capita* household consumption for household h in cluster c. y_{ch} is related to household-level variables through the following regression model:

$$\log y_{ch} = \mathbf{x}_{ch}^{\mathrm{T}}\beta + \eta_c + \varepsilon_{ch},$$

$\mathbf{x}_{ch}$ includes variables that are common between the census and survey datasets, and location-specific variables that can be linked to both datasets. Since residual terms may be correlated within the cluster, we have a village-level random component η_c and a household-level random component ε_{ch}. We allow for the heteroskedasticity of ε_{ch}. We estimate the empirical distribution of η_c and ε_{ch} using the residuals from an ordinary least-squares regression, and find the point estimate and variance–covariance matrix of $\hat{\beta}$ by a generalized least-squares regression.

We then predict consumption for each census record. We include the model errors associated with parameter estimation, and the idiosyncratic errors that arise from unobserved error terms. That is, in the rth simulation, we randomly draw $\tilde{\beta}^{(r)}$, $\tilde{\varepsilon}_{ch}^{(r)}$, and $\tilde{\eta}_c^{(r)}$ in accordance with the estimated distribution of $\hat{\beta}$, η_c, and ε_{ch}. The predicted consumption is $\tilde{y}_{ch}^{(r)} = \exp(\mathbf{x}_{ch}^{\mathrm{T}}\tilde{\beta}^{(r)} + \tilde{\eta}_c^{(r)} + \tilde{\varepsilon}_{ch}^{(r)})$. The point estimate $\widehat{P}$ and its associated standard error s.e.$(\widehat{P})$ of poverty rate $P(\{y_{ch}\}) \equiv \frac{\#\{y_{ch}|y_{ch}<\zeta\}}{\#\{y_{ch}\}}$ is given by the mean and standard deviation of $P(\{y_{ch}^{(r)}\})$ taken over r, where $\#(\cdot)$ is the counting measure and ζ the poverty line. In general, s.e.$(\widehat{P})$ tends to be smaller when $\widehat{P}$ is produced at a spatially more aggregated level, because the household-level and village-level random components tend to cancel out with each other. Hence, there is a trade-off between the precision of the point estimate and the level of disaggregation.

In Cambodia, Fujii (2006) produces a poverty map using this SAE methodology, and discusses its application to the targeting of education programs. He combines the Cambodia Socio-Economic Survey (CSES) 1997 and the National Cambodia Population Census for 1998. CSES 1997 is a sample survey that contains detailed information on consumption and other indicators such as housing information, demographic composition of the household, and education of each household member. The sample size of the survey is 6,010 and it is representative at the stratum level of CSES 1997, or the level of Phnom Penh, Other Urban and Rural areas. The details of the CSES 1997 dataset are given in NIS (1998).

The National Cambodia Population Census contains information for over 2.1 million households in Cambodia. The information was collected in March 1998 on a *de facto* basis. The census dataset covers virtually all households in Cambodia, except for an estimated population of 45,000 that was not interviewed due to the military operations. It contains a number of variables that are common with the survey dataset. The details of this dataset are given in NIS (2000). Fujii (2006) also used a compilation of geographic variables.

The SAE estimates of poverty indicators are produced at the level of communes, where a commune is the second smallest administrative unit in Cambodia after village and before district and province. Each commune contains on average about 1,300 households, and the average standard errors at the commune level are 7.4%. At the stratum level, the SAE estimates were not significantly different from the survey estimates, suggesting that the SAE estimates are consistent with what is observed in the survey. The poverty map based on the SAE estimates is included in the appendix of the National Poverty Reduction Strategy (Council for Social Development, 2002) and has been used by a variety of stakeholders for targeting aid resources. [1]

3. MWBI POVERTY MAPPING

MWBI is calculated by weighting multiple basic needs indicators at the community level. Davis (2003) identifies three weighting schemes based on statistical techniques—principal component analysis, factor analysis, and ordinary least squares. Principal component analysis reduces the dimensionality of a dataset by finding linear combinations that best explain the variations of variables in a dataset. Factor analysis is similar to principal component analysis, but critical assumptions are different. In factor analysis, we try to decompose the total variance in data into common factor variance and unique factor variance. Davis (2003) argues that, while factor analysis is more elaborate, the method is subjective because we need to interpret the factors to give them meaning, which relies on previous knowledge and intuition about underlying relationships. In ordinary least-squares approach, we use the coefficients derived from a regression analysis of the determinants of poverty as weights to create an MWBI.

In this study, we create an MWBI map based on the first principal component. This is because the procedure of principal component analysis is straightforward and involves less subjective judgments. This does not, however, necessarily ensure that weighting scheme obtained from the principal component analysis is "better" than other statistical weighting schemes or even subjectively determined weighting schemes. One could argue, for example, the assumptions for factor analysis seem more appropriate than those for principal component analysis, because lack of different sorts of basic needs can be explained by a common factor, "poverty." Also, the weights calculated by the principal component analysis in one dataset are not readily applicable to other datasets. Hence, the weights for this year may not be appropriate for the dataset collected next year. On the other hand, subjectively predetermined weights could at least warrant fair inter-temporal comparisons.

Let us suppose that there are L community-level variables for N communities. Let z_{nl} be normalized observations for community $n(=1, 2, \ldots, N)$ and variable $l(=1, 2, \ldots, L)$, where each variable has a mean zero and unit standard deviation. We define $\mathbf{Z} \equiv (z_{nl})$. Let us consider the linear combination of the variables $\mathbf{v} = \mathbf{Z}\mathbf{w}$, where the weights are $\mathbf{w} = (w_1, w_2, \ldots, w_L)^{\mathrm{T}}$ with $\|\mathbf{w}\| = 1$. We can find the first principal component by calculating such $\mathbf{w}$ that maximizes the variance of $\mathbf{v}$.

$$\max_{\mathbf{w}}(\mathbf{v}^{\mathrm{T}}\mathbf{v}) \quad \text{s.t.} \quad \|\mathbf{w}\| = 1.$$

Letting the Lagrange multiplier be λ and taking first order conditions, we get $(\mathbf{Z}^{\mathrm{T}}\mathbf{Z} - \lambda\mathbf{I})\mathbf{w} = \mathbf{O}$. Hence, a straightforward calculation shows $\mathbf{v}^{\mathrm{T}}\mathbf{v} = \lambda$. The variance of $\mathbf{v}$ is maximized when $\mathbf{w}$ is an eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$. Our MWBI is the principal component score $S = \frac{\mathbf{v}}{\sqrt{\lambda}}$, which has a mean zero and unit standard deviation. This score gives us the ranking of the welfare levels of communities. However, unlike the standard measures provided by the SAE, there is no obvious interpretation of the principal component score in general.

We use the Seila Commune Database (CDB) to create an MWBI map. The Seila CDB is a comprehensive database that includes basic socio-economic information at the village level. The Seila CDB is managed by the Provincial Departments of Planning (PDoP) under the technical supervision of the Ministry of Planning, and the information is collected firsthand by village leaders. A set of guidelines were developed in consultation with the PDoP to standardize data collection practices and provide a quality control check. Data collection was first conducted in five provinces in 1998, and annual updates have been carried out since then. The Seila CDB is supposed to be flexible and specific provincial questions could be added per province-specific situations or needs. The list of collected variables is also reviewed regularly. Hence, the geographic coverage and the variables included in the Seila CDB vary from year to year.

We used the fifth round of the Seila CDB, which we call the Seila CDB5. The Seila CDB5 was collected during November 2002–January 2003. Unlike the previous rounds, the Seila CDB5 covers all the provinces in Cambodia. Moreover, following the review of CDB by UNOPS and UNDP (2002), the quality control of Seila CDB5 is supposed to be significantly better than previous rounds. The Seila CDB5 includes demographic characteristics, housing characteristics, asset holdings, education information, and rice production of the village. Some variables are collected at the commune level.

Despite the efforts for the quality control, rapidly collected data like the Seila CDB5 have limitations. Unlike the socio-economic surveys and census, for example, the village leaders are not given sufficient training to collect data. Furthermore, the questionnaire covers a wide range of variables, some of which the village leaders may not be familiar with. This does not necessarily mean that the Seila CDB5 is useless, but that we need to use it with caution.

Since it is only at the commune level that the SAE poverty estimates are created and that some Seila CDB5 indicators are available, we aggregated village-level indicators to the commune level in order to create the MWBI poverty map. Different questionnaires were used in the urban and rural areas. This study focuses on rural areas, because the design of the urban questionnaire makes analysis difficult, and because the majority of poor people live in rural areas. [2]

Table 1. *List of variables, their summary statistics and their scoring coefficients for MWBI and MWBI-ADJ*

(1) Variable	MWBI (2) Mean	MWBI (3) Standard deviation	MWBI (4) First eigenvector	MWBI-ADJ (5) Mean	MWBI-ADJ (6) Standard deviation	MWBI-ADJ (7) First eigenvector
Dependency ratio	0.425	0.043	0.399	0.480	0.027	0.322
% HHs with thatched roof and no latrine	0.470	0.211	0.481	0.552	0.200	0.511
Literacy ratio	0.747	0.183	−0.485	0.611	0.166	−0.508
Motorcycles per capita	0.033	0.022	−0.399			−0.439
% HHs with access to water at home	0.401	0.310	−0.239			−0.219
Distance to closest market in minutes	57.2	84.8	0.397			0.370

Unit of analysis is commune, and each commune has a unit weight. $N = 1473$ for MWBI and $N = 1447$ for MWBI-ADJ.

We picked six indicators that cover different aspects of living standards, which are listed in Table 1. While the choice of indicators is arbitrary, we included important dimensions of welfare such as education, asset, market access, housing conditions, and demographic characteristics. Moreover, the choice of indicators does not seem to heavily influence the results of our analysis so long as these dimensions of welfare are covered.

The means and standard deviations of each indicator are reported in Columns (2) and (3) in Table 1. Here, we are interested in the commune-level ranking of poverty and thus the unit of analysis is a commune. Therefore, we used a unit weight for each commune.

The largest λ was 2.69, whereas the rest of eigenvalues are less than one. Thus, it indeed seems appropriate to use only the first principal component in this study. The first principal component explains about 45% of the total variation. The MWBI scoring coefficient, or the eigenvector associated with the largest eigenvalue, is given in Column (4) of Table 1. The absolute value of the scoring coefficient for each indicator is moderately large, suggesting that every indicator included in this analysis is indeed relevant. Positive coefficients are attached to the indicators for which a higher value suggests lower levels of welfare, and negative coefficients are attached to the indicators that have an opposite meaning. Therefore, higher values of MWBI represent lower levels of living standards.

4. COMMUNE CLASSIFICATION DATABASE

We also used the CCDB in this study. District chiefs are asked to subjectively rank the communes in their district from the poorest. Hence, each commune is given a number from one to the number of communes in the district. We reversed the ranking for our analysis so that the smallest numbers signify the worst poverty situations in the district. The information in the CCDB was collected by PDoP during the same period as the Seila CDB5. The CCDB is the first attempt to collect district chiefs' opinions on poverty systematically at the national level.

The reason we used the CCDB is because the comparison between the SAE poverty map and MWBI poverty map gives us no idea as to why the two maps are different. While CCDB does not provide us with an answer, it helps us to determine which map is more likely to reflect the situation accurately. If the CCDB ranking and the SAE or MWBI ranking of poverty are not positively correlated, the SAE or MWBI ranking is not consistent with the perception of village leaders. We can then investigate the plausible causes of counter-intuitive rankings. Having explanations for counter-intuitive rankings is essential when we use poverty maps for policy-making.

There are three important points we need to make about the use of CCDB. First, different people look at poverty differently. Hence, the subjective ranking for the same district would differ from person to person. We may observe counter-intuitive rankings in some districts simply because the CCDB ranking involves subjective judgments. Second, district chiefs may have given us false opinions in order to prioritize some communes over others for, say, political reasons, even though we do not have the evidence. We believe that, if anything, this would not affect our results systematically. Finally, the CCDB allows us to compare the communes only if they are in the same district. Therefore, unlike the SAE or MWBI poverty maps, we cannot rank all the communes in Cambodia only with the CCDB.

5. COMPARISONS

We first looked at the correlations between the SAE poverty indicators and MWBI at the ecozone level. An ecozone is a group of several provinces, [3] and the SAE poverty indicators include poverty rate (P_0), poverty gap (P_1), and poverty severity (P_2). Because the SAE poverty indicators and MWBI obviously measure different things, we would expect some differences between them.

Some of the important sources of differences are worth noting. First, the SAE and MWBI are based on different sets of indicators in this study. While we restricted the number of indicators used for MWBI to allow for clear interpretation, SAE estimates are based on a rich set of variables. For example, MWBI uses only literacy ratio to reflect education level, SAE uses more detailed description of education. Second, the SAE reflects the poverty situation in the census year of 1998, whereas MWBI reflects the poverty situation in 2003 when the Seila CDB5 was collected.

Further, even if both SAE and MWBI are produced at the same point in time from the same set of indicators, they do not necessarily result in the same rankings for at least three reasons. First, the weights used for SAE and MWBI are in general different so that the weights implicit in MWBI do not necessarily reflect consumption poverty. Second, even if the weights are identical, MWBI can only capture the average consumption of the commune. Thus, the poverty measures, which reflect the bottom tail of distribution, may still be different even if two communes have identical commune-level averages. Third, SAE takes into account idiosyncratic errors at the household and cluster levels by Monte-Carlo simulation. This in turn means that the SAE ranking is subject to statistical errors.

Despite these sources of differences, we expect some positive correlations because higher values of the SAE poverty indicators and MWBI indicate a lower level of welfare in both cases. Indeed, we observe moderately positive correlations in Plain, Tonle Sap, and Coastal ecozones as shown in the upper half of Table 2. However, this is not the case for the Plateau/Mountain ecozone. This point may be more clearly seen by the scatter plots of rankings by P_0 and MWBI given in Figure 1. For example, very few communes are in the top-left or bottom-right corners in Figure 1a–c. However, many dots appear in these corners in Figure 1d, showing the ranking reversal in this ecozone. These observations warrant closer examination of the Plateau/Mountain ecozone.

CCDB helps us identify what is possibly wrong. We compared the rankings based on the SAE poverty rate, the MWBI, and CCDB data. Table 3 shows the proportion of districts for which the sign of Spearman rank correlation coefficient for the welfare rankings of two different indicators is positive. To ensure that each district has a positive or negative sign, we gave a positive sign with probability of one half for a very small number of districts where the Spearman coefficient is exactly zero.

Table 2. *Correlation of MWBI and SAE poverty indicators at the commune and district levels by ecozone*

Ecozone	MWBI			MWBI-ADJ			Obs
	P_0	P_1	P_2	P_0	P_1	P_2	
Commune-level comparison							
Plain	0.34	0.32	0.31	0.36	0.34	0.32	591
Tonle Sap	0.34	0.31	0.29	0.43	0.37	0.34	426
Coastal	0.32	0.33	0.34	0.42	0.38	0.35	142
Plateau/Mountain	−0.35	−0.35	−0.35	−0.23	−0.23	−0.23	288
District-level comparison							
Plain	0.19	0.24	0.27	0.31	0.33	0.34	55
Tonle Sap	0.36	0.30	0.26	0.48	0.42	0.38	53
Coastal	0.17	0.30	0.46	0.55	0.49	0.46	21
Plateau/Mountain	−0.39	−0.43	−0.48	−0.30	−0.36	−0.42	41

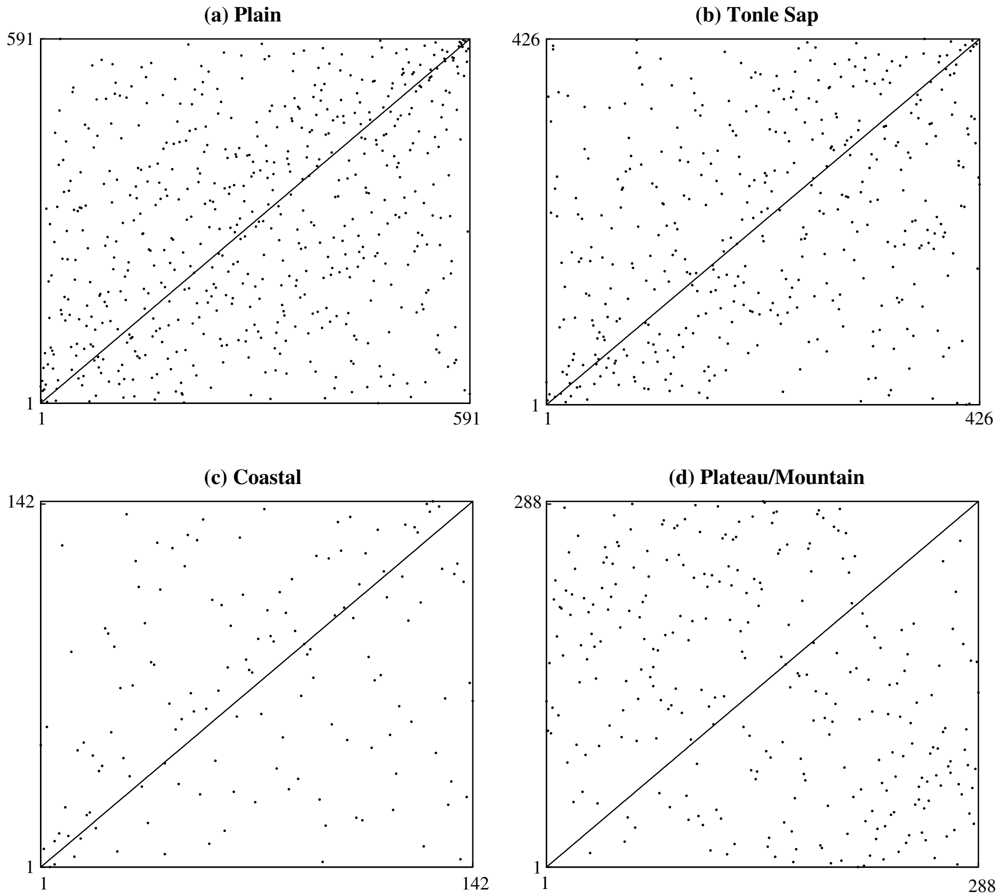

Figure 1. *Comparison of poverty rankings by ecozones. Horizontal axis measures the ranking by P_0 and vertical axis by MWBI. Each dot corresponds to a commune.*

Under the null hypothesis that the two rankings are uncorrelated, each district gets a positive sign with a probability of one half. The asterisk after the number indicates that the probability of the event that the proportion of district with a positive sign is at least the observed value is less than 0.01 under the null hypothesis. Because each cell has a number greater than one half and most cells have a significant value, the rankings from three different sources of information seem consistent with each other overall.

Comparison of the proportion of positive correlations in Table 3 suggests that the MWBI and CCDB rankings are the closest pair. There are two possible explanations for this. First, both CCDB and MWBI are based on the observations in 2003, whereas the SAE estimates are referenced to the census year of 1998. Hence, SAE estimates do not reflect the welfare changes that took place during 1998–2003. Second, it is likely that the district chiefs' rankings given in CCDB are based on the easily observable indicators. Because consumption is not as easy to observe as the indicators used to calculate the MWBI, it is not surprising that the CCDB ranking is closer to the MWBI ranking than to the ranking of SAE poverty rate.

In order to account for the changes during 1998–2003, we have also constructed an adjusted version of MWBI (MWBI-ADJ) by taking dependency ratio, percentage of households with thatched roof and no latrine, and literacy ratio from the census instead of the CDB5 (other three variables do not exist in the census). Columns (5) and (6) in Table 1 provide

Table 3. *Proportion of districts where the sign of Spearman rank correlation coefficient for the welfare rankings based on two indicators is positive*

Ecozone	P_0 MWBI	P_0 CCDB	MWBI CCDB	P_0 MWBI-ADJ	MWBI-ADJ CCDB	Obs
Plain	0.79*	0.74*	0.85*	0.85*	0.91*	53
Tonle Sap	0.72*	0.70*	0.91*	0.74*	0.94*	54
Coastal	0.83*	0.61	0.83*	0.89*	0.94*	18
Plateau/Mountain	0.56	0.54	0.93*	0.49	0.90*	41

* The percentage is different from one half at the 1% significance level.

their summary statistics. The first eigenvector for MWBI-ADJ is given in Column (7) and its associated eigenvalue is 2.51, which is the only eigenvalue greater than one. By comparing Columns (2) and (5), we see that some improvements were made during 1998–2003. Further, Table 2 shows that the correlation between P_0 and MWBI-ADJ is higher than that for P_0 and MWBI. A similar comparison for Table 3 yields the same conclusion, except for Plateau/Mountain ecozone.

The correlations between P_0 and other indicators in the Plateau/Mountain ecozone are not significant. This suggests that the poverty estimates in this ecozone may be problematic, and it is consistent with the comments we received from various government officials and local experts about the SAE estimates (Fujii, 2007). While most of them thought the map looks reasonable, many of them thought that the SAE estimates for the four northeastern provinces of Kracheh, Mondol Kiri, Rotanak Kiri, and Stueng Treng appear to understate poverty. We looked at the correlation between SAE poverty measures and MWBI for each province in the Plateau/Mountain ecozone. While the four northeastern provinces had a negative correlation, the rest of the provinces in the Plateau/Mountain ecozone had a positive correlation.

It is worth pointing out that the SAE estimates are consistent with what is observed in the CSES 1997 in the northeastern provinces. We calculated poverty rates at the provincial level. The poverty rates for Kracheh, Rotanak Kiri, and Stueng Treng are 28.1%, 0.0%, and 12.1%, respectively, much lower than the national average of 36.1%. While CSES 1997 is not representative at the provincial level and there is no observation in Modol Kiri, the apparent discrepancy between the SAE ranking and MWBI ranking seems to come from the nature of the CSES 1997 dataset. CSES 1997 is representative at the level of Phnom Penh, Other Urban and Rural strata, and thus only one price system is assumed within each stratum. However, some of the essential goods in the rural areas of the northeastern provinces are much more expensive than other rural areas. This is in part because of the inaccessibility to these provinces. [4]

Higher prices in the northeastern provinces may have lead to the apparent underestimate of poverty in the northeast because the nominal consumption is deflated by a price index that is lower than the true price index that people in the northeastern provinces face. This in turn may have caused downward bias for the SAE poverty estimates. Much of the negative correlations found in the Plateau/Mountain ecozone may be indeed driven by the variations in prices.

One reality check for this argument is to look at different levels of aggregation. Since the SAE estimates are subject to smaller standard errors at more aggregated levels, the correlation tends to be larger in absolute value when we aggregate the data. Hence, we have also constructed MWBI and MWBI-ADJ at the district level by first aggregating the commune-level indicators and carrying out the principal component analysis at the district level. The lower half of Table 2 reports the correlations between poverty indicators and MWBI indicators at the district level.

As we can see from Table 2, when we look at the correlation between the poverty measures and MWBI-ADJ (which we prefer to MWBI here because it allows us to eliminate the "time effect" at least partially), the district-level correlations are indeed higher than the commune-level correlations in absolute value for all the ecozones except for Plain ecozone. In Plain ecozone, a higher correlation is observed for P_2. Further, the commune-level and district-level correlations of MWBI-ADJ with P_0 and P_1 are not very different. Overall, our explanations of the possible causes of the discrepancy between

MWBI and SAE rankings are consistent with our observations.

6. IMPLICATIONS FOR GEOGRAPHIC TARGETING

We have shown that the SAE, MWBI, and CCDB rankings give us a reasonably consistent picture with an exception of northeastern provinces. Still, the choice of the map matters because the extent to which the potential gains from targeting are captured depend upon the indicators we use for targeting. In this section, we evaluate the efficiency gains when we use an alternative poverty map. For this purpose, we would like to know the true distribution of poverty in Cambodia. However, because we do not know the true distribution, we use the SAE estimate of poverty $\widehat{P}$ instead of P.

To measure the extent to which the potential gains from targeting are captured when we use MWBI instead of SAE estimates, we use the concentration curve of poverty. The concentration curve is a generalization of the Lorenz curve. To draw a concentration curve, instead of the ranking of individual incomes, we use a ranking by the group to which individuals belong. Formally, the concentration curve is defined as follows. Let $\Gamma \equiv \{1, 2, \ldots, G\}$ be the index set for groups such as communes. Each individual belongs to exactly one group. Let us denote the population and poverty rate for group $g \in \Gamma$ by N_g and P_g, and define $a_g \equiv \frac{\sum_{c=1}^{g} N_c}{\sum_{c=1}^{G} N_c}$ and $b_g \equiv \frac{\sum_{c=1}^{g} N_c P_c}{\sum_{c=1}^{G} N_c P_c}$. Then, a_g and b_g are the cumulative share of the population and the cumulative share of poor people contained in groups 1 through g. The concentration curve $C(s)$ is a piecewise linear function of the share of population s and its graph connects $(0, 0)$ and (a_g, b_g) for $\forall g \in \Gamma$. In a special case where there is only one group (i.e., "Cambodia"), the concentration curve is just the 45-degree line.

A sample concentration curve is given in Figure 2. The horizontal and vertical axes represent the cumulative share of the population and poor people, respectively, and the lengths of OA and OC are unity. At the individual level, "poverty rate" can only take zero or one. Hence, if we sort by the individual-level poverty rate, poor people come first. The individual-level concentration curve looks like the bold line OIB, and OE represents the poverty rate at the national level. Now, suppose there are three zones in the country. The zonal-level concentration curve would look like the kinked line $OLMB$. Note that the slope of the concentration curve represents the poverty rate of the group relative to the national poverty rate. For example, $\overline{OF}$ is the share of population in the poorest commune, and $\overline{OK}$ is the ratio of the poor people in the poorest commune to the total number of poor people. Thus, the poverty rate in the poorest commune is $\overline{OE} \cdot \frac{\overline{OK}}{\overline{OF}}$.

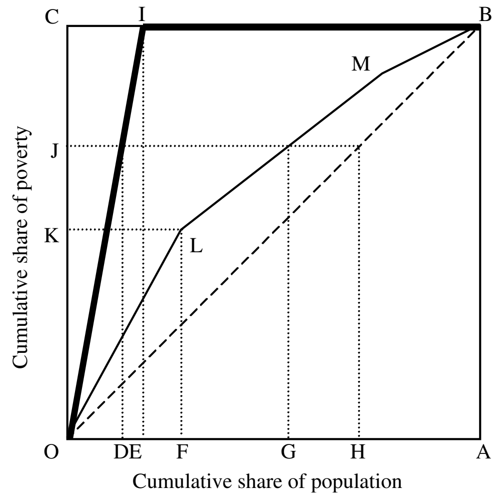

Figure 2. *Sample concentration curve.*

With a few additional assumptions, we can interpret the concentration curve in the context of targeting. First, we assume that the cost for bringing a poor person out of poverty is fixed. While this is a strong assumption, it is not entirely unrealistic for certain development programs. For example, the cost for direct food aid may be approximated by fixed cost *per capita*, because the amount of food delivered to people is in practice often fixed. Second, we assume that people do not migrate in order to capture the benefits of programs in their neighborhood. Because the cost of changing the location of residence is generally high especially for poor people, this is not usually a problematic assumption. Yet, for certain programs such as public construction works, this may well be a problem. This issue is explored in Fujii and Jack (2005). Third, everyone in the same group is treated equally. That is, if a person in a group receives aid, everyone else in the group receives aid. If the resources are not enough to cover everyone in the group, everyone receives the aid with an equal probability. We can take the 45-degree line as a reference case of no information, because everyone in the country receives the aid with the same probability.

Under these assumptions, the share of the population covered in the program is proportionate to the cost of the program as the *per capita* cost of the program is fixed. Hence, the horizontal axis can be interpreted as the cost of the program standardized so that the cost of eliminating poverty is unity. Here, the assumption of fixed *per capita* cost is convenient because we do not have to worry about how the MWBI corresponds to the poverty gap. The vertical axis measures how much poverty can be reduced given the total cost of the program. Alternatively, it can be interpreted as the goal of poverty reduction. In this case, the horizontal axis measures the cost for achieving that goal.

In the reference case of no information, we need to cover the proportion s of the population in order to reduce the poverty rate by the same proportion s. Given the goal of poverty reduction, we can reduce the cost of program by spatial targeting. We define budgetary gains as the amount of the cost reduced by targeting. On the graph, this corresponds to the horizontal distance between the concentration curve for a particular targeting scheme and the 45-degree line. For example, when we want to reduce poverty by $\overline{OJ}$ in Figure 2, the cost is $\overline{OH}(=\overline{OJ})$ without any information. If resources are targeted at the individual level, the cost is $\overline{OD}$. Hence, the budgetary gains from individual-level targeting are $\overline{DH}(=\overline{OH}-\overline{OD})$. Likewise, the budgetary gains from commune-level targeting are $\overline{GH}(=\overline{OH}-\overline{OG})$. Formally, the budgetary gains can be written as $B(g)\equiv g - C^{-1}(g)$, where g is the goal of poverty reduction.

It would be convenient to have an overall measure of budgetary gains. We define the average budgetary gain as $2\cdot\int_0^1 B(g)\mathrm{d}g$. Hence, the average budgetary gain is zero for the national-level targeting. The average budgetary gain for the commune-level targeting corresponds to the ratio of the area OLMB to the area OBC.

Now suppose that the groups are communes sorted in the descending order of poverty rates so that we have $P_i \geqslant P_j$ for $i \leqslant j$. We call this concentration curve the SAE concentration curve $C^{SAE}(s)$. By construction, $C^{SAE}(s)$ is concave. We can also draw the MWBI concentration curve $C^{MWBI}(s)$ by sorting the communes in descending order of the MWBI so that we have $S_i \geqslant S_j$ for $i \leqslant j$. This may not be concave. Because we always take the poorest group first for $C^{SAE}(s)$, we have $C^{SAE}(s) \geqslant C^{MWBI}(s)$ for $\forall s \in [0,1]$ by construction.

Figure 3 shows the SAE and MWBI concentration curves in Cambodia. There are two points to be noted here. First, $C^{MWBI}(s) > s$ is satisfied for most values of s except for very small values of s. This means that targeting based on the MWBI is still better than no targeting, provided that the amount of resources for targeting is reasonably large. Second, the

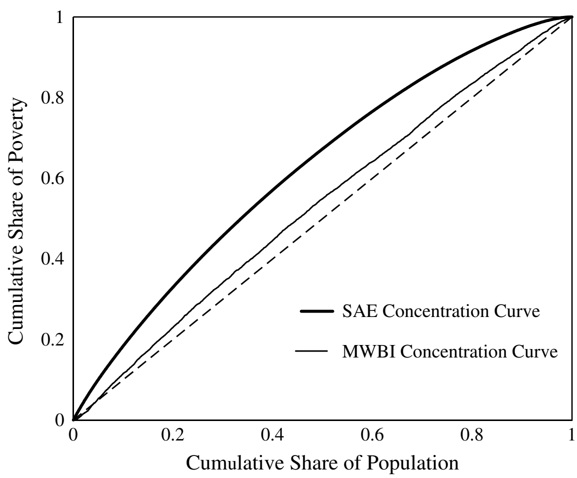

Figure 3. *Poverty concentration curves in Cambodia.*

budgetary gains vary substantially by the goal of poverty reduction. As Figure 4 shows, the budgetary gains for both SAE and MWBI are higher when the goal is moderate and they are small when the goal is very high or very low; $B(g)$ takes a maximum value when it is around 0.5 for MWBI targeting and 0.6 for SAE targeting.

This should make sense because, when the goal is low, there is little to gain in absolute terms. However, the gain in relative terms may be large. That is, the ratio of the amount of budget needed to achieve a given level of poverty reduction with some poverty estimates may be smaller than that without any information. This contrasts with the case when the goal is high. The gains are low in both relative and absolute terms when the goal of poverty reduction is very high. Unless you can perfectly distinguish the poor from the non-poor, one would have to virtually give everyone enough resources to get out of poverty. Thus, the commune-level information is not very useful to bring the last poor person in the country out of poverty.

One interesting point to note is that the ratio of budgetary gains $\left(\frac{B^{MWBI}(g)}{B^{SAE}(g)}\right)$ is stable over a range of values of g as Figure 5 shows. Hence, in our empirical results, we can expect reasonably well how much efficiency is lost by using the MWBI poverty map regardless of the goal of poverty reduction.

The average budgetary gains for SAE and MWBI concentration curves are 0.247 and 0.064, and the ratio of these two is 0.26. We also conducted this analysis excluding the Plateau/Mountain ecozone because this area was deemed problematic. In this case, the ratio of average budgetary gains is 0.37, a slightly

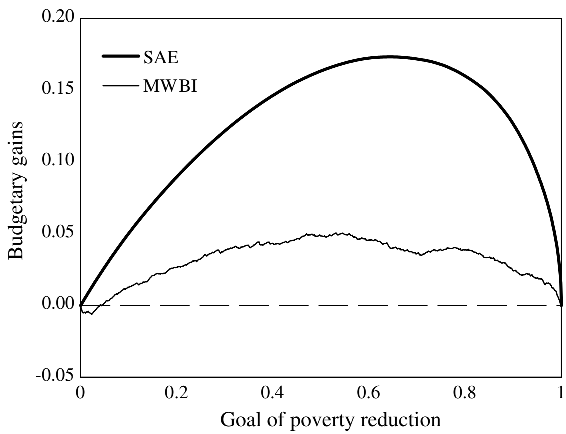

Figure 4. *Budgetary gains for SAE and MWBI concentration curves.*

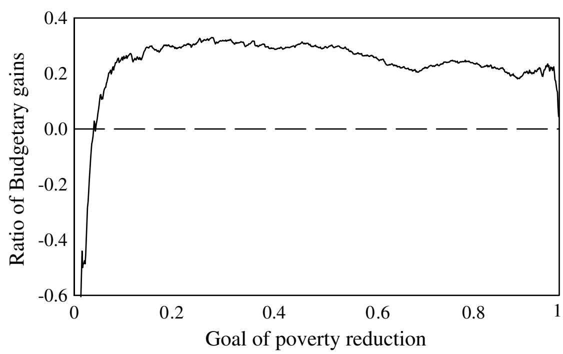

Figure 5. *Ratio of budgetary gains.*

higher value. This reflects the high positive correlations between the SAE and MWBI rankings outside the Plateau/Mountain ecozone.

It is not immediately clear whether 0.37 is high or low. If SAE and MWBI do not produce a consistent ranking, this number could have been close to zero or even negative. Given that 0.37 seems high. On the other hand, it is far below one, which is what we would get if SAE and MWBI give rise to an identical ranking. Thus, while MWBI targeting captures a sizable portion of the potential gains from commune-level targeting, there will remain a large amount of potential gains that MWBI targeting fails to capture.

7. CONCLUSIONS

This study provided empirical results of different methods of poverty mapping. We constructed commune-level poverty indicators based on the principal component scores and compared them with SAE poverty estimates using Cambodian data. While the SAE and MWBI estimates correspond reasonably well, there were notable differences. We explored the differences by ecozones with the help of the tertiary dataset (CCDB), and pointed out several plausible causes of discrepancies, including the changes that have taken place during 1998–2003 and the issues of price index in CSES 1997. These causes of discrepancies are not inherent weaknesses of SAE. Yet, the SAE estimates do not necessarily reflect the *current* poverty situation accurately because of the time-lag between the census year and the production of the estimates as well as of any problems inherent in the consumption survey. Thus, ground truthing (field verification) is particularly important when we apply SAE estimates to policy making.

We also studied the implication of using the MWBI estimates instead of the SAE estimates for geographic targeting. We showed a way to employ the concentration curve for the analysis of the geographic targeting. MWBI is likely to capture one-third to one-fourth of the potential gains from commune-level targeting, depending on the goal of targeting. Since such a ratio has never been known, it helps us to form an idea of the cost of resorting to rapid data collection.

While our calculations are based on a number of assumptions and may not hold for other countries or at other points in time, it provides a useful reference point. It is essential for policymakers to have some ideas about how much one can gain from additional information. For example, if no census data were available and census data were to be used only for the purpose of poverty mapping, it would certainly make sense to use MWBI instead of SAE estimates because collection of census data is costly. Indeed, collection of census data can easily cost, say, hundred times more than community-level data such as CDB5. On the other hand, even though the efforts and resources required to produce SAE estimates are generally higher, SAE has advantages over MWBI in terms of interpretability and methodological rigorousness.

When the goal of poverty reduction is low, the differences in the budgetary gains from targeting based on SAE and MWBI estimates are small. Hence, when additional data collection is required, it would make sense to go for MWBI estimates using a rapidly collected dataset. In practice, this means that MWBI is most appropriate to achieve a short-term goal of poverty reduction. MWBI would be suited for monitoring changes in poverty.

When the budget is sufficiently large, on the other hand, the budgetary gains may be large enough to justify SAE estimates with some additional data collection. If the consumption surveys have a panel component, we can update poverty maps without a new census. Otherwise, it means that collection of census or a large-sample data may be justified once in every few years (instead of, say, a decade) to enable more accurate targeting. Rapidly collected datasets may not possess the most ideal quality, but they can offer invaluable timely information, and complement well with household surveys and census.

NOTES

1. Fujii (2007) discusses how the SAE poverty map has been used in Cambodia.

2. The definitions of rural areas for National Cambodia Population Census and CDB5 are slightly

different. This study employs the definition of CDB5, but we excluded 13 communes in Phnom Penh that are classified as a rural commune in CDB5. This is because we do not have rural communes in Phnom Penh in the CSES 1997.

3. The definitions of the ecozones are as follows. Plain: Kampong Cham, Kandal, Prey Veng, Svay Rieng, and Takeo. Tonle Sap: Banteay Mean Chey, Battambang, Kampong Chhnang, Kampong Thom, Pursat, Siem Reap, and Pailin. Coastal: Kampot, Koh Kong, Krong Preah Sihanouk, and Krong Keb. Plateau/Mountain: Kampong Speu, Kracheh, Mondol Kiri, Preah Vihear, Rotanak Kiri, Stueng Treng, and Otdar Mean Chey.

4. We are unable to construct a reliable price index for the northeastern provinces because there are so many items for which price data are missing in these provinces.

REFERENCES

CIESIN (2006). *Where the poor are: An atlas of poverty*. Center for International Earth Science Information Network (CIESIN), Columbia University.

Coaby, D., Grosh, M., & Hoddinott, J. (2004). Targeting outcomes redux. *World Bank Research Observers, 19*(1), 61–85.

Council for Social Development (2002). *National poverty reduction strategy 2003–2005.* Council for Social Development, Royal Government of Cambodia.

Davis, B. (2003). Choosing a method for poverty mapping. Food and Agriculture Organization.

Elbers, C., Fujii, T., Lanjouw, P., Özler, B., & Yin, W. (2007). Poverty alleviation through geographic targeting. *Journal of Development Economics, 88*(1), 198–213.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2002). Welfare in villages and towns: Micro-level estimation of poverty and inequality. Policy research working paper 2911, The World Bank.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica, 71*(1), 355–364.

Elbers, C., Lanjouw, P., Mistiaen, J., Özler, B., & Simler, K. (2003). Are neighbours equal? Estimating local inequality in three developing countries. FCND discussion paper no. 147, Food Consumption and Nutrition Division, International Food Policy Research Institute.

Fujii, T. (2005). Micro-level estimation of child malnutrition indicators and its application in Cambodia. Policy research working paper 3662, The World Bank.

Fujii, T. (2006). Commune-level estimation of poverty measure and its application in Cambodia. In R. Kanbur, A. J. Venables, & G. Wan (Eds.), *Spatial disparities in human development: Perspectives from Asia* (pp. 289–314). United Nations University Press.

Fujii, T. (2007). To use or not to use? Poverty Mapping in Cambodia. In T. Bedi, A. Coudouel, & K. Simler (Eds.), *More than a pretty picture: Using poverty maps to design better policies and interventions* (pp. 125–142). The World Bank.

Fujii, T., & Jack W. (2005). Optimal geographic targeting. Mimeo, Georgetown University.

Henninger, N., & Snel, M. (2002). Where are the poor? Experiences with the development and use of poverty maps. World Resources Institute.

NIS (1998). *Cambodia socio-economic survey 1997: Technical report on survey design and implementation.* National Institute of Statistics (NIS), Ministry of Planning, Royal Government of Cambodia.

NIS (2000). *General Population Census of Cambodia 1998: Village Gazetteer.* National Institute of Statistics, Ministry of Planning, Royal Government of Cambodia.

Suharyo, W., Akhmadi Hastuti Filaili, R., Budiati, S., & Munawar, W. (2005). *Developing a poverty map for Indonesia: A tool for better targeting in poverty reduction and social protection programs. Book 4: Field verification.* SMERU Research Institute.

UNOPS & UNDP (2002). Final report: Resourcing secondary data and enhancing the commune database. rural investment and local governance. CMB/01/R72/RILG. United Nations Office for Project Services (UNOPS) and United Nations Development Programme (UNDP).