2-2006

How Well Can We Target Resources with Quick-and-Dirty Data?: Empirical Results from Cambodia

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

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

How Well Can We Target Resources with "Quick-and-Dirty" Data?: Empirical Results from Cambodia

Tomoki Fujii

February 2006

Paper No. 03-2006

How Well Can We Target Resources with "Quick-and-Dirty" Data?: Empirical Results from Cambodia

Tomoki FUJII*

January 27, 2006

1 Introduction

Poverty reduction is a top priority for international organizations, governments and non-governmental 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.

*email: tfujii@smu.edu.sg.

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.

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. It has also been used to analyze the relationship between poverty and other geographic factors.

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 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.

An MWBI is constructed from more than one indicators of interest that are related to the human basic needs. Such an indicator would 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 very easy to interpret because the weights are bound to be arbitrary.

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 seems to reflect the situation of poverty more accurately in various parts of the country. It also helps to explore why there may be differences between the two results.

The second purpose of this paper is to evaluate how much we could fail to capture the efficiency gains if we use a "quick-and-dirty" poverty map. 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. However, because the poorest area in the MWBI poverty map may not be the poorest in the SAE poverty map, we may not gain much from geographic targeting if MWBI is very different the SAE poverty map. We evaluate to what extend the MWBI would capture the potential gains from geographic targeting.

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 data set. Elbers et al. (2002, 2003a) analyze statistical properties of 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. They first applied the SAE to Ecuador. Their methodology has been 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 et al., 2003b), child malnutrition (Fujii, 2005) and crime (Demombynes and Özler, 2002). It is also applied to the geographic targeting (Elbers et al., 2004)

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:

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

$\mathbf{x}_{ch}$ 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 and a household-level random component η_c. We allow for the heteroskedasiticty of ϵ_{ch}. We estimate the empirical distribution of η_c and ϵ_{ch} using the residual 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 r-th simulation, we random draw $\tilde{\beta}^{(r)}$, $\tilde{\epsilon}_{ch}^{(r)}$ and $e\tilde{t}a_c^{(r)}$ in accordance with the estimated distribution of $\hat{\beta}$, η_c and ϵ_{ch}. The predicted consumption is $\tilde{y}_{ch}^{(r)} = \exp(\mathbf{x}_{ch}^T \tilde{\beta}^{(r)} + \tilde{\eta}_c^{(r)} + \tilde{\epsilon}_{ch}^{(r)})$. The point estimate $\hat{P}$ and its associated standard error $s.e.(\hat{P})$ of poverty rate $P(\{y_{ch}\}) \equiv \frac{\#\{y_{ch} | y_{ch} < \zeta\}}{\#\{y_{ch}\}}$ is given by

$$\hat{P} = \frac{1}{R} \sum_{r=1}^{R} P(\{y_{ch}\}), \quad s.e.(\hat{P}) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} [P(\{y_{ch}\}) - \hat{P}]^2}$$

where $\#(\cdot)$ is the counting measure, ζ the poverty line and R the number of simulations. In general, $s.e.(\hat{P})$ tends to be smaller when $\hat{P}$ is produced at a spatially more aggregated level, because the idiosyncratic errors 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 (2004) 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. 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 6010 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 (1998). Fujii (2004) 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 1300 households, and the average standard errors at the commune-level are 7.4 percent. 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.

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 data set. 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. Arguably, the assumptions for factor analysis may be more appropriate than those for principal component analysis, because lack of different sorts of basic-needs can be explained by a common factor, "poverty." Moreover, 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 can 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 $n(= 1, 2, \cdots, N)$ and variable $l(= 1, 2, \cdots, N)$ 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 $\mathbf{v} = \mathbf{Zw}$ where the weights are $\mathbf{w} = (w_1, w_2, \cdots, w_l)^T$ with $\|\mathbf{w}\| = 1$. We can find the first principal component by calculating that maximizes the variance of $\mathbf{v}$.

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

Letting the Lagrange multiplier be λ and taking first order conditions, we get

$$(\mathbf{Z}^T\mathbf{Z} - \lambda\mathbf{I})\mathbf{w} = \mathbf{O}$$

Hence, a straightforward calculation shows $\mathbf{v}^T\mathbf{v} = \lambda$. The variance of $\mathbf{v}$ is maximized when $\mathbf{w}$ is an eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{Z}^T\mathbf{Z}$. 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, the interpretation of the principal component score can be arbitrary.

We use the Seila Commune Database (CDB) to create a MWBI map. The Seila CDB is a comprehensive database that includes basic socioeconomic 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 CDB was collected between November 2002 and 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. Since the SAE poverty estimates are created and some indicators are available only at the commune level, we aggregated village-level indicators to the commune level in order to create the MWBI poverty map. Different questionnaires were used in the rural areas and urban 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.[1]

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

[1]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 Phnom Penh in the CSES 1997

Table 1: List of variables, their summary statistics and their scoring coefficients for MWBI poverty mapping. Unit of analysis is commune, and each commune has equal weight (N=1458).

Variable	Mean	Standard Deviation	First eigenvector
Dependency ratio	.425	.043	0.401
% HHs with thatched roof & no latrine	.467	.209	0.482
Motorcycles per capita	.033	.022	-0.398
% HHs with access to water at home	.402	.311	-0.236
Literacy ratio	.747	.183	-0.485
Distance to closest market in minutes	57.2	84.8	0.397

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 our analysis so long as these dimensions of welfare are covered.

The means and standard deviations of each indicator are reported in the second and third column of Table 1. Here, we are interested in the commune-level ranking of poverty and thus the unit of analysis is a commune. Therefore, we did not apply weights here. For the same reason, we did not apply weights to the principal component analysis.

The largest λ was 2.69, whereas the rest of eigenvalues are less than one. Thus, we only use the first principal component in this study. The first principal component explains about 45 percent of the total variation. The MWBI scoring coefficient, or the eigenvector associated with the largest eigenvalue, is given in the fourth column 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, a high value of MWBI suggests a low level 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, with the smallest numbers signifying the worst poverty situation in the district. The information in the CCDB was collected during the same period by PDoP. 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 intuitive. 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 Comparison

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[2], 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 and are referenced to different points in time, we do not expect perfect correlations. Yet, we expect some positive correlations because higher values indicate a lower level of welfare in both cases, Indeed, we observe moderately positive correlations in Plain, Tonlesap, and Coastal ecozones as shown in Table 2. However, this is not the

[2]The definition of ecozones are as follows. Plain: Kampon Cham, Kandal, Prey Veng, Svay Rieng and Takeo. Tonle Sap: Banteay Mean Chey, Battambang, Kampong Chhnang, Kampong Thmo, 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.

Table 2: Correlation of MWBI and SAE poverty indicators at the commune level for each ecozone. SAE estimates are not available for 11 communes because the census data is missing.

Ecozone	P_0	P_1	P_2	Obs
Plain	0.35	0.33	0.30	591
Tonlesap	0.32	0.26	0.20	425
Coastal	0.38	0.39	0.35	143
Plateau/Mountain	-0.30	-0.25	-0.21	288

Table 3: Proportion of districts where the sign of Spearman rank correlation coefficient for the welfare rankings based on two indicators is positive. An asterisk denotes that the percentage is different from one half at the 0.01 level of significance.

Ecozone	P_0 and MWBI	P_0 and CCDB	MWBI and CCDB	Obs
Plain	0.79*	0.74*	0.85*	53
Tonlesap	0.72*	0.70*	0.91*	54
Coastal	0.83*	0.61	0.83*	18
Plateau/Mountain	0.56	0.54	0.93*	41

case for the Plateau/Mountain ecozone. This suggests that something may be wrong with this 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.

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 observation in 2003, whereas the SAE estimates are referenced to the census year of 1998. Hence, SAE estimates do not reflect the changes in welfare that took place between 1998 and 2003. Second, it is likely that the district chief's rankings given in CCDB are based on easily observable indicators. Because consumption is not easy to observe compared with 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.

The correlations between P_0 and other indicators in the Plain/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. 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 Plain/Mountain ecozone. While the four northeastern provinces had a negative correlation, the rest of the provinces in the Plain/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%, 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.[3] This may have lead to the apparent underestimate of poverty in the northeast.

6 Implications for geographic targeting

We have shown that the SAE, MWBI and CCDB rankings give us a 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

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

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 simulated value $\hat{P}$ 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 $\Gamma \equiv \{1, 2, \cdots, G\}$ be the index set for groups such as communes. Each individual belongs to exactly one group. Let us denote the population and poverty rate of $g \in \Gamma$ by N_g and poverty rate P_g, and define and $a_g \equiv \frac{\sum_{c=1}^{g} N_c}{\sum_{c=1}^{G} N_c}$ and $b_g \equiv \frac{\sum_{c=1}^{g} N_c P_c}{\sum_{c=1}^{G} N_c P_c}$. 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 1. The horizontal and vertical axes represent the cumulative share of the population and poor people respectively, and the length of OA and OC and is 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. Hence, the individual-level concentration curve looks like the bold line OIB, and OE

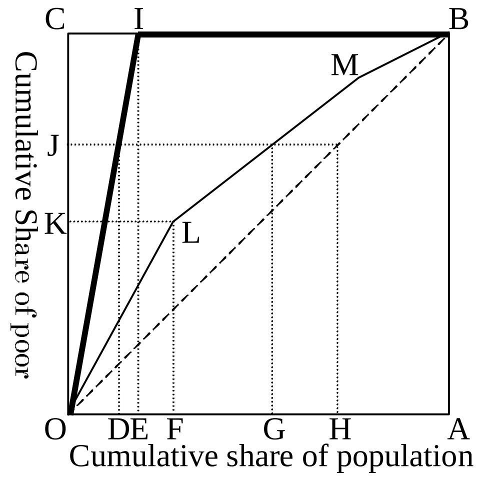

Figure 1: Sample concentration curve

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 OF, is the share of population in the poorest commune, and OK is the ratio of the poor people in the poorest commune to the total number of poor people. Hence, the poverty rate in the poorest commune is $OE\frac{OK}{OF}$.

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 an unrealistic assumption for certain 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 is fixed as *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, fixed *per capita* cost is a convenient assumption because we do not have to worry about how the MWBI corresponds to the poverty gap. The vertical axis tells us how much poverty is reduced when the cost of the program is given. Therefore, it can be interpreted as the goal of poverty reduction. By reading off the graph of concentration curve below, we have the cost for achieving the 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 proportion. 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 $\overline{\mathrm{OJ}}$ in Figure 1, the cost $\overline{\mathrm{OH}}(=\overline{\mathrm{OJ}})$ is without information. If resources are targeted at the individual level, the cost is $\overline{\mathrm{OD}}$. Hence, the budgetary gains from individual-level targeting are $\overline{\mathrm{DH}}(=\overline{\mathrm{OH}}-\overline{\mathrm{OD}})$. Likewise, the budgetary gains from commune-level targeting are $\overline{\mathrm{GH}}(=\overline{\mathrm{OH}}-\overline{\mathrm{OG}})$. Formally, the budgetary gains can be written as $B(g) \equiv g - C^{-1}(g)$, where g is the goal of poverty reduction. It is convenient to have an overall measure. We define the average budgetary gain as $2 \cdot \int_0^1 B(g)\,\mathrm{d}\,g$, which is the ratio of the area OLMB to the area OBC.

Now suppose that the groups are communes sorted from the ones with the highest poverty rates so that we have $P_i \geq P_j$ for $i \leq j$. We call this concentration curve $C^{SAE}(s)$ the SAE concentration curve. 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 $S_i \geq S_j$ for $i \leq j$. This may not be concave. By construction, we have $C^{SAE}(s) \geq C^{MWBI}(s)$ for $\forall s \in [0, 1]$.

Figure 2 shows the SAE concentration curve and MWBI concentration curve. There are two points to note here. First, $C^{MWBI}(s) > s$ is satisfied for most of the values of 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 budgetary gains vary substantially by the goal of poverty reduction. As Figure 3 shows, the budgetary gains for both SAE and MWBI are higher

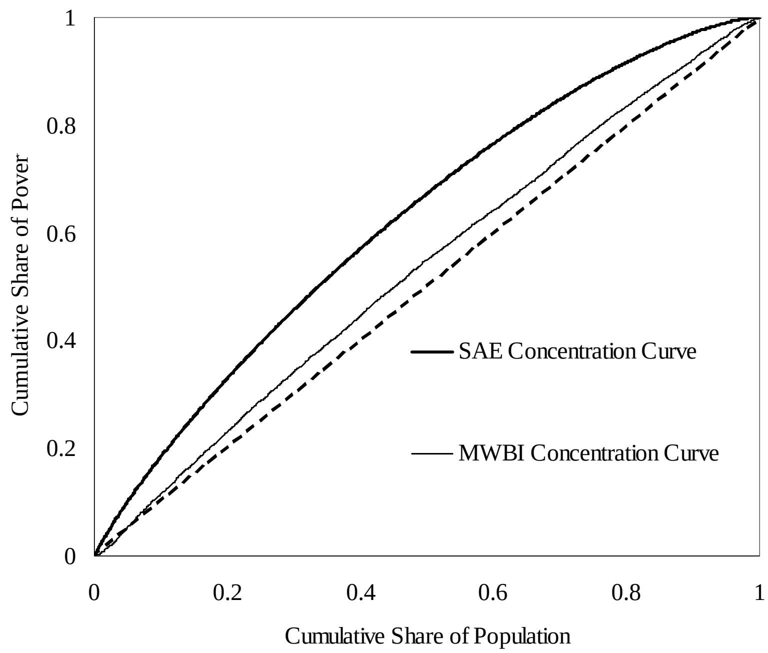

Figure 2: Poverty concentration curves in Cambodia.

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 is around 0.5 for MWBI targeting and 0.6 for SAE targeting.

As Figure 4 shows, the ratio of budgetary gains ($\frac{B^{MWBI}(g)}{B^{SAE}(g)}$) is stable over a range of values of g. Hence, in this case, we can expect reasonably well how much efficiency is lost by using MWBI poverty map. 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 ecozones, and the ratio of average budgetary gains in this case is 0.37. Hence, while MWBI targeting captures a sizable of potential gains from commune-level targeting, there will remain a large amount of

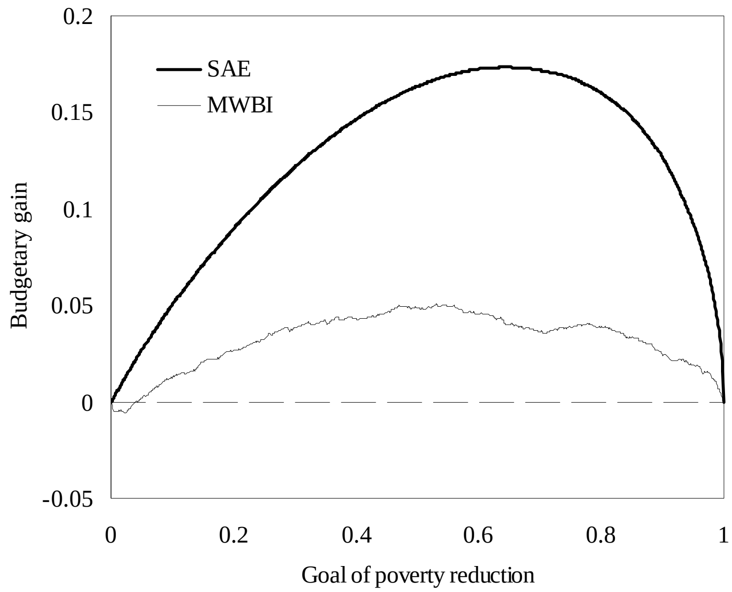

Figure 3: Budgetary gains for SAE and MWBI concentration curves.

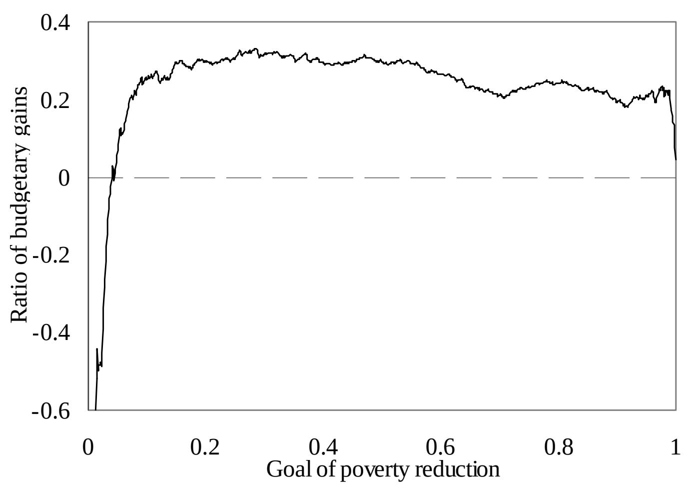

Figure 4: Ratio of budgetary gains.

potential gains MWBI targeting fails to capture.

7 Conclusion

This study provides empirical results of the poverty maps based on different indicators. We constructed a commune-level poverty indicators based on the principal component scores and compared them with poverty estimates based on small area estimation. While the SAE and MWBI estimate correspond reasonably well, there were notable differences. We explored the differences by region with the help of commune classification database, and pointed out several plausible causes of discrepancies.

We then explored the implications for calculated multiple commune-level. We studied the implications for geographic targeting using the concentration curve. 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 have an idea of the cost of going for quick-and-dirty method. Of course, this ratio depends on a number of assumptions. There is no reason to assume that it should be similar in other countries and at other points in time. Accumulation of empirical results would tell us if this is the case.

Whether this amount is large would depend upon the cost of producing these estimates. If no census data were available and census data were to be used for mapping only, it would certainly make sense to use MWBI instead of SAE estimates because collection of census data can easily cost, say, hundred times more than collection of community level data such as CDB5.

However, even though the efforts and resources required to SAE estimates are higher, SAE has advantages over MWBI in terms of interpretability and methodological rigorousness. As this paper shows, even when SAE estimates are available, MWBI estimates could still be useful for the purpose of checking the irregularity of SAE estimates.

References

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'

Demombynes, G., and G. Özler (2002) 'Crime and local inequality in South Africa.' Policy Research Working Paper 2925, World Bank

Elbers, C., J.O. Lanjouw, and P. Lanjouw (2002) 'Welfare in villages and towns: Micro-level estimation of poverty and inequality.' Working Paper 2911, World Bank

_ (2003a) 'Micro-level estimation of poverty and inequality.' *Econometrica* 71(1), 355–364

Elbers, C., P. Lanjouw, J. Mistiaen, B. Özler, and K. Simler (2003b) '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

Elbers, C., T. Fujii, P. Lanjouw, B. Özler, and W. Yin (2004) 'Poverty alleviation through geographic targeting.' Policy Research Working Paper 3419, The World Bank

Fujii, T. (2004) 'Commune-level estimation of poverty measure and its application in Cambodia.' Research Paper 2004/48, United Nations University, World Institute for Development Economics Rsearch

_ (2005) 'Micro-level estimation of child malnutrition indicators and its application in cambodia.' Policy Research Working Paper, World Bank

Fujii, T., and W. Jack (2005) 'Optimal geographic targeting.' mimeo, Georgetown University

Henninger, N., and M. Snel (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, Ministry of Planning, Royal Government of Cambodia)

UNOPS, and 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)