Commune Level Poverty Estimates and Ground Truthing

April, 2003

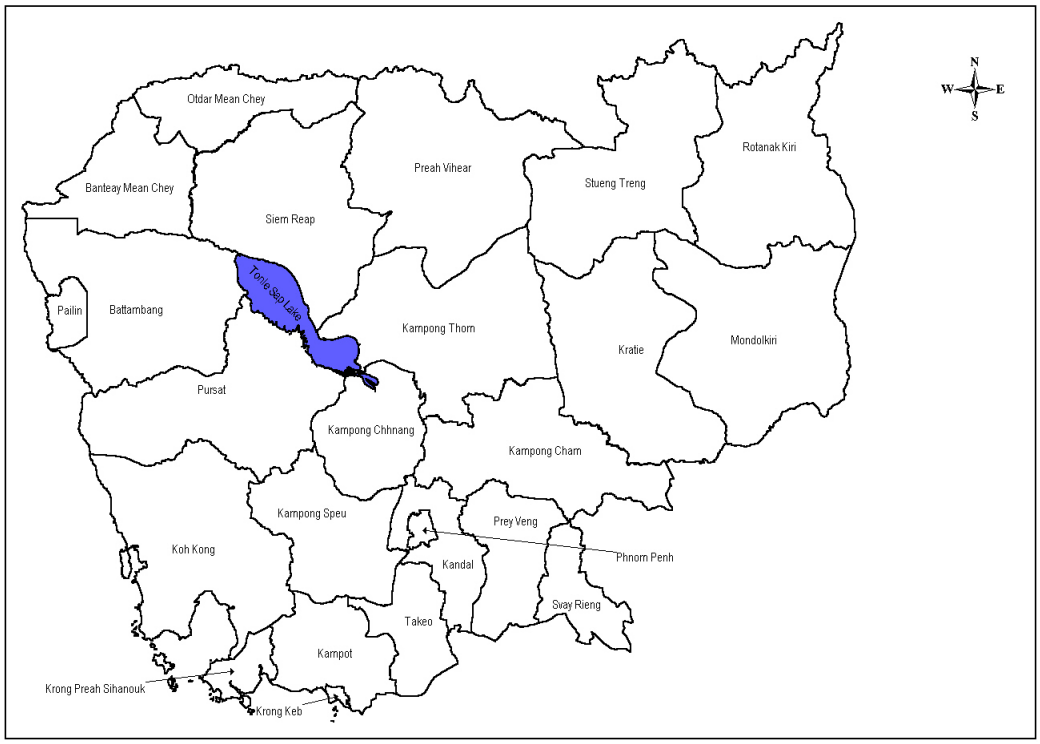
Report Prepared by Tomoki FUJII

Table of Contents

List of Tables

List of Figures

Map of Cambodia



SECTION I. Introduction

Background

The United Nations World Food Programme (WFP) Cambodia, in close collaboration with various government institutions, has been conducting poverty analysis and mapping refinement. The main objectives of the refinement have been (i) to refine core poor areas in Cambodia, (ii) to identify priority areas for social sector interventions such as those targeted to adult and child education, health and nutrition interventions, and (iii) to identify priority areas for assisting variable population groups such as those in flood and drought prone areas, forest and fishing concession areas.

To refine the core poor areas in the country, WFP worked closely with the Ministry of Planning to combine Cambodia Socio-Economic Survey 1997 (CSES 1997), Population Census 1998 and other GIS databases by using the small area estimation technique recently developed by Elbers, Lanjouw and Lanjouw (2003). Conceptually, the methodology imputes the consumption expenditure for each of over 2.1 million households in the census data through the regression model coefficients of which are estimated with the CSES 1997 data while explicitly taking into account the residuals. This exercise yielded the report titled, "Estimation of Poverty Rates at Commune Level in Cambodia: Using the Small-Area Estimation Technique to Obtain Reliable Estimates" (MoP and WFP, 2002).

The results presented in MoP and WFP (2002) are a significant improvement over previous poverty maps produced by WFP Cambodia. However, as stated in MoP and WFP (2002), it reflects the poverty situation as of 1998. This implies that the poverty situation may have changed significantly since then, and using the poverty map as a sole basis for the formulation of targeting policies is likely to be misleading and inappropriate. In particular, those areas struck by repeated natural disasters like drought and flood may have been severely impoverished since 1998.

Another matter that should be given careful consideration is the nature of the estimates in MoP and WFP (2002). The estimates of poverty rates presented in MoP and WFP (2002) are, as with most of other poverty estimates, subject to errors. Suppose that a commune has an estimated poverty rate of 55.4% and its associated standard error is 11.2%, then the seemingly high poverty rate is by chance and the actual poverty rate may well be below the national average of 36.1%. Although the magnitude of standard error in comparison with the point estimates was found to be low enough for the estimates to be useful, there are communes for which the standard errors are quite high. It is, therefore, desirable to check if such communes should be eligible for the assistance of WFP Cambodia.

Also, MoP and WFP (2002) point out that the communes outside the sampling frame of CSES 1997 may not have been estimated accurately. This stems from the fact that some parts of Cambodia

were not visited due to safety issues when the CSES 1997 survey was conducted. Hence, the estimates of coefficients in the consumption regression were based only on the data within the sampling frame of CSES 1997.

Given aforementioned issues, it is desirable and necessary to conduct an exercise that checks the validity of the poverty estimates in light of easily observable indicators. The exercise will be referred to as ground-truthing in this report, but it should be emphasized that it is not a statistical validation of the poverty estimates provided in MoP and WFP (2002). It is an exercise that is carried out to have better understanding of where the estimates are likely to be off, and, more importantly, where people are really in need and where WFP should target its resources.

WFP Cambodia Strategic Direction 2003-2005

To put the ground-truthing in the context of the WFP Cambodia's activities, it should be noted that WFP Cambodia has undertaken reprogramming exercise to determine how it can best provide appropriate assistance in the near future. The product of most recent reprogramming exercise is the report entitled "WFP Cambodia Strategic Directions 2003-2005" (WFP, 2002a). It identifies eight key concepts that are relevant to the work of WFP, which are food security, people in crisis, sustainable livelihoods, human rights based approach, Millennium Development Goals (MDGs), WFP global policies, UN Development Assistance Framework (UNDAF) and National Policies and Plans.

Though the readers are referred to WFP (2002a) to understand the significance of each term, it would be worth reviewing the concepts most relevant to this report. Food security is defined as "all people, at all times, have physical and economic access to sufficient, safe, and nutritious food to meet their dietary needs and food preferences for an active and health life." To better articulate situations where food aid is necessary and appropriate, the concept of "people in crisis" is of help. People in crisis are those who face basic food security problems, because i) they have suffered from a traumatic change, which poses sudden shocks to family and community food security, ii) because they are at a critical turning point when long term trends create a new threat to basic food security or iii) because they are left out or fall behind a development process that is occurring, which will lead to families' inability to remain food secure.

The significance of the ground-truthing exercise can be best understood as an effort to understand and identify the people in crisis. WFP Cambodia has adopted the mixed geographic and sectoral approach to reach the people in crisis (WFP, 2002a). In this approach, the aid is targeted to the poorest and most food insecure communes, but a specialization in certain types of activities is required. While the ground-truthing exercise will not be able to distinguish very well people in different types of crisis and hence the policy that is appropriate to a specific area should be determined on the case by case basis, it provides useful information on how resources should be

targeted geographically. In other words, the ground-truthing exercise helps greatly the "geographic" part of the mixed approach. For the "sectoral" part, WFP Cambodia has selected three sectors in which to concentrate its work: education, health and nutrition, and disaster and mitigation. Five strategic objectives fitting into these three sectors have been also identified (*See* Box 1).

Box 1. Focal Sectors and five mid-term objectives of WFP Cambodia (WFP, 2002a).

Education

1. Decreased repetition and higher retention rates are achieved, and short term hunger is reduced, particularly among girls, in basic education, through targeted school feeding and complementary support in disadvantaged areas.
2. Higher rates of functional literacy are achieved, through support to coordinated literacy and life skills training programs, particularly to vulnerable women and adolescent girls in disadvantaged communes.

Health and Nutrition

3. Child nutritional status is improved through food aid including nutritional supplements and training in caring and basic health practices and complementary technical support.
4. People with HIV/AIDS and TB who seek assistance through existing programs are provided food supplements to enable their participation in care and treatment. Special attention is given to integrated home care programs serving families affected by AIDS.

Disaster and Mitigation

5. Communities in crisis due to sever floods or drought are assisted through activities that reduce vulnerability, promote sustainable livelihoods and restore or develop community assets.

At the national level, the Royal Government of Cambodia (RGC) has recently adopted the second Socio-Economic Development Plan of Cambodia (SEDP II), and the first National Poverty Reduction Strategy (NPRS), which includes the poverty map presented in MoP and WFP (2002). This is expected to have significant impacts on the development community in Cambodia as it provides a common ground for the cooperation and coordination among all the stakeholders, including the governmental institutions, non-governmental organizations, bilateral and multilateral

donor agencies, academic and research institutions, and private sector, and agrees with the spirit of the UNDAF. The ground-truthing exercise brings about value-added to WFP as well as other stakeholders by providing useful additional insights on the use of the poverty map given the changes since 1998 and the current situation.

At the international level, the MDGs, which were agreed upon at the Millennium Summit in 2000 by world leaders, identify eight global goals and eighteen targets. As one of the signatories of the MDGs, the RGC has adopted an interim set of targets. For example, the RGC sets the target to reduce the percentage of population living below the national poverty line from 36 percent to 31 percent by 2005. Achievement of such a target will require appropriate allocation of resources, which in turn is closely linked to the ground-truthing exercise. In a similar manner, the ground-truthing exercise is in accordance with the five programming principles identified in WFP global policy (*See*, Box 2).

Box 2. Programming Principles of WFP (WFP, 2002b).

a) Enable young children and expectant and nursing mothers to meet their special nutritional and nutrition-related health needs.

b) Enable poor households to invest in human capital through education and training.

c) Help poor families to gain and preserve assets.

d) Mitigate the effects of recurring natural disasters in vulnerable areas.

e) Helphouseholds which depend on degraded natural resources to shift to more sustainable livelihoods.

Objective, Scope and Structure

There are three main objectives of the ground-truthing exercise. The first objective is to assess the current situation of poverty using the Commune Database (CDB) collected by the Seila Programme. The Seila CDB collects a number of relatively easily observable variables, which provide WFP and other stakeholders with more current information on the poverty situation in Cambodia[1] at the commune or village level by asking questions to village or commune leaders. The latest round of Seila CDB contains data from 1,621 communes, including 1,471 rural communes, covering

[1] It is also possible to compare the current situation with the past as the Seila CDB has records since 1998. However, it is beyond the scope of this paper since the extent to which the comparison can be made is severely limited by the fact that the geographical coverage and collected variables differ from year to year.

Cambodia almost completely. The Seila CDB allows fully exploiting the knowledge of local leaders, and collecting information in a cost effective manner without compromising the geographical coverage.

The second objective is to explore the possibility of deriving useful poverty indices using the Seila CDB. While the small area estimation technique employed in MoP and WFP (2002) is considered to produce reliable estimates of poverty measures, its data requirement is demanding. The survey and census datasets are required, both of which can be carried out usually only occasionally. In particular, census is carried out at most only once in a decade in many countries. Hence, if it is possible to construct useful indicators of poverty from easily observable indicators, poverty can be monitored more closely and in a more timely manner at the commune level.

As briefly discussed above, testing the validity of poverty estimates given in MoP and WFP (2002) in the statistical sense is out of the scope of this report. In fact, if such validation were to be carried out, it would be necessary to carry out a new socioeconomic survey to derive consumption measure. Even if such data collection were possible, it would be already difficult to call it a statistical validation as the difference between poverty estimates in MoP and WFP (2002) and those from a new socioeconomic survey may be due to errors in the estimate, but may also be due to the changes that have taken place since 1998.

The third objective is to try to assess some of the concerns expressed about the poverty map. For that purpose, the Commune Classification Database (CCDB) was constructed. The CCDB has systematically collected for the first time the opinion of district chiefs on the relative status of poverty situation of the communes within each district in Cambodia. Using this data together with the poverty indices derived from the Seila CDB and poverty estimates in MoP and WFP (2002), the concerns were evaluated.

This report is structured as follows. In SECTION II, the Seila CDB will be described and its quality is discussed. The description of the CCDB is also provided. In SECTION III, the current situation of poverty will be assessed with descriptive statistics of a number of socio-economic indicators. This section corresponds to the first objective of the paper. In SECTION IV, the possibility of poverty index with easily observable index will be pursued and the second objective is fulfilled. SECTION V discusses the implications of the analysis of preceding sections with a particular emphasis on the concerns about the poverty maps to address the third objective., followed by conclusion.

SECTION II. Seila Commune Database and Commune Classification Database

Seila Commune Database

The Seila CDB is a comprehensive database, containing basic socioeconomic data at the village level with its objective being to provide data and input for planning situation analysis and local development planning, for decision making on the allocation of resources, and for measuring impacts of local development activities. It was originally introduced with the assistance of UNDP, UNOPS and CARERE as a project database. The Seila CDB is managed by the Provincial Departments of Planning (PDoP) under the technical supervision of the MoP. The initial design of the database program, which is based on the standard village data book (VDB), was completed in 1998. The PDoP in the original five Seila provinces first collected data for the Seila CDB and annual update has been carried out.

There have been five rounds of data collection. The data used in this study is collected from the latest round, and will be called the Seila CDB5. The collected variables and geographical coverage vary from year to year as noted in footnote 1 in page 4. For example, in the earlier version of the Seila CDB described in Seila Programme (2001), the list of 77 questions for the Seila CDB issued by the MoP is included. These 77 questions were assumed to be the minimum information to be collected annually in every village. However, the database was supposed to be flexible and specific provincial questions can be added as per the province-specific situation or needs. The province is responsible for the formulation of the questions and the clear definitions of these questions. The answers to these questions are for provincial use only.

In the 77 questions, there are 13 different categories, including village statistics, housing, education, health, water and sanitation, transportation, agriculture, production, animal husbandry, rural economy, household assets, labor migration and community based organizations. For further details on each of the 77 questions, readers are referred to the VDB of May 2001 version found in Seila Programme (2001). In each province selected for data collection, PDoP is responsible for annually collecting and checking the data in the VDB.

While the Seila CDB has been a cost-effective way of collecting the commune level data useful for planning and decision making, the reviews on the earlier rounds of Seila CDB have suggested that the quality control have not yet been systematic. The latest review on the Seila CDB (UNOPS and UNDP, 2002) points out the following issues which should be addressed to further enhance the CDB as a national database:

- *Quality Control.* One of the strengths of the CDB is that information is collected firsthand by village leaders, either the Village Development Committees, or village chiefs assisted by local resource persons associated with one of the service sectors. A

set of guidelines were developed in consultation with provincial planning departments in the past year to standardize data collection practices and provide a quality control check. What is required is a field check to determine the extent to which local collection procedures reflect these standards, what the understanding of village data collectors about the CDB and the guidelines is, and what actions are taken to verify the data once it is collected at provincial level. Once findings are made of field practices, improved data verification methods can be established to ensure a stronger introduction of the CDB at the national level.

- *Analytical Outputs.* Significant improvements have been made to the development of analytical products from the data. It is recognized that the process to enhance the analysis should continue. Further work on the index used for resource allocations, linking it to secondary sources revisiting the issue of impact analysis, and further refinement of the commune profiles are considered here.

- *Optimizing linkages with other national databases.* Preliminary work has been accomplished to consider how the CDB is related to other datasets, but so far this has been limited to cross checks of the CDB poverty rank results, and identification of other sources. It is appropriate to consider how key information systems, such as those belonging to the Ministry of Health and Education, can be better coordinated with the CDB.

UNOPS and UNDP (2002) provide recommendation regarding revisions to each of the 77 questions, and also recommend that the CDB include a component on the in-village prices for essential items for consumption, to help with the poverty index. As a result of UNOPS and UNDP (2002) and further discussion on the usefulness and reliability of different indicators, the Seila CDB5 collected the following variables during the period between November, 2002 and January, 2003. The list of variables collected in this round is provided in Appendix B. Some variables were taken at the village level while others are taken at the commune level. Also, different questionnaires were used in the rural areas and urban areas. This study focuses on rural areas.

Commune Classification Database

The special feature of this study is the inclusion of district chiefs' opinions on poverty ranking in Cambodia called the CCDB. The data for the CCDB was collected during the same period of as the Seila CDB5 and those in charge of the Seila CDB5 data collection were given the list of the commune in a district, made an appointment with the district chief and asked the chief to rank the communes in the terms of poverty. They were asked to provide their "subjective" judgments, instead of basing them on any "objective" indicator. Hence, in CCDB, each commune is given a

number from one to the number of communes in the district, and the communes with smaller numbers are judged to have worse poverty situations.

The CCDB is of importance in two different ways. Firstly, this is the first attempt to collect their opinions systematically at the national level. Compared with household or individual level surveys, district chiefs' opinions are easier and cheaper to collect. Secondly, there has been no systematic study on the comparison between poverty estimates provided in MoP and WFP (2002) and experts' opinion. Hence, the district chiefs' opinions collected for this study are useful for that purpose.

There are several reasons why it is meaningful to know the nature of district chiefs' opinions. For example, though district chiefs are in a better position to know about each commune in the district, they may have incentives to prioritize some communes over others due to political and other reasons. Even if district chiefs have good understanding of the poverty situation, it is not clear how they perceive poverty. Questions arise whether their poverty ranking are based on, or can be approximated by, poverty rate, poverty gap, poverty severity or completely other indicators.

As will be discussed below, there are some methodological difficulties in using the CCDB. However, it provides a way to check how the poverty estimates are correlated with the intuition of the experts. While disagreement between the rankings by the poverty estimates and opinions of district chiefs may require additional investigation into the district, targeting decision may be confidently formulated when agreement exists in the district.

SECTION III. Univariate Descriptive Statistics from the Seila Database

Why Bother Univariate Descriptive Statistics?

Before trying to derive poverty indices from the Seila CDB5, it is useful to look at univariate descriptive statistics. This is particularly true when the derivation is based on the principal component or factor analysis as these depend upon *ad hoc* weights as discussed in MoP and WFP (2002). Readers are reminded that there is a critical difference between the welfare scores derived in the next section and the poverty indices derived in MoP and WFP (2002). While the latter is based upon consumption and thus has a clear meaning and analytical foundations, the former is more susceptible to the researchers' subjective interpretation. Although the fact that the welfare scores are based on the researchers' subjective interpretation does not necessarily mean that they are of no use, the users of welfare scores should exercise their caution and consider carefully what is implied by them.

To make the argument more concrete, consider a "naïve welfare score" at the village level using the principal component analysis[2] derived from the following four indicators; 1) the distance to the nearest market, 2) the net enrollment rate at the primary level, 3) percentage of houses with a thatched roof and 4) the number of motor cycles *per capita*. One could argue that it will be able to capture such important dimensions of poverty as economic, education, housing and asset dimensions. Such an argument would be valid, but the question is to what extent each dimension is covered as the score derived from this exercise is a scalar.

Though the data selects the weights when principal component analysis is applied, it is *ad hoc* in the sense that the welfare scores can be very misleading if the choice of variables is inappropriate. To be more specific, we must assume that the four indicators are correlated with the *hidden* welfare. However, there is no good reason to assume *a priori* that the net enrollment rate at the primary level has anything to do with poverty, especially if very effective primary education programs are implemented only in limited parts of the country. This contrasts with the consumption regression approach used in MoP and WFP (2002) as the coefficients on irrelevant variables are likely to be insignificant and are not included in the regression model. In other words, it is unnecessary to assume *a priori* that the four indicators above are correlated with the poverty measures in the regression approach.

This is of great importance especially if there are geographical factors that prescribe the pattern of certain indicators. For example, the number of motor boats *per capita* is not likely to be high in mountain areas such as Rotanak Kiri even in relatively rich villages unless they are located near a

[2] For those not familiar with the principal component analysis, more explanation will be provided in the next section.

river. On the other hand, in provinces like Kandal, Prey Veng and Svay Rieng, having a motor boat is likely to be a good indication of non-poor. Hence, it is unlikely that the number of motor boats alone is able to explain the poverty situation well. But the judgment of this sort is up to the researchers and cannot escape from their bias. This is the reason it is important to assess the current situation with descriptive statistics from a single indicator.

It should also be noted that some of the indicators presented below are of importance in itself. For example, the low school attendance rate may be the sole reason for a school-sector intervention; even when the targeted area has a relatively low poverty rate, it may be argued that the investment in human development is of great importance nevertheless as the area is likely to be left behind in development in the future. Family planning programmes may be targeted to the areas where the birth rate is high.[3]

Derivation of Indicators

It is important to note first that most of the variables listed in Appendix B are not appropriate in their original forms. Hence, they were converted to an appropriate measure. For example, the number of bicycles in the village does not provide much information about asset poverty. Instead, if it is expressed in terms of *per capita*, then it is more meaningful. Many of the village level or commune level indicators derived in this way are averages of some sort, and their usefulness is limited by this fact.

To illustrate this point, consider two villages A and B, and both of the villages have 100 people. In Village A, there is a rich bicycle collector who has 200 bicycles, and the remaining 99 people have no bicycle. In Village B, 50 persons have a bike and the rest has none. In this case, it would be fair to say that Village B is wealthier than village A as more people have a bicycle. But, on average, one person has two bicycles in Village A while one person has only 0.5 bicycles in Village B. Though this example may look too extreme, it should be noted that unequal allocation within the commune is not reflected in the indicators presented in this report.[4]

In this study, a number of indicators have been derived and mapped out. The indicators are grouped into five categories; i) demographic indicators, ii) housing indicators, iii) asset indicators,

[3] However, there is a separate issue that the decision makers should be aware of. It should be reminded that the data is heavily dependent on the ability of the village leaders to understand and answer the questions, which may limit the quality of the data despite the efforts to control the data quality. Given the quantity and nature of the questions, it would be reasonable to suspect that there may be some communes for which data error is significant. Maps are useful for describing the overall spatial pattern, but there may be some communes for which some or all of the variables seem unusually high or low. When possible and appropriate, the figures are compared with the ones from other available data sources. Though this does not eliminate the problem, it helps to identify potential sources of concern.

[4] MoP and WFP (2002) have advantage in this regard, too. The regression was carried out at the household level and hence the heteroskedasticity among households in the village are accounted for in MoP and WFP (2002).

iv) education indicators and v) other indicators. While the indicators employed in this study are far from exhaustive, they capture some of the most important aspects of poverty. In what follows, selected indicators of interest are presented category by category and the relevance of each indicator to poverty is explained using the CSES 1997 data whenever possible. The map for each indicator is presented in Appendix F.

Demographic Indicators

Six demographic indicators are presented in this report. The definition of each demographic indicator is provided in Table 1, and its summary statistics is given in Table 2. In the following analysis, the summary statistics were calculated only from rural communes, and each commune was given an equal weight. Hence, the mean of FEMRATIO in Table 2 is *not* the ratio of females in rural areas as such computation must be carried out using the population with the weight. This is because the unit of analysis is the commune in this study, and no weight is used in the following analyses using the Seila CDB5 unless otherwise noted.

All of the indicators in Table 1 potentially have some relevance to poverty. Firstly, female ratio may be affected by social, cultural and economic status of female. In general, however, the spatial variation of female ratio is not expected to vary much. In fact, no systematic relationship between poverty and female ratio was found from the CSES 1997 data.[5] Average female ratio for poor people is 0.522 while it is 0.526 for non-poor people. However, communes with exceptionally high or low female ratio may require closer examination. For example, high female ratio may be an indication of the movement of male labor force into urban areas for better job opportunities, which in turn may imply higher vulnerability of the people left in the commune.

Secondly, dependency ratio is often associated with poverty since each member of the household is more likely to be worse off when there are more dependents to feed. CSES 1997 supports this as the average dependency ratio is 40.8 percent while it is 48.9 percent for non-poor. Readers should note that the definition used in this study should be thought of as a proxy, since people aged under 15 or above 64 are automatically considered to be a dependant in this study.

Thirdly, birth rate is often positively correlated with poverty for a similar reason to dependency ratio. One should note, however, that the recollection error may result in severe bias. As with birth rate, family size is also often correlated positively with poverty, which is clear from Table 3.

[5] In what follows, all the calculations based on CSES 1997 data are made by the author. Some of the numbers presented in this report may not seem consistent with the Poverty Profile 1997 (MoP, 1998). This is due to the data problems described in MoP (2000) and MoP and WFP (2002). Poverty lines are set at 1629 Riels for Phnom Penh, 1214 Riels for Other Urban and 1,036 for Rural Stratum for the reasons fully discussed in MoP and WFP (2002). One should note that the unit of each record is an individual or household in CSES 1997 while it is a village or commune in the Seila CDB.

Table 1. Demographic Indicators. Capital name in the bracket is the short name.

a) *Female ratio*	(FEMRATIO)
Ratio of females in the population.	
b) *Dependency ratio*	(DEPRATIO)
Ratio of dependants in the population. People aged under 15 or above 64 are considered as a dependant.	
c) *Birth rate*	(BRATE)
The number of babies born last year over the population.	
d) *Family size*	(FAMSIZE)
Family size is defined as population over the number of families.	
e) *Ratio of female headed household*	(F_HHH_RATIO)
Ratio of female headed household is the number of households headed by female divided by the total number of households.	
f) *Ratio of female headed household with children under five*	(F_HHH_UD5_RATIO)
Ratio of female headed household is the number of households headed by female with one or more children under five divided by the total number of households.	

Table 2 Univariate Summary Statistics of Demographic Indicators.[6]

Variable	Obs	Mean	S.D.	Min	Max
FEMRATIO	1471	0.5158	0.0151	0.4411	0.5629
DEPRATIO	1471	0.4252	0.0430	0.2626	0.6320
BRATE	1471	0.0169	0.0091	0.0000	0.1198
FAMSIZE	1471	5.0082	0.4401	2.6175	8.8345
F_HHH_RATIO	1471	0.1602	0.0702	0.0000	0.7742
F_HHH_UD5_RATIO	1471	0.0313	0.0290	0.0000	0.3113

[6] Obs means the number of observations (*i.e.* the number of rural communes). S.D. means the standard deviation. Min and max are the minimum value and maximum value of all the observations, and give the range of the variable. This applied to the subsequent tables similar to Table 2.

Table 3. Household size and poverty rate. CSES 1997 is calculated by the author and CSES 1999 Round 2 is after MoP (2001).

Household Size	CSES 1997	CSES 1999 Round2
1-2 persons	9.7%	9.5%
3-4 persons	24.3%	23.6%
5-6 persons	37.0%	34.9%
7-8 persons	44.5%	43.9%
9-10 persons	51.7%	44.9%
More than 10 persons	51.1%	27.8%
Cambodia	**36.1%**	**35.9%**

Since experiences from a number of countries suggest that female headed households are likely to be more vulnerable, ratio of female headed households can be a candidate indicator to identify poor areas. However, in Cambodia, female headed households have lower poverty rate than male headed household. The picture is, however, quite different if existence of very young children in the household is taken into consideration. In this case, the poverty rate for female headed households with one or more children under five is 42.3 percent while the poverty rate for other households is 35.7 percent using CSES 1997.

Table 4 Household size and poverty rate. CSES 1997 is calculated by the author and CSES 1999 Round 2 is after MoP (2001).

Household Size	CSES 1997	CSES 1999 Round2
Female headed household	33.3%	33.6%
Male headed household	36.8%	36.4%
Cambodia	**36.1%**	**35.9%**

Housing Indicators

Four housing indicators are presented in this report. The definition of each housing indicator is given in Table 5 and its summary statistics in Table 6. The type of roof has been frequently used to predict the existence of poverty in Cambodia as has been the case in many other countries. Using the CSES 1997 data, the poverty rate for each roof type was estimated. The rate for thatch, bamboo

and grass was estimated at 48.9 percent, which is much greater than the national average of 36.1 percent. The rate for tile, concrete/brick/stone, galvanized iron/aluminum and other were 30.1 percent, 3.6 percent, 24.9 percent and 18.6 percent respectively.

While the thatched roof houses seem to be an indication of poverty, even a casual observation in the field tells us that there is a great heterogeneity in terms of living standards among the families living in a thatched house. While the thatched roof is perhaps cheap and weak, some families even possess a TV and live in a spacious house. Such families are probably not poor, if not rich. This observation may partly reflects the fact that the transportation costs of materials required for other types of roof such as tile and galvanized iron is higher in remote areas. On the other hand, poorest families often seem to live in a small house with no sanitary facility, and the roof and wall of their houses have a number of holes. It would be, therefore, meaningful to look at poverty rate for the thatched roof houses with no latrine.

Analysis of CSES 1997 uncovers intriguing facts. The poverty rate for thatched roof houses is higher when a latrine exists in the house. However, this order is reversed for poverty gap and poverty severity. In fact, poverty severity is twice as much for thatched roof households. Though there are only 41 households living in a thatched roof house with a toilet, as opposed to 2476 households without a latrine, it may well be the case that extremely poor people often reside in a thatched roof house with no latrine.

Table 5. Housing Indicators. Capital name in the bracket is the short name.

a) *Ratio of houses with a thatched roof*	(R_THATCH_RATIO)
The number of houses with a thatched roof over the total number of houses	
b) *Ratio of thatched roof houses with no latrine*	(R_THATCH_NO_LAT_RATIO)
The number of houses with a thatched roof minus the total number of latrines in thatched roof houses divided by the total number of houses.[7]	
c) *Ratio of houses with a latrine*	(RAT_TOILET)
The total number of toilets over the total number of houses.	
d) *Ratio of families with poor water access*	(FAM_H2O_OTHE_RATIO)
The number of families with no private or communal piped water, pump well or ring well usable year round over the total number of families.	

Table 6 Summary statistics of housing indicators

Variable	Obs	Mean	S.D.	Min	Max
R_THATCH_RATIO	1471	0.4769	0.2109	0.0310	1.0000
R_THATCH_NO_LAT_RATIO	1471	0.4643	0.2104	0.0265	1.0000
RAT_TOILET	1471	0.0813	0.1160	0.0000	0.9547
FAM_H2O_OTHE_RATIO	1471	0.4164	0.3537	0.0000	1.0000

Table 7 Comparison of poverty measures for people living in a thatched roof house using CSES 1997 data.

	Poverty Rate	Poverty Gap	Poverty Severity	Share
Thatched roof with a latrine	49.7%	8.6%	2.4%	1.6%
Thatched roof without a latrine	48.9%	12.8%	4.8%	98.4%
Thatched roof	48.9%	12.8%	4.7%	100.0%

[7] It is assumed that one house has at most one latrine. While exception may exist, this seems to be a reasonable assumption

Existence of latrine reflects the standards of housing conditions, and is also considered as a good indicator of poverty. The poverty rate for people living in a house with toilet is 10.2 percent whereas it is 40.0 percent for those without a toilet. Access to water measures the standards of housing conditions. As Table 8 shows, people who get water from the pipe or buy water are less likely to be poor. The Seila CDB5 has a quite different format and thus the data we have may not be very useful. However, poor water access can be a reason for intervention in itself and the map is useful for targeting projects to improve water access.

Table 8. Water source and poverty measures using CSES 1997 data.

Main source of water	Poverty Rate	Poverty Gap	Poverty Severity	Share
Piped Water	9.4%	1.5%	0.4%	8.8%
Tube/Pipe well	41.5%	11.0%	4.2%	15.8%
Dug well	40.4%	10.2%	3.7%	37.1%
Spring/river/stream/lake/pond/rain	37.1%	8.5%	2.9%	28.6%
Bought	18.6%	3.9%	1.2%	5.9%
Other	53.0%	14.2%	5.9%	3.8%
Cambodia	36.1%	8.9%	3.2%	100.0%

Asset Indicators

Five asset indicators, including indicators for bicycles, carts, motorcycles, televisions and cars, are presented in this report. The definition of each indicator is presented in Table 9 and its summary statistics in Table 10. In general, individual is better off with more assets. Hence, it may first seem reasonable to assume that all the assets indicators presented here are negatively correlated with poverty. However, this is not necessarily the case as Table 11 shows.

One should first note that the poverty rate for those in a household with one or more carts is higher than that for those without. Moreover, the order is reversed for the poverty severity. This suggests that people are more likely to be "shallow poor" if they have carts. This situation can be better understood if one recognizes that carts are an inferior good.[8] Obviously, extremely poor

[8] An inferior good is consumed less when the individual gets wealthier. Of course, whether a cart is inferior good depends on the preference of the individual.

people are unlikely to possess a cart. However, once they get less poor, they may be able to possess a cart, even if they are living under the poverty line. Now, let us consider the situation where they get much richer so that they are now above the poverty line. They may think that carts are not convenient enough and purchase a truck instead. Though it is not clear from Table 11, bicycle may also be an inferior good.

Table 9. Asset Indicators. Capital name in the bracket is the short name.

a) *Number of bicycles per capita*	(PC_BICYCLE)
The number of bicycles over the population	
b) *Number of carts per capita*	(PC_CART)
The number of carts over the population	
c) *Number of motorcycles per capita*	(PC_MOTO)
The number of motorcycles over the population	
d) *Number of televisions per capita*	(PC_TV)
The number of televisions over the population	
e) *Number of cars per capita*	(PC_CAR)
The number of cars over the population	

Table 10 Summary statistics of asset indicators

Variable	Obs	Mean	S.D.	Min	Max
PC_BICYCLE	1471	0.0977	0.0585	0.0000	0.4106
PC_CART	1471	0.0822	0.8856	0.0000	33.9830
PC_MOTO	1471	0.0333	0.0222	0.0000	0.1884
PC_TV	1471	0.0521	0.0333	0.0000	0.1741
PC_CAR	1471	0.0036	0.0053	0.0000	0.0405

Table 11. Existence of asset in the household and poverty measures using CSES 1997 data.

Asset	Individual in a household with the asset (A)			Individual in a household without the asset (B)			Share of (A)
	Poverty rate	Poverty gap	Poverty severity	Poverty rate	Poverty gap	Poverty severity	
Bicycle	35.9%	8.6%	3.0%	36.5%	9.3%	3.5%	59.8%
Cart	40.2%	9.4%	3.1%	34.2%	8.6%	3.2%	32.4%
Motorcycle	18.0%	3.8%	1.2%	41.7%	10.4%	3.8%	23.5%
Television	18.0%	3.6%	1.1%	41.5%	10.4%	3.8%	23.0%
Car	4.2%	0.2%	0.0%	36.6%	9.0%	3.2%	1.5%

On the other hand, motorcycles, televisions and cars seem to be a clearer indicator of non-poor. Given cars are by far the most expensive asset among the five assets presented here, it is natural that the poverty measures for those in a household with one or more cares are quite low. In this regard, the indicator for cars is likely to be a good indicator. But the fact that only 1.5 percent of the population live in a household with a car undermines the usefulness of the indicator.

Motorcycles and televisions are likely to be a good indicator of poverty in many parts of the country in this regard. For example, motorcycles are more useful where the road infrastructure is good, and television is of little use when the reception is very bad. Hence, these indicators, too, are vulnerable to the geographic conditions, which do not have direct causal relationship with poverty.

Education Indicators

For education indicators, ten indicators have been derived. The definition of each indicator is provided in Table 12 and its summary statistics in Table 13. The first two indicators in Table 12 provide the standards of current education. A few cautions are in order. Firstly, no distinction was made in the Seila CDB5 for the primary level education and lower secondary level education. This is presumably due to the fact that education at the primary and lower secondary (PLS) level is compulsory whereas the upper secondary (US) level education is not. PLS level corresponds to the grades one to nine, and ages six to fourteen. US level corresponds to the grades ten to twelve, and ages fifteen to seventeen.

Secondly, as one can see from the definition in Table 12, the definition of school attendance rate is different from more commonly used measures such as the net enrollment rate and gross enrollment rate. The reason school attendance rate is used in this report is due to the questionnaire design, but school attendance rate does have some advantages over the net enrollment rate and gross enrollment rate. The net enrollment rate, defined as the ratio of the number of school age children in

school at the specified level to the number of all school age children at that level, only increases when the child of target age is in school. For example, a child aged 16 in primary school is not included in the net enrollment rate. In a country like Cambodia where it is common to delay schooling for a variety of reasons, this is a clear disadvantage. Gross enrollment rate, defined as the ratio of the number of people of any age group in school at the specified level to the number of all school age children at that level, but this measure does not have a clear correspondence between the denominator and numerator. Increased gross enrollment rate can result from increased rate of repetition, which is clearly undesirable. On the other hand, school attendance rate cannot increase without having a new child in school. Finally, it should be noted that the net enrollment rate and the school attendance rate should be close at the lowest level of education as the difference emerges only when the student go to upper level of education.

Table 14 presents the school attendance rates by poverty status and gender, and the gender gap in school attendance rate by poverty status at the PLS and US levels calculated from the CSES 1997 data. Both at the PLS and US levels, the school attendance rate is substantially higher for non-poor children. In other words, poor children are less likely to attend school. The gender gap in school attendance rate is slightly higher for the poor at the PLS level than that for the non-poor, but the difference vanishes at the US level. For the sake of comparison, Table 15 presents similar results for the net enrollment rate. It should be noted that the net enrollment rate and school attendance rate are close at the PLS level. Secondly, at the secondary level, these two are substantially different, which in turn means that most of the children aged between15 and 17 in school are not in secondary school but in primary school.

The fact that the school enrollment rate is substantially lower for poor children is of great concern from an intergenerational perspective. As was discussed in Fujii and Ear (2002), poverty is likely to be reproduced when poor children cannot receive education as much as non-poor children can. In particular, the school attendance rate for poor female children is very low. Targeting education programs such as school feeding programs for poor female children is, therefore, of great importance in Cambodia.

Table 12. Education Indicators. Capital name in the bracket is the short name.

a) *School attendance rate at PLS level*	(SAR_TSCH_6_14)
The number of children aged 6 to 14 in school over the number of children aged 6 to 14	
b) *School attendance rate at US level*	(SAR_TSCH_15_17)
The number of children aged 15 to 17 in school over the number of children aged 15 to 17	
c) *Gender gap at PLS level*	(GEN_GAP_6_14)
The net enrollment rate for boys aged 6 to 14 minus that for girls aged 6 to 14	
d) *Gender gap at US level*	(GEN_GAP_15_17)
The net enrollment rate for boys aged 15 to 17 minus that for girls aged 15 to 17	
e) *Rate of adult literacy*	(TILT_15OV)
The ratio of the number of literate people aged above 14 to the number of people above 14	
f) *Rate of female adult literacy*	(FILT_15OV)
The ratio of the number of literate female aged above 14 to the number of female above 14	
g) *Number of classrooms per school age child at PLS level*	(PC_PCLSRM)
The number of classrooms at PLS level per children aged 6 to 14	
h) *Number of classrooms per school age child at US level*	(PC_PTCH)
The number of classrooms at the US level per children aged 15 to 17	
i) *Number of teachers per school age child at PLS level*	(PC_SCLSRM)
The number of teachers at PLS level per children aged 6 to 14	
j) *Number of teachers per school age child at US level*	(PC_STCH)
The number of teachers at US level per children aged 15 to 17	

Table 13 Summary statistics of education indicators

Variable	Obs	Mean	S.D.	Min	Max
SAR_TSCH_6_14	1471	0.7683	0.1601	0.0000	1.0000
SAR_TSCH_15_17	1471	0.5721	0.2116	0.0000	1.0000
GEN_GAP_6_14	1471	0.0119	0.0770	-0.5752	0.4574
GEN_GAP_15_17	1471	0.0418	0.1049	-0.6122	0.5513
TILT_15OV_RATIO	1471	0.2512	0.1824	0.0000	0.9712
FILT_15OV_RATIO	1471	0.2799	0.1942	0.0000	0.9884
PC_PCLSRM	1471	0.0147	0.0065	0.0000	0.0447
PC_PTSCH	1471	0.0176	0.0085	0.0000	0.0693
PC_SCLSRM	1471	0.0063	0.0136	0.0000	0.1358
PC_STCH	1471	0.0140	0.0311	0.0000	0.2771

Table 14. School attendance rate and its gender gap at the PLS and US levels using CSES 1997 data.

		School attendance rate			Gender gap	Share
		Male (A)	Female (B)	Total	(A)-(B)	
PLS level	Non-Poor	71.8%	68.3%	70.1%	3.6%	58.0%
	Poor	59.5%	53.7%	56.6%	5.8%	42.0%
	Total	66.6%	62.2%	64.4%	4.5%	100.0%
US level	Non-poor	70.8%	45.8%	59.1%	25.0%	64.7%
	Poor	56.4%	31.5%	42.7%	24.9%	35.3%
	Total	66.2%	40.2%	53.3%	26.0%	100.0%

Table 15. Net enrollment rate and its gender gap at the PLS and US levels using CSES 1997 data.

		Net enrollment rate			Gender gap	Share
		Male (A)	Female (B)	Total	(A)-(B)	
PLS level	Non-Poor	71.3%	67.7%	69.5%	3.6%	58.0%
	Poor	59.5%	53.6%	56.6%	5.9%	42.0%
	Total	66.3%	61.8%	64.1%	4.5%	100.0%
US level	Non-poor	7.4%	6.1%	6.8%	1.2%	64.7%
	Poor	0.6%	1.1%	0.9%	-0.5%	35.3%
	Total	5.2%	4.2%	4.7%	1.1%	100.0%

Table 16. Adult literacy rates by poverty status and gender using CSES 1997 data.

	Male	Female	Total
Non-poor	82.3%	60.0%	70.1%
Poor	73.9%	50.9%	61.2%
Total	79.6%	57.0%	67.2%

Table 16 compares the adult literacy rate for each group of poverty status and gender. Poverty is negatively correlated with the adult literacy rate. As with the school attendance rate, the poor female has the lowest adult literacy. This again implies that adult literacy programs should be designed to include poor females as much as possible.

CSES 1997 data does not have data that directly corresponds to PC_PCLSRM, PC_SCLSRM, PC_PTCH and PC_STCH. In particular, there is no information in CSES 1997 about the number of classrooms. However, PC_PCLSRM and PC_SCLSRM can be considered as a good proxy for measuring the "quality" of education. PC_PTCH and PC_STCH can also be considered as such a proxy. It should be noted that PC_PTCH and PC_STCH are different from the more commonly used student-teacher ratio.[9] Since the village questionnaire of CSES 1997 contains information on the number of teachers at each level of education, it is possible to derive crude estimates of PC_PTCH and PC_STCH once the number of corresponding school age children is known. Though

[9] Due to the design of the Seila CDB, it was not possible to derive the student-teacher ratio. For example, the Seila CDB does not contain information how many children attend the upper secondary school.

the village questionnaire does not contain such information, it is possible to derive crude estimates of the number of corresponding school age children by the following procedure.

First, the village questionnaire has the number of children under eighteen. To derive the estimates, some reasonable assumumptions must be made. Assuming that the age distribution in the village is the same as that for the stratum, it is possible to derive the ratio of the number of children in the age group of the PLS or US levels to the number of children under eighteen.[10] Using the numbers in NIS (1998; p53), the ratio was derived and then the ratio was multiplied to the number of children under eighteen to arrive at the estimate of the number of children in the age group at the respective level. Using the CSES 1997 data, the estimates of average PC_PTCH and PC_STCH for the given poverty status are provided in Table 17.

Table 17. The number of teachers per school age children at the PLS and US levels using CSES 1997.

	PLS level	US level
Non-poor	0.0319	0.0110
Poor	0.0218	0.0081
Total	0.0277	0.0100

Other Indicators

In this paper, four indicators that do not fall in the above-mentioned categories are presented. The definition of each indicator is given in Table 18. PC_TOT_PROD tries to capture the relationship between the production and poverty. Since the overwhelming majority of the farmers produce rice, this is likely to be a good indicator. The relationship between poverty and crime is not clear, but one may speculate that poverty may cause increased social insecurity partly out of necessity and partly out of frustration. One could also argue that the cause of crimes is more likely to be inequality or other reasons. In any case, to the best of the author's knowledge, there have not been a map of crime rate at the commune level in Cambodia, and thus this indicator is already of interest to look at. PC_LANDCONFLICT is also included for a similar reason. It is often pointed out that poor market access often leaves the poor behind in development. Analysis of CSES 1997 data supports this point. The average distance to the nearest permanent market was 5.7 kilometers for the non-poor while it

[10] For three villages which did not have information on the number of children under 18, the total population was used instead.

was 7.6 kilometer for the poor with combined average of 6.4 kilometers.[11] While the distance data is also available in the Seila CDB5 data, the decision was made to use the data in minutes for the analysis as the time to the market is likely to be more relevant and the distance data in kilometers seems less reliable because there are many outliers.

Table 18. Other Indicators. Capital name in the bracket is the short name.

a) *Rice Production Per Capita*	(PC_TOT_PROD)
Total Rice Yield divided by the population	
b) *Number of Heinous Crimes Per Capita*	(PC_INSECURITY)
The number of murder, robbery and theft cases last year divided by the population	
c) *Number of Land Conflicts Per Capita*	(PC_LANDCONFLICT)
The number of land conflicts last year divided by the population	
d) *Average Time to the Nearest Market*	(MIN_MARKET)
Average time from each village to the nearest market by motor or motorboat weighed by the village population	

Table 19 Summary statistics of other indicators

Variable	Obs	Mean	S.D.	Min	Max
PC_TOT_PROD	1471	1.9623	19.2874	0.0000	403.5088
PC_INSECURITY	1471	0.0013	0.0023	0.0000	0.0450
PC_LANDCONFLICT	1471	0.0025	0.0039	0.0000	0.1141
MIN_MARKET	1471	56.86	84.58	0.00	1440.00

[11] There are 50 villages for which the data on the distance to the nearest permanent market is not available. Hence, the 615 observations in these villages, out of the total of 6,010 observations were not used to compute the distance.

SECTION IV. Deriving Overall Indicator of Poverty: A Factor Analysis

Does factor analysis matter for poverty analysis?

While the univariate descriptive statistics are useful to understand the nature of each indicator, it is not easy to understand the "overall" situation of poverty when there are a number of indicators that have some relevance to poverty. To present the information in a more comprehensible form, it is necessary to reduce substantially the dimensions, or the number of variables. This can be done by giving weights to different indicators in a completely arbitrary manner. In this study, however, factor analysis has been applied to avoid arbitrariness.

The idea of factor analysis is to derive the "hidden" factors common in the set of variables. Appendix E provides a concise overview of the factor analysis. In the case of poverty analysis, our goal is to derive welfare scores that are negatively correlated with poverty measures using a set of variables in the Seila CDB5. For that purpose, a number of indicators were derived in SECTION III and most of the indicators were correlated with the poverty rate. It should be noted that bivariate correlation is only suggestive and is not necessary to be a candidate for the variables in factor analysis. While application of principal component analysis seems common among practitioners, application of factor analysis does not seem very common at least in more academic literature. The reason may be that factor analysis entails a number of sources of arbitrariness. However, as Sahn and Stifel (2000) showed empirically in Africa, it is possible to employ factor analysis and successfully apply it to data sets with limited economic information to come up with meaningful results.

While the weights are not arbitrary in factor analysis, it should be noted that the *choice* of variables is arbitrary. Although the author tried to choose reasonable set of variables and make sure that the results are relatively robust with respect to the choice of variables, one can always carry out similar analysis with different set of variables. Moreover, the weights found in this exercise do not have strong theoretical foundations. In fact, it is natural to assume that different institutions should use different weights if they have different objectives. Hence, while meaningful and useful, the results presented in this section should be understood as one of the various convenient ways of expressing the information.

Deriving factor scores

In this study, factor analysis was carried out using the principal component method, and factor score was then derived using the regression. As will be argued below, the factor score allows us to compare the poverty situation across the rural communes. The benchmark set of variables in this analysis is the variables presented in SECTION III and the results are provided in Table 20. Additional results of factor analysis carried out with different sets of variables are presented in

Appendix G. In this analysis, the number of factors is set at two. While this choice is arbitrary, it seemed that the first two have large enough eigenvalue[12] and meaningful interpretation. The main point of the analysis did not change much when the number of assumed common factors was slightly changed.

Table 20. The results of a commune-level factor analysis. Principal component method was used and no weight was applied. First two eigenvalues of reduced correlation matrix were 7.12 and 1.92 explaining 24.55% and 6.61% of the total variance and the number of observations is 1471.

	Factor Loading Matrix		Uniqueness	Standard Scoring Coefficients	
	Factor 1	Factor 2		Factor 1	Factor 2
FEMRATIO	0.213	-0.394	0.800	0.030	-0.205
DEPRATIO	-0.541	0.041	0.706	-0.076	0.022
BRATE	-0.404	0.150	0.814	-0.057	0.078
FAMSIZE	-0.021	0.004	1.000	-0.003	0.002
F_HHH_RATIO	0.371	-0.548	0.561	0.052	-0.286
F_HHH_UD5_RATIO	0.040	-0.475	0.773	0.006	-0.248
R_THATCH_RATIO	-0.809	-0.086	0.338	-0.114	-0.045
R_THATCH_NO_LAT_RATIO	-0.818	-0.123	0.316	-0.115	-0.064
RAT_TOILET	0.492	0.391	0.605	0.069	0.204
FAM_H2O_OTHE_RATIO	-0.262	0.405	0.767	-0.037	0.211
PC_BICYCLE	0.593	-0.302	0.557	0.083	-0.158
PC_CART	0.049	-0.040	0.996	0.007	-0.021
PC_MOTO	0.602	0.159	0.612	0.085	0.083
PC_TV	0.741	0.060	0.447	0.104	0.031
PC_CAR	0.184	0.371	0.829	0.026	0.194
SAR_TSCH_6_14	0.691	-0.095	0.514	0.097	-0.049
SAR_TSCH_15_17	0.682	-0.072	0.529	0.096	-0.038
GEN_GAP_6_14	-0.161	0.083	0.967	-0.023	0.043
GEN_GAP_15_17	-0.025	-0.179	0.967	-0.004	-0.093
TILT_15OV_RATIO	-0.841	0.088	0.285	-0.118	0.046
FILT_15OV_RATIO	-0.836	0.080	0.295	-0.117	0.041
PC_PCLSRM	0.381	0.208	0.812	0.054	0.108
PC_PTSCH	0.575	0.212	0.624	0.081	0.111
PC_SCLSRM	0.449	0.326	0.692	0.063	0.170
PC_STCH	0.453	0.321	0.692	0.064	0.167
PC_TOT_PROD	0.092	-0.006	0.992	0.013	-0.003
PC_INSECURITY	-0.100	0.365	0.857	-0.014	0.190
PC_LANDCONFLICT	-0.105	0.322	0.885	-0.015	0.168
MIN_MARKET	-0.506	0.104	0.734	-0.071	0.055

[12] See footnote 24 in page 52.

Factor 1 and Factor 2 explain over thirty percent of total variance, but there remain significant amount of variance unexplained by these two factors. This suggests that poverty is, in fact, multi-dimensional, and is not necessarily the weakness of this analysis.[13] Factor analysis allows us to compare the poverty situation taking many indicators into account, which otherwise would have been difficult. Of course, one should also bear in mind that there is a limitation to the comparison enabled by this sort of analysis since, as noted above, there are a number of arbitrary choices available to the analysts, and the ranking is in part dependent upon such choices.

Now we are in a position to interpret the results. Factor 1 is positively correlated most of "good" indicators and negatively correlated with most of "bad" indicators. Perhaps, the exceptions are FEMRATIO, F_HHH_RATIO and F_HHH_UD5_RATIO. In particular, the fact that F_HHH_UD5_RATIO is a little bit surprising because it is contrary to what would be expected from the results in SECTION III. Yet, there is no compelling reason to assume that these demographic variables should have a negative sign and the coefficient for F_HHH_UD5_RATIO is small. Therefore, Factor 1 can be considered as the "general welfare score".

Factor 2 requires more careful interpretation. Signs on the asset indices provide us with some insights. The signs on relatively cheap assets such as bicycles and carts are negative, while the signs on relatively expensive assets including motorcycles, televisions and cars are positive. This seems to imply that Factor 2 is correlated negatively with shallow poverty but positively with deep poverty and wealth. While some signs seem counterintuitive, the signs for most of the factors can be understood as such. Hence Factor 2 can be called as the "negative shallow poverty score".

To see if this interpretation makes sense, the correlation between poverty measures and factor scores are correlated. Table 21 provides the results. As one can see, the absolute value of correlation between Factor 1 and poverty severity is relatively high in comparison with Factor 2, whereas Factor 2 is more strongly correlated with poverty rate. This observation is consistent with the above conjecture. The picture becomes clearer when the population is used as weight for the calculation of correlation, though population is not an appropriate weight for some of the indicators used in this analysis.[14]

One should also note that the correlation between the poverty measures and factor scores are only weakly correlated. There are at least four reasons for this. Firstly, the poverty measures represent consumption poverty whereas factor scores are a composite index of arbitrarily chosen indicators, which are supposed to be related to poverty. Secondly, the poverty measures are not

[13] If poverty were one-dimensional, we would not have had to use multiple indicators in the first place.
[14] For example, while the population is an appropriate weight for DEPRATIO, it is not for R_THATCH_RATIO as the number of houses in the commune is the appropriate weight. It is possible to apply different weights for different variables, but the benchmark is no weight as the unit of analysis in this report is communes.

responsive to the rich, whereas factor scores are responsive to the rich.[15] Thirdly, the poverty measures reflect the situation as of 1998 whereas the factor scores as of 2002/03. Fourthly, the poverty measures are estimates and contain statistical errors. Hence, the absolute value of the observed correlation between poverty measures and factor scores is, in expectation, always lower than the *true* value.

Table 21 Correlation between poverty measures and factor scores at the commune level (obs=1460).

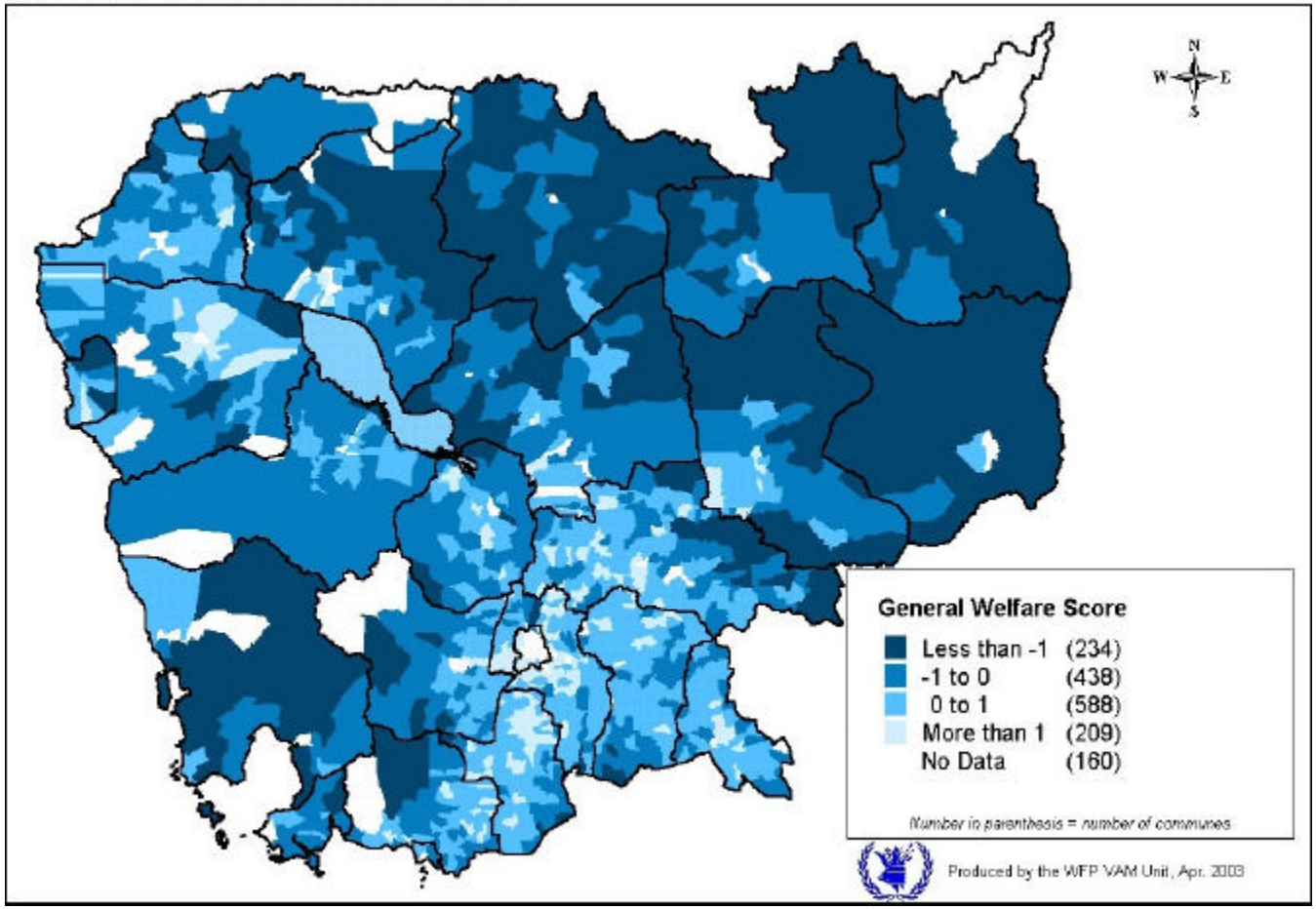
Poverty Measure	Without Weights		With Weights	
	Factor 1	Factor 2	Factor 1	Factor 2
Poverty Rate	-0.0494	-0.2192	-0.2582	-0.2175
Poverty Gap	-0.1247	-0.1001	-0.2774	-0.1047
Poverty Severity	-0.1389	-0.0311	-0.2585	-0.0381

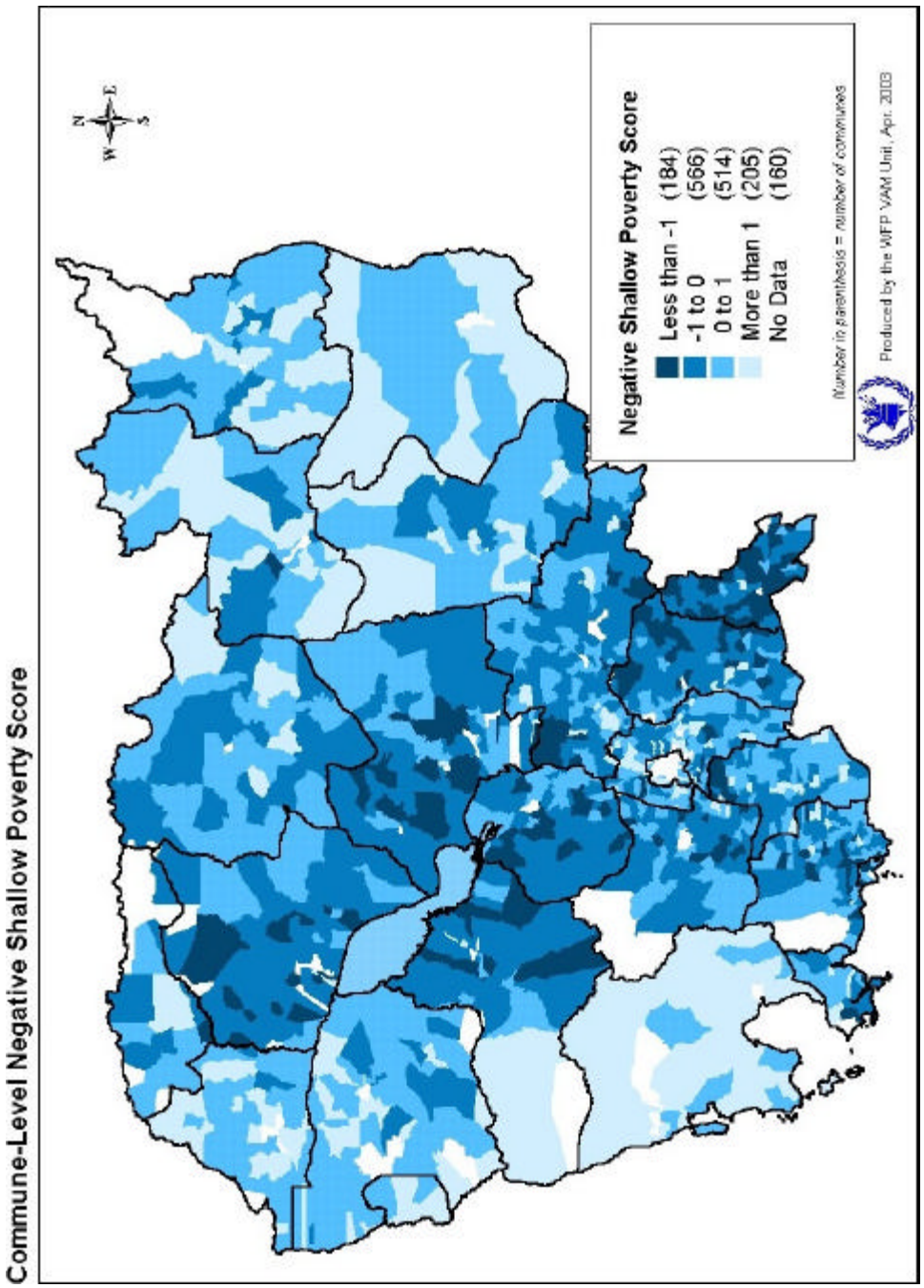
Mapping the Results

Once the factor scores are derived for each commune, it is straightforward to show it on the map. Using a geographic information system, it is possible to map the factor scores. The figures in page 29 and page 30 are the maps of the factor scores for Factor 1 and Factor 2 respectively. The readers are reminded that the area of the commune does not represent the significance of the commune, though larger communes obviously show up more conspicuously in the map. Provided that our interpretation is right, the darker areas in the general welfare score map in page 29 are poorer. The interpretation of the negative shallow poverty score map in page 30 is more complicated. Communes with darker colors have more moderately poor people, whereas communes with lighter colors have either extremely poor people or non-poor people. While the map is intuitively appealing, it is not clear how it should be used, especially in comparison with poverty maps in MoP and WFP. In SECTION V, the implications are discussed with a special focus on the map of general welfare score.

[15] For example, even when the rich people increase consumption, poverty measures do not change. On the other hand, if a rich gets a car in the village, Factor 1 (and Factor 2) increases.

Commune-Level General Welfare Score




Commune-Level Negative Shallow Poverty Score
Negative Shallow Poverty Score
Less than -1 (184)
-1 to 0 (566)
0 to 1 (514)
More than 1 (205)
No Data (160)
(Number in parenthesis = number of communes
Produced by the WFP VAM Unit, Apr. 2003

SECTION V. Discussion and Conclusion

Do the Results Make Sense?

The fact that there exists negative correlation between the general welfare score and poverty measures is reassuring as the poverty measures are expected to be lower when the commune enjoys a higher level of welfare. Of course, poverty measures and general welfare score are different measures and should not be expected to automatically yield the same, or even similar, maps. One should also bear in mind that the poverty situation has changed since 1998.

Since the ground-truthing exercise is not a statistical validation exercise, it is not possible to provide the firm evidence that can prove the validity of poverty maps. However, it is possible to provide a circumstantial evidence. The negative correlation is a first step. But, it is possible to go a step further by taking into consideration the opinions of district chiefs collected in CCDB, which, we argue, provides further circumstantial evidence on the validity of the poverty map.

The difficulty with dealing with the CCDB is, however, it is taken only at the district level, and the commune level ranking is available only within each district. Therefore, unlike the poverty measures and general welfare scores, it cannot be used to make commune level comparisons between districts. Also, it does not contain the only rankings, and thus the data must be treated as an ordinal measure. A natural candidate for the methodology to test whether the ranking from the CCDB is independent of that from the poverty measures or general welfare score is to use the Spearman's rank-order correlation.

However, the subjective judgment of district chiefs cannot be expected to have exactly the same ranking as that obtained from the general welfare score or poverty measures. This is of great relevance when the Spearman's rank-order correlation is used at the district level, as the number of communes in one district is quite small to apply the test of the independence of ranking. Moreover, different district chiefs would have different emphasis on the aspects of poverty they think are important. In order to be able to reject that the null hypothesis that two rankings are independent, the two rankings must be extremely close when the sample size is small (*See* Table 22). Moreover, what is of interest from the viewpoint of the ground-truthing of the nationwide poverty map is not the ranking correlation of each district, but the overall correlation between the different rankings.

Before proceeding, the negative poverty measures are used hereafter to make the ranking comparisons more intuitive. For example, the negative poverty rate for a commune with the poverty rate of 25% is -25%. The reason for doing this is that as greater the value of the poverty rate, the poverty situation is better. The ranking in CCDB and general welfare score are expressed in a similar manner in that the situation is better when the number is greater.

In this paper, an approach is taken that would be fairly weak if applied to each district, but powerful if applied to all the districts. The null hypothesis H_0 is that the rankings from two measures are uncorrelated all over the country, with the alternative being that the rankings from two measures are positively correlated. Under the null hypothesis, the probability that the Spearman's rank-order correlation takes a negative value and a positive value is both 0.5, provided that a positive or negative sign is assigned with a 50 percent chance when the rank-order correlation is zero. Hence, it is possible to assume that the distribution of the signs of the Spearman's rank-order correlation for all the districts follows the binomial distribution with the number of trials equal to the number of district, the number of success equal to the number of positive signs, and the success probability equal to 0.5.

In less technical terms, the situation can be understood as follows: suppose that the ranking are arbitrary and that district chief's ranking, general welfare score and poverty measures are independent of each other, then the sign of rank-order correlation is like flipping a fair coin. If there are sufficiently many positive signs, then it is reasonable to conclude that there exist a positive correlation between the two rankings and the two rankings are not arbitrary rankings.

To carry out the test, the communes for which any of poverty measures, district chief's ranking or general welfare score is not available have been dropped from the analysis. Also, the districts in which the data is available for only one commune were dropped. After these procedures, 167 districts remained in the analysis. In Table 22, the first two columns are the measures used to compare the rankings. The third column is the number of districts for which the Pearson's rank-order correlation is positive, and the fourth the percentage of them out of all the 167 districts. The fifth column is the P-value under H_0. Hence the numbers in the third columns are all significant even at 0.1% level. The last two columns are provided for those who are interested in the test of independence of ranking for each district. The sixth column shows the number of districts for which the null hypothesis of independence was rejected at 5% level and the seventh column its ratio to the number of total districts. While the percentage is relatively small, they are much higher than 5%.

Table 22 Test of rank-order correlation at the district level (obs=167).

Measure 1	Measure 2	Number of Positive Correlations		P-Value	Significantly Positive Correlations at 5%	
Negative Poverty Rate	General Welfare Score	116	69.4%	0.000	40	24.0%
Negative Poverty Rate	District Chief's Ranking	110	65.9%	0.000	26	15.6%
General Welfare Score	District Chief's Ranking	147	88.0%	0.000	73	43.7%

The observations above say that the rankings from negative poverty rate, general welfare score and district chief's ranking are positively correlated. In particular the general welfare score and district chief's ranking seem to be most closely correlated among the three pairwise correlations. This may be because both of these two measures are taken at the same time and because district chief's ranking is likely to be at least in part based upon some easily observable indicators. Also, it may be the case that the statistical errors associated with poverty rate make it difficult to provide a sharp ranking in a district, especially because the poverty rates in the same district tend to be similar. Overall, the results obtained from the three measures are, broadly speaking, consistent and make sense. Yet, this does not by any means imply the poverty maps are infallible. To further evaluate the poverty maps, let us now turn to more specific issues.

Poor Northeast: Truth or Myth?

When the poverty maps were presented before and after the publication of MoP and WFP (2002), a number of people, including people working in the government, NGOs and international organizations, have been consulted to assess how the poverty map seems to reflect the poverty situation in their eyes. While most of the people said that the overall picture seems reasonable, there have been three common remarks about the poverty maps. To rapidly assess these remarks, we have conducted a number of field interviews in various provinces. While these interviews are not methodologically rigorous and cover only limited number of villages, comments from the villagers have provided us with invaluable insights.

The first common remark was that the poverty estimates in Pailin and Western Battambang seemed to be too high for many people. Secondly, many thought the estimates in northern Kampong Speu are too low. Our experience through informal interviews with villagers in these areas also matches these observations. However, there is a good reason why this may be the case. As emphasized in MoP and WFP (2002), these areas are outside the sampling frame of CSES 1997 and the economic system is quite different from other parts of the country.[16] Hence, the estimates in these areas may not accurately reflect the situation as of 1998.

However, there is the third point which remains to be explained. Many people also mentioned that the poverty estimates in northeastern provinces, especially Mondol Kiri and Rotanak Kiri, seemed too low given the current situation. A number of indicators presented in Appendix F and child malnutrition indicators presented in NIS, DGH and ORC Macro (2001) also provide circumstantial evidence that these provinces are poor. Our experience in these provinces also seems to suggest that the estimates are lower than what they should be. **エラー! 参照元が見つかりません。** also shows that Mondol Kiri and Rotanak Kiri are relatively worse off. Katz reportedly said "Mondol Kiri appears as a well-off province. If policymakers look at this and ignore Mondol Kiri, it

[16] A typical example is the fact that Thai Baht is widely used in Pailin and eastern Battambang.

would be a disaster" (McDonald-Gibson, 2003). Hence, it is critical to explore why the estimates appear too low.

There are at least five possible answers for this question. While it is not possible to provide a definitive answer, it is the author's opinion that all of these possibilities have some relevance. Let us now turn to the first possibility, which is that the observation of people, including ours, is wrong. It may sound outrageous, but policymakers should understand what implied by consumption poverty before applying the estimates to policy. It seems clear that the people in northeastern provinces generally have very low cash income. However, cash income is not the same as consumption. In fact, consumption is likely to be very different from income in provinces like Rotanak Kiri and Mondol Kiri. In these provinces, many people go to forest to obtain what they eat, which is counted toward consumption but not toward income.

A number of people have mentioned that northeastern provinces should be one of the poorest provinces. Our experience, however, does not agree with it. In our experience, most of the people had means of production. While they were suffering from food shortage, many, if not most, of them were surprisingly optimistic about their future consumption. This optimism is mainly because of the fact that they can get something to eat from the forest. This optimism contrasts sharply with villagers we interviewed in Prey Veng for whom the situation seems desperate with no means of production after they have been hit by floods and droughts repeatedly.

Table 23. Provincial estimates of mean consumption and poverty rate using CSES 1997 data. Mean consumption is expressed in terms of per day *per capita* in Riels. The number in the bracket is the ranking. The estimates are *not* based on a representative sample at the provincial level.

No	Province	Mean Consumption		Poverty Rate		Obs
1	Banteay Mean Chey	1410.0	(16)	54.7%	(18)	235
2	Battambang	1887.6	(6)	26.2%	(6)	375
3	Kampong Cham	1385.3	(17)	39.5%	(16)	715
4	Kampong Chhnang	1486.4	(13)	37.9%	(15)	165
5	Kampong Speu	1742.5	(9)	26.3%	(7)	205
6	Kampong Thom	1439.9	(15)	35.3%	(13)	270
7	Kampot	1547.3	(12)	36.1%	(14)	240
8	Kandal	1798.8	(7)	28.8%	(10)	505
9	Koh Kong	9153.4	(1)	0.0%	(1)	60
10	Kracheh	1902.2	(5)	28.1%	(9)	145
11	Mondol Kiri	-	-	-	-	0
12	Phnom Penh	4614.3	(2)	11.0%	(3)	1200
13	Preah Vihear	-	-	-	-	0
14	Prey Veng	1148.6	(19)	61.6%	(20)	555
15	Pursat	1265.6	(18)	44.2%	(17)	165
16	Rotanak Kiri	3515.7	(3)	0.0%	(1)	40
17	Siem Reap	1129.0	(20)	60.7%	(19)	305
18	Krong Preah Sihanouk	1772.4	(8)	28.1%	(8)	140
19	Stueng Treng	1706.7	(11)	12.1%	(4)	50
20	Svay Rieng	1715.7	(10)	35.0%	(12)	245
21	Takeo	1472.7	(14)	34.2%	(11)	375
22	Otdar Mean Chey	-	-	-	-	0
23	Krong Keb	2356.4	(4)	23.0%	(5)	20
24	Pailin	-	-	-	-	0
	Cambodia	**1906.1**		**36.1%**		**6010**

A closer look at CSES 1997 data set also supports the argument above. While the CSES 1997 data is not representative at the provincial level, it is possible to see if the poverty map is consistent with what was observed in CSES 1997 data. While four provinces are not covered in CSES 1997, including Mondol Kiri[17], Rotanak Kiri is the least poor province in terms of poverty rate. Stueng Treng and Kracheh have lower poverty rate than the national average. All of the observations above seem to suggest that how people perceive poverty is not necessarily closely related to consumption poverty. CSES 1999 Round 2 gives similar results. In Plateau/Mountain Rural stratum, the poverty

[17] The reason Mondol Kiri was not covered is that its population was too small, but it was in the sampling frame of CSES 1997.

rates were estimated at 18.0 percent (MoP, 2001), which is substantially lower than the national average of 35.9 percent.[18]

In relation to this point, one should not confuse consumption poverty measure with malnutrition measure such as the prevalence of stunting and underweight. The fact that there is high prevalence of stunting and underweight in Rotanakiri and Mondul Kiri (NIS, DGH and ORC Macro, 2001) does not necessarily imply that these areas are consumption poor.[19] Nutritional status of children is affected not only by their food consumption but also by child care, prevailing practices and disease. Stunting and wasting are considered to reflect the cumulative effects and short-term effects of malnutrition respectively, with underweight in between. While consumption measure is considered to be more stable than income, it is essentially a short-term measure and, unlike underweight and stunting, cumulative effects are not captured. Although international evidence suggests that income increases imply similar rates of reduction in malnutrition, there are many exceptions (*See*, Haddad et al., 2002). Hence, it should not be assumed that the spatial pattern of child malnutrition should resemble that of consumption poverty.

The second possible answer to the seemingly low poverty estimates in the Northeast is that, even though the map accurately reflected the situation as of 1998, the massive changes might have taken place so that the poverty situation in these provinces is now very different. While we do not have concrete evidence why this may be the case, a number of respondents in our interview took place in several parts of Rotanak Kiri have told us that the situation has worsened since 1998. One of the reasons they mentioned is that ethnic minorities have been driven to more and more marginalized land, and they have a more difficult time finding food in the forest.

The third possibility is the problem with census data. While there is no evidence to support the claim, it is plausible that some ethnic minorities have been systematically undercounted in the northeastern provinces in the census data. To conduct the census, it is crucial to know the location of people before the census is carried out. However, a number of people in the northeast move frequently. This makes it more difficult to keep track of the location of people. If those not counted in the census happen to be poor people, the poverty estimates are biased downwards.

The fourth possibility is the problem with the CSES 1997 data. The argument in MoP and WFP (2002) take the reliability of CSES 1997 and the ordinary assumptions to compute poverty measures at the stratum level as given. However, if there was systematic overstatement of consumption in these provinces, the poverty estimates are biased downwards. This is an important

[18] Of course, the caveat for using CSES 1999 data applies here. In particular, the estimate is 73.3 percent if only Round 1 is used. While results from Round 2 is often more favored, this may suggest that there is a substantial seasonal variability in consumption.

[19] Note also that the prevalence of wasting is only slightly above the national average in Ratanak Kiri and Mondol Kiri.

point when there is a considerable price differences within the stratum. In northeastern Cambodia, considerably higher prices for essential goods such as rice have been observed[20] and the results are presented in Table 24 though such comparison may be questioned on the basis of the representativeness of the market prices.

Table 24 Price comparisons for selected items between four northeastern provinces and other provinces. Prices for all items but glutamate/MSG are higher for the northeastern provinces.

	Northeast		Other Provinces		Cambodia	
Item	Mean	Obs	Mean	Obs	Mean	Obs
Ordinary Paddy	733	3	408	173	413	176
Rice	1000	2	748	244	750	246
Broken Rice	685	8	579	265	582	273
Salt	771	8	568	352	573	360
Fish sauce	1233	9	1215	299	1215	308
Glutamate/MSG	4625	4	4814	281	4812	285
Sugar (Refined)	1718	11	1603	332	1607	343

Finally, the fifth possibility is that the poverty estimates themselves have been biased downwards because of the inappropriateness of the model. It is possible that there were too few models for the poverty maps. Because the CSES 1997 data was representative only at the stratum level, or the level of Phnom Penh, Other Urban and Rural, the consumption regression models were created at this level. This implicitly assumes that the structure of correlation between logarithmic consumption and other socioeconomic variables is the same throughout each stratum. However, Rural stratum is in particular heterogeneous. The pattern of consumption in Rotanakiri may not be the same as that in Kampot.

The consequence of such an assumption can be better understood with Figure 1. Let us consider that there are two areas, AREA 1 and AREA 2, and average years of education are known for all the communes in these two areas. Also assume that data on poverty rates are available for randomly selected communes in these areas. Suppose our goal here is to find the relationship between poverty and average years of education, and impute the poverty rates for all the communes in AREA 1 and AREA 2 using the average years of education. If we have enough observations so that the observations are representative in each area, we can run regression for AREA 1 and AREA 2 separately and obtain two dotted lines. However, if we need to pool the data, then the resulting regression line is the bold line. When this happens, the estimates of poverty for AREA 1 are

[20] The average of all the (non-missing) observations for each item in the village was computed. Then the average was taken for each item over the villages without weights. For this calculation, data from both urban and rural villages were used as there were no observations for many items in rural northeastern villages, and people often go to the urban areas to buy goods.

downward biased. In AREA 2, they are upward biased when the average years of education is low and downward biased when the average years of education is high.

It should be reminded that the above example is hypothetical and may not necessarily apply to the poverty mapping exercise carried out in MoP and WFP (2002). Indeed, the fact that provincial level estimate of poverty ranking match quite well with that from poverty mapping (*cf* MoP and WFP, 2002: p33) suggests that Figure 1 is not very likely to be the case. Still, we can not statistically reject this possibility and policymakers should bear in mind this may be the case.

The five possibilities mentioned above seem to provide reasonable accounts at least partly on the seeming discrepancy between the poverty estimates and the perception of poverty by many, including ourselves. However, since the evidence at hand is too limited, it is difficult to derive meaningful policy implications from the discussion above. Hence our next step is to utilize fully the results from the previous sections.

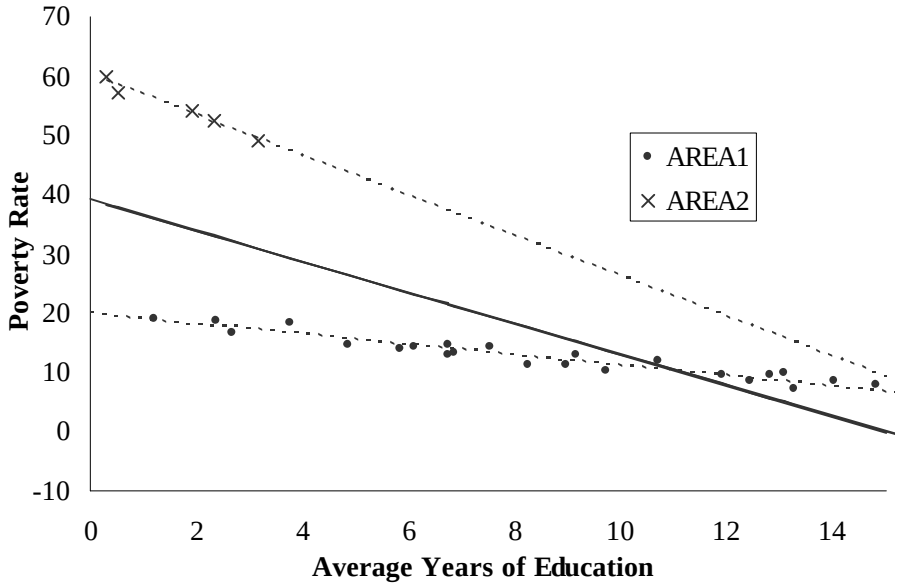


Figure 1. Hypothetical example of biased regression. The dotted lines represent the "true" regression lines for AREA 1 and AREA 2, and the bold line is the "pooled" regression line.

General welfare score and poverty estimates in the northeast

As we have shown in the previous section, the general welfare scores are weakly and negatively correlated with poverty estimates. However, this is an argument at the national level, and may not apply to some parts of the country. To probe into the problem of northeastern regions, it is useful to look at the scatter plot of the general welfare scores and poverty rates. As is clear from Figure 2, the

overall correlation is negative. However, when we only look at the Plateau/Mountain regions[21], the opposite correlation is observed. Table 25 elucidates this point. Poverty measures are negatively correlated with the general welfare score (Factor 1) in all but Plateau/Mountain regions.

To further dissect the problem, Figure 2 was further disaggregated into provinces. Figure 3 is the same scatter plot as Figure 2, but colored by provinces and focused on the Plateau/Mountain region. As shown in Table 26, the correlation between the general factor score and the poverty rate takes a positive value in Kracheh, Mondol Kiri, Rotanak Kiri and Stueng Treng, while it takes a negative value in other provinces. This seems to suggest that the four northeastern provinces are exceptional.

[21] The definitions of ecozones are provided in Appendix C.

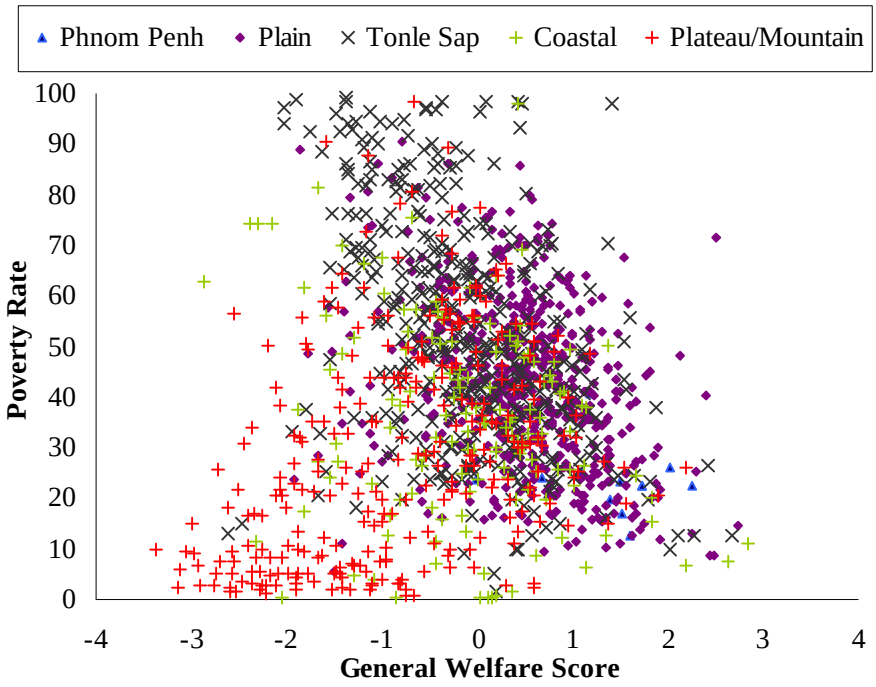


Figure 2 General welfare score versus poverty rate scatter plot. Each dot represents a commune.

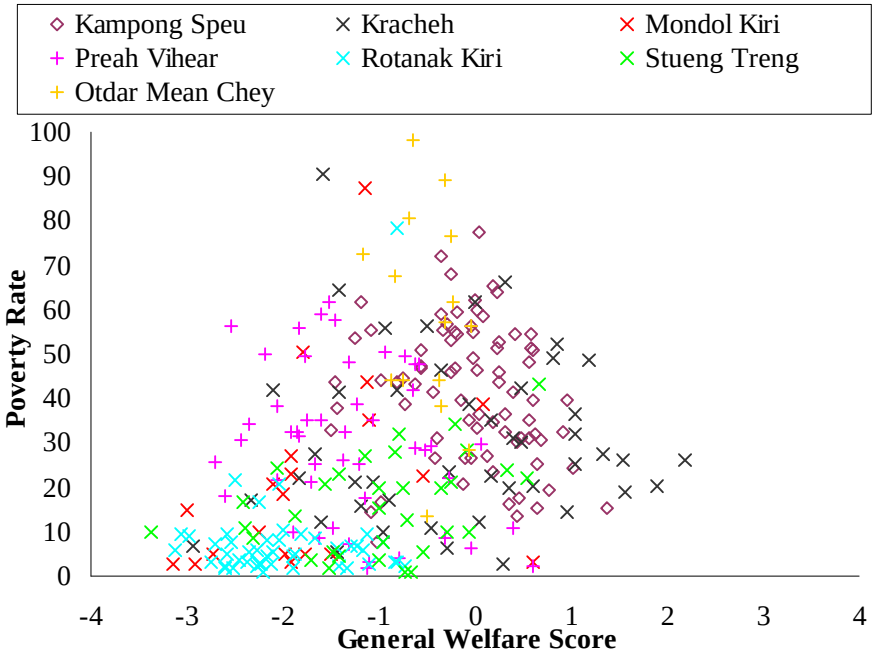


Figure 3 General welfare score versus poverty rate scatter plot in the Plateau/Mountain region. Each dot represents a commune.

Table 25 Correlation between the poverty measures and factor scores by ecozone (obs=1460).

Ecozone	Poverty Measure	Factor 1	Factor 2
Phnom Penh (obs=13)	Poverty Rate	-0.1750	-0.0432
	Poverty Gap	-0.0951	-0.0070
	Poverty Severity	-0.0276	0.0097
Plain (obs=591)	Poverty Rate	-0.3403	-0.2178
	Poverty Gap	-0.3074	-0.1464
	Poverty Severity	-0.2727	-0.1042
Tonle Sap (obs=425)	Poverty Rate	-0.4341	-0.1000
	Poverty Gap	-0.3779	-0.0198
	Poverty Severity	-0.3249	0.0837
Coastal (obs=143)	Poverty Rate	-0.2725	-0.0585
	Poverty Gap	-0.3236	0.0957
	Poverty Severity	-0.3009	0.1541
Plateau/Mountain (obs=288)	Poverty Rate	0.3471	-0.2359
	Poverty Gap	0.2628	-0.1926
	Poverty Severity	0.2062	-0.1540

Table 26 Correlation between the poverty measures and general welfare scores (Factor 1) by province in Plateau/Mountain.

Poverty Measure	Poverty Rate	Poverty Gap	Poverty Severity	obs
Kampong Speu	-0.1862	-0.1946	-0.1937	80
Kracheh	0.0247	0.0014	-0.0056	46
Mondel Kiri	0.3313	0.2696	0.2361	20
Preah Vihear	-0.2846	-0.2911	-0.2948	48
Rotanak Kiri	0.1976	0.2411	0.2536	46
Stueng Treng	0.3925	0.4137	0.4074	33
Otdar Mean Chey	-0.2060	-0.2698	-0.2724	15

Table 27 Test of rank-order correlation at the district level in Kracheh, Mondol Kiri, Rotanak Kiri and Stueng Treng (obs=23).

Measure 1	Measure 2	Number of Positive Correlations		P-Value	Significantly Positive Correlations at 5%	
Negative Poverty Rate	General Welfare Score	8	34.8%	0.895	2	8.7%
Negative Poverty Rate	District Chief's Ranking	10	43.5%	0.661	2	8.7%
General Welfare Score	District Chief's Ranking	19	82.6%	0.000	7	30.4%

To verify this, the null hypothesis that the rankings from two measures are independent was tested with only the observations from these four provinces. As the results in Table 27 show, District Chief's ranking and the ranking from the general welfare score are still positively correlated, but in these four provinces, the null hypothesis that the rankings from the poverty rate and general welfare score were not rejected. Similarly, the null hypothesis that the district chief's ranking and the ranking from the poverty rate are independent was not rejected. Hence, the ranking obtained from the poverty rate is indeed irregular, and this also suggests that the poverty estimates in these four provinces should be treated with greater caution.

Let us now consider the policy implications. A policy question is whether one should carry out interventions in these provinces. Whether or not consumption poverty is low, the fact that the education indicators and nutrition indicators is very bad in these provinces alone can be a rationale for social intervention. In particular, the lack of the human capital in these provinces is likely to have significant intergenerational consequences. Hence, to prevent these provinces left out of development, measures should be taken to build human capital in these provinces.

In relation to poverty eradication, one may need more careful thoughts. What is recommendable from what is known is to conduct an independent survey focusing on the particular aspect of poverty, possibly including consumption, in these provinces. If one believes that the consumption estimates are accurate, then the question is how to reconcile the rankings. One possible interpretation is that the multidimensionality of poverty is very complex in these four provinces so that the ranking from consumption poverty is very different from that from other indicators. Then, one should choose the ranking that seems to reflect most the aspects of poverty one would like to deal with.

Conclusion

The poverty map is consistent with what was observed in CSES 1997, CDB5 and CCDB, and this provides a good circumstantial evidence of the validity of the poverty map. However, when the rankings from the poverty rate and general welfare and district chief's ranking are compared, the ranking from the poverty rate was least closely related with the other two. We have argued that the

statistical errors of poverty estimates and the timing of the data collection, *inter alia*, can account for this. The question arises how these different rankings should be employed.

As discussed in MoP and WFP (2002), when a number of communes are targeted, the statistical errors are not likely to be a big issue. However, a small number of communes are to be targeted or ranked using the poverty map, the statistical error is more likely to be a problem. Hence, while it is admittedly arbitrary, one possibility is to use the poverty map for the selection of a relatively large number of communes, which may be a number actually larger than the final selection of communes, and modify it with other data sets. In particular, when the rankings from the general welfare score and district chief's ranking match, the targeting may be carried out with more confidence.

There are three important lessons to be learned from this ground-truthing exercise. Firstly, the poverty maps are found to provide reasonable and useful estimates indeed. Secondly, while the meaning of welfare scores is not as clear as consumption, it can provide some useful results. One should note that it was not possible to verify whether northeastern provinces are exceptional without using welfare scores. Thirdly, poverty rankings can vary depending on the timing and indicator. Therefore, while poverty maps in MoP and WFP (2002) provide a useful basis for targeting, one should also try to combine more up-to-date information and indicators of poverty that reflect the aspect of poverty one would like to deal with. One possible way of doing this was suggested in the previous paragraph.

To conclude, the analysis of CDB5 has provided useful insights into the current poverty situation in a cost effective manner. The CCDB data has helped to compare the reliability of the poverty ranking for the current situation obtained from the poverty measures and general welfare measure. While the interpretation of general welfare measure is not as straightforward as consumption, and there are a number of sources of arbitrariness in the analysis, it is still possible to assess the poverty situation. Even though factor analysis is not the best available way of assessing poverty situation at the commune level, given that the resources available for the data collection required for targeting are also limited, the power of cost-effective data set should not be underestimated.

Appendix A. References

CSD (2002) *National Poverty Reduction Strategy 2003-2005: Executive Summary*. December, 2002. Council for Social Development: Phnom Penh, Cambodia.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003) "Micro-level Estimation of Poverty and Inequality", *Econometrica* 71(1): 355-364.

Fujii, T. and Ear, S. (2002) "How does Spousal Education Matter?: Some Evidence from Cambodia," *Asian Development Review* 19(1):117-138.

Haddad, L., Alderman, H., Appleton, S., Song, L. and Yohannes, Y. (2002) "Reducing Child Undernutrition: How Far Does Income Growth Takes Us?," *FCND Discussion Paper* No. 137. Food Consumption and Nutrition Division, International Food Policy Research Institute.

McDonald-Gibson, C. (2003) "Warning Over Maps as Guide to Poverty Reduction," *Phnom Penh Post* 12(8), April 11-24, 2003.

MoP (1998) "A Poverty Profile of Cambodia – 1997," Ministry of Planning, Cambodia.

MoP (2001) "A Poverty Profile of Cambodia – 1999," Ministry of Planning, Cambodia.

MoP and WFP (2002) *Estimation of Poverty Rates at Commune-Level in Cambodia: Using the Small-Area Estimation Technique to Obtain Reliable Estimates.* October, 2002. Ministry of Planning and World Food Programme: Phnom Penh, Cambodia.

NIS (1998) "Report on the Cambodia Socio-Economic Survey 1997", National Institute of Statistics, Ministry of Planning, Cambodia.

Sahn, D.E. and Stifel, D.C. (2000) "Poverty Comparisons Over Time and Across Countries in Africa," *World Development* 28(12): 2133-2155.

Seila Programme (2001) *Commune Database User Manual*. June 2001. Version 1.1. Seila Programme: Phnom Penh, Cambodia.

UNOPS and UNDP (2002) *Final Report: Resourcing Secondary Data and Enhancing the Commune Database*. Rural Investment and Local Governance (CMB/01/R72/RILG). Consultancy Number 3. UNOPS and UNDP: Kuala Lumpur, Malaysia and Phnom Penh, Cambodia.

WFP (2002a) *WFP Cambodia Strategic Directions 2003-2005*. December, 2002. World Food Programme: Phnom Penh, Cambodia.

WFP (2002b) *Consolidated Framework of WFP Policies: An Updated Version.* September, 2002. WFP/EB.3/2002/INF/9: Rome, Italy.

Appendix B. List of Questions Available in the Seila CDB5.

The following table provides the list of the questions available in the Seila CDB5. The first column is the question identification number. The second column is the short name for the question and also used to refer to the variable. The third column provides the question asked to the enumerator. The fourth column is the level at which the data was taken. If this column is V, the data was taken at the village level, whereas C means commune level variable. The fifth column tells whether the data was taken only in the Urban Areas only, denoted by U, in the Rural Areas only, denoted by R or both, denoted by B. The short names given in the table are different from the original form, as the original form was confusing and inconsistent in the naming rule.

QID	Short Name	Question	V/C	R/U
1	FAMILY	Total number of families	V	B
2	FEM_TOT	Total number of females	V	B
3	MAL_TOT	Total number of males	V	B
4	FEM_0_5	# Girls 0-5 years old (under 6's)	V	B
5	MAL_0_5	# Boys 0-5 years old (under 6's)	V	B
6	FEM_6_14	# Girls 6 to 14 years old	V	B
7	MAL_6_14	# Boys 6 to 14 years old	V	B
8	FSCH_6_14	# Girls 6 to 14 who go to school	V	B
9	MSCH_6_14	# Boys 6 to 14 who go to school	V	B
10	FEM_15_17	# Women 15 to17 years old	V	B
11	MAL_15_17	# Men 15 to17 years old	V	B
12	FEM_18_64	# Women 18 to 64 years old	V	B
13	MAL_18_64	# Men 18 to 64 years old	V	B
14	FEM_65OV	# Women over 65 years of age	V	B
15	MAL_65OV	# Men over 65 years of age	V	B
16	FILT_15OV	# Illiterate women over 15 years old	V	B
17	MILT_15OV	# Illiterate men over 15 years old	V	B
18	R_THATCH	# Houses with thatched roof	V	B
19	R_TILE	# Houses with tiled roof	V	B
20	R_FIBRO	# Houses with fibro roof	V	B
21	R_ZINC	# Houses with zinc roof	V	B
22	R_CONC	# Houses with concrete roof	V	B
23	TOILET	# Latrines	V	R
24	FAM_H2O_HOME	# Families with piped water, private pump well or private ring well, usable year round, at their house, less then 150m.	V	R
25	FAM_H2O_150M	# Families with a communal tap, pump well or ring well, usable year round, within 150m of their house.	V	R
26	FAM_H2O_OTHE	Most common source of water for other families: pond, river, rain water, other.	V	R
27	FAM_COW	# families with cattle and buffalo in village	V	R
28	FAM_PIG	# families with pigs in the village	V	R
29	RICE_PRICE	Average farm gate price of paddy in Riel for this month, December	V	R
30	N_MOTO	# Motorcycles	V	R

QID	Short Name	Question Name	V/C	R/U
31	N_CAR	# Tractors/koyons/cars	V	R
32	N_OXCART	# Horse carts and ox carts	V	R
33	N_BICYCLE	# Bicycles	V	R
34	N_ROWBOAT	# Row boats	V	R
35	N_MOTORBOAT	# Boats with motor	V	R
36	N_TV	# TVs	V	R
37	FAM_TBA	# families who used a traditional birth attendant in the past year	V	R
38	FAM_MIDWIFE	# families who used a trained midwife in the past year	V	R
39	N_TBA	# traditional birth attendants in the village	V	R
40	N_MIDWIFE	# government trained midwives in the village	V	R
41	FAM_IRRIGATION	# family who have some irrigated rice land	V	R
42	FAM_FERTILIZER	# family using chemical fertilizer in the past year	V	R
43	FAM_PESTICIDE	# family using pesticide in the past year	V	R
44	N_INSECURITY	# murder, robbery, theft cases in the past year	V	B
45	N_LANDCONFLICT	# land conflict case in the past year	V	B
46	F_HHH_UD5	# female headed household/families, where the head is a mother with one or more children with under 5 yrs old	V	B
47	FAM_VIOLENCE	# families having problems with violence in home	V	B
48	LOC_MARKET	Name of the nearest market villager frequently go to buy goods	V	R
49	MIN_MARKET	Time taken to get from village to this nearest market by motor or motorboat	V	R
50	KM_ROAD	Distance in Km to nearest year-round road (4 wheel motor vehicles)	V	R
51	MIN_ROAD	Time taken to get from village to nearest year-road by motor or motorboat	V	R
52	H_GARBAGE	# Houses which have access to garbage collection by a garbage collector?	V	U
53	H_PIPEWATER	# Houses which have access to piped water	V	U
54	H_ELECTRICITY	# Houses with electricity	V	U
55	FAM_TOILET	# Families with latrines	V	U
56	FAM_MOTO	# families with motorcycles	V	U
57	FAM_CAR	# families with tractors/koyons/cars	V	U
58	FAM_BICYCLE	# families with bicycles	V	U
59	FAM_TV	# families with TVs	V	U
60	N_TRAFFICK	# trafficking cases reported in the past year	V	U
61	PRI_CLSRM	# primary school classrooms in the commune	C	B
62	SEC_CLSRM	# secondary school classrooms in the commune	C	B
63	PRI_TCH	# primary school teachers in the commune	C	B
64	SEC_TCH	# secondary school teachers in the commune	C	B
65	WET_RICELAND	Area wet season rain fed rice land in Ha	C	R
66	WET_IRRI	Area wet season supplemental irrigated rice land in Ha	C	R
67	WET_PROD	Rice production in wet season, MT	C	R
68	DRY_IRRIGATION	Area of full-irrigated dry season rice land in Ha	C	R
69	DRY_RECESSION	Area of recession dry season rice land in Ha	C	R
70	DRY_PROD	Rice production in dry season, MT	C	R

QID	Short Name	Question Name	V/C	R/U
71	KIN_CLSRM	# kindergarten classrooms	C	U
72	KIN_TCH	# kindergarten teachers in the commune	C	U
73	F_HHH	# Female household headed	V	B
74	MILT_15_17	# illiterate men from 15-17ys	V	B
75	MILT_18_64	# illiterate men from 18-64ys	V	B
76	MILT_65OV	# Illiterate men over 65 years old	V	B
77	MSCH_15_17	# boy 15 to 17ys who go to school	V	B
78	MSCH_0_17	# boy 6 to 17ys who go to school	V	B
79	FILT_15_17	# illiterate women from 15-17ys	V	B
80	FILT_18_64	# illiterate women from 18-64ys	V	B
81	FILT_65OV	# Illiterate women over 65 years old	V	B
82	FSCH_15_17	# girl 15 to 17ys who go to school	V	B
83	FSCH_0_17	# girl 6 to 17ys who go to school	V	B
84	FAM_THATCH	# family living in thatched roof	V	R
85	FAM_TILE	# family living in tiled roof	V	R
86	FAM_FIBRO	# family living in fibro-cement roof	V	R
87	FAM_ZINC	# family living in zinced roof	V	R
88	FAM_CONC	# family living in concrete roof	V	R
89	LAT_THATCH	# Latrine in total thatch house	V	R
90	LAT_TILE	# Latrine in total tiled house	V	R
91	LAT_FIBRO	# Latrine in total fibro-cement house	V	R
92	LAT_ZINC	# Latrine in total zinc house	V	R
93	LAT_CONC	# Latrine in total concrete house	V	R
94	TV_THATCH	# TVs in total thatch house	V	R
95	TV_TILE	# TVs in total tiled house	V	R
96	TV_FIBRO	# TVs in total fibro-cement house	V	R
97	TV_ZINC	# TVs in total zinc house	V	R
98	TV_CONC	# TVs in total concrete house	V	R
99	BABY_BORN	# Women deliver baby in village	V	R
100	PRI_CLS	# Primary classes in commune	C	B
101	SEC_CLS	# Secondary classes in commune	C	B
102	KIN_CLS	# Kindergarten classes in commune	C	U
103	RICE_AREA	Rice land area in commune	C	R
104	MSCH_0_5	# boy 0 to 5ys who go to school	V	U
105	FSCH_0_5	# girl 0 to 5ys who go to school	V	U

Appendix C. Definition of Ecozones

In this study, the definition of ecozones in MoP (2001) was employed and it is shown in . The only modification made in this study is that Pailin is now classified as Tonle Sap zone instead of Plateau/Mountain zone as Pailin was a part of Battambang before and does seem to have closer socio-economic system to Battambang province than any other provinces in Plateau/Mountain zone. In any case, the main findings of this paper will not be changed by this.

Table 28 Definition of ecozones.

Ecozone	**Province**
Phonm Penh	Phnom Penh
Plain	Kampon Cham, Kandal, Prey Veng, Svay Rieng, Takeo
Tonle Sap	Banteay Mean Chey, Battambang, Kampong Chhnang, Kampong Thom, Pursat, Siem Reap, Pailin
Coastal	Kampot, Koh Kong, Krong Preah Sihanouk, Krong Keb
Plateau/Mountain	Kampong Speu, Kracheh, Mondol Kiri, Preah Vihear, Rotanak Kiri, Stueng Treng, Otdar Mean Chey,

Appendix D. List of Additional Indicators

The following indicators are also used in the analysis.

a) *Distance to the main road*	(KM_ROAD)
The average distance to the main road weighted by the population.	
b) *Time taken to the main road*	(MIN_ROAD)
The average time taken to the main road weighted by the population.	
c) *Latrine zinc roof house ratio*	(LAT_ZINC_RATIO)
Average number of latrines in houses with zinc roof. If there is no zinc roof, it is set at zero.	
d) *Television thatched roof house ratio*	(TV_THATCH_RATIO)
Average number of televisions in thatched roof house ratio.	
e) *Ratio of houses with a tile roof*	(R_TILE_RATIO)
The number of houses with a tile roof over the total number of houses	
f) *Ratio of houses with a zinc roof*	(R_ZINC_RATIO)
The number of houses with a zinc roof over the total number of houses	
g) *Number of traditional birth attendant per capital*	(PC_TBA)
The number of traditional birth attendant per capita	

Variable	Obs	Mean	S.D.	Min	Max
KM_ROAD	1471	6.02	14.31	0.00	195.63
MIN_ROAD	1471	31.97	78.16	0.00	1440.00
LAT_ZINC_RATIO	1471	0.0916	0.1377	0.0000	0.9564
TV_THATCH_RATIO	1471	0.1395	0.1248	0.0000	0.9336
R_TILE_RATIO	1471	0.2470	0.2022	0.0000	0.8769
R_ZINC_RATIO	1471	0.2345	0.1677	0.0000	0.8420
PC_TBA	1471	0.0023	0.0030	0.0000	0.0290

Appendix E. An Overview of Factor Analysis[22]

What is the factor analysis used for?

The idea of factor analysis is to explore the "hidden" common factors that contribute to at least two observable variables. Before explaining the procedure of factor analysis, it is instructive to consider a situation where the factor analysis is typically applied. Suppose the teacher wants to measure the student's scholastic abilities. For that purpose, he has the grades of algebra, geometry, physics, classics and history. The first problem he encountered is that each grade was expressed in different formats as classics and history used letter grades, algebra and geometry numbers 0 to 100 and physics 0 to 10. He decided to translate letter grades into numbers and then standardized it so that the grade of each subject has the mean 0 and standard deviation 1. He observed each pair of the five normalized grades has a positive correlation. Also, algebra, geometry and physics are strongly positively correlated, and so are classics and history. He, therefore, thought that there should be two important common factors hidden in the data, which explain well the performance of each student, even though there are some subject-specific abilities. He called them the general academic ability and scientific tendency respectively. In the terminology of factor analysis, the proportion of the variance of each subject explained by the two common factors is communality and the rest is called the uniqueness. The communality may be different from subject to subject. Factor analysis provides a way to measure such common factors by what is called a factor score.

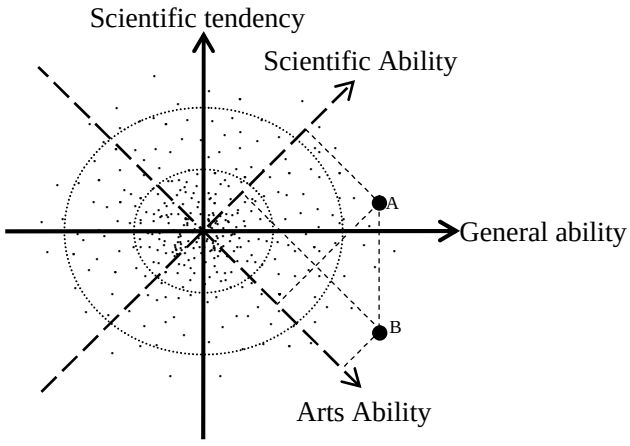


Figure 4 An example of the distribution of factor scores.

[22] This appendix is provided for those who would like to intuitively understand the gist of the analysis carried out in the main text.

Using the factor analysis, he can derive the factor scores for the general ability and scientific tendency for each student, which are plotted in Figure 4. The horizontal axis measures each student's general ability and the vertical each student's scientific tendency. The two concentric dotted circles provide information on how far each student is away from the average. Now let us consider two persons A and B on the graph. A and B are excellent students as their general ability is exceptionally high. But their "types" are different as A is good at science, whereas B's factor score for general ability originates mainly from arts subjects. He can also observe that A is more "balanced" student than B because A is closer to the horizontal axis.

Considering the meaning of the general ability and scientific tendency, he noticed that this combination may be expressed as "scientific ability" and "arts ability" instead. By rotating the coordinates, it is indeed possible to do so. The dashed coordinates provides a way to present the same information in different ways. For example, A is much better than B in scientific abilities as A is "more northeastern" than B. In fact, A has an excellent scientific ability as it is outside the larger circle. On the other hand, B's scientific ability is just slightly better than the average as it is within the smaller circle. Similarly, A's arts ability is quite good, but B's arts ability is much better. Rotation like this can provide a way to express the same information in a different way, but whether this is useful depends on whether it makes interpretation of results easier, and rotation is not always necessary.

The example above describes the typical situation where the factor analysis is used. Factor analysis has a very wide application such as literature, marketing, sociology, economic, education and psychology. It is particularly useful when the underlying common factor is not directly measurable, but a number of relevant variables can be observed. Poverty analysis has a similar situation. While consumption and income are most commonly used as it has more strong theoretical foundations, the burden of data collection can be substantial. Consumption data often contains a long list of items and accurate income data is difficult to obtain when a substantial proportion of the population is a non wage earner, or when a large informal sector exists. Besides, people tend to understate their income. Measures used in this report depend upon more easily observable indicators.[23]

Procedures of the factor analysis

There are several ways to estimate the common factors, such as maximum likelihood method, unweighted least squares method and principal component method. In this paper, the principal component method was employed as it is one of the most commonly used methods and can be implemented easily. Principal component method literally uses the techniques of the principal

[23] While this is the case in Cambodia, it may not be so in other countries where there is no village leaders who can collect information with ease.

component analysis to extract the common factors. Now let us turn to the actual procedure of the factor analysis.

The statistical factor model assumes that the observation can be decomposed into the unobservable common factors and unobservable unique factors. Let us assume that there are p observable variables, $q(\leq p)$ unobservable common factors and n observations. Then, in a matrix form, the statistical factor model can be written as follows:

$$Y = XB + E,$$

where $Y = (y_{ij})$, $X = (x_{ik})$, $B = (b_{kj})$ and $E = (e_{ij})$ are an $n \times p$, $n \times q$, $q \times p$ and $n \times p$ matrix respectively. The subscripts $i(\in [1, n])$, $j(\in [1, p])$ and $k(\in [1, q])$ are the indices for the identification number of observation, observable variable and unobservable common respectively. X is called the factor score matrix, and B^T the factor loading matrix. E is the matrix of unique factors. Hereafter, it is assumed that each observable variable is standardized to have a mean zero and a unit standard deviation, and thus $V \equiv YY^T$ represents the correlation matrix of observed variables.

In the above statistical factor model, the two critical assumptions are that the unique factors are uncorrelated with each other and that the unique factors are uncorrelated with the common factors. The first assumption implies that a $1 \times p$ vector $(e_{i1}, e_{i2}, \ldots, e_{ip})^T$ has the covariance matrix of $p \times p$ diagonal matrix $D = diag\{d_1^2, d_2^2, \ldots, d_p^2\}$, in which each diagonal element is the variance of the unique factor. It should be noted that X, B and E must be estimated and hence there are an infinite number of solutions.

In the principal component method, the initial estimate, also called the prior, Δ of D is set at $p \times p$ zero matrix, and the first q principal components are extracted as the common factor. However, sometimes other initial estimates are used. Another commonly used initial estimate Δ of D is the squared multiple correlations. That is, each of the j-th observable variable is regressed on all the other observable variables and one minus the R-squared is used for the estimate of d_j^2, and this specification is sometimes called principal factor method. Once Δ is obtained, the eigenvalues $l_1 \geq \cdots \geq l_p$ and the corresponding eigenvectors[24] $\mathbf{a}_1, \cdots, \mathbf{a}_p$ for $W \equiv (V - \Delta)$ are calculated. To derive the factor loading matrix, which expresses the correlation of each factor with the observable variables, is computed as $\hat{B} = [\sqrt{l_1}\mathbf{a}_1, \cdots, \sqrt{l_q}\mathbf{a}_q]$. The diagonal elements of W provides an estimate of d_j^2. To derive factor scores, the regression method was used and its formula is given by $\hat{X} = Y\hat{B}^T(\hat{B}\hat{B}^T)^{-1}$.

[24] For the matrix Δ, a nonzero vector $\mathbf{a}$ is called an eigenvector of Δ when it satisfies $\Delta\mathbf{a} = l\mathbf{a}$ for some scalar l. l is called an eigenvalue of Δ.

Appendix F. Maps for Selected Indicators

In this appendix, the maps for the indicators discussed in Section SECTION III are presented. All the maps presented here are based on Seila CDB5 and accompanied by brief observations. The map should only be used to grasp the overall pattern of each indicator and it is not appropriate to try to pick up a few communes based on these maps. It should be reiterated that Seila CDB is collected through village leaders' assessment and thus, despite the efforts to maintain the quality of data, there are likely to be some communes, for which the error is substantially high. One should bear in mind that this is the downside of the cost-effective dataset like Seila CDB.

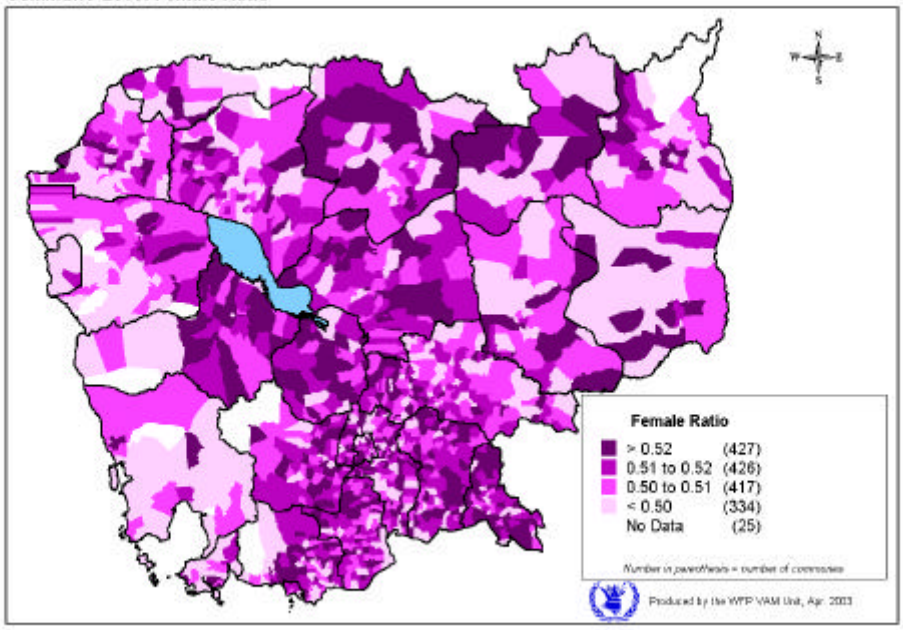
Having said these, the maps can provide us with some useful insights. The fact that it is possible to create commune-level maps for a number of indicators is of great advantage. While one cannot tell exactly why people are poor in different parts of the country, the combination of different indicators can at least suggest the underlying mechanism of poverty. One should also note that some of the indicators are available only in rural areas and this is the reason some maps have white areas in the center of Phnom Penh. The decision on the break points for categorization is arguably arbitrary, but based on the roughly equal count in each category.

Demographic Indicators

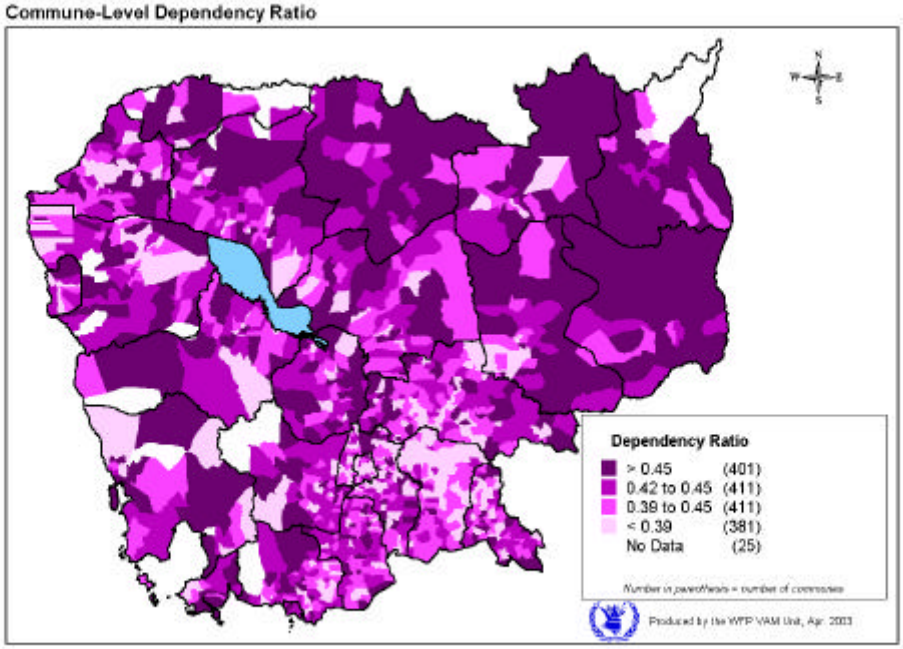
The maps for FEMRATIO, DEPRATIO, BRATE, FAMSIZE, F_HHH_RATIO and F_HHH_UD5_RATIO are presented. The following is the map for FEMRATIO. While the geographical variation of FEMRATIO seems relatively small, there are some parts of country, most notably some parts in Kampong Chhnang, Prey Veng and Svay Rieng, where FEMRATIO is quite high. This may be because of the fact that a sizable part of male labor force go to Phnom Penh for work..

Not surprisingly, the maps for DEPRATIO and BRATE have some resemblance. The rate of birth is particularly high for the northeastern part of the country. This may be one of the reasons for child malnutrition there. The map of FAMSIZE, however, seems to have quite different pattern from DEPRATIO and BRATE. F_HHH_RATIO and F_HHH_UD5_RATIO have some similarities, and they are geographically scattered.

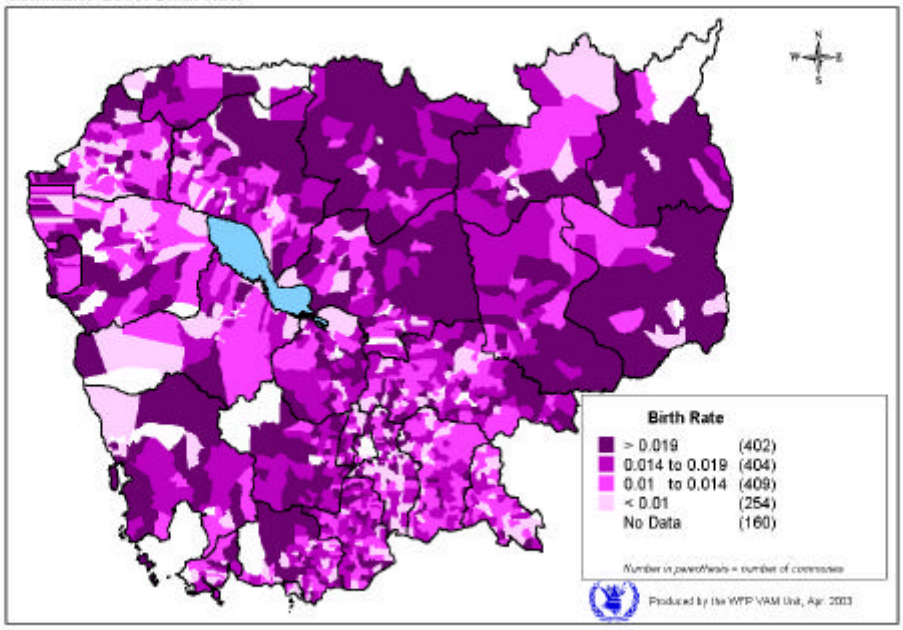
Commune-Level Female Ratio



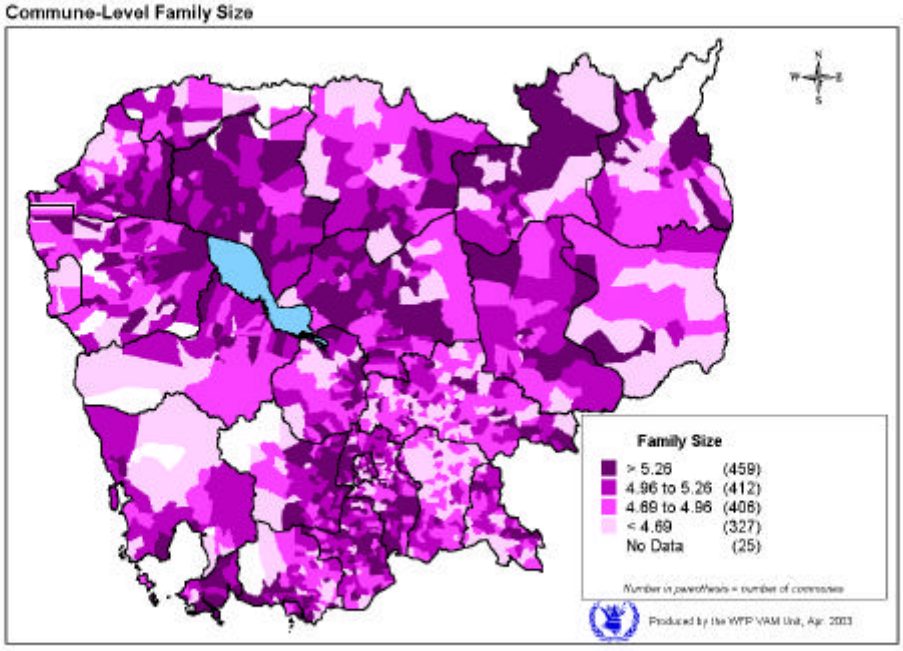
Commune-Level Dependency Ratio



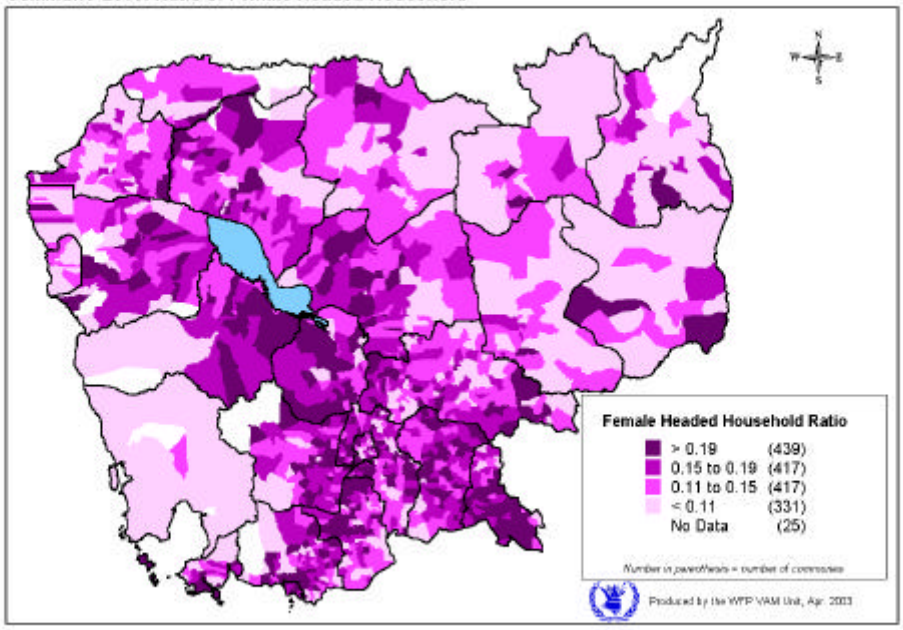
Commune-Level Birth Rate



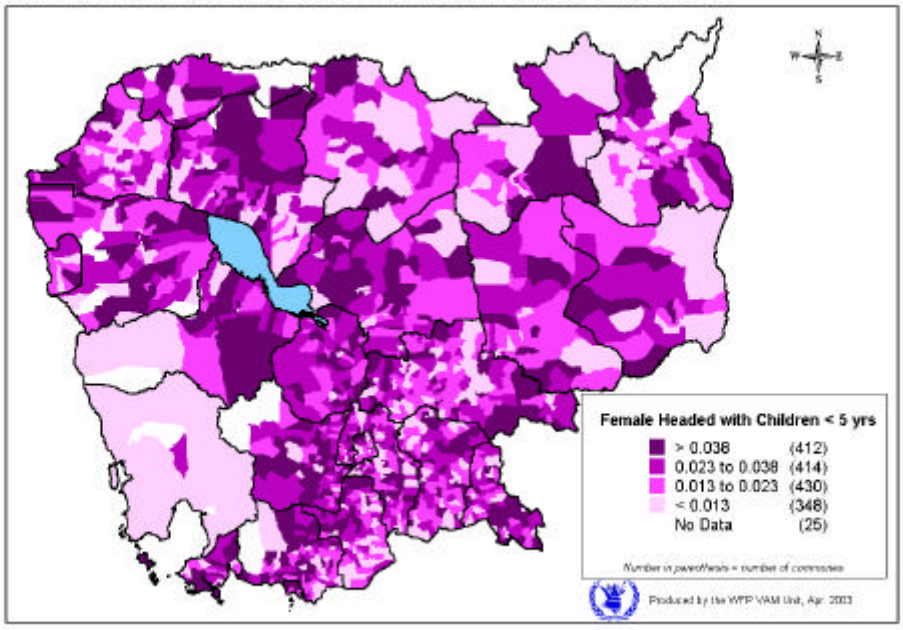
Commune-Level Family Size



Commune-Level Ratio of Female Headed Household



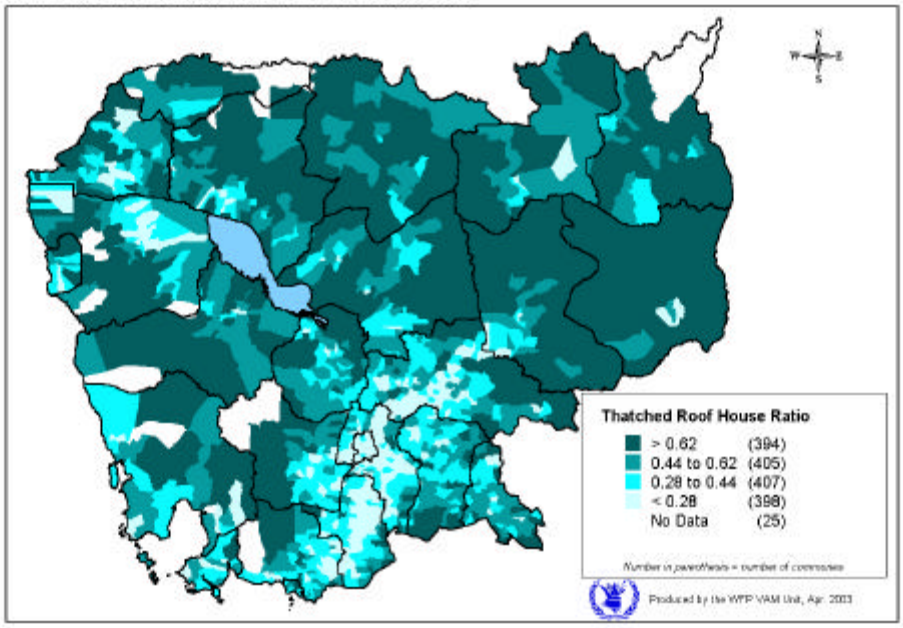
Commune-Level Ratio of Female Headed Household with Children Under 5 Years



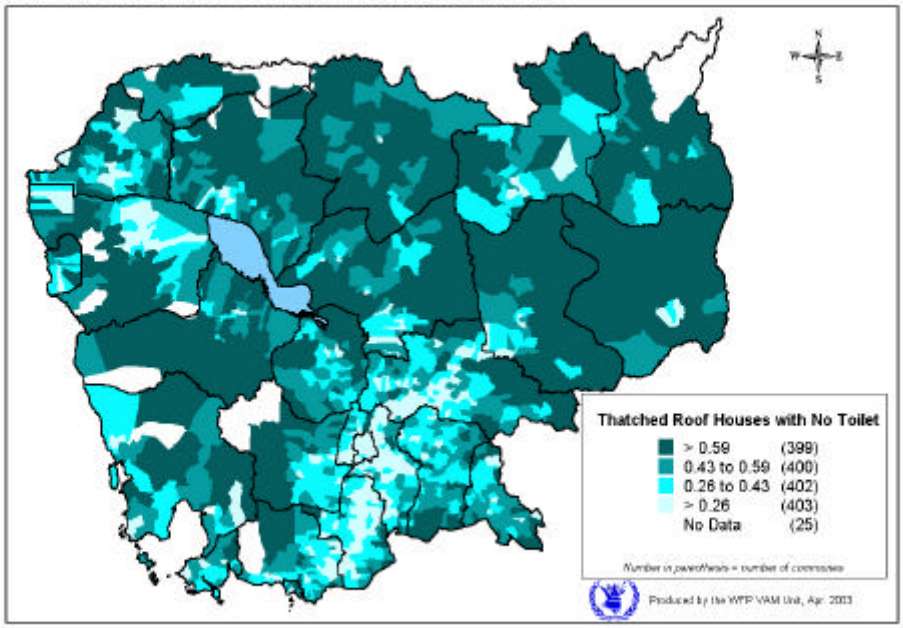
Housing Indicators

For housing indicators, the maps for R_THATCH_RATIO, R_THATCH_NO_LAT_RATIO, RAT_TOILET, FAM_H2O_OTHE_RATIO are presented. These indicators have somewhat similar patterns. Mondol Kiri and northern Siem Reap are generally bad in these indicators.

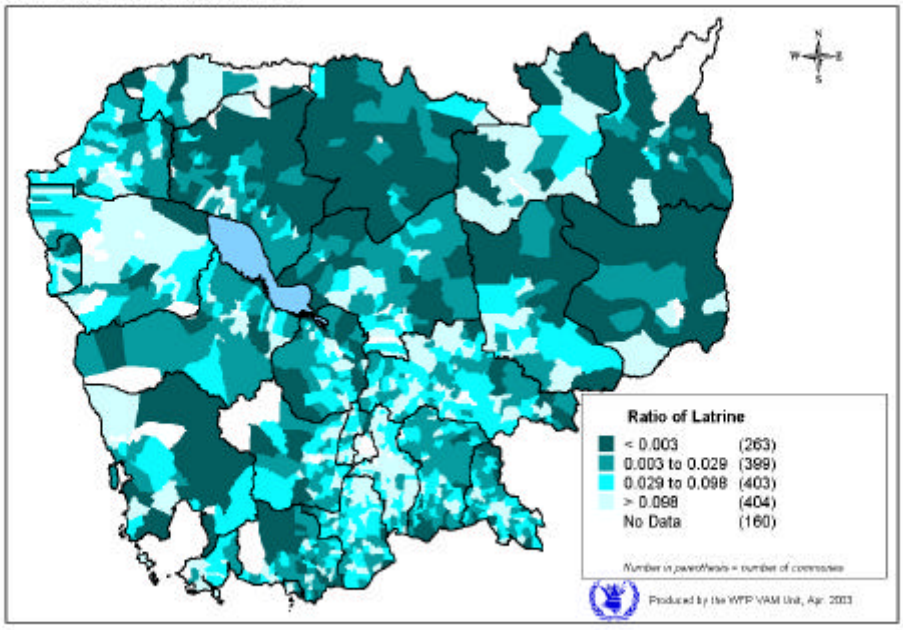
Commune-Level Ratio of Houses with Thatched Roof



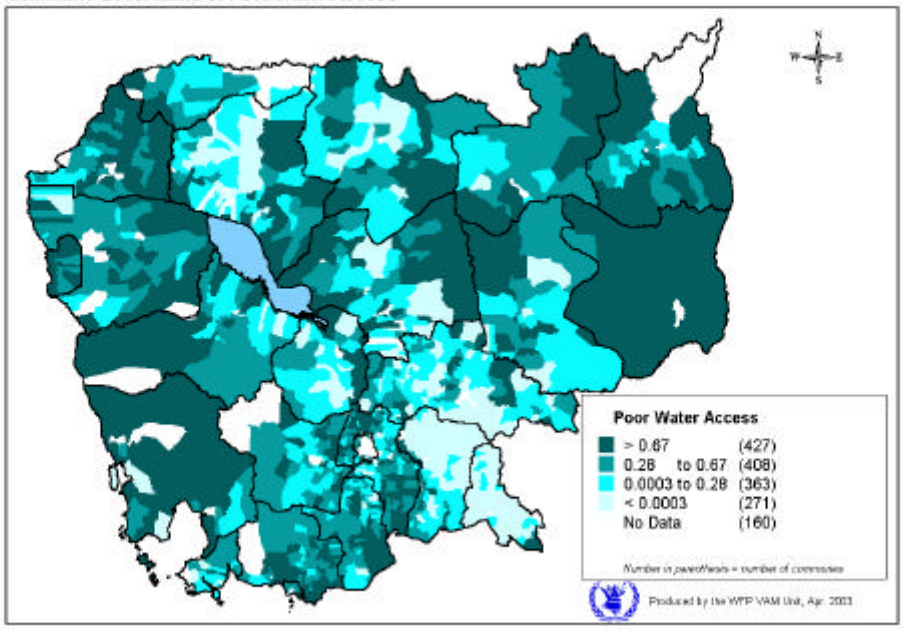
Commune-Level Ratio of Thatched Roof Houses with No Toilet



Commune-Level Ratio of Latrine



Commune-Level Ratio of Poor Water Access

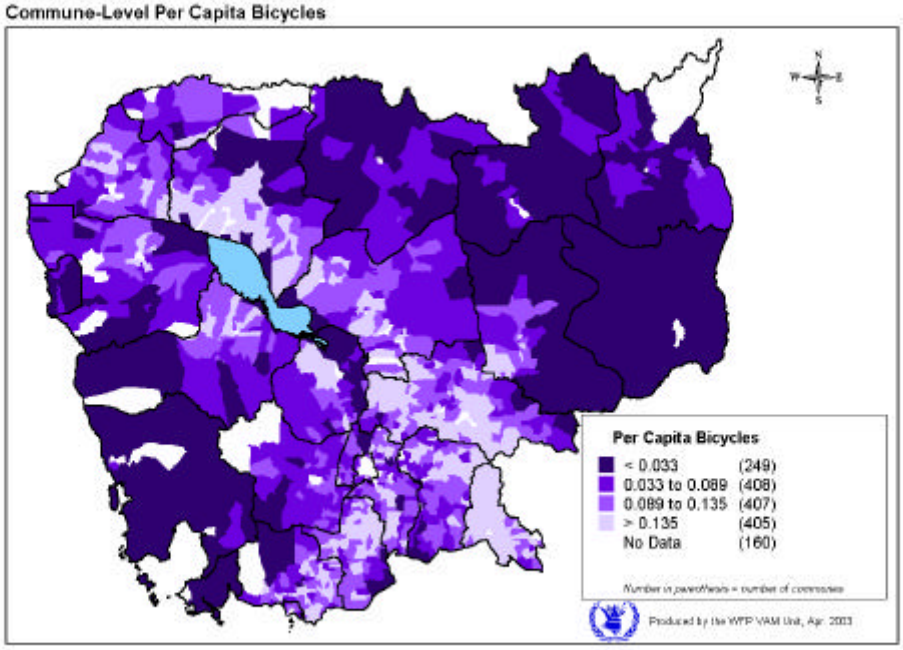


Asset Indicators

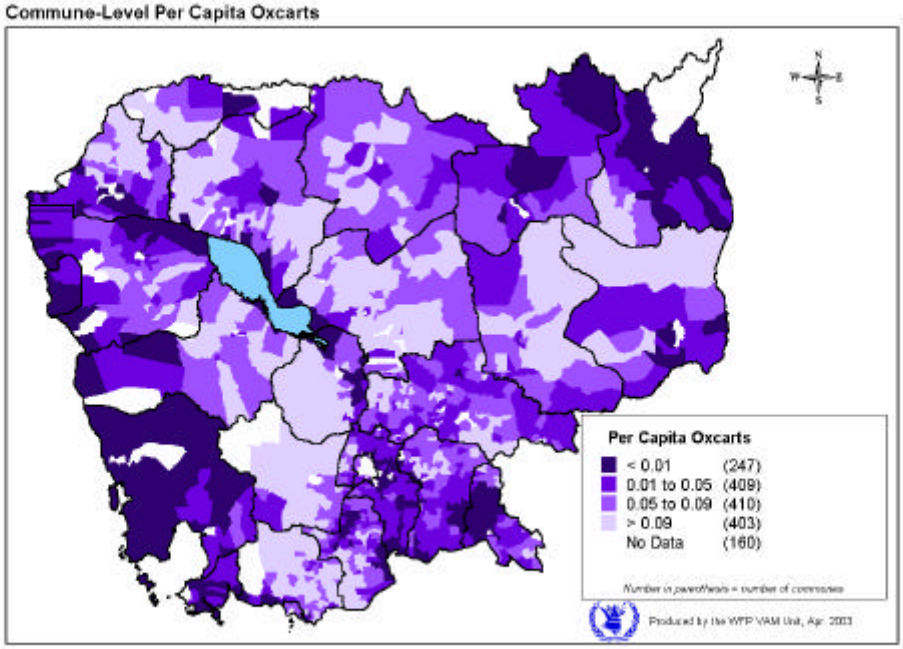
For asset indicators, PC_BICYCLE, PC_CART, PC_MOTO, PC_TV and PC_CAR are presented. It should be noted that these variables are particularly vulnerable to the inequality that exists among

households. However, it is in general unlikely to be a big problem as there are a few people, if any, who can afford to possess a multiple number of these goods. Except for TV, these assets are for transport, and related not only to wealth of the commune but also to the road infrastructure. Similarly, the number of TV *per capita* is related to the reception as well.

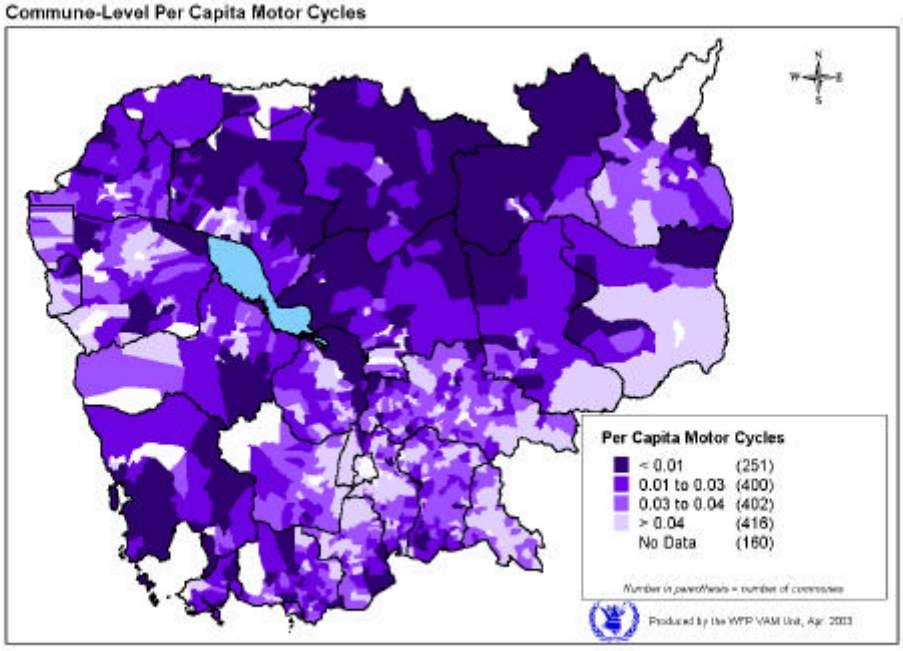
Commune-Level Per Capita Bicycles



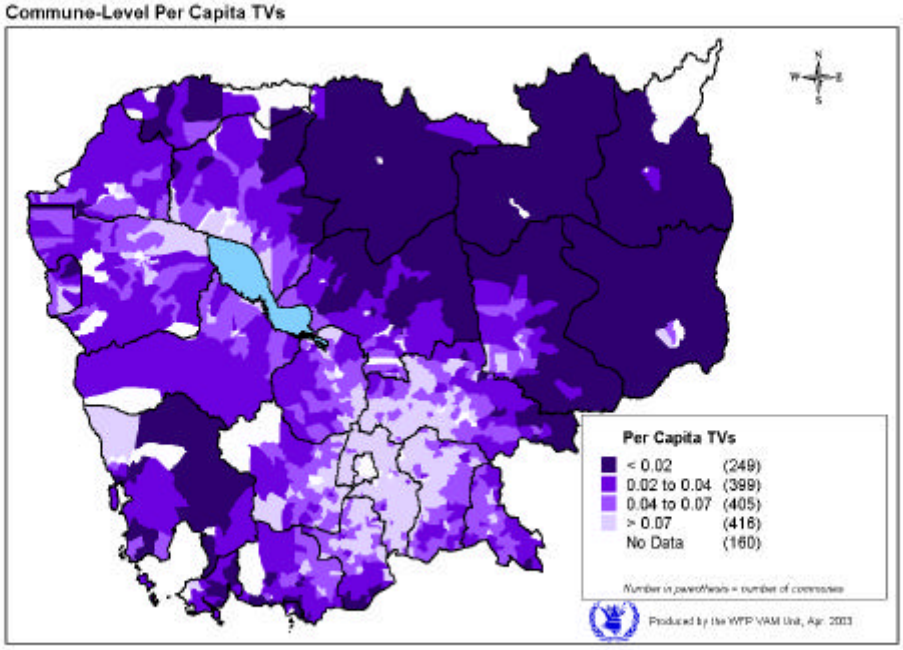
Commune-Level Per Capita Oxcarts

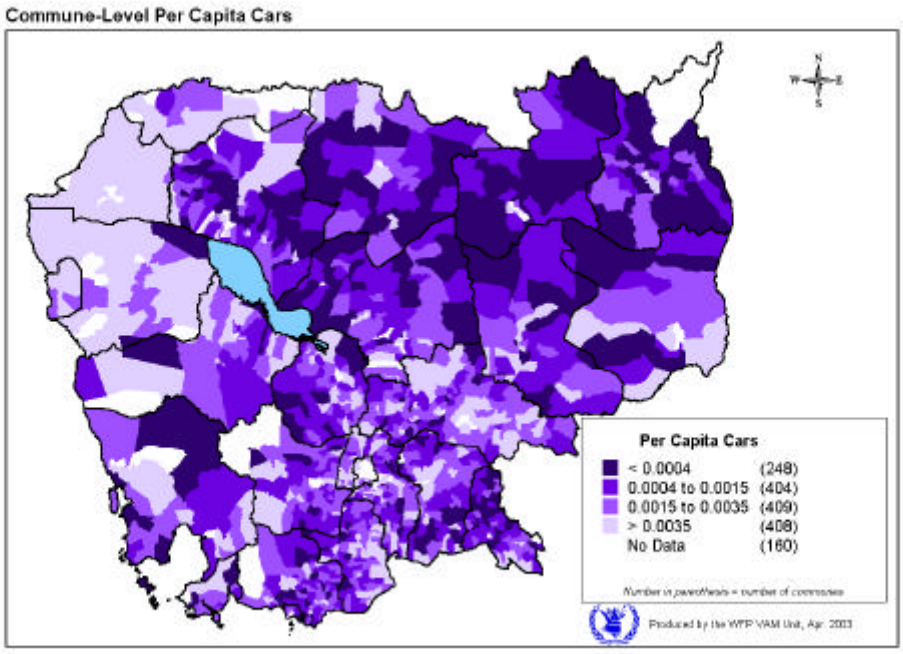


Commune-Level Per Capita Motor Cycles



Commune-Level Per Capita TVs

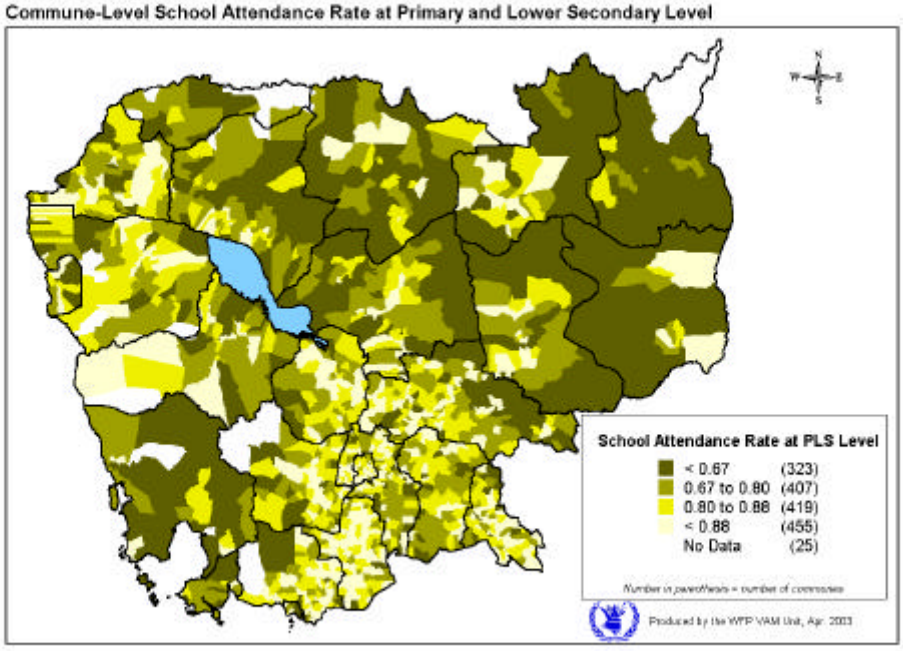




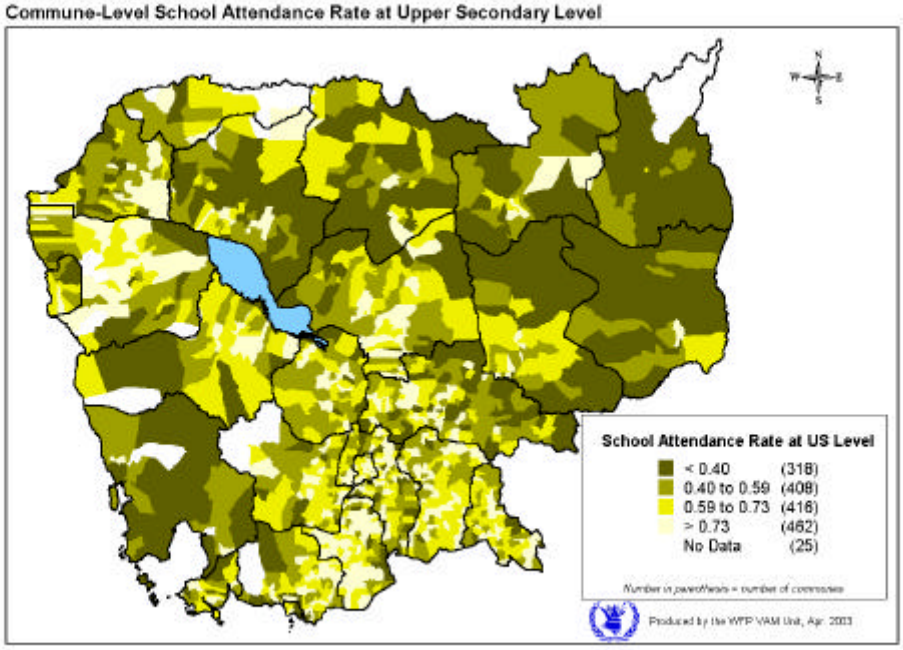
Education Indicators

For education indicators, SAR_TSCH_6_14, SAR_TSCH_15_17, GEN_GAP_6_14, GEN_GAP_15_17, TILT_15OV, FILT_15OV, PC_PCLSRM, PC_PTCH, PC_SCLSRM and PC_STCH are presented. School attendance is particularly low in the northeastern areas. It should be noted that gender gap is of secondary importance in comparison with school attendance rate. To elucidate this point, let us consider two communes A and B. Suppose school attendance rates for boys and girls at the upper secondary level are respectively 65 percent and 45 percent in commune A, and 5 percent and 1percent in commune B. In this case, the gender gap for commune A and commune B is 20 percent and 4 percent respectively. However, it does not make much sense to say that commune B is better than commune A. It is only useful when communes with similar school attendance rates are compared at issue. As one would expect, the spatial pattern of adult literacy and education have similarities. The "quality" of education, including PC_CLSRM, PC_PTCH, PC_SCLSRM and PC_STCH as measured by the number of classrooms and teachers per school age children at a given level is, on the contrary, does not seem to be closely related with other education indicators.

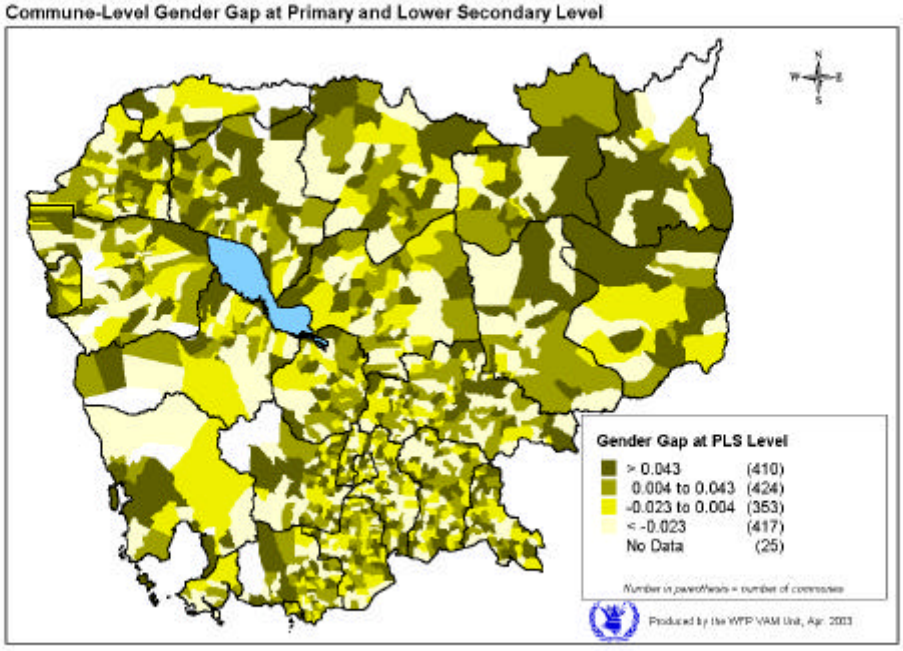
Commune-Level School Attendance Rate at Primary and Lower Secondary Level



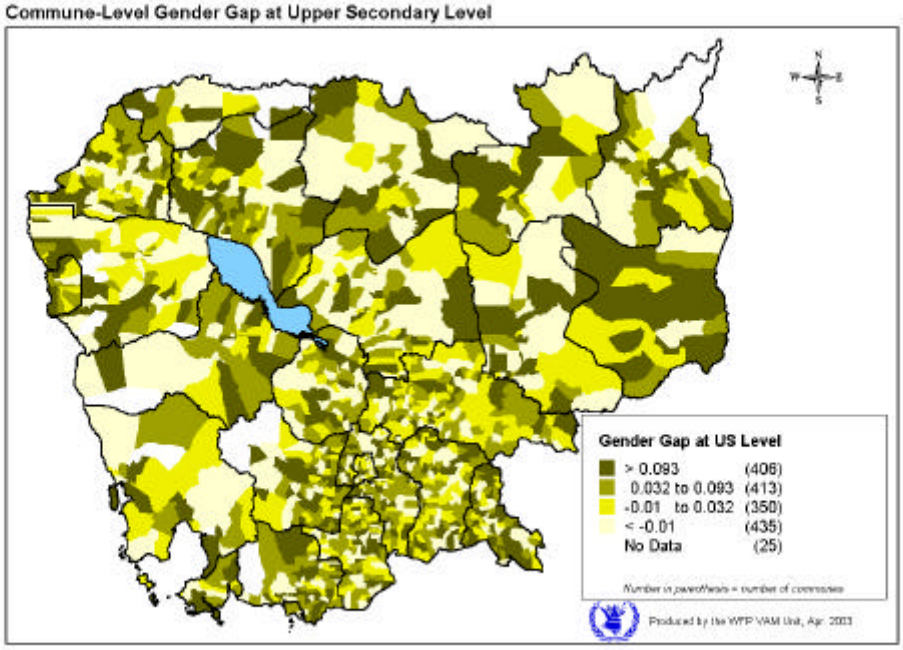
Commune-Level School Attendance Rate at Upper Secondary Level



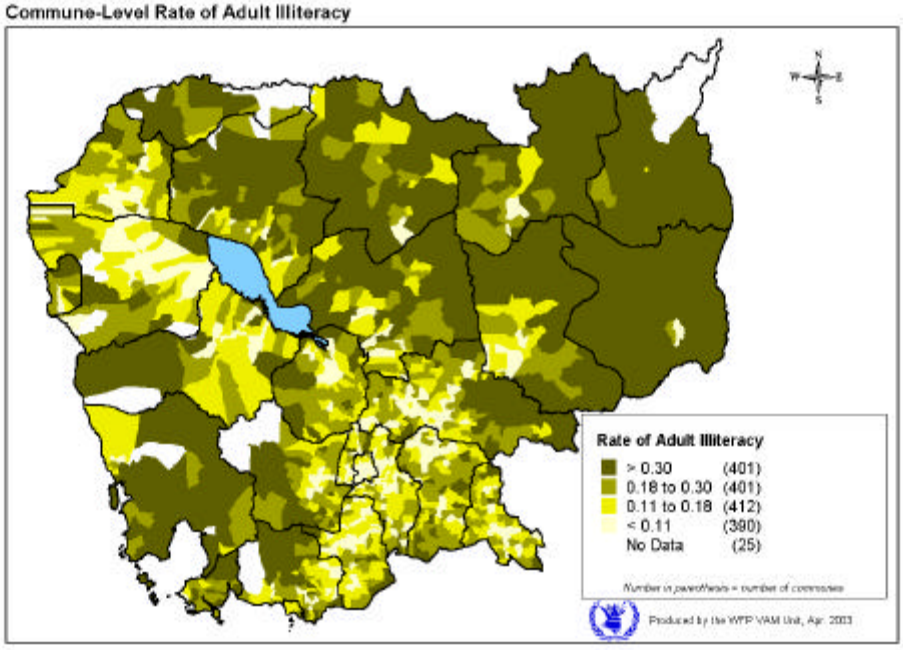
Commune-Level Gender Gap at Primary and Lower Secondary Level



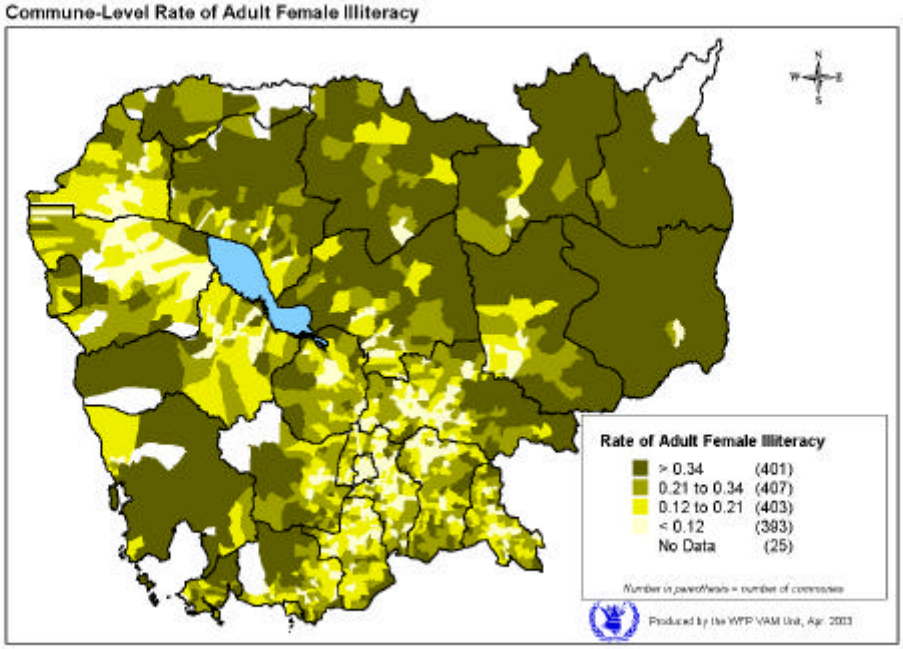
Commune-Level Gender Gap at Upper Secondary Level



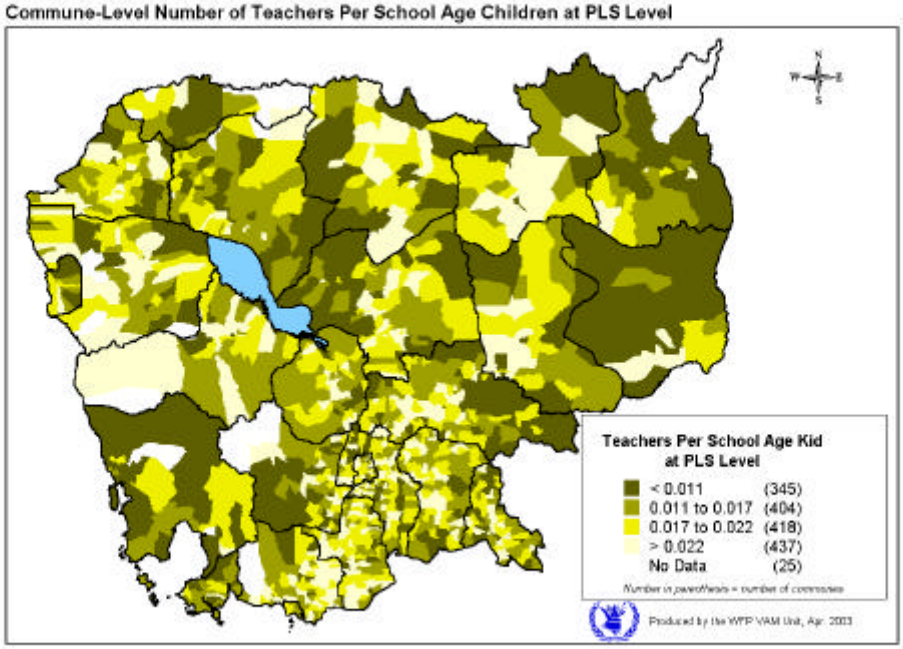
Commune-Level Rate of Adult Illiteracy



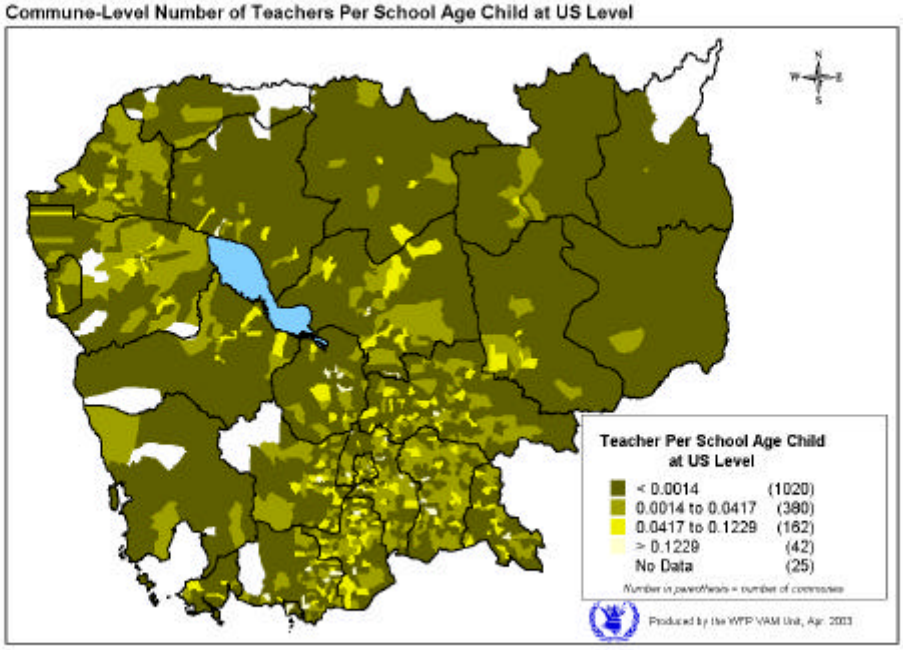
Commune-Level Rate of Adult Female Illiteracy



Commune-Level Number of Teachers Per School Age Children at PLS Level

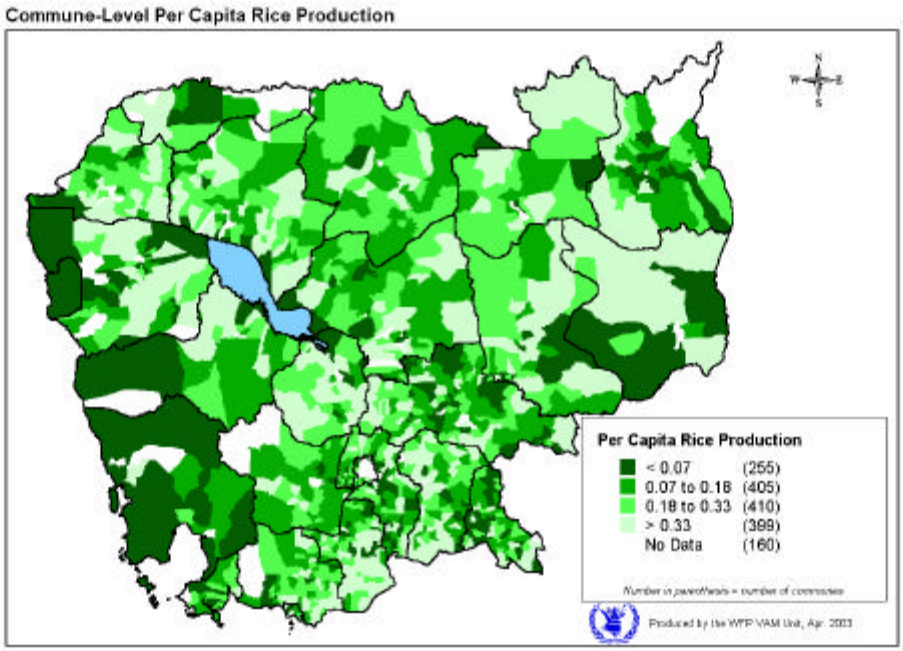


Commune-Level Number of Teachers Per School Age Child at US Level

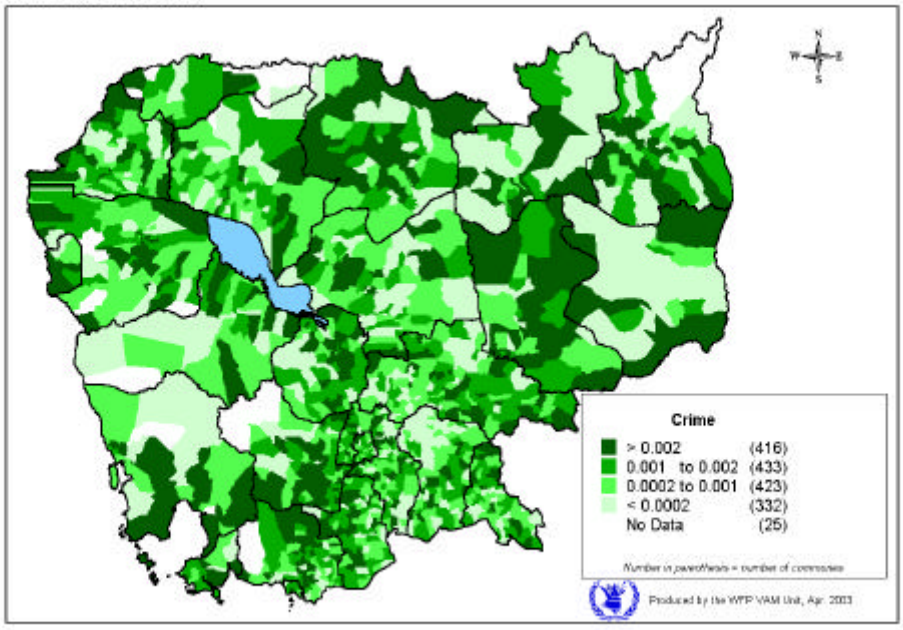


Other Indicators

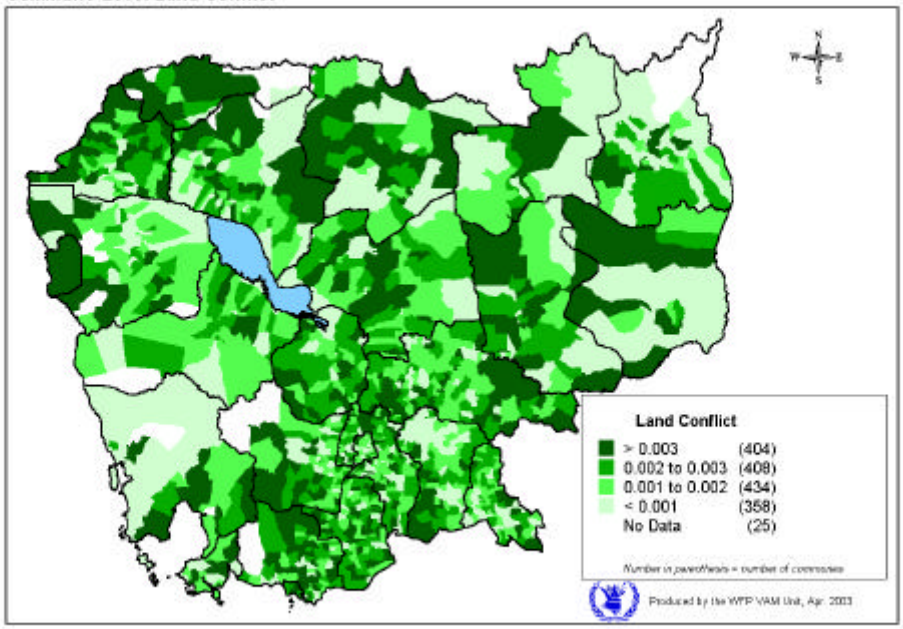
As with Section SECTION III, PC_TOT_PROD, PC_INSECURITY, PC_LANDCONFLICT and MIN_MARKET are presented below. MIN_MARKET seems to have a spatial pattern similar to many of previously presented maps. It would be interesting to note that PC_INSECURITY and PC_LANDCONFLIC have somewhat similar spatial pattern.



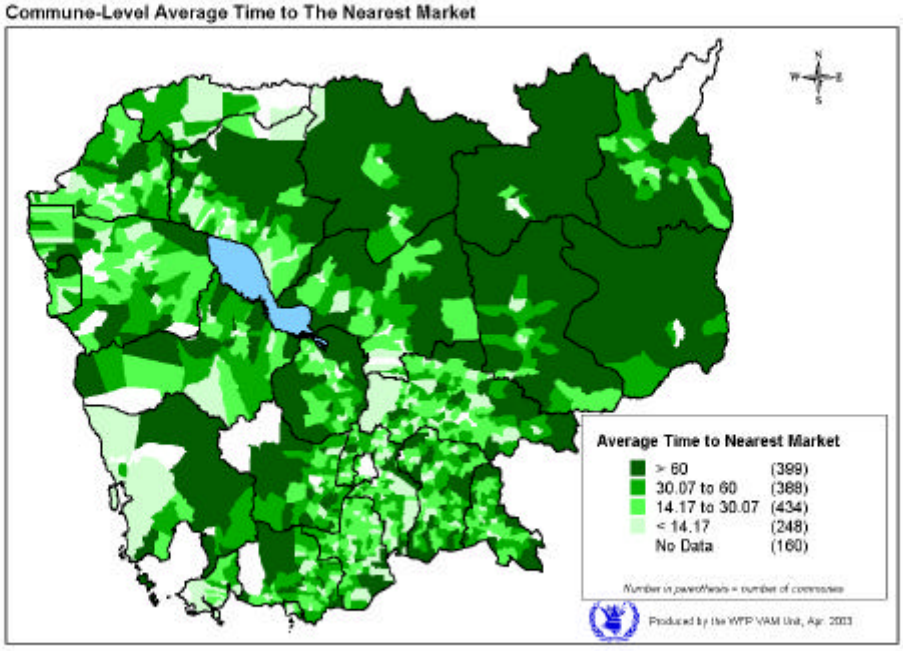
Commune-Level Crime



Commune-Level Land Conflict



Commune-Level Average Time to The Nearest Market



Appendix G. Additional Results on Factor Analysis

Rotation

As discussed in Appendix E, it is possible to present the equivalent results in a different way by applying rotation. Table 29 presents the results of varimax rotation. It does not seem to be the case that rotation makes the interpretation easier, and no further discussion will be made on rotation.

Table 29 The factor loading matrix after varimax rotation, and the correlation of factor scores with poverty measures when varimax rotation is applied.

	Factor Loading Matrix	
	Factor 1	Factor 2
FEMRATIO	0.015	-0.207
DEPRATIO	-0.074	0.027
BRATE	-0.051	0.082
FAMSIZE	-0.003	0.002
F_HHH_RATIO	0.032	-0.289
F_HHH_UD5_RATIO	-0.012	-0.248
R_THATCH_RATIO	-0.117	-0.037
R_THATCH_NO_LAT_RATIO	-0.119	-0.056
RAT_TOILET	0.083	0.199
FAM_H2O_OTHE_RATIO	-0.022	0.213
PC_BICYCLE	0.072	-0.163
PC_CART	0.005	-0.021
PC_MOTO	0.090	0.077
PC_TV	0.106	0.024
PC_CAR	0.039	0.191
NER_TSCH_6_14	0.093	-0.056
NER_TSCH_15_17	0.093	-0.044
GEN_GAP_6_14	-0.020	0.045
GEN_GAP_15_17	-0.010	-0.093
TILT_15OV_RATIO	-0.115	0.054
FILT_15OV_RATIO	-0.114	0.050
PC_PCLSRM	0.061	0.104
PC_PTSCH	0.088	0.105
PC_SCLSRM	0.075	0.165
PC_STCH	0.075	0.162
PC_TOT_PROD	0.013	-0.004
PC_INSECURITY	-0.001	0.191
PC_LANDCONFLICT	-0.003	0.169
MIN_MARKET	-0.067	0.059
Correlation With Poverty Rate	-0.065	-0.215
Correlation With Poverty Gap	-0.132	-0.091
Correlation With Poverty Severity	-0.141	-0.021

Using population as the weight

Communes are heterogeneous in the size and population. It may be argued that larger communes should be treated with a higher weight. Though there is no theoretical justification for using the population weight as the weight for the commune, it is possible to look at the population-weighted results. The map for the first factor (SC_PC_WW1) is presented at the end of this appendix.

Table 30 The results of a commune-level factor analysis. Principal component method was used and population was used as the weight. First two eigenvalues of reduced correlation matrix were 6.61 and 1.91 explaining 22.79% and 6.58% of the total variance and the number of observations is 1471.

	Factor Loading Matrix		Uniqueness	Standard Scoring Coefficients	
	Factor 1	Factor 2		Factor 1	Factor 2
FAMRATIO	0.150	-0.476	0.751	0.023	-0.249
DEPRATIO	-0.503	0.086	0.739	-0.076	0.045
BRATE	-0.307	0.109	0.894	-0.046	0.057
FAMSIZE	-0.054	0.155	0.973	-0.008	0.081
F_HHH_RATIO	0.261	-0.570	0.607	0.040	-0.299
F_HHH_UD5_RATIO	-0.009	-0.422	0.821	-0.001	-0.221
R_THATCH_RATIO	-0.796	-0.072	0.361	-0.120	-0.037
R_THATCH_NO_LAT_RATIO	-0.811	-0.108	0.330	-0.123	-0.056
RAT_TOILET	0.538	0.395	0.555	0.081	0.207
FAM_H2O_OTHE_RATIO	-0.151	0.466	0.760	-0.023	0.244
PC_BICYCLE	0.482	-0.338	0.653	0.073	-0.177
PC_CART	0.061	-0.067	0.992	0.009	-0.035
PC_MOTO	0.646	0.087	0.575	0.098	0.045
PC_TV	0.701	0.069	0.503	0.106	0.036
PC_CAR	0.167	0.483	0.739	0.025	0.253
NER_TSCH_6_14	0.656	-0.138	0.551	0.099	-0.072
NER_TSCH_15_17	0.631	-0.100	0.591	0.096	-0.053
GEN_GAP_6_14	-0.121	0.038	0.984	-0.018	0.020
GEN_GAP_15_17	-0.038	-0.193	0.961	-0.006	-0.101
TILT_15OV_RATIO	-0.809	0.032	0.344	-0.122	0.017
FILT_15OV_RATIO	-0.798	0.016	0.363	-0.121	0.009
PC_PCLSRM	0.422	0.045	0.820	0.064	0.023
PC_PTSCH	0.591	0.088	0.643	0.089	0.046
PC_SCLSRM	0.470	0.240	0.721	0.071	0.126
PC_STCH	0.474	0.235	0.720	0.072	0.123
PC_TOT_PROD	0.080	-0.048	0.991	0.012	-0.025
PC_INSECURITY	-0.088	0.315	0.893	-0.013	0.165
PC_LANDCONFLICT	-0.176	0.292	0.884	-0.027	0.153
MIN_MARKET	-0.486	0.050	0.761	-0.074	0.026

Using squared multiple correlations for the prior communality

As discussed in Appendix E, it is common to use squared multiple correlations for the prior communality. The results are in Table 31 and the map of the first factor score (SC_FM_NW1) is given at the end of this appendix.

Table 31 The results of a commune-level factor analysis. Principal component method was used and no weight was applied. First two eigenvalues of reduced correlation matrix were 6.81 and 1.50. The number of observations is 1471.

	Factor Loading Matrix		Uniqueness	Standard Scoring Coefficients	
	Factor 1	Factor 2		Factor 1	Factor 2
FAMRATIO	0.190	-0.098	0.954	0.017	-0.038
DEPRATIO	-0.497	0.054	0.751	-0.031	0.014
BRATE	-0.368	0.098	0.855	-0.008	0.016
FAMSIZE	-0.016	0.010	1.000	-0.018	0.003
F_HHH_RATIO	0.349	-0.244	0.819	0.042	-0.138
F_HHH_UD5_RATIO	0.036	-0.221	0.950	0.005	-0.082
R_THATCH_RATIO	-0.830	0.015	0.310	-0.140	0.291
R_THATCH_NO_LAT_RATIO	-0.840	-0.003	0.295	-0.195	-0.287
RAT_TOILET	0.484	0.185	0.732	0.052	0.085
FAM_H2O_OTHE_RATIO	-0.233	0.128	0.929	-0.030	0.050
PC_BICYCLE	0.552	-0.157	0.671	0.049	-0.072
PC_CART	0.043	-0.005	0.998	0.002	-0.004
PC_MOTO	0.574	0.065	0.667	0.075	0.018
PC_TV	0.713	-0.015	0.492	0.043	0.010
PC_CAR	0.168	0.059	0.968	0.002	0.034
NER_TSCH_6_14	0.656	-0.131	0.552	0.047	-0.026
NER_TSCH_15_17	0.645	-0.049	0.581	0.057	-0.020
GEN_GAP_6_14	-0.143	0.073	0.974	-0.002	0.008
GEN_GAP_15_17	-0.023	-0.057	0.996	0.003	-0.014
TILT_15OV_RATIO	-0.859	0.232	0.208	-0.237	0.297
FILT_15OV_RATIO	-0.854	0.225	0.220	-0.147	0.102
PC_PCLSRM	0.359	0.096	0.862	0.048	0.043
PC_PTSCH	0.550	0.131	0.680	0.054	0.065
PC_SCLSRM	0.449	0.734	0.259	0.071	0.433
PC_STCH	0.454	0.733	0.257	0.073	0.449
PC_TOT_PROD	0.080	0.014	0.993	0.009	0.001
PC_INSECURITY	-0.091	0.151	0.969	-0.010	0.057
PC_LANDCONFLICT	-0.096	0.112	0.978	-0.011	0.049
MIN_MARKET	-0.464	0.110	0.773	-0.027	0.021

With more variables

The set of variables chosen in SECTION III is not only one of many possible choices. Hence it is important to see how this choice has affected the results. In this subsection, a number of other variables are added. The description of added variables is given in Appendix D, and the results are presented in Table 32. The map of the first factor score (SC_MV_NW1) is given at the end of this appendix.

Table 32 The results of a commune-level factor analysis. Principal component method was used and no weight was applied. First two eigenvalues of reduced correlation matrix were 8.59 and 2.68 explaining 23.89% and 7.44% of the total variance and the number of observations is 1471.

	Factor Loading Matrix		Uniqueness	Standard Scoring Coefficients	
	Factor 1	Factor 2		Factor 1	Factor 2
FAMRATIO	0.203	-0.334	0.847	0.024	-0.125
DEPRATIO	-0.530	0.078	0.713	-0.062	0.029
BRATE	-0.400	0.097	0.830	-0.047	0.036
FAMSIZE	-0.005	0.076	0.994	-0.001	0.028
F_HHH_RATIO	0.367	-0.308	0.770	0.043	-0.115
F_HHH_UD5_RATIO	0.042	-0.189	0.962	0.005	-0.071
R_THATCH_RATIO	-0.807	-0.087	0.341	-0.094	-0.032
R_THATCH_NO_LAT_RATIO	-0.817	-0.154	0.309	-0.095	-0.057
RAT_TOILET	0.495	0.646	0.338	0.058	0.241
FAM_H2O_OTHE_RATIO	-0.258	0.359	0.805	-0.030	0.134
PC_BICYCLE	0.593	-0.352	0.524	0.069	-0.131
PC_CART	0.045	-0.051	0.995	0.005	-0.019
PC_MOTO	0.595	0.263	0.576	0.069	0.098
PC_TV	0.764	0.065	0.412	0.089	0.024
PC_CAR	0.184	0.497	0.719	0.021	0.186
SAR_TSCH_6_14	0.679	-0.105	0.528	0.079	-0.039
SAR_TSCH_15_17	0.661	-0.083	0.556	0.077	-0.031
GEN_GAP_6_14	-0.161	0.062	0.970	-0.019	0.023
GEN_GAP_15_17	-0.016	-0.154	0.976	-0.002	-0.058
TILT_15OV_RATIO	-0.841	0.044	0.291	-0.098	0.016
FILT_15OV_RATIO	-0.834	0.031	0.303	-0.097	0.012
PC_PCLSRM	0.344	0.097	0.872	0.040	0.036
PC_PTSCH	0.534	0.108	0.703	0.062	0.041
PC_SCLSRM	0.418	0.112	0.813	0.049	0.042
PC_STCH	0.421	0.124	0.808	0.049	0.046
PC_TOT_PROD	0.083	-0.011	0.993	0.010	-0.004
PC_INSECURITY	-0.099	0.263	0.921	-0.012	0.098
PC_LANDCONFLICT	-0.094	0.189	0.955	-0.011	0.071
MIN_MARKET	-0.565	0.157	0.656	-0.066	0.059
KM_ROAD	-0.443	0.126	0.788	-0.052	0.047
MIN_ROAD	-0.458	0.176	0.759	-0.053	0.066
LAT_ZINC_RATIO	0.338	0.647	0.467	0.039	0.242
TV_THATCH_RATIO	0.512	0.049	0.736	0.059	0.018
R_TILE_RATIO	0.553	-0.499	0.445	0.064	-0.186
R_ZINC_RATIO	0.245	0.655	0.512	0.028	0.244
PC_TBA	-0.671	0.136	0.532	-0.078	0.051

With less variables

Similarly, it is possible to reduce the number of variables. The smaller set of variables was created from the previous subsection using a stepwise regression. The results are presented in Table 33 and the map of the first factor score (SC_LV_NW1) is given at the end of this appendix

Table 33 The results of a commune-level factor analysis. Principal component method was used and no weight was applied. First two eigenvalues of reduced correlation matrix were 3.89 and 1.52 explaining 27.82% and 10.84% of the total variance and the number of observations is 1471.

	Factor Loading Matrix		Uniqueness	Standard Scoring Coefficients	
	Factor 1	Factor 2		Factor 1	Factor 2
RAT_TOILET	0.535	0.560	0.400	0.137	0.369
PC_TBA	-0.637	0.148	0.572	-0.164	0.098
R_THATCH_RATIO	-0.840	0.060	0.291	-0.216	0.040
PC_TV	0.809	0.040	0.343	0.208	0.027
PC_MOTO	0.668	0.309	0.459	0.172	0.203
MIN_MARKET	-0.552	0.035	0.694	-0.142	0.023
PC_PTCH	0.517	0.094	0.723	0.133	0.062
PC_CAR	0.175	0.705	0.472	0.045	0.465
R_TILE_RATIO	0.602	-0.553	0.332	0.155	-0.364
FAMSIZE	-0.036	0.002	0.999	-0.009	0.001
PC_LANDCONFLICT	-0.094	0.225	0.940	-0.024	0.148
FAM_H2O_OTHE_RATIO	-0.276	0.432	0.737	-0.071	0.285
F_HHH_UD5_RATIO	0.016	-0.129	0.983	0.004	-0.085
DEPRATIO	-0.587	0.121	0.641	-0.151	0.080
RAT_TOILET	0.535	0.560	0.400	0.137	0.369

Discussion on the results

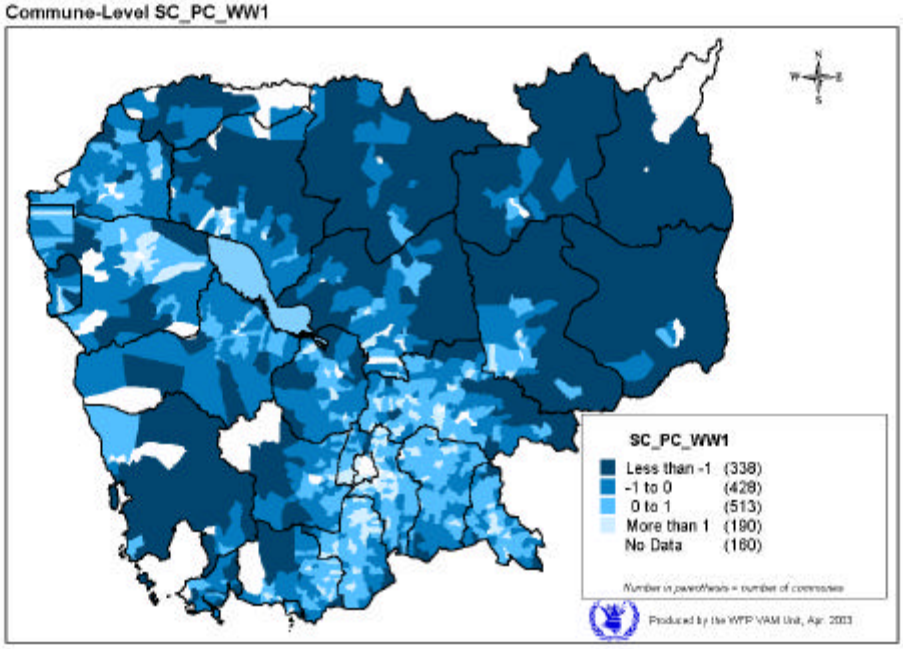
As can be seen from the previous tables, different methods of factor analysis create different results. But the results are qualitatively similar. With some exceptions, the patterns of signs in the factor loading matrix are similar, especially for factor 1. Table 34 provides the correlation between the first factor scores derived with different methods. The results seems to suggest that the results presented in the main text are relatively robust with regard to different methods.

Table 34 Correlation between the first factor scores derived with different methods and poverty measures.

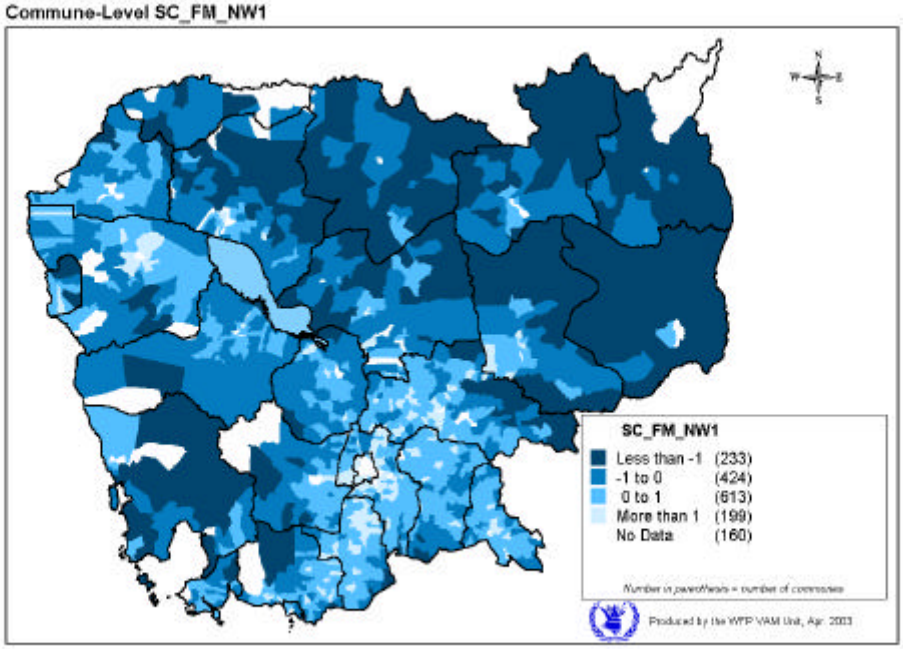
	SC_PC_NW1	SC_PC_WW1	SC_FM_NW1	SC_MV_NW1	SC_LV_NW1
SC_PC_NW1	1.0000				
SC_PC_WW1	0.9991	1.0000			
SC_FM_NW1	0.9904	0.9911	1.0000		
SC_MV_NW1	0.9911	0.9914	0.9803	1.0000	
SC_LV_NW1	0.9449	0.9449	0.9298	0.9580	1.0000
Pov Rate	-0.0494	-0.0532	-0.0698	-0.0211	-0.0418
Pov Gap	-0.1247	-0.1268	-0.1427	-0.1015	-0.1022
Pov Severity	-0.1389	-0.1397	-0.1537	-0.1197	-0.1125

Maps of first factor scores

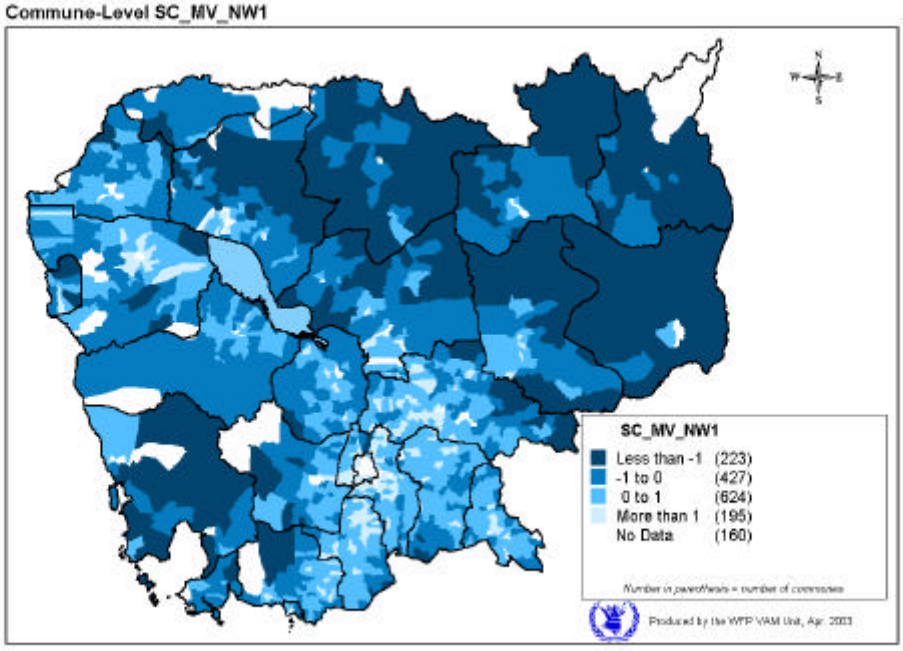
Commune-Level SC_PC_WW1



Commune-Level SC_FM_NW1



Commune-Level SC_MV_NW1



Commune-Level SC_LV_NW1