Kingdom of Cambodia
Nation Religion King

World Food Programme




# Estimation of Poverty Rates at Commune-Level in Cambodia

## Using the Small-Area Estimation Technique to Obtain Reliable Estimates

CAMBODIA
OCTOBER 2002

The Ministry of Planning
and
The United Nations World Food Programme

# Estimation of Poverty Rates at Commune-Level in Cambodia

**Using the Small-Area Estimation Technique to Obtain Reliable Estimates**

CAMBODIA
OCTOBER 2002

The Ministry of Planning
and
The United Nations World Food Programme

# FOREWORD

It is our pleasure to share with you a copy of the recently completed report "Estimation of Poverty Rates at Commune-Level in Cambodia." This study is the result of a joint collaboration between the United Nations World Food Programme, the Ministry of Planning, and the World Bank.

In addition, many others have been involved in the process. In particular we would like to thank the following partners for providing data, helpful inputs and constructive comments: National Institute of Statistics, Ministry of Education, Ministry of Health, UNDP, UNICEF, UNESCO, WHO, UNFPA, USAID, European Commission, Asian Development Bank, and IFAD.

This study represents a refinement of the poverty mapping exercise undertaken by WFP in 2000. In order to ensure the greatest accuracy possible this exercise involved combining, through rigorous statistical analysis, information from three data sets: the 1998 census, the 1997 socio-economic survey (CSES 1997) and geographical information systems (GIS). In fact, this is the first time that the recently developed small area estimation technique has been successfully used in Asia to derive poverty estimates at such a disaggregated level as the commune.

While WFP has been pleased to facilitate the process, it is important to note that ownership of the results rests with the Ministry of Planning which fully endorses the report and the maps. We believe that these maps will prove useful to assist not only government, but also international organizations, donors and NGOs to better target their respective programmes of assistance within Cambodia.

We would like to reiterate that the maps serve as an important tool for planners and decision-makers. However, users need to keep in mind that while the information presented is an accurate reflection of poverty in 1998, the situation may have subsequently changed in the intervening years in some areas. In order to account for this, efforts are currently underway to verify the information at commune-level through a comprehensive ground truthing exercise.

We would like to thank all of you in helping to better understand poverty in Cambodia. By knowing where the poor are we hope that all of us can be more effective in better meeting their specific needs.

R. Hansen

Rebecca Hansen,
WFP Representative

H.E. Kim Saysamalen
Under Secretary of State, Ministry of Planning

ACKNOWLEDGEMENTS

This study would not have been possible without the cooperation and dedicated assistance of various institutions and individuals. The United Nations World Food Programme Cambodia (WFP) and the Ministry of Planning (MOP) of Royal Government of Cambodia are indeed very grateful to H.E. Chhay Than, Minister of Planning for his support and encouragement from the beginning stage of study design through to finalization of the report. The continuous support and guidance of H.E. Kim Saysamalen, Under-Secretary of State of the Ministry of Planning, are gratefully acknowledged. Special thank is also due to H.E. San Sythan, Director of National Institute of Statistics, for providing access to the unit record of data of the General Population Census 1998.

WFP and MOP are particularly indebted to the author of the report, Tomoki Fujii, for his enthusiastic efforts in gathering data, developing analytical models, conducting analysis, running simulations, and finally writing this report.

The author is particularly indebted to Jean Olson Lanjouw (Brookings Institution), Peter Lanjouw (World Bank, Washington DC) and Berk Ozler (World Bank, Washington DC) for their analytical and conceptual support of this project. Livia Montana (Macro International) provided him with logistical support and helped assemble the GIS data set. The author also thanks Mahadevan Ramachandran (WFP, Regional Bureau for Asia), Steve Schonberger (World Bank, Cambodia), Sophal Ear (UNDP, Cambodia) and Venkatesh Sundararaman (University of Washington) for their support.

The author is grateful to Has Bunton and Meng Kimhor at the National Institute of Statistics of the Ministry of Planning for their generous assistance in preparing the census data set used in this study. Many thanks are also due to Khiev Bory, Chou Putheany, Hour Dariddh, and Saing Lundy, counterparts from the Ministry of Planning for their participation in training sessions on poverty mapping methodology.

Special thanks also go to Jacob Hook for editing this report.

# ABSTRACT

The CSES 1997 data and the 1998 census data were combined to produce estimates of poverty measures at the commune-level in Cambodia. Using the small-area estimation technique developed by Elbers et al. (2001), poverty rates, poverty gaps, and poverty severity were estimated at the commune-level. Although there are a number of communes for which the standard errors associated with the estimates are too high, it was found that, on average, the standard errors associated with estimates are small enough to make them useful. The estimates are expected to provide invaluable information for policy-analysts and decision-makers.

# TABLE OF CONTENTS

**FOREWORD** ..... III

**ACKNOWLEDGEMENTS** ..... IV

**ABSTRACT** ..... V

**SECTION I INTRODUCTION** ..... 1

**SECTION II METHODOLOGY** ..... 8

**SECTION III DATASETS** ..... 13

**SECTION IV IMPLEMENTATION** ..... 20

**SECTION V RESULTS** ..... 27

**SECTION VI CONCLUSION** ..... 35

**REFERENCES** ..... 36

**APPENDIX A GLOSSARY** ..... 38

**APPENDIX B LIST OF VARIABLES** ..... 39

**APPENDIX C DEFINITION OF POVERTY AND INEQUALITY MEASURES** ..... 43

**APPENDIX D REGRESSION RESULTS** ..... 44

**APPENDIX E FIRST STAGE DIAGNOSTICS** ..... 52

**APPENDIX F COMMUNE-LEVEL POVERTY MAPS** ..... 53

**APPENDIX G COMMUNE-LEVEL POVERTY RATES IN COMPARISON WITH THE NATIONAL POVERTY RATE** ..... 57

**APPENDIX H COMMUNE-LEVEL POVERTY MAPS WITH MINISTRY OF PLANNING (2001) POVERTY LINE** ..... 58

**APPENDIX I COMMUNE-LEVEL INEQUALITY (GINI INDEX)** ..... 62

**APPENDIX J COMMUNE-LEVEL POVERTY RATES FOR WOMEN** ..... 63

**APPENDIX K COMMUNE-LEVEL POVERTY RATES FOR CHILDREN** ..... 64

**APPENDIX L COMPILED LISTS OF POVERTY ESTIMATES** ..... 65

# SECTION I INTRODUCTION

## Background

The purpose of this report is to provide refined poverty maps for Cambodia using the recently developed small-area estimation technique. The United Nations World Food Programme (WFP) has created poverty maps that identify poor areas and used them for targeting. In particular, WFP has published a poverty map that combines the census data and socio-economic survey data (WFP, 2001) as the preliminary results. The *final* poverty maps presented in this report have been produced as part of our continued efforts to produce a set of more accurate maps.

The maps will serve as a basis for formulating targeting plans for various social interventions carried out by WFP. It is also expected that they will be used as a valuable tool to formulate plans, programmes and strategies to combat poverty for other organisations. The Royal Government of Cambodia (RGC) has placed poverty eradication at the top of its agenda, and has expressed a desire to employ the final poverty maps generated through this exercise as a tool to guide the allocation of resources at the provincial, district, and commune-levels. In order to build the capacity to understand and effectively use poverty maps for anti-poverty policy formulation, several key government officials were involved during the course of this poverty mapping exercise, and a number of counterparts from a various organisations were invited to the workshop held upon the completion of this exercise. Besides the RGC, numerous other organisations have expressed an interest in using the final poverty maps as a basis for deciding upon the nature and locations of their own programmes.

The focus of the report is to present the methodology employed in the mapping exercise in as clear and detailed a fashion as possible. This report is intended to serve as a reference for those who would like to appreciate the possibilities and limitations of the poverty maps, and are interested in the technical aspects of poverty mapping. Implications for the formulation of targeting plans for WFP and other organisations will be mentioned throughout the report for the purpose of illustration. However, specific issues associated with the use of poverty maps for policy formulation will not be discussed fully in this document.

**Targeting Poverty**

The multi-dimensionality of poverty has been increasingly emphasised by international aid organisations, academia, and non-governmental organisations. In the Human Development Report 1997, it is maintained that human poverty is more than income poverty—it is the denial of choices and opportunities for living a tolerable life (UNDP, 1997). The World Development Report 2001/2002 accepts a definition of poverty encompassing not only material deprivation but also low achievements in education and health. It also broadens the notion of poverty to include vulnerability and exposure to risk, voicelessness, and powerlessness (World Bank, 2001). Besides being multi-dimensional, poverty manifests itself in a variety of ways in different spatial and temporal contexts. For example, Jalan and Ravallion (1998, 2000) distinguish between chronic poverty, a long-term low standard of living, and transient poverty, which can be attributed to intertemporal variability in consumption. They find that the patterns determining these two types of poverty are significantly different. Therefore, for any kind of targeted intervention, one should carefully consider the dimensions of poverty that are at issue and decide which kind of poverty is to be targeted.

A number of governmental bodies, local and international NGOs, and international organisations have made eradicating poverty a priority and have established many social programmes to this end. In designing such interventions, efficient allocation of resources is essential for making poverty alleviation effective. Targeting often helps to make an intervention cost-effective because one can avoid wasting resources on the non-poor, which would occur in the absence of targeting. If targeting were costless and consumption poverty were at issue, it would be desirable to formulate a targeting policy such that the gap between current consumption and the poverty line is just fulfilled.

In reality, such a social intervention is unlikely to be possible for two reasons. First, targeting is not costless as there are administrative, political and other costs. If the targeting moves below a certain geographical level, for example villages, households or individuals, it becomes increasingly costly. Second, there are two types of errors in targeting. One is the error of exclusion, in which intended beneficiaries cannot benefit from the intervention. The other is the error of inclusion, in which an intervention reaches individuals who were not intended to be beneficiaries (Hoddinott, 1999). These errors are commonly known as Type I and Type II errors, respectively. If the costs of targeting outweigh the benefits, it may be more efficient to distribute rations to everyone without targeting.

Food security is one of the issues for which targeting may be useful. A person is food-secure if the number of calories available for him/her to eat exceeds his/her requirements. If one village is significantly more food-insecure than other, it may be most efficient to deliver food directly to the village concerned. The targeted people are determined by project staff, and such interventions are called administratively targeted interventions. On the other hand, if there are relatively rich people and poor people in the village, food-for-work programme may be more efficient. This enables exclusion of relatively rich people as the opportunity cost from work is relatively high for richer people. A targeting scheme like this is called self-targeting.

The question policy-makers face is whether and what targeting policy may make anti-poverty programmes more efficient. Poverty maps give useful information for answering such a question, and can be combined with other maps to derive even more valuable information by overlaying maps using geographical information systems (GIS). For example, the targeting policy of a basic medical care programme may be formulated by choosing poor areas where indicators show low health status. Irrigation systems may be most beneficial in poor areas that are drought-prone. Such targeting would be difficult or impossible without poverty maps.

**Poverty Reduction Strategy Papers (PRSP) and Poverty Monitoring and Analysis (PMA)**

The RGC has developed strategies to combat poverty in a series of policy papers including the National Programme to Rehabilitate and Develop Cambodia (1994), the first Five-Year Socio-Economic Development Plan (1996-2001), the Royal Government Platform for the Second Term (1998-2003), and the Policy Framework Paper. In 2000, the RGC prepared the Interim Poverty Reduction Strategy Paper (I-PRSP) and the full PRSP is forthcoming in October 2002.

The I-PRSP states that identifying the geographic distribution of the poor will allow anti-poverty interventions. The commune-level estimates of poverty rates produced in this report are expected to help donors and implementing organisations to identify target areas and formulate efficient and effective programmes and policies to reduce poverty. The maps in this report are also expected to provide a basis for promoting coordination within and between donors and implementing organisations in addressing poverty.

Policy-analysts are reminded that poverty rates are not the only measure for analysing poverty. For the purpose of targeting it may be more useful to check, for example, the absolute number, or density of poor people. Since different types of programmes are likely to give rise to different impacts on the poor both in terms of magnitude and pattern, policy-analysts should consider what can be derived from the maps.

Poverty maps are also expected to contribute to the development of the Cambodian Poverty Monitoring and Analysis System (CPMAS). For example, the commune-level poverty rates may be used as the baseline poverty rates for monitoring. In addition, poverty maps can be helpful for identifying the areas for which certain types of data should be collected. For example, areas of extreme poverty can be highlighted for close monitoring. When new poverty maps are created in the future, the progress of poverty reduction in Cambodia can be assessed in conjunction with information from the various programmes implemented by the government, non-governmental organisations, donor agencies and international organisations.

**Poverty Maps**

Socio-economic surveys have been widely used to analyse poverty in developing countries. Cambodia is no exception. Three successive Cambodian Socio-Economic Surveys (CSES), carried out in 1993/94, 1997 and 1999 respectively, have been used to generate poverty estimates. They have provided valuable information for policy-analysts and decision-makers.[1] However, the sampling designs of these surveys impose severe limitations on the geographical level of disaggregation at which the poverty estimates are reliable. None of the three surveys can provide a reliable estimate of poverty rates even at the provincial level.

However, it is often the case that what policy-makers really need is information that is geographically disaggregated. They may want poverty estimates at the district or even commune-level. If policy-makers want to deliver food to poor people, knowing poverty rates at the provincial level may not be very useful as too many non-poor may benefit due to the error of inclusion. Pockets of poverty, or poor areas surrounded by non-poor areas, cannot be identified from the socio-economic surveys alone. Therefore, poverty maps significantly improve the ability of policy-makers to efficiently and equitably allocate scarce resources to fight poverty.

The poverty maps presented in this report employ a methodology which is similar to, but more sophisticated than, the one used by the World Food Programme in 2000 and published as a preliminary report in 2001 entitled: "Identifying Poor Areas in Cambodia: Combining Census and Socio-Economic Survey Data to Trace the Spatial Dimensions of Poverty" (WFP, 2001). As with the previous poverty mapping exercise, this report combines survey data and census data to produce poverty estimates. Prior to 2000, poverty maps prepared by WFP Cambodia were based on a basic needs approach. The basic needs approach assigns weights to key indicator variables in an *ad hoc* manner, whereas the method presented in WFP (2001) and in this report uses a weighting scheme that has a stronger analytical footing.

The reason the survey data and census data are combined is to take advantage of the strengths of both and to help overcome the weaknesses of each. The socio-economic survey data has a detailed consumption component and detailed household-level characteristics. However, the sample size is too small to derive poverty rates for relatively small geographical areas. On the other hand, the census data covers almost the entire population. However, household-level characteristics are very limited and there is no

[1] For details of the poverty analysis, readers are referred to the three poverty profiles (Prescott and Pradhan, 1997; Ministry of Planning, 1998; and Ministry of Planning, 2001).

information on consumption. By imputing the survey data into the census data, the strengths of each, namely the consumption information and large sample size respectively, can be combined. The methodology will be presented in detail in SECTION II.

The main difference between WFP (2001) and this report is the explicit treatment of the standard errors associated with the poverty estimates. The previously published poverty maps, including WFP (2001), did not have information on the reliability of the estimates. If the standard errors associated with the estimates are very large, the estimate is in effect just a random number that contains almost no information on actual poverty rates. On the other hand, if the poverty estimates have small standard errors, a sharp comparison, between communes for example, is possible.

Even when the poverty estimates are the same for two communes, the associated standard errors could be very different. If there are two poor communes with the same estimated poverty rates but very different standard errors, the policy-maker may prefer to direct resources to the commune with the smaller standard error as it is less likely to be a non-poor commune in reality. Time and resources permitting, the policy-maker may also want to check, in one way or another, the state of poverty in the commune with the larger standard error.

Explicit inclusion of standard errors has two benefits. Firstly, the standard error *per se* has valuable information for policy-makers as it helps to reduce the errors associated with targeting and thus increases the efficiency of targeting. Secondly, it is a good reminder for policy-makers about the nature of the numbers they deal with. All estimates in this report are subject to errors because of nature of the statistical analysis employed.[2] The overall pattern of poverty as of 1998 is reflected in the maps presented in this report, but a specific commune may have a high standard error. The implications of such a situation depend upon the nature of the targeting at issue. When only a few communes in the whole country are chosen for targeting for a particular intervention, then policy-makers should certainly look at the confidence interval computed from the estimates of poverty rate and standard error. When a relatively large number of communes are targeted, then standard errors for

[2] It is often the case that standard errors are not estimated. This *does not*, however, imply that there is no error.

each commune may not be a concern.

Estimation of standard errors is possible through a computer simulation. By repeatedly applying the empirical distribution of error terms obtained through the consumption regression using the survey data to the imputed consumption data in census records, the poverty rates can be estimated in each simulation. By taking the standard deviation of the estimates, it is possible to estimate the standard error associated with the poverty rate estimates. The estimates of poverty rates and standard errors are based on fairly weak assumptions. Thus the methodology used in this report has wide applicability.[3] The details of this procedure will be explained in SECTION IV.

This report is structured as follows. SECTION II explains the methodology. The econometric theory of poverty mapping is explained briefly. This section will be of particular interest to those who are interested in the technical aspects of poverty mapping. In SECTION III, the datasets used in this analysis are explained. Some issues related to use of the datasets are also discussed. SECTION IV provides a description of the implementation of the methodology outlined in SECTION II. This section will be useful for those who would like to generate similar maps in the future. SECTION V explains the results. The focus is on the precision of the estimates. Finally, concluding remarks are given in SECTION VI.

[3] This does not necessarily ensure *small* standard errors.

## SECTION II METHODOLOGY

### Consumption as a measure of welfare

As with WFP (2001), this paper uses household consumption expenditure to measure the welfare of people. Consumption is not the only possible measure and can capture only certain aspects of poverty. For example, consumption measures cannot capture health and nutrition poverty or education poverty.[4] Subjective perceptions of poverty may not correspond perfectly with consumption. However, there are certain advantages to using consumption measures. For example, consumption is expressed in monetary terms and its meaning is easily understood. It should be noted this does not imply that consumption reflects poverty *better* than other indicators.

Consumption is not the only monetary measure. Income is often used as a welfare measure. However, income is often subject to under-reporting and seasonal variation. When a large informal sector exists or a large fraction of production is for self-consumption, income is unlikely to reflect welfare very well. Consumption tends to be a less problematic indicator in those respects. Therefore, consumption may be thought of as a good proxy for measuring true welfare, although it is by no means the perfect measure. Policy-analysts should keep this in mind when using poverty maps.

Consumption data is based on the CSES data sets, which were based on surveys taken at the household level. Consumption is defined as goods and services bought on the market, received in kind, or produced by the household. All of the consumption items in the CSES questionnaire, including food items and non-food items, are aggregated to arrive at the consumption aggregate for the household. When the *per capita* consumption is derived, the household consumption is simply divided by the number of individuals in the household. This exercise is questionable in that consumption items may be public goods within the household. There are also economies of scale within households. Moreover, children need less than adults to adequately sustain themselves. Treating adults and children with the same weight is of debatable validity. Although some researchers prefer

[4] See World Bank (1999) for an assessment of different aspects of poverty.

to use adult-equivalence measures, the simple average is reported for two reasons. First, the weights used to derive adult-equivalence measures have been controversial. Second, adult-equivalence measures were not used in previous poverty profiles. Therefore, to ensure comparability with the previously published figures, *per capita* household consumption measures have been used.

As with other poverty profiles, the poverty incidence, poverty gap and poverty severity are reported. To derive these poverty measures, the poverty line must be defined. One is considered to be poor when one's level of consumption is below the poverty line. The poverty rate in this report is synonymous with the incidence of poverty, or the head count index. It refers to the proportion of the population living below the poverty line. The poverty line, in terms of *per capita* per day consumption, used in this report is described in SECTION IV. The poverty gap gives information on how far off people are from the poverty line. This measures the average amount of resources per capita required to bring all the poor to the level of the poverty line. Poverty severity takes into account the inequality among the poor. A higher weight is placed on those people who are further away from the poverty line. The formulas used for these measures are found in APPENDIX C.

**Theory of Poverty Mapping**

The concepts behind poverty mapping are straightforward. First, the survey data is used to estimate a consumption model. This model describes the relationship between consumption and right-hand-side (independent) variables. Right-hand-side variables are restricted to those variables that can also be found in the census or in a tertiary data set that can be linked to both the census and the survey. A geographical information system (GIS) data set is used as tertiary data. The census data is then fed into the model with the parameter estimates to derive the statistics of interest.

The critical assumption behind this step is that the models estimated from the survey data apply to census records. If the census and survey data were taken at the same time, it would be reasonable to assume that this is the case. The census and the survey data used in this report were not taken at the same time. However, the time difference between the census year 1998 and survey year 1997 is small, and it is still reasonable to assume that the relationship between consumption and the right-hand-side variables holds for the census year. The poverty estimates correspond to the census year because the explanatory variables for predicting consumption come from the census data.

The fact that the census data and survey data were collected by different organisations may cast doubt on the comparability of the two data sets. Hence, a check of comparability between the census data and survey data has been carried out. Some key summary statistics have been compared at the stratum level of CSES 1997 with decisions being made as to whether or not these statistics are close enough to be comparable. Variables found to be incomparable have not been used in the consumption model.

The theoretical underpinnings of this methodology are given in detail in Elbers, Lanjouw and Lanjouw (2001). It has been applied in several countries and has been successful in creating poverty maps at a low level of geographical aggregation, which would not have been possible otherwise. Alderman et al. (2001) have examined how low levels of geographical aggregation can be achieved using this methodology in the case of South Africa. Demombynes et al. (2001) have compared empirical evidence taken from Ecuador, Madagascar, and South Africa. The methodology has been recently extended to combine two surveys, one with a detailed but small sample and the other with a much larger sample (Elbers, Lanjouw, Lanjouw and Leite, 2001).

In what follows, a brief summary of the discussion in Elbers, Lanjouw and Lanjouw (2001) is provided. Per-capita household consumption, $y_h$, for household $h$ is related to a $k$-vector of observable characteristics, $\mathbf{x_h}$, through the following model.

$$\ln y_h = \mathbf{x}_h^T \beta + u_h$$

where $\beta$ is a $k$-vector of parameters and $u_h$ is a disturbance term. $u_h$ satisfies $E[u_h \mid \mathbf{x}_h] = 0$. In application, the disturbance term is decomposed into the location, or cluster-specific, effect and the household-specific effect to allow for spatial autocorrelation and heteroscedasticity among households. The parameter $\beta$ is estimated through regression using the household survey data. This regression will be referred to as the first-stage regression.

For the purposes of the poverty maps, what is of interest is not the consumption of each household but various welfare measures at a certain level of aggregation. In this report, commune-level aggregation was chosen because such a level of aggregation is *useful* and the estimate at that level is *acceptable*. Welfare estimates at a more aggregated level such as the district or provincial level are more accurate. Hereafter, the welfare measure for the commune $c$ with $M_c$ households is denoted as $W(\mathbf{m}_c, \mathbf{X}_c, \beta, \mathbf{u}_c)$, where $\mathbf{m}_c$ is a $M_c$-vector of household size. $\mathbf{X}_c$ and $\mathbf{u}_c$ are a $M_c \times k$ matrix of observable characteristics, and a $M_c$-vector of disturbances respectively.

Because the vector of disturbances for the target population, $\mathbf{u}_c$, is always unknown, the expected value $\mu_c = E[W \mid \mathbf{m}_c, \mathbf{X}_c, \zeta_c]$ of the welfare measure $W$ given the observable characteristics in the commune is estimated. $\zeta_c$ is the vector of model parameters, including those which describe the disturbances. To construct an estimator of $\mu_c$, $\zeta_c$ is replaced by its consistent estimator $\hat{\zeta}_c$. This yields an estimator of the form $\hat{\mu}_c = E[W \mid \mathbf{m}_c, \mathbf{X}_c, \hat{\zeta}_c]$. This expectation is often analytically intractable, so computer simulation is used to arrive at the estimator $\tilde{\mu}_c$ presented in this report.[5]

The difference between $\tilde{\mu}$,[6] the estimator of the expected value of $W$ in this report, and the actual level of welfare $W$ can be written as:

$$W - \tilde{\mu} = (W - \mu) + (\mu - \hat{\mu}) + (\hat{\mu} - \tilde{\mu})$$

The first term on the right-hand-side of the equation is called the idiosyncratic error, which is due to the presence of a disturbance term in the consumption model. The second term, the model error, is due to variance in the first-stage estimates of the parameters of the consumption model. The last term, the computation error, is due to using an inexact method to compute $\hat{\mu}$.

The variance in $\tilde{\mu}$ due to idiosyncratic error falls approximately proportionately with the

[5] Analytical results for some welfare measures are found in Elbers, Lanjouw and Lanjouw (2000).
[6] For the sake of notational simplicity, the subscript $c$ will be dropped.

size of the population of households in the commune. In other words, since the component of the prediction error grows as the target population becomes smaller, there is a practical limit to the degree of disaggregation possible. This is precisely the reason village-level estimates were not produced for this report.

The model error is determined by the properties of $\hat{\boldsymbol{\zeta}}_c$ and hence it does not increase or fall systematically as the size of the target population changes. Its magnitude depends, in general, only on the standard errors of the first-stage coefficients and the sensitivity of the indicators to deviations in household consumption. For a given commune, its magnitude will also depend on the distance of the explanatory variables for households in that commune from the level of those variables in the sample data.

The computation error depends upon the computational method used. Using simulation methods with sufficient computational resources and time, this error can be made as small as one desires. When the distribution of $\hat{\boldsymbol{\zeta}}_c$ is known or can be estimated, a Monte-Carlo simulation can be designed to capture both the idiosyncratic error and the model error. The simulated disturbance term $\hat{\mathbf{u}}_c^R$ and the simulated consistent estimator $\hat{\boldsymbol{\zeta}}_c^R$ are drawn for the $R$-th simulation to generate the $R$-th welfare estimate $\hat{W}^R$. The estimator $\tilde{\mu}$ is found by taking the mean of $\hat{W}^R$ over $R$ and the associated standard error can also derived by taking the standard deviation of $\hat{W}^R$.

## SECTION III DATASETS

### The CSES Data Sets and Sampling Frames

To produce the poverty maps, four distinct data sets—two socio-economic survey data sets, a census data set and a GIS data set—were used. The consumption model is built upon the two socio-economic surveys, namely, the CSES 1997 and the CSES 1999. For reasons discussed later, the CSES 1999 was used only for auxiliary purposes. As already explained, the strengths of the surveys are that they have detailed information on consumption. From these data sets, a welfare index of the standard of living of each household can be derived.

The CSES 1997 was conducted by the National Institute of Statistics of the Ministry of Planning under the Capacity Development for Socio-Economic Surveys and Planning Project. The project was funded by the United Nations Development Programme (UNDP) and the Swedish International Development Agency, and executed by the World Bank. The CSES 1997 was the first multi-subject household survey conducted in Cambodia. The questionnaires for CSES 1997 included three substantive components: a village questionnaire, a core questionnaire for households, and a social sector household module. The village questionnaire collected information at the village level, whereas the other two questionnaires were targeted at household-level. The social sector household module focused on health and education.

The sample design for the CSES 1997 treated villages as the primary sampling units and households as secondary sampling units. A sampling frame that was developed for the Socio-Economic Survey of Cambodia 1996 was updated with newly available information for use as the sampling frame for the 1997 survey. Due to security and other considerations, some parts of the country were excluded from the frame. In the CSES 1997, there were three sampling strata: Phnom Penh, Other Urban, and Rural. The total sample size of the CSES 1997 was 6010 households in 474 villages. In the Phnom Penh stratum, a sample was taken from 120 sample villages with 10 households from each village, while in the Other Urban stratum, 10 households from each of 100 villages were sampled. In the Rural stratum, 15 households were sampled in each of 254 villages. For each of the three sampling strata, a consumption model for small-area estimation was constructed.[7]

The CSES 1999 is the second survey conducted under the Capacity Development for Socio-Economic Surveys and Planning Project. The interviews were carried out in two rounds between January and March 1999 and between June and August 1999. As with the CSES 1997, the CSES 1999 had three substantive components. However, instead of the social sector module, the CSES 1999 had an income and employment module. The core questionnaire of the CSES 1999 is similar to that of the CSES 1997. The CSES 1999 has ten sampling strata defined from the urban and rural sectors within each of five zones (Phnom Penh, Plain, Tonle Sap, Coastal and Plateau).

The CSES 1999 was designed to capture seasonal variations in consumption and to reduce the number of field staff involved in data collection and supervision in order to provide them with more intensive training and to exercise more intensive control of field operations (NIS, 2000). Subsequent analysis, however, uncovered a large discrepancy in consumption aggregates between the two rounds. The inconsistencies in measured consumption between the two rounds of the survey indicate the potential presence of widespread and systematic measurement error (Ministry of Planning, 2001). As was discussed in WFP (2001), it was believed, given this concern, that there were four available options for the purposes of this research project: (1) use both rounds of the CSES 1999, (2) use the CSES 1999 round 1 data only, (3) use the CSES 1999 round 2 data only, or (4) use the CSES 1997. In the preliminary analysis presented in WFP (2001), the third alternative was chosen to maintain comparability with the Poverty Profile 1999, which was not yet published. However, after Poverty Profile 1999 (Ministry of Planning, 2001) was published, the forth alternative became the most reasonable option because the poverty estimates from the CSES 1997 seemed more reliable than those from the CSES 1999.

Due to security issues, some parts of Cambodia were not covered in the sampling frame of the CSES 1997. In terms of the number of households, 11.6% of the rural areas and 2.6% of the urban areas were not covered in CSES 1997. Although there remained some areas

[7] National Institute of Statistics (1998) provides further details on the CSES 1997 data set.

not covered in the sampling frame, the corresponding figures for the CSES 1999 were 3.8% for the rural areas and 0.3% for the urban areas. Those numbers are relatively small but not negligible. Hence, it was decided to take advantage of the better geographical coverage of the sampling frame for CSES 1999 to see if the consumption model holds for those households which are located outside the CSES 1997 sampling frame, but inside the CSES 1999 sampling frame.

To do so, the following steps were taken. Firstly, the villages in the CSES 1999 that were excluded from the CSES 1997 sampling frame were identified using data from National Institute of Statistics (1997). Then the parameters of the consumption model were estimated using the CSES 1999 data. Secondly, each record in the CSES 1999 data set was assigned the corresponding stratum code of the CSES 1997. Ideally, two regressions with the same set of regressors should be run separately for the areas inside and outside the sampling frame of CSES 1997 so that the hypothesis that the estimated parameters for those two areas are the same can be tested. However, the sample sizes for the excluded areas were too small to allow one to generate meaningful results. Instead, an alternative approach was taken. The regression was first run without excluded areas and the coefficient $\beta^0$ was estimated. Then another regression was run with excluded areas and the coefficient $\beta^1$ was estimated. The hypothesis $\beta^0 = \beta^1$ was tested. The rejection of this hypothesis would suggest that, if the CSES 1997 had included the excluded areas, the consumption model would have been different.

It should be noted that, in the procedure described above, the CSES 1999 data set does not affect the estimated parameters used in the simulation. This is because there is serious concern about the quality of the CSES 1999 data set. However, as was observed in Ministry of Planning (2001), there were some common patterns between the two rounds. This seems to suggest that the overall pattern of consumption was not altered to the extent that the test described above is unlikely to be invalid. The tests were carried out with CSES 1999 Round 1, CSES 1999 Round 2 and both rounds pooled.

**The Census Data Set and Definition of Household**

The Cambodian National Population Census was conducted over a period of ten days in

March 1998. It was the first population census to be conducted in Cambodia since 1962 and was done on a *de facto* basis. The census covered all persons staying in Cambodia, including foreigners, at the reference time, which was the midnight of March 3, 1998. Foreign diplomatic corps and their families were, however, excluded. Special arrangements were made to enumerate homeless populations. Prior to the census, a complementary project to increase publicity and obtain the cooperation of the population was executed by UNESCO with UNFPA funding. There was also a preliminary house-listing operation before the census was conducted.

The census questionnaire consisted of two forms, Form A (the house list) and Form B (the household questionnaire). The materials of walls, roof and floors for each house were observed by the enumerator and recorded in Form A together with other information. Form B had four parts. Part 1 collected information on usual household members present and absent on census night as well as visitors present on the census night. Part 2 gathered specific information on each usual household member and visitors present on the reference night, including full name, relationship to the head of household, sex, age, marital status, mother tongue, religion, birth place, migration, literacy, education and employment. Part 3 contains questions on fertility of females ages 15 and over. Part 4 contained housing characteristics, conditions and other facilities.

The geographical frame for the census followed the defined structure of province, district, commune, and village in descending order of aggregation. There are 24 provinces in Cambodia, including the municipality of Phnom Penh, and the towns of Kep, Sihanoukville, and Krong Pailin. A few areas were not covered during the census due to military operations, which were (i) the entire districts of Anlong Veaeng in Otdar Mean Chey province, Samlot in Bat Dambang province and Veal Veaeng in Pousat province and (ii) Ou Bei Choan village of Ou Chro district in Banteay Mean Chey province. The population in these excluded areas is estimated to be about 45,000 (National Institute of Statistics, 2000b). Table 1 summarises the coverage of the census.[8] Since it is not possible to estimate poverty measures for the excluded areas, only the 1,594 communes included in the census are analysed in this report.

---

[8] Further information on the implementation of the census can be found in National Institute of Statistics (2000b)

**Table 1 The number of administrative units included and excluded in the census.**

| | Included in Census | Excluded from Census | Total |
|---|---|---|---|
| Province | 24 | 0 | 24 |
| District | 180 | 3 | 183 |
| Commune | 1594 | 15 | 1609 |
| Village | 13339[9] | 67 | 13406 |

Prior to the computer simulation, two treatments were applied to the census data set. Firstly, special settlements were excluded from the data set. Special settlements are groups of people who were found together on the census night. They are transitional and may not necessarily live in the commune. Hence, they were not included for the calculation of the poverty estimates.

Secondly, there is a practical inconsistency between the definitions of household used in the census and survey data sets. Even though the census data set distinguishes between usual members of household and visitors in Form A, Form B Part 2 includes both as long as they were present on the reference night, and makes no distinction between them. This means that the data user has to take the usual members of the household as well as visitors present on the census night as the household. The survey, however, asks questions about the usual members of the household, including those absent at the time of survey. Moreover, there were households that did not appear to be regular households. For example, there were households with more than 100 people.

Hence a practical decision was made to take care of this issue. Only those households for which the household size is less than 16 and the number of visitors is less than 10 were used for the analysis. The original data set contained 2,162,086 regular households, and it was reduced to 2,150,235 households as a result of those treatments. Admittedly, the decision may have been somewhat arbitrary. However, it seemed more reasonable to make such a distinction than to ignore the issue. More importantly, the exclusion does not affect the main findings of the report significantly because it was much less than 1% of the data set that was dropped.

[9] The number of villages does not include special settlements. There were 411 special settlements.

**The GIS Data Set**

A set of geographically derived indicators were also used in this analysis. These indicators included distance calculations, land use and land cover information, climate indicators, vegetation, agricultural production, and flooding. A number of data sets from various sources were compiled into a GIS and the geographic indicators were generated for all villages and communes in Cambodia. Very coarse resolution data was summarised at the commune-level, while high resolution data was attributed to individual villages. Distances from villages to roads, other towns, health facilities, and major rivers were calculated. Indicators based on satellite data with varying temporal resolutions included land use within the commune (agricultural, urban, forested, etc.), a vegetation greenness indicator to proxy agricultural productivity, and the degree to which the area was lit by nighttime lights as a proxy of urbanization. Relatively stable indicators including soil quality, elevation, and various 30-year average climatological variables were also generated from other composite data sets.

The sources as well as the spatial and temporal dimensions of the data sets vary. Some datasets were assumed to not have changed greatly over time. Others, where multi-temporal data was available, included both yearly and monthly indicators as well as change and long term average indicators. Road, river, village location, and administrative boundary data were obtained under a UNTAC project and updated in 1996 by the Department of Geography under a UNDP sponsored CARERE project. Health facility latitude and longitude locations were provided by the World Health Organisation. Land use and land cover data were obtained from Landsat Thematic Mapper (TM) satellite for 1993 and 1997 at 50 meter resolution. Agricultural production data at the commune-level was taken from the commune-level crop assessment database prepared by WFP. NASA's Advanced Very High Resolution Radiometer (AVHRR) satellite data at 7km resolution was used to generate the Normalized Differential Vegetation Index (NDVI). A 19-year monthly series of AVHRR-derived NDVI data, covering 1981 - 2000 and compiled by Clark Labs, was used to generate the NDVI values. NDVI indicators included monthly values, 19 year average and standard deviation, and coefficient of variation (19 year standard deviation divided by the average).

A global digital elevation model at 1km resolution, GTOPO30, was used for elevation values. GTOPO30 was developed under the coordination of the US Geological Survey, in collaboration with NASA, UNEP/GRID, USAID and others. City lights satellite data at 1km resolution was collected during 1994-95 by the Defense Meteorological Satellite Programme and obtained from the National Geophysical Data Center. The soil quality data is based on a reclassification of the FAO/UNESCO Soils Map of the World, which contains 106 soil type classes. The USDA Natural Resources Conservation Service and the University of Puerto Rico overlaid the FAO/UNESCO map with a global climate dataset, and using the combined climate and soils data, reclassified the FAO/UNESCO soils map according to suitability for food production. The University of East Anglia Climatic Research Unit's Global Climate Dataset was obtained from the Intergovernmental Panel on Climate Change. These 30-year monthly averages, interpolated into five degree grids, are based on daily weather station data collected from 1961 to 1990.

## SECTION IV IMPLEMENTATION

### Comparing Survey Data with Census Data

As already discussed, the basic idea of the small area estimation is to find the consumption for each household in the census so that it can be used to estimate the poverty rates. Although, in principle, any level of aggregation is possible, reliable estimates cannot be derived without aggregating up to a certain level. This is because the disturbance terms cancel each other out as the level of aggregation goes up and the estimate becomes less subject to random errors.

The consumption aggregate is estimated for each household in the census using the location-effect variables and household-level variables. The former refers to the variables that are common among the households in the same village or commune. The census mean and GIS data are in this category. An example of a census mean variable is the literacy rate of the head of household in the village. An example of a GIS variable is the distance to a main road. Household-level variables refer to variables that are specific to a household. For example, the age of the household head and the material of the walls are in this category. Inclusion of the location-effect variable is straightforward as the same data set is merged into the survey and census. In other words, there is no need to be concerned about the comparability between the census and survey with regard to location-effect variables.

The census and the survey were designed and implemented by different groups of people for different purposes. Therefore, it cannot be taken for granted that the data are directly comparable. Moreover, the census and the survey often use different questionnaire formats for the same questions, leading to possible data inconsistencies. Therefore, care should be taken when the variables based upon such questions are used for poverty mapping.

Before checking comparability, the census and survey questionnaires were carefully examined to find out which question in the census corresponds to which question in the survey. The exact phrasing of questions often differed between the two, making it necessary to assess the impact of question phrasing on data set comparability. Summary statistics of the two data sets were examined for this purpose.

In what follows, quantitative variables and qualitative variables are distinguished. Qualitative variables are those variables that are measured by the nominal scale, whereas quantitative variables are measured by the ordinal, interval or ratio scale. For example, the sex of the head of household is coded as 1 for male and 2 for female. These numbers in themselves carry no meaning. They may have been 4 and –3 respectively instead. On the other hand, the age of the head of household ranges from under 20 to over 90, and the number does have a meaning.

For qualitative variables, recoding is often needed as different questionnaires have different coding schemes. For example, in both the census and the survey, there is a question about toilets. However, the way the question was asked was considerably different in each. In the census, the respondents were asked whether or not a toilet is available within the premises. On the other hand, the CSES questionnaire asked what toilet facility the household has, with options including a septic tank, a pit latrine, other without septic tank, a public toilet and nothing. In such cases, recoding is required.

Once recoding is completed, the two data sets can be compared. For qualitative variables, the frequency with which each value occurred is compared. For quantitative variables, the mean, maximum, minimum and standard deviation are compared. The comparison was carried out for each stratum in the CSES 1997 sampling frame. Through this process, the above-mentioned peculiarities regarding household size and other minor recoding problems were discovered. These problems were then fixed. It is very important to carry out a thorough check of data set consistency since the derived estimates may be very different from what is predicted from the survey.

**The Choice of Consumption Aggregate and Poverty line**

The recent socio-economic surveys conducted in Cambodia were in 1997 and 1999. Given that the census was taken in 1998 and hence both survey years are one year apart from the census year and given that the CSES 1999 had a better sampling frame, it seemed more desirable to use CSES 1999 as was done in WFP (2001). However, as discussed already, the CSES 1999 had serious data problems, which did not come to light until after the publication of Ministry of Planning (2001). Hence, in this report, the CSES 1997 data is used.

Using the CSES 1997 to define the consumption aggregate was not as straightforward as it initially seemed. There were two possible alternatives. One alternative was to use the adjusted consumption aggregate derived by Knowles (1998) while the other alternative was the unadjusted consumption aggregate defined in Ministry of Planning (2001). When Ministry of Planning (1998) was published, the data set contained errors which necessitated the use of the adjusted consumption aggregate. The mistakes were subsequently corrected and hence the adjustments made by Knowles are unnecessary for this report. Therefore, this report follows the definition of unadjusted consumption given in Ministry of Planning (2001).

To ensure comparability with the publicised benchmark national poverty rate of 36.1%, the poverty line was redefined so that the same poverty rates could be reproduced using the unadjusted consumption aggregate for each of the three strata. As a result, the poverty lines, in terms of per capita per day consumption, employed in this report are 1,629 Riels for Phnom Penh, 1,214 Riels for Other Urban and 1,036 Riels for Rural.[10] Table 2 compares poverty lines given in different published reports. Poverty maps using the poverty lines for unadjusted consumption given in Ministry of Planning (2001) (CSES 1997 Unadjusted in Table 2) are also reported.[11]

[10] The average official exchange rate in 1998 was USD 1=3807.8 Riel. The purchasing power parity (PPP) conversion factor to official exchange rate ratio in 1998 was 0.189 (Based on the World Development Indicators). When domestic price differences are ignored, the poverty lines for Phnom Penh, Other Urban and Rural translate into 2.30, 1.72 and 1.46 in terms of PPP USD respectively.

[11] Poverty maps corresponding to the poverty lines in the last column are given in 0. Poverty maps with CSES 1997 Unadjusted poverty lines are provided in APPENDIX H.

**Table 2 Comparison of poverty lines across different years**

| Total Poverty line | SESC 1993/94 | CSES 1997 Adjusted | CSES 1997 Unadjusted | CSES 1999 Round 2 | **This report** |
|---|---|---|---|---|---|
| Phnom Penh | 1,578 | 1,819 | 1,923 | 2,408 | **1,629** |
| Other Urban | 1,264 | 1,407 | 1,398 | 2,008 | **1,214** |
| Rural | 1,117 | 1,210 | 1,195 | 1,751 | **1,036** |

*Note :* Figures for SESC 1993/94 were taken from Prescott and Pradhan (1997), CSES 1997 Adjusted from Ministry of Planning (1998), CSES 1997 Unadjusted and CSES 1999 Round 2 from Ministry of Planning (2001). All figures are in Riels and in terms of per day *per capita* consumption.

By construction, the poverty rate for each stratum in this report is the same as given in Ministry of Planning (1998). However, there is no guarantee that the poverty gap and the poverty severity are the same. Hence checking these indices provides an indication of how important the choice between the adjusted and unadjusted measures is. The poverty gap was estimated at 8.9% and the poverty severity at 3.2% in this report. The corresponding figures in Ministry of Planning (1998) were 8.7% and 3.1% respectively. Although these numbers are not exactly the same, the differences are small enough to be considered random errors. This seems to suggest that the analysis presented in this report will be robust with respect to the choice of consumption aggregate. Table 3 compares the poverty measures from different years.

**Table 3 Comparison of poverty measures across different years.**

| Poverty measures | SESC 1993/94 | CSES 1997 Adjusted | CSES 1997 Unadjusted | CSES 1999 Round 2 | **This report** |
|---|---|---|---|---|---|
| Poverty Incidence | 39.0 | 36.1 | 47.8<br>(1.5) | 35.9<br>(2.4) | **36.1**<br>**(1.6)** |
| Poverty Gap | 9.2 | 8.7 | 13.7<br>(0.7) | 6.5<br>(0.7) | **8.9**<br>**(0.6)** |
| Poverty Severity | 3.1 | 3.1 | 5.3<br>(0.3) | 2.0<br>(0.4) | **3.2**<br>**(0.3)** |

*Note :* For sources, see Table 2. The figures in parenthesis are the standard errors, which take into account the stratification, sampling weights and clustering. Standard errors were not reported in Prescott and Pradhan (1997) and Ministry of Planning (1998). All measures are percentages.

## The Consumption Model

As was discussed above, the CSES 1997 has three strata and is intended to be representative at that level. A predetermined number of villages were randomly chosen in each stratum, and 10 to 15 households were sampled in each village. Hereafter, subscripts

$v$ and $h$ are used to denote a village and a household respectively. For example, the expansion factor will be denoted as $l_{vh}$.

The first step in creating a poverty map is developing an accurate empirical model of household consumption. The following consumption model is estimated:

$$\ln y_{vh} = E[y_{vh} \mid \mathbf{x}_{vh}^T] + u_{vh} = \mathbf{x}_{vh}^T \beta + \eta_v + \varepsilon_{vh}.$$

where $y_{vh}$ is the *per capita* consumption and $u_{vh}$ is the disturbance term, which is the sum of the common component $\eta_v$ and the idiosyncratic component $\varepsilon_{vh}$. These two components, $\eta_v$ and $\varepsilon_{vh}$, are assumed to be independent of each other and uncorrelated with observable household characteristics $\mathbf{x}_{vh}$. This specification allows for an intra-cluster correlation in the disturbances and heteroscedasticity in $\varepsilon_{ch}$. Explicit treatment of the location effects is important as some of the effects of location may remain unexplained even with a rich set of regressors. The household characteristics $\mathbf{x}_{vh}$ in this model are not limited to variables that are specific to the household. They can also include the characteristics of the village in which the household is located. For example, $\mathbf{x}_{vh}$ can include the village-level means of the census data and the GIS data, which capture a part of the location effects. The details of this regression are found in APPENDIX D. Cross-terms between a household-level variable (*e.g.* age of the household head) and a GIS variable were also included. For notational convenience, the variance of a random variable will be hereafter denoted as $\sigma_{\bullet}^2 \equiv Var[\bullet]$. When $\bullet$ has a subscript $s$, it is expressed using a comma as $\sigma_{\bullet,s}^2 \equiv Var[\bullet_s]$.

Elbers, Lanjouw and Lanjouw (2001) point out that, for any given disturbance variance $\sigma_{u,vh}^2 = \sigma_{\eta,vh}^2 + \sigma_{\varepsilon,vh}^2$, the greater the fraction due to the common component, the less one enjoys the benefits of aggregating over more households within a country. To assess the performance of the consumption model, a number of diagnostic statistics are checked.[12] Since unexplained location effects reduce the precision of poverty estimates, the first goal is to explain the variation in consumption due to location as far as possible with the choice and construction of $\mathbf{x}_{vh}$. Location means of household-level variables derived from the census data are particularly useful for this purpose.

[12] See APPENDIX E for details.

The next step is to estimate each component of the disturbance term. First, the residual term $\hat{u}$ was derived from the OLS regression. The common component $\eta_v$ was estimated non-parametrically at the average of $\hat{u}$ in the cluster as follows:

$$\hat{u} = \hat{u}_{v\cdot} + (\hat{u}_{vh} - \hat{u}_{v\cdot}) = \hat{\eta}_v + e_{vh}$$

where $\hat{u}_{v\cdot}$ is the average of $\hat{u}$ over the households in the same village. To model heteroscedasticity in the idiosyncratic part of the residual, a restricted number of household characteristics, $z_v$, that best explain variation in $e_{vh}$ out of potential explanatory variables, their squares and interactions were chosen. The following logistic model of the variance of $\varepsilon_{vh}$ conditional on $z_{vh}$, bounding the prediction between zero and a maximum, $A \equiv 1.05 \times \max_{v,h}\{e_{vh}^2\}$, was estimated:

$$\ln\left[\frac{e_{vh}^2}{A - e_{vh}^2}\right] = z_{vh}^T\alpha + r_{vh}$$

Letting $B_{vh} \equiv \exp\{z_{vh}^T\hat{\alpha}\}$ and using the delta method, the model implies a household specific variance estimator for $\varepsilon_{vh}$ of

$$\hat{\sigma}_{\varepsilon,vh}^2 = \left[\frac{AB_{vh}}{1 + B_{vh}}\right] + \frac{1}{2}\hat{\sigma}_r^2\left[\frac{AB_{vh}(1 - B_{vh})}{(1 + B_{vh})^3}\right]$$

Detailed results for this residual regression are again found in APPENDIX D. Once $\hat{\sigma}_{\varepsilon,vh}^2$ is computed, the household residuals are standardised as follows:

$$e_{vh}^* = \frac{e_{vh}}{\hat{\sigma}_{\varepsilon,vh}} - \left[\frac{1}{H}\sum_{vh}\frac{e_{vh}}{\hat{\sigma}_{\varepsilon,vh}}\right]$$

where $H$ is the number of households in the survey. Before proceeding to conduct the simulation, the estimated variance-covariance matrix, $\hat{\Sigma}$, was weighted by $l_{vh}$ to obtain GLS estimates of the first-stage parameters, $\hat{\beta}_{GLS}$, and their variance $Var(\hat{\beta}_{GLS})$.

**Simulations**

From the consumption model, $\hat{\alpha}$ , $\hat{\beta}_{GLS}$ , and their associated variance-covariance matrices as well as the empirical distribution of $e^{*}_{vh}$ and $\hat{\eta}_v$ are obtained. Assuming multivariate normal distribution, for the $R$-th simulation $\tilde{\alpha}^R$ and $\tilde{\beta}^R$ are drawn. Once $\tilde{\alpha}^R$ is drawn, the household-specific variance of the household component of disturbance, $(\tilde{\sigma}^2_{\varepsilon,vh})^R$, is estimated for each census household. Then, the error terms are drawn in two stages to take clustering into account. The location-specific error $\tilde{\eta}^R_v$ is drawn from the empirical distribution of $\hat{\eta}_v$. Then the household component $\varepsilon^R_{vh}$ is obtained with a draw from the empirical distribution of $e^{*}_{vh}$ in the corresponding cluster (*i.e.* village) and $(\tilde{\sigma}^2_{\varepsilon,vh})^R$. The simulated value of consumption $\hat{y}^R_{vh}$ for household $h$ in village $v$ is, therefore,

$$\hat{y}^R_{vh} = \exp(\mathbf{x}^T_{ch}\tilde{\beta}^R + \tilde{\eta}^R_v + \tilde{\varepsilon}^R_{vh})$$

The full set of simulated $\hat{y}^R_{vh}$ is used to compute the $R$-th estimate of poverty measures for each commune except for some outliers. For example, the $R$-th estimate of poverty incidence for commune $c$, $\hat{I}^R_c$, is computed as follows:

$$\hat{I}^R_c = \frac{1}{n_c}\sum_{v\in V_c}\sum_{h\in H_v}\mathrm{Ind}(\hat{y}^R_{vh} < z)\cdot n_{vh}$$

where $V_c$ denotes the set of villages in commune $c$, $H_v$ the set of households in village $v$, $n_{vh}$ the size of household $h$ in village $v$, $z$ the poverty line, $n_c$ the population of commune $c$, and $\mathrm{Ind}(\bullet)$ is an indicator function. For this paper, the simulation was repeated 100 times. The mean and standard deviation of the estimates of poverty measures from each simulation were computed to arrive at the commune-level estimates of poverty measures and their associated standard errors. In a similar manner, poverty measures at more aggregated levels, such as district, province and stratum, can be estimated.

# SECTION V RESULTS

## Creating Poverty Maps

Once the commune-level estimates of poverty measures are computed, it is straightforward to create poverty maps. The polygon data for communes are combined with poverty estimates by the GIS. The map presented on the following page is the refined version of the previous poverty mapping exercise (WFP, 2001). A number of other maps are also presented in the appendices to this report.

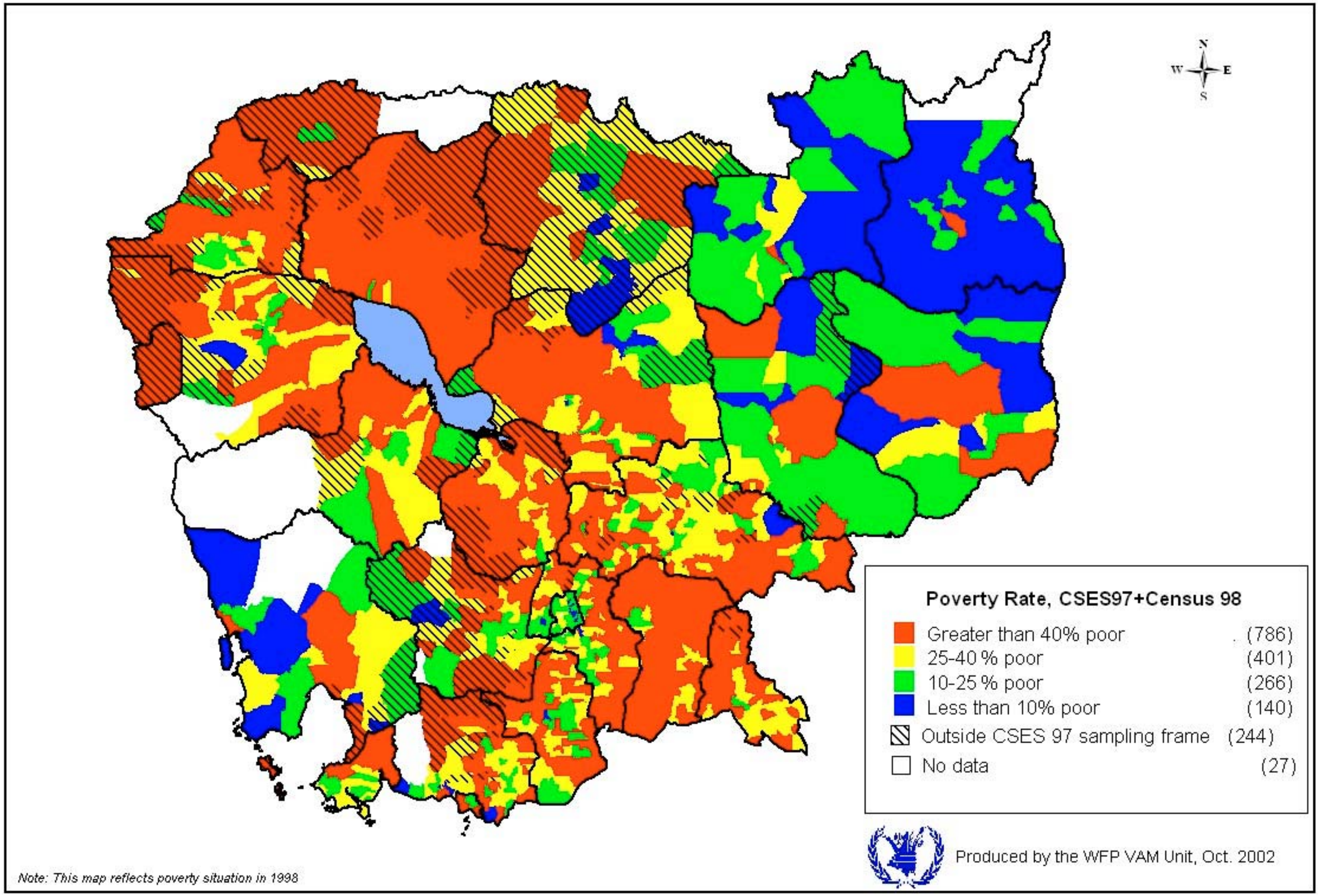
## Are the Villages Excluded from the CSES 1997 Sampling Frame Different?

As noted before, the sampling coverage of the CSES 1997 is smaller than that of the CSES 1999. Hence, the CSES 1999 data was used to check if the consumption model applies to those areas, which were excluded from the CSES 1997 sampling frame but included in the CSES 1999 sampling frame. Table 4 provides a summary of the sampling frame of 1997. Unfortunately the number of samples from those excluded areas was too small to meaningfully compare the equality of the coefficients for the included and excluded areas. Hence, as discussed before, the hypothesis that the estimated regression coefficients with and without excluded areas are the same was tested.

**Table 4 Summary of the number of villages in CSES 1999 included and excluded from the sampling frame of CSES 1997**

| Stratum | Round 1 | | Round 2 | | Total | |
|---|---|---|---|---|---|---|
| | Excluded | Included | Excluded | Included | Excluded | Included |
| Phnom Penh | 0 | 600 | 0 | 600 | 0 | 1200 |
| Other Urban | 50 | 810 | 10 | 850 | 60 | 1660 |
| Rural | 180 | 1360 | 110 | 1430 | 290 | 2790 |
| Cambodia | 230 | 2770 | 120 | 2880 | 450 | 5650 |

Poverty Mapping Refinement 2002, CSES 1997 and Census 1998



If the relationship between the right-hand-side variables and consumption is kept intact in each round, in principle, the same conclusion should be derived. However, the results obtained in this study are mixed. For the Other Urban stratum, the hypothesis was rejected at the significance level of 1% when Round 1 data or pooled data (*i.e.* Round 1 and Round 2) was used. However, the hypothesis could not be rejected even at the significance level of 10% when Round 2 data was used. For the Rural stratum, the pooled sample could not reject the hypothesis but the Round 1 and Round 2 data both sets rejected it when used separately. Although these results are inconclusive, there is a good reason for caution when using the estimates for the excluded areas.[13]

**How Accurate are the Estimates?**

The preliminary poverty map presented in WFP (2001) uses a different combination of data sets from that used in this report, and does not record standard errors. It is, therefore, not possible to compare the commune-level estimates directly. It makes more sense to compare at the stratum level, because it is possible to use the poverty measures in poverty profiles as benchmarks for comparison. Table 5 compares the stratum-level poverty rates obtained in WFP (2001) and in this report for survey data estimates (CSES only) and for estimates obtained by combining the survey and census (CSES + census). The latter data are the stratum-level poverty estimates that are consistent with the poverty maps. For the sake of comparison, the WFP (2001) Model 2 was used as it also has three consumption models for each stratum. It should be noted that the WFP (2001) Model 1 was used to create the poverty maps in WFP (2001), and poverty rates were substantially underestimated for all three strata.

[13] For example, Krong Pailin was the poorest province in the map. It is, however, generally considered to be a non-poor area. This may be because Krong Pailin was out of the sampling frame of the CSES 1997. To the best of our knowledge, the hatched areas in the map in page 28 were outside the sampling frame of the CSES 1997.

**Table 5 Stratum level comparison of estimates of poverty rates in percentage**

| | Stratum | CSES Only | | CSES + census | |
|---|---|---|---|---|---|
| This report | Phnom Penh | 11.1 | (1.8) | 11.9 | (1.2) |
| | Other Urban | 29.9 | (3.3) | 30.0 | (1.0) |
| | Rural | 40.1 | (2.0) | 43.0 | (1.1) |
| WFP (2001) Model 2 | Phnom Penh | 9.7 | | 12.5 | |
| | Other Urban | 25.2 | | 26.7 | |
| | Rural | 40.1 | | 49.4 | |

*Note:* CSES 1997 was used for this report and CSES 1999 Round 2 for WFP (2001). The figures in brackets are standard errors. Standard errors was calculated by the author. The standard errors for CSES Only take into account clustering and expansion factors.

There are two observations that can be made. Firstly, the patterns of poverty rates in this report and in WFP (2001) are quite similar regardless of the data sets used. The poverty rate in Phnom Penh is around 10-12%, Other Urban is 25-30% and Rural is 40-50%. Secondly, when the differences between CSES Only and CSES + census for this report and for WFP (2001) are compared, there are much smaller discrepancies for this report. This suggests that the accuracy of estimates has improved substantially since WFP (2001). The difference between CSES Only and CSES + census for this report is small enough to be attributed to the random error.

The level of accuracy of the commune-level estimates varies from commune to commune. For example, the standard errors associated with commune-level estimates range between 0.1% and 22.6%. Table 6 provides summary statistics on the standard errors. The first column (Mean S.E.) is the simple average of the standard errors. Urban areas have lower standard errors. The median of standard errors is presented in the second column (Median S.E.). The third column (S.E. Ratio) is the average of the ratio of the standard error to the point estimate. The fourth (Avg # HH) and fifth columns (# Commune) provide the average number of households in the commune and the number of communes in the stratum respectively.

**Table 6 Summary statistics of the standard errors associated with commune-level poverty rate estimates**

| | Mean S.E. | Median S.E. | S.E. Ratio | Avg # HH | # Commune |
|---|---|---|---|---|---|
| Phnom Penh | 4.0 | 3.5 | 35.7 | 2169 | 76 |
| Other Urban | 5.0 | 4.9 | 23.8 | 1345 | 159 |
| Rural | 7.9 | 7.6 | 27.4 | 1289 | 1359 |
| Cambodia | 7.4 | 7.2 | 27.4 | 1337 | 1594 |

The first three columns provide a general picture of the levels of accuracy. The standard errors are low enough for the results to be useful as proxies, but are high for a number of communes, so policy-analysts should take the estimates with caution. At the same time, it should be noted that none of the summary statistics above are perfect. For example, a relatively high level of standard error may not matter if the point estimates are high enough. A commune with the point estimate of the poverty rate of 95% and standard error of 15% is clearly a very poor commune. On the other hand, even if the ratio of the standard error to the point estimates is high, it does not matter when the absolute value of the standard error is low. If the point estimate and standard error were 0.1%, then the commune is not a poor commune. In practical terms, the size of the commune is also important. Provided that the cost of a programme increases in proportion to the size of the commune, mis-targeting for small communes is a relatively minor issue in terms of efficient use of resources. The statistics above do not incorporate the size of the commune.

One way to address these issues is to define a poor commune and non-poor commune by the ratio of the difference between the poverty estimate and a reference level to the standard error. If a poor commune is defined as a commune whose point estimate is higher than the national poverty rate by at least two times the standard error, and if a non-poor commune has the opposite definition, then 48% of all communes can be classified as either poor or non-poor. When a commune cannot be classified, the communes can be aggregated to make the standard error smaller. A poverty map that incorporates both the standard error and the point estimate is presented in APPENDIX G. It should be noted that the standard errors tend to be higher for communes with smaller populations.

Although the magnitudes of the standard errors are not small enough to be ignored, and can be quite high for some communes, the commune-level estimate is accurate enough to make a sharp comparison with the national poverty rate for half of the communes. Even for other communes, the estimates provide useful information for targeting, especially when multiple communes are taken together. It is likely, for example, that net gains from targeting poorer-than-national-average communes are positive. Although the usefulness of the estimates depends upon the purpose to which they are put, given that reliable poverty estimates have previously been produced only at the stratum level, commune-level estimates with this level of accuracy are still very useful. Even when the estimates need to be made at a more aggregated level such as district or even province to reduce standard errors, the usefulness of the estimates from this exercise will not be undermined as no other reliable estimates are available at this level. Table 7 shows that the estimates of poverty measures at provincial level are comparable with the stratum-level estimates obtained from the survey data alone.

## Table 7 Provincial level estimates of poverty measures.

| No. | Province Name | Poverty Rate | Poverty Gap | Poverty Severity | % of Total Population | Poverty Share | Est. # of Poor People |
|---|---|---|---|---|---|---|---|
| 1 | Banteay Mean Chey | 40.88 | 12.79 | 5.63 | 5.09 | 7.55 | 228.8 |
| | | (2.79) | (1.23) | (0.65) | | | |
| 2 | Battambang | 26.41 | 7.34 | 2.93 | 6.84 | 6.55 | 198.7 |
| | | (1.68) | (0.70) | (0.36) | | | |
| 3 | Kampong Cham | 12.07 | 3.06 | 1.14 | 14.33 | 6.27 | 190.1 |
| | | (1.71) | (0.62) | (0.32) | | | |
| 4 | Kampong Chhnang | 44.60 | 12.54 | 4.97 | 3.67 | 5.93 | 179.9 |
| | | (2.43) | (1.01) | (0.50) | | | |
| 5 | Kampong Speu | 18.18 | 4.26 | 1.52 | 5.27 | 3.47 | 105.3 |
| | | (2.28) | (0.77) | (0.35) | | | |
| 6 | Kampong Thom | 29.07 | 7.89 | 3.11 | 4.97 | 5.24 | 158.9 |
| | | (2.22) | (0.91) | (0.47) | | | |
| 7 | Kampot | 18.67 | 4.68 | 1.72 | 4.74 | 3.21 | 97.3 |
| | | (2.42) | (0.95) | (0.48) | | | |
| 8 | Kandal | 18.40 | 4.62 | 1.72 | 9.51 | 6.34 | 192.3 |
| | | (2.01) | (0.74) | (0.37) | | | |
| 9 | Koh Kong | 8.16 | 2.52 | 1.14 | 0.97 | 0.29 | 8.7 |
| | | (1.28) | (0.56) | (0.33) | | | |
| 10 | Kracheh | 38.59 | 11.68 | 4.94 | 2.31 | 3.23 | 97.8 |
| | | (2.44) | (1.03) | (0.57) | | | |
| 11 | Mondol Kiri | 19.87 | 5.60 | 2.29 | 0.28 | 0.20 | 6.2 |
| | | (5.23) | (1.85) | (0.88) | | | |
| 12 | Phnom Penh | 11.92 | 2.88 | 1.16 | 8.35 | 3.61 | 109.4 |
| | | (1.19) | (0.36) | (0.16) | | | |
| 13 | Preah Vihear* | 29.06 | 7.34 | 2.72 | 1.03 | 1.08 | 32.9 |
| | | (4.21) | (1.41) | (0.60) | | | |
| 14 | Prey Veng | 53.14 | 15.85 | 6.51 | 8.45 | 16.28 | 493.6 |
| | | (1.84) | (0.94) | (0.52) | | | |
| 15 | Pursat | 40.74 | 11.75 | 4.79 | 3.13 | 4.63 | 140.2 |
| | | (2.50) | (1.01) | (0.50) | | | |
| 16 | Rotanak Kiri | 8.81 | 2.86 | 1.41 | 0.83 | 0.27 | 8.1 |
| | | (2.52) | (0.70) | (0.36) | | | |
| 17 | Siem Reap | 53.73 | 19.13 | 9.05 | 6.04 | 11.77 | 356.8 |
| | | (2.03) | (1.17) | (0.74) | | | |
| 18 | Krong Preah Sihanouk | 34.12 | 10.67 | 4.81 | 1.34 | 1.66 | 50.4 |
| | | (2.56) | (1.15) | (0.68) | | | |
| 19 | Stueng Treng | 16.37 | 3.85 | 1.42 | 0.70 | 0.42 | 12.6 |
| | | (2.75) | (0.75) | (0.31) | | | |
| 20 | Svay Rieng | 43.49 | 11.81 | 4.62 | 4.30 | 6.78 | 205.5 |
| | | (3.21) | (1.26) | (0.62) | | | |
| 21 | Takeo | 15.22 | 4.29 | 1.92 | 7.05 | 3.89 | 117.9 |
| | | (1.98) | (0.64) | (0.38) | | | |
| 22 | Otdar Mean Chey* | 39.05 | 13.55 | 6.26 | 0.57 | 0.81 | 24.5 |
| | | (3.01) | (1.68) | (1.01) | | | |
| 23 | Krong Keb | 48.97 | 17.67 | 8.79 | 0.18 | 0.32 | 9.6 |
| | | (5.17) | (2.92) | (1.88) | | | |
| 24 | Pailin* | 97.24 | 61.89 | 42.43 | 0.06 | 0.20 | 6.0 |
| | | (2.26) | (6.84) | (7.46) | | | |

*Note:* The figures in brackets are standard errors. All the figures except for # of poor people are expressed as percentages. # of poor people is expressed in thousands. Poverty share is the ratio of the number of poor people in the province to the total number of poor people in the country. The provinces marked with an asterisk are completely excluded from the sampling frame of CSES 1997. See pages 27-29 for details.

## Extensions

The focus so far has been on poverty measures for the entire population. It is also possible to create maps with other measures that can be derived from consumption. For example, policy-makers may be interested in inequality measures. An inequality map is given in APPENDIX I. It is also possible to derive poverty maps for specific target groups.[14] Poverty maps for women and children under the age of five years are presented in APPENDIX J and APPENDIX K respectively. These were derived by using the number of females in the household or the number of children under the age of five years instead of the total household size as the census weight.

[14] Elbers et al. (2001) do not discuss this possibility. The results should be taken as experimental.

## SECTION VI CONCLUSION

The poverty maps provide invaluable information to policy-analysts and decision-makers. When there is no reliable information for identifying the poor, targeting policies, if they exist at all, are likely to be inefficient and subject to arbitrary political influences. To deliver assistance to those most in need, policy should be formulated based upon reliable information. The poverty maps presented in this report give such information at the commune-level. Moreover, the power of poverty maps is multiplied when they are combined with other maps such as education and nutrition maps using GIS.[15] For example, by overlaying poverty, education and nutrition maps, a policy analyst can identify areas in which poor children are suffering from malnutrition and cannot go to school because of poverty. School feeding programs, for instance, are most likely to be successful if targeted toward such areas.[16]

The commune-level estimates of poverty rates presented in this report are reliable enough to be useful. However, it should also be noted that there are errors associated with the estimates and they may be very large for a number of communes. Moreover, the picture depicted here reflects the conditions as of 1998.[17] It should be recalled that analysis of poverty is never static and thus efforts to acquire up-to-date information and monitor changes in poverty will be indispensable for enabling the efficient, effective and timely delivery of assistance. Hence, policy-makers should not be misled by the intuitive appeal of poverty maps. The poverty maps presented here can and will serve as a sound basis to formulate targeting policy, but cannot and should not be taken as the sole basis. Whenever possible, other maps and data sources as well as observations from the field should be incorporated in the analysis. This is particularly true for the areas outside the sampling frame of CSES 1997, which are indicated by hatch marks on the maps.

[15] There are projects in progress for an education map and a nutrition map in Cambodia.
[16] Bigman, D. and Fofack, H. (2000) give an excellent overview of geographical targeting.
[17] Update using the small area estimation technique will not be possible until new survey data *and* new large-sample data are available.

## REFERENCES


Alderman, H. Babita, M., Demombynes, G., Makhatha, N. and Özler, B. (2001) "How Low Can You Go? Combining Census and Survey Data for Mapping Poverty in South Africa." mimeo. The World Bank.

Bigman, David and Fofack, Hippolyte (2000) "Geographical Targeting for Poverty Alleviation: An Introduction to the Spatial Issue." *World Bank Economic Review*. 14(1) January 2000 pp129-45.

Demombynes, G., Elbers, C., Lanjouw, J., Lanjouw, P., Mistiaen, J. and Özler, B. (2001) "Producing a Better Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries." mimeo. The World Bank.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2000) "Welfare in Villages and Towns." *Timbergen Institute Discussion Paper*, TI 2000-029/2.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2001) "Welfare in Villages and Towns: Micro-Level Estimation of Poverty and Inequality." mimeo. The World Bank.

Elbers, C., Lanjouw, J. O., Lanjouw, P. and Leite (2001) "Poverty and Inequality in Brazil' New Estimates from Combined PPV-PNAD Data." mimeo, The World Bank.

Foster, J.E., Greer, J. and Thorbecke, E. (1984) "A Class of Decomposable Poverty Indices." *Econometrica* 52: 761-766.

Greene, W. H. (2000) "Econometric Analysis Fourth Edition." Prentice Hall.

Hoddinott, John (1999) "Targeting: Principles and Practice." Technical Guide #9 International Food Policy Research Institute. Washington D.C.

Jalan, Jyotsuna and Ravallion, Martin (1998) "Transient Poverty in Postreform Rural China." *Journal of Comparative Economics*. 26 338-357.

Jalan, Jyotsuna and Ravallion, Martin (2000) "Is Transient Poverty Different? Evidence for Rural China." *Journal of Development Studies*. 36(6) 82-99.

Knowles J. (1998) "An Updated Poverty Profile for Cambodia, 1997: Technical Report." Capacity Development for Socio-Economic Surveys and Planning Project. Ministry of Planning, Cambodia.

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

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

National Institute of Statistics (1997) "Selection of Sample Villages Prepared for Cambodia Socio-Economic Survey 1997." National Institute of Statistics, Ministry of Planning.

National Institute of Statistics (1998) "Cambodia Socio-Economic-Survey 1997: Technical Report on Survey Design and Implementation." National Institute of Statistics, Ministry of Planning, Cambodia.

National Institute of Statistics (2000a) “Cambodia Socio-Economic-Survey 1999: Technical Report on Survey Design and Implementation.” National Institute of Statistics, Ministry of Planning, Cambodia.

National Institute of Statistics (2000b) “General Population Census of Cambodia 1998: Village Gazetteer.” National Institute of Statistics, Ministry of Planning, Cambodia.

Prescott, N. and Pradhan, M. (1997) “A Poverty Profile of Cambodia.” *World Bank Discussion Paper*, No. 373, The World Bank.

United Nations Development Programme (1997). “Human Development Report 1997.” Oxford University Press: New York.

WFP (2001) “Identifying Poor Areas in Cambodia: Combining Census and Socio-Economic Survey Data to Trace the Spatial Dimensions of Poverty.” Phnom Penh, Cambodia.

World Bank (1999) “Cambodia Poverty Assessment.” Report No. 19858-KH, The World Bank.

World Bank (2001) “World Development Report 2001: Attacking Poverty.” Oxford University Press: New York.

## APPENDIX A Glossary

Definitions of acronyms and simple explanations of econometric terms used in this report are provided below. Readers are referred to standard econometrics textbooks for more precise definitions and meanings of econometric terms (*e.g.* Green, 2000).

| | |
|---|---|
| chronic poverty | Poverty due to long-term low levels of living. |
| CARERE | Cambodian Area Rehabilitation and Regeneration Project |
| CSES | Cambodian Socio-Economic Survey |
| disturbance term | Random error term that cannot be explained by right-hand-side variables |
| GIS | Geographical Information System |
| GLS | Generalized Least Squares. GLS is often used when heteroscedasticity exists. |
| Heteroscedasticity | When the variance of the disturbance term is not constant, the model is called heteroscedastic |
| idiosyncratic error | Component of the disturbance term that is specific to a household |
| OLS | Ordinary Least Squares |
| point estimate | A point estimate is a single number that approximates the true parameter of interest. |
| PRSP | Poverty Reduction Strategy Papers |
| residual term | Unexplained part of the variation in left-hand-side variable |
| RGC | Royal Government of Cambodia |
| right-hand-side variable | Variables put on the right hand side of a regression equation. They are also called independent or exogenous variables. |
| Simulation | Construction of a model which describes the structure and processes of a real-world situation to be studied. The model is often created on a computer when experiment is not feasible. |
| standard error | A measure of the accuracy of an estimate. When the standard error is smaller, the estimate is more accurate. |
| transitional poverty | Poverty due to seasonal variation in consumption |
| WFP | World Food Programme |
| MOP | Ministry of Planning |

## APPENDIX B List of variables

The following table is a list of variables used in the model. For the definition of "other", readers are referred to the census and survey questionnaires.[18] To derive max_* and min_*, the 30 year average between 1961 and 1990 was taken for each month and the maximum and minimum of the indicator * was chosen. The corresponding month is given in the table.

| Variable Name | Description |
|---|---|
| AV_ROOM | average number of rooms in the village |
| DEC1996 | NDVI value for December 1996 |
| DEPRATIO | dependency ratio |
| DIST_RIV | distance to the river |
| FEMRATIO | ratio of females in the household |
| FLOOR2 | floor material is wood/plywood |
| FLOOR3 | floor material is cement/brick/stone |
| FLOOR5 | floor material is parquet/polished wood |
| FLOOR6 | floor material is mosaic/ceramic tiles |
| FLOOR_2P | ratio of households with wood/plywood floor in village |
| FLOOR_3P | ratio of households with cement/brick/stone floor in village |
| FLOOR_6P | ratio of households with mosaic/ceramic floor in village |
| FLOOR_7P | ratio of households with other floor in village |
| HEDU_R2P | ratio of household head not completed primary education in village |
| HEDU_R4P | ratio of household head less than secondary school/diploma education in village |
| HESTA_2P | ratio of household head in village whose employment status is an employer |
| HESTA_3P | ratio of household head in village whose employment status is own account |
| HESTA_4P | ratio of household head in village whose employment status is family worker |
| HESTA_5P | ratio of household head in village whose employment status is other |
| HHACTN2 | household head is unemployed/student/other |
| HHACTN3 | household head is a home-maker |
| HHACT_1P | ratio of household head in village whose activity is employed |
| HHACT_3P | ratio of household head in village whose activity is home-maker |
| HHAGE | age of household head |
| HHEMPSTA3 | employment status of household head is own account |
| HHEMPSTA4 | employment status of household head is family worker |
| HHFEMU_P | ratio of households in village headed by female under 40 |
| HHFEM_P | ratio of households in village headed by female |
| HHINDN1 | household head working in agricultural and fishery industry |
| HHINDN2 | household head working in manufacturing industry |
| HHINDN3 | household head working in service industry |
| HHIN_6P | ratio of households in village whose head is employed in textile/apparel/leather sector |
| HHIN_8P | ratio of households in village whose head is employed in utility sector |
| HHIN_13P | ratio of households in village whose head is employed in financial intermediation sector |
| HHIN_14P | ratio of households in village whose head is employed in real estate/renting/business activities sector |
| HHIN_15P | ratio of households in village whose head is employed in public administration/defense sector |
| HHIN_16P | ratio of households in village whose head is employed in education sector |

[18] See National Institute of Statistics (1998) and National Institute of Statistics (2000b)

| Variable Name | Description |
|---|---|
| HHLIT2 | head of household is literate |
| HHLIT_P | ratio of households in village headed by literate household head |
| HHMAR3 | household head widowed |
| HHMRRD | ratio of households headed by married household head |
| HHOCC8 | household head's occupation craft and related products |
| HHOC_10P | ratio of households in village whose head's occupation is elementary occupation |
| HHSIZE | number of individuals in household |
| LDF_CH | change of low-density forest in the commune 1993-97 |
| LIGHT2 | source of light is battery |
| LIGHT3 | source of light is kerosene lump/pump lantern |
| MAXEDUC1 | maximum level of education in household no education |
| MAXEDUC2 | maximum level of education in household primary school not completed |
| MAXEDUC4 | maximum level of education in household lower secondary school |
| MAXEDUC5 | maximum level of education in household secondary school/diploma |
| MAX_DTR | 30 year average maximum diurnal temperature range (March) |
| MAX_PRE | 30 year average maximum precipitation (September) |
| MAX_TMN | 30 year average maximum temperature range (May) |
| MAX_VAP | 30 year average maximum vapour pressure (May) |
| MAX_WND | 30 year average maximum wind speed (December) |
| MDF_97 | medium density forest in commune in 1997 |
| MDF_CH | change in medium density forest in commune 1993-97 |
| MEDU_R5P | ratio of households in village whose members' maximum level of education is less than undergraduate |
| MEM0_4P | number of people in household aged 0-4 |
| MEM5_14P | number of people in household aged 5-14 |
| MEMEMP_P | employment rate in village |
| MEMSTDNT | number of students in household |
| MEMSTD_P | ratio of students to population in village |
| MIN_CLD | 30 year average minimum cloud cover (January) |
| MIN_PRE | 30 year average minimum precipitation (January) |
| MIN_RAD | 30 year average minimum radiation (August) |
| MIN_TMN | 30 year average minimum daily minimum temperature (January) |
| MIN_VAP | 30 year average minimum vapour pressure (January) |
| MIN_WET | 30 year average minimum wet day frequency (January) |
| MIN_WND | 30 year average minimum wind speed (May) |
| MORT_M | crude mortality rate for male in village |
| MORT_T | crude mortality rate for male and female in village |
| NONAG_P | ratio of people in village working in non-agricultural sector |
| NOV1996 | NDVI value for November 1996 |
| NOV_COV | Coefficient of variation of 19 year average NDVI for November |
| OCCUP_3P | ratio of people in village who are professionals |
| OTHF_97 | ratio of other forest in commune |
| OTHF_CH | change in other forest in commune 1993-97 |
| OTHNF_97 | ratio of other non-forest in commune |
| ROOF2 | roof made of tile |
| ROOF3 | roof made of concrete/brick/stone |
| ROOF4 | roof made of galvanized iron/aluminum |
| ROOF5 | roof made of other materials |
| ROOF_1P | ratio of households in village whose roof is made of bamboo/thatch/grass |
| ROOF_3P | ratio of households in village whose roof is made of concrete/brick/stone |

| Variable Name | Description |
|---|---|
| ROOF_5P | ratio of households in village whose roof is made of other |
| SESTA_1P | ratio of households in village whose employment status of spouse is employer |
| SESTA_2P | ratio of households in village whose employment status of spouse is paid employee |
| SESTA_3P | ratio of households in village whose employment status of spouse is own account |
| SESTA_4P | ratio of households in village whose employment status of spouse is family worker |
| SOILS | soil quality indicator in commune |
| SOIL_Q | soil quality indicator in village |
| SPACT3 | main activity of spouse last 12 month is home maker |
| SPACT5 | main activity of spouse last 12 month is other |
| SPACT_1P | ratio of households in village whose activity of spouse is employed |
| SPACT_2P | ratio of households in village whose activity of spouse is unemployed |
| SPEDUC1 | education of spouse is no educational level |
| SPEDUC5 | education of spouse is secondary school or higher |
| SPEMPSEC2 | employment sector of spouse is not government |
| SPEMPSTA3 | employment status of spouse is own account |
| SPIN_5P | ratio of households in village in which spouse is employed in food products/beverage/tobacco sector |
| SPIN_6P | ratio of households in village in which spouse is employed in textile/apparel/leather sector |
| SPIN_10P | ratio of households in village in which spouse is employed in sales sector |
| SPIN_11P | ratio of households in village in which spouse is employed in hotel and restaurants sector |
| SPLIT2 | spouse is literate |
| SPOCC6 | spouse is a service worker |
| SPOC_1P | ratio of households in village in which spouse works for armed forces |
| SPOC_2P | ratio of households in village in which spouse is legislator/senior official/manager |
| SPOC_4P | ratio of households in village in which spouse is technician/associate professional |
| SPOC_9P | ratio of households in village in which spouse is plant and machine operators/assemblers |
| TOILET2 | toilet not available within the premise |
| V81_97 | coverage of cropping mosaic with cropping area <30% in commune in 1997 |
| V82_97 | coverage of cropping mosaic with cropping area >30% in commune in 1997 |
| V91_97 | coverage of agricultural land in commune in 1997 |
| V91_CH | change in coverage of agricultural land in commune 1993-1997 |
| V94_97 | coverage of urban areas in commune in 1997 |
| WALL2 | material of wall is wood/plywood |
| WALL3 | material of wall is concrete/brick/stone |
| WALL5 | material of wall is other |
| WALL_4P | ratio of households in village whose material of wall is galvanized iron/aluminum |
| WALL_5P | ratio of households in village whose material of wall is other |
| WATER4 | drinking water is from pond, river, stream or rainwater |
| WATER5 | drinking water is bought |
| WATER6 | drinking water is from other sources |
| WATER_2P | ratio of households in village whose supply of drinking water is tube/piped well |
| WATER_97 | coverage of water/wetland in commune in 1997 |
| WATER_CH | change of the ratio of water/wetland areas in commune 1993-97 |
| _YHAT | predicted logarithmic consumption |

The following variables are derived from the variables above.

| Variable Name | Description |
|---|---|
| LOGH | log(HHSIZE) |
| HHSIZE2 | HHSIZE squared |
| HHSIZE3 | HHSIZE cubed |
| HHAGE2 | HHAGE squared |
| WALLHH1 | WALL3*HHSIZE |
| WALLHH2 | WALL3*HHSIZE2 |
| LIGHTHH1 | LIGHT3*HHSIZE |
| ROOFHH1 | ROOF3*HHSIZE |
| ROOFHH3 | ROOF3*HHSIZE3 |
| DEPRATIO2 | DEPRATIO squared |
| CROSS261 | MDF_97*WATER5 |
| CROSS444 | MDF_CH*SPEMPSEC2 |
| CROSS661 | LDF_CH*SPACT3 |
| CROSS832 | OTHF_97*ROOFHH3 |
| CROSS873 | OTHF_CH*SPEMPSTA3 |
| CROSS1029 | V81_97*FLOOR5 |
| CROSS1242 | V82_97*HHOCC8 |
| CROSS1460 | V91_97*WATER4 |
| CROSS1581 | V91_CH*WATER5 |
| CROSS1641 | V91_CH*MAXEDUC5 |
| CROSS1661 | V91_CH*DEPRATIO |
| CROSS1686 | WATER_97*LIGHT2 |
| CROSS1850 | WATER_CH*SPOCC6 |
| CROSS1861 | WATER_CH*SPACT3 |
| CROSS1933 | V94_97*SPEDUC1 |
| CROSS1997 | V94_97*MAXEDUC1 |
| CROSS2001 | V94_97*MAXEDUC5 |
| CROSS2017 | V94_97*HHSIZE3 |
| CROSS2240 | OTHNF_97*MAXEDUC4 |
| CROSS2780 | MIN_CLD*WATER4 |
| CROSS3950 | MIN_RAD*ROOFHH1 |
| CROSS4081 | MAX_TMN*ROOF2 |
| CROSS4626 | MAX_VAP*FLOOR2 |
| CROSS4658 | MAX_VAP*LOGH |
| CROSS4801 | MAX_WND*ROOF2 |
| CROSS4807 | MAX_WND*LIGHT3 |
| CROSS4867 | MAX_WND*FLOOR3 |
| CROSS4990 | NOV1996*FLOOR6 |
| CROSS5023 | NOV1996*FEMRATIO |
| CROSS5068 | DEC1996*HHEMPSTA3 |
| CROSS5134 | DEC1996*HHAGE2 |

## APPENDIX C Definition of Poverty and Inequality Measures

Poverty rate, poverty gap, and poverty severity all belong to the FGT class (Foster, Greer and Thorbecke, 1984). An FGT measure $P$ with parameter $\alpha$, denoted as $P_\alpha$, is defined as:

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{z-y_i}{z}\right)^\alpha \cdot \mathrm{Ind}(y_i < z)$$

where $n$ is the population, $y_i$ is the consumption of the $i$-th individual, and $z$ is the poverty line. Poverty rate, gap and severity correspond to $P_0$, $P_1$ and $P_2$ respectively. In other words, they correspond to the FGT measure with the parameter being 0, 1 and 2 respectively.

Inequality is measured by the Gini coefficient $G$ in this report. $G$ is computed as:

$$G = \frac{1}{2n^2\mu}\sum_i\sum_j\left|y_i - y_j\right|, \text{ where } \mu \equiv \frac{1}{n}\sum_i y_i$$

$\mu$ is the average per capita consumption. The Gini coefficient is a measure of the degree to which a population shares the resources unequally. Geometrically, it corresponds to the ratio of the area between the $45°$ line and the Lorenz curve to the area of the triangle under the $45°$ line. The index varies between zero and one, zero corresponding to no inequality and one corresponding to the maximum possible inequality.

# APPENDIX D Regression Results

Detailed regression results are presented in this appendix. See APPENDIX B for the description of each variable. For each stratum, the OLS and GLS results are presented together, followed by the results for the associated residual regression. Regression diagnostics are found in APPENDIX E.

## Phnom Penh Stratum

| Variable | OLS Results | | | | GLS Results | |
|---|---|---|---|---|---|---|
| | Coef. | Std. E | t | P>\|t\| | Coef. | Std. E |
| HHINDN3 | -0.393 | 0.105 | -3.75 | 0.000 | -0.260 | 0.068 |
| HHMAR3 | -0.221 | 0.068 | -3.24 | 0.001 | -0.162 | 0.058 |
| HHACTN3 | -0.366 | 0.117 | -3.14 | 0.002 | -0.285 | 0.083 |
| HHINDN1 | -0.514 | 0.120 | -4.29 | 0.000 | -0.337 | 0.076 |
| HHINDN2 | -0.412 | 0.120 | -3.44 | 0.001 | -0.270 | 0.075 |
| SPACT5 | 0.179 | 0.066 | 2.72 | 0.007 | 0.117 | 0.056 |
| HHACTN2 | -0.324 | 0.107 | -3.02 | 0.003 | -0.216 | 0.071 |
| LIGHTHH1 | -0.033 | 0.008 | -3.95 | 0.000 | -0.031 | 0.006 |
| WALL3 | 1.129 | 0.154 | 7.35 | 0.000 | 1.136 | 0.143 |
| MAXEDUC5 | 0.235 | 0.050 | 4.70 | 0.000 | 0.193 | 0.055 |
| TOILET2 | -0.110 | 0.043 | -2.55 | 0.011 | -0.100 | 0.033 |
| FLOOR3 | 5.670 | 1.914 | 2.96 | 0.003 | 4.506 | 1.493 |
| MAXEDUC2 | -0.099 | 0.041 | -2.41 | 0.016 | -0.111 | 0.026 |
| ROOFHH1 | -6.500 | 1.890 | -3.44 | 0.001 | -4.198 | 2.688 |
| WATER5 | -0.109 | 0.037 | -2.94 | 0.003 | -0.051 | 0.030 |
| WALL2 | 0.207 | 0.049 | 4.26 | 0.000 | 0.191 | 0.032 |
| WALLHH1 | -0.251 | 0.041 | -6.16 | 0.000 | -0.243 | 0.041 |
| SPLIT2 | -0.124 | 0.036 | -3.45 | 0.001 | -0.065 | 0.028 |
| WALLHH2 | 0.016 | 0.003 | 5.95 | 0.000 | 0.015 | 0.003 |
| MEMSTDNT | 0.052 | 0.012 | 4.41 | 0.000 | 0.052 | 0.010 |
| DEPRATIO2 | -0.273 | 0.095 | -2.86 | 0.004 | -0.243 | 0.075 |
| LOGH | -0.462 | 0.056 | -8.25 | 0.000 | -0.451 | 0.038 |
| ROOF_3P | -0.293 | 0.069 | -4.25 | 0.000 | -0.136 | 0.086 |
| FLOOR_6P | 0.234 | 0.062 | 3.79 | 0.000 | 0.195 | 0.082 |
| HHACT_1P | 1.774 | 0.239 | 7.42 | 0.000 | 1.887 | 0.264 |
| HHIN_16P | 2.904 | 0.854 | 3.40 | 0.001 | 2.582 | 0.926 |
| SPIN_10P | 0.388 | 0.096 | 4.05 | 0.000 | 0.396 | 0.095 |
| MEDU_R5P | -1.505 | 0.296 | -5.09 | 0.000 | -1.732 | 0.357 |
| HHIN_8P | 5.133 | 1.658 | 3.10 | 0.002 | 4.817 | 2.003 |
| HEDU_R2P | 0.755 | 0.184 | 4.11 | 0.000 | 0.852 | 0.194 |
| CROSS1933 | 1.13E-06 | 2.36E-07 | 4.78 | 0.000 | 1.142E-06 | 3.027E-07 |
| CROSS1997 | -6.33E-07 | 2.30E-07 | -2.75 | 0.006 | -7.267E-07 | 2.237E-07 |
| CROSS4867 | -0.261 | 0.086 | -3.03 | 0.003 | -0.207 | 0.067 |
| CROSS832 | 1.27E-08 | 2.55E-09 | 4.99 | 0.000 | 1.783E-08 | 2.488E-09 |
| CROSS2001 | -1.79E-07 | 6.70E-08 | -2.67 | 0.008 | -1.412E-07 | 6.189E-08 |
| CROSS3950 | 0.042 | 0.012 | 3.45 | 0.001 | 0.027 | 0.017 |
| CROSS5023 | -1.109 | 0.400 | -2.77 | 0.006 | -0.954 | 0.377 |
| SPOC_4P | 0.707 | 0.311 | 2.27 | 0.023 | 0.944 | 0.393 |
| CONS | 8.387 | 0.348 | 24.07 | 0.000 | 8.247 | 0.407 |

| Variable | Residual Regression Results | | | |
|---|---|---|---|---|
| | Coef. | Std. E | t | P>\|t\| |
| MAXEDUC2 | -1.172 | 0.350 | -3.35 | 0.001 |
| ROOFHH1 | 0.907 | 0.200 | 4.54 | 0.000 |
| WALLHH1 | -1.368 | 0.366 | -3.74 | 0.000 |
| WALLHH2 | 0.114 | 0.031 | 3.62 | 0.000 |
| LOGH | -0.596 | 0.213 | -2.80 | 0.005 |
| HHINDN3*ROOFHH1 | -0.861 | 0.201 | -4.29 | 0.000 |
| HHINDN3*WALL2 | 0.664 | 0.191 | 3.48 | 0.001 |
| HHINDN3*WALLHH1 | 1.569 | 0.371 | 4.23 | 0.000 |
| HHINDN3*WALLHH2 | -0.133 | 0.032 | -4.18 | 0.000 |
| HHMAR3*WATER5 | -0.857 | 0.318 | -2.69 | 0.007 |
| HHACTN3*WALL3 | 2.769 | 1.285 | 2.15 | 0.031 |
| HHACTN3*ROOFHH1 | -0.838 | 0.221 | -3.79 | 0.000 |
| HHACTN3*LOGH | 0.595 | 0.198 | 3.00 | 0.003 |
| HHINDN1*ROOFHH1 | -1.149 | 0.381 | -3.01 | 0.003 |
| HHINDN1*WALLHH1 | 1.149 | 0.351 | 3.28 | 0.001 |
| HHINDN2*WALL3 | 3.626 | 1.200 | 3.02 | 0.003 |
| HHINDN2*MAXEDUC5 | -1.962 | 0.899 | -2.18 | 0.029 |
| HHINDN2*ROOFHH1 | -0.675 | 0.216 | -3.13 | 0.002 |
| SPACT5*MAXEDUC5 | -1.578 | 0.517 | -3.05 | 0.002 |
| SPACT5*ROOFHH1 | 0.265 | 0.065 | 4.05 | 0.000 |
| HHACTN2*ROOFHH1 | -0.632 | 0.217 | -2.91 | 0.004 |
| HHACTN2*WALLHH1 | 1.257 | 0.406 | 3.10 | 0.002 |
| HHACTN2*WALLHH2 | -0.110 | 0.035 | -3.13 | 0.002 |
| WALL3*MAXEDUC2 | 0.934 | 0.399 | 2.34 | 0.019 |
| MAXEDUC5*WALLHH1 | 0.102 | 0.035 | 2.90 | 0.004 |
| MAXEDUC5*SPLIT2 | 1.747 | 0.445 | 3.93 | 0.000 |
| MAXEDUC2*SPLIT2 | 0.908 | 0.370 | 2.45 | 0.014 |
| ROOFHH1*SPLIT2 | -0.298 | 0.074 | -4.03 | 0.000 |
| WALLHH1*SPLIT2 | 0.137 | 0.057 | 2.39 | 0.017 |
| CONS | -5.549 | 0.375 | -14.82 | 0.000 |

## Other Urban Stratum

| Variable | OLS Results | | | | GLS Results | |
|---|---|---|---|---|---|---|
| | Coef. | Std. E | t | P>\|t\| | Coef. | Std. E |
| ROOF4 | 0.210 | 0.034 | 6.16 | 0.000 | 0.233 | 0.033 |
| ROOF5 | 0.245 | 0.073 | 3.35 | 0.001 | 0.306 | 0.099 |
| TOILET2 | -0.125 | 0.039 | -3.23 | 0.001 | -0.125 | 0.044 |
| LIGHT3 | -0.344 | 0.037 | -9.22 | 0.000 | -0.325 | 0.042 |
| HHEMPSTA4 | 0.290 | 0.084 | 3.44 | 0.001 | 0.302 | 0.076 |
| FLOOR2 | -0.138 | 0.029 | -4.70 | 0.000 | -0.165 | 0.030 |
| LOGH | 2.550 | 0.915 | 2.79 | 0.005 | 1.729 | 0.849 |
| DEPRATIO | -0.300 | 0.068 | -4.45 | 0.000 | -0.272 | 0.062 |
| HHOC_10P | -2.027 | 0.242 | -8.38 | 0.000 | -2.058 | 0.234 |
| SPOC_9P | 13.852 | 1.178 | 11.76 | 0.000 | 14.044 | 1.181 |
| WATER_2P | -0.474 | 0.052 | -9.17 | 0.000 | -0.434 | 0.049 |
| WALL_5P | 1.527 | 0.198 | 7.73 | 0.000 | 1.524 | 0.181 |
| ROOF_1P | -0.337 | 0.086 | -3.92 | 0.000 | -0.329 | 0.084 |
| FLOOR_3P | -1.004 | 0.122 | -8.22 | 0.000 | -0.958 | 0.117 |
| HHACT_3P | -3.441 | 0.476 | -7.23 | 0.000 | -3.569 | 0.508 |
| SESTA_1P | 17.809 | 3.659 | 4.87 | 0.000 | 15.705 | 3.707 |
| MIN_CLD | -0.027 | 0.009 | -3.08 | 0.002 | -0.022 | 0.008 |
| MAX_DTR | -0.017 | 0.004 | -4.36 | 0.000 | -0.015 | 0.004 |
| MAX_PRE | -0.009 | 0.002 | -5.31 | 0.000 | -0.009 | 0.002 |
| MIN_TMN | -0.022 | 0.005 | -3.94 | 0.000 | -0.022 | 0.006 |
| MAX_WND | 0.069 | 0.013 | 5.29 | 0.000 | 0.081 | 0.011 |
| ROOF_5P | -2.052 | 0.531 | -3.87 | 0.000 | -2.050 | 0.515 |
| HHIN_14P | 18.656 | 4.172 | 4.47 | 0.000 | 21.589 | 4.629 |
| MEMEMP_P | -2.488 | 0.460 | -5.41 | 0.000 | -2.581 | 0.418 |
| HHLIT2 | -0.094 | 0.032 | -2.90 | 0.004 | -0.090 | 0.030 |
| SPLIT2 | -0.081 | 0.028 | -2.86 | 0.004 | -0.057 | 0.029 |
| CROSS1242 | -1.610E-07 | 2.870E-08 | -5.62 | 0.000 | -1.635E-07 | 3.523E-08 |
| CROSS1850 | -7.410E-07 | 8.600E-08 | -8.61 | 0.000 | -6.808E-07 | 1.126E-07 |
| CROSS1861 | -3.260E-07 | 6.590E-08 | -4.94 | 0.000 | -2.883E-07 | 7.347E-08 |
| CROSS5134 | -3.574E-04 | 8.860E-05 | -4.03 | 0.000 | -3.481E-04 | 8.225E-05 |
| CROSS1581 | 6.110E-08 | 1.760E-08 | 3.48 | 0.001 | 6.098E-08 | 1.918E-08 |
| CROSS1641 | -1.380E-07 | 3.920E-08 | -3.53 | 0.000 | -1.695E-07 | 5.600E-08 |
| CROSS444 | -1.410E-07 | 2.210E-08 | -6.39 | 0.000 | -1.547E-07 | 2.283E-08 |
| CROSS1029 | -7.040E-08 | 1.620E-08 | -4.34 | 0.000 | -7.626E-08 | 1.798E-08 |
| CROSS4081 | 0.001 | 1.627E-04 | 8.03 | 0.000 | 0.001 | 1.574E-04 |
| CROSS4658 | -0.010 | 0.003 | -3.32 | 0.001 | -0.007 | 0.003 |
| CROSS4990 | -7.601 | 1.066 | -7.13 | 0.000 | -8.098 | 1.339 |
| CROSS5068 | -1.201 | 0.264 | -4.55 | 0.000 | -1.297 | 0.251 |
| SPACT_1P | 0.430 | 0.147 | 2.93 | 0.004 | 0.400 | 0.141 |
| MORT_M | -1.278 | 0.350 | -3.65 | 0.000 | -0.983 | 0.310 |
| CONS | 16.923 | 1.329 | 12.73 | 0.000 | 16.100 | 1.451 |

| Variable | Residual Regression Results | | | |
|---|---|---|---|---|
| | Coef. | Std. E | t | P>\|t\| |
| ROOF4 | -0.654 | 0.184 | -3.56 | 0.000 |
| TOILET2 | 2.004 | 0.487 | 4.11 | 0.000 |
| FLOOR2 | -1.150 | 0.287 | -4.01 | 0.000 |
| HHLIT2 | -0.777 | 0.216 | -3.6 | 0.000 |
| ROOF4*HHLIT2 | 0.811 | 0.370 | 2.19 | 0.029 |
| ROOF5*HHLIT2 | 2.994 | 1.337 | 2.24 | 0.025 |
| TOILET2*LIGHT3 | -0.779 | 0.220 | -3.54 | 0.000 |
| TOILET2+LOGH | -0.970 | 0.240 | -4.05 | 0.000 |
| FLOOR2*DEPRATIO | 1.761 | 0.537 | 3.28 | 0.001 |
| CONS | -3.177 | 0.186 | -17.04 | 0.000 |

## Rural Stratum

| Variable | OLS Results | | | | GLS Results | |
|---|---|---|---|---|---|---|
| | Coef. | Std. E | t | P> (t) | Coef. | Std. E |
| LIGHT3 | -0.924 | 0.155 | -5.96 | 0.000 | -0.899 | 0.210 |
| FLOOR2 | 4.244 | 0.749 | 5.67 | 0.000 | 3.065 | 0.857 |
| WALL2 | 0.090 | 0.017 | 5.41 | 0.000 | 0.081 | 0.016 |
| SPLIT2 | -0.072 | 0.013 | -5.50 | 0.000 | -0.056 | 0.012 |
| WALL5 | 0.169 | 0.024 | 7.09 | 0.000 | 0.118 | 0.027 |
| HHINDN1 | -0.072 | 0.017 | -4.27 | 0.000 | -0.054 | 0.016 |
| MAXEDUC4 | 0.079 | 0.023 | 3.41 | 0.001 | 0.076 | 0.023 |
| ROOF2 | 0.734 | 0.175 | 4.19 | 0.000 | 0.657 | 0.170 |
| SPEDUC5 | 0.288 | 0.060 | 4.81 | 0.000 | 0.247 | 0.169 |
| HHEMPSTA4 | 0.141 | 0.043 | 3.27 | 0.001 | 0.074 | 0.028 |
| TOILET2 | -0.164 | 0.036 | -4.62 | 0.000 | -0.147 | 0.037 |
| ROOF4 | 0.062 | 0.020 | 3.08 | 0.002 | 0.060 | 0.020 |
| ROOF3 | 0.614 | 0.152 | 4.03 | 0.000 | 0.568 | 0.074 |
| WATER6 | -0.084 | 0.033 | -2.49 | 0.013 | -0.082 | 0.038 |
| LOGH | -0.448 | 0.020 | -22.86 | 0.000 | -0.448 | 0.014 |
| MEMSTDNT | 0.029 | 0.006 | 4.54 | 0.000 | 0.025 | 0.006 |
| DEPRATIO2 | -0.325 | 0.041 | -8.00 | 0.000 | -0.287 | 0.036 |
| MEMSTD_P | 1.183 | 0.159 | 7.42 | 0.000 | 0.977 | 0.326 |
| MORT_T | 1.338 | 0.134 | 9.99 | 0.000 | 1.287 | 0.278 |
| FLOOR_6P | 4.928 | 0.578 | 8.53 | 0.000 | 4.334 | 1.184 |
| FLOOR_2P | -0.073 | 0.028 | -2.60 | 0.009 | -0.121 | 0.057 |
| FLOOR_7P | -13.020 | 3.461 | -3.76 | 0.000 | -19.370 | 7.198 |
| HHACT_3P | -2.699 | 0.307 | -8.78 | 0.000 | -2.814 | 0.611 |
| HHIN_13P | -156.384 | 28.334 | -5.52 | 0.000 | -162.317 | 59.682 |
| MEM0_4P | 1.430 | 0.366 | 3.91 | 0.000 | 1.142 | 0.747 |
| HESTA_5P | 42.911 | 4.848 | 8.85 | 0.000 | 37.725 | 9.628 |
| MEM5_14P | -1.429 | 0.269 | -5.31 | 0.000 | -1.261 | 0.584 |
| SPIN_5P | -4.418 | 0.740 | -5.97 | 0.000 | -5.550 | 1.630 |
| NONAG_P | 0.278 | 0.111 | 2.50 | 0.012 | 0.442 | 0.228 |
| AV_ROOM | -0.052 | 0.015 | -3.36 | 0.001 | -0.047 | 0.029 |
| SPOC_2P | 5.423 | 2.921 | 1.86 | 0.063 | 7.695 | 6.105 |
| HHFEM_P | -1.553 | 0.193 | -8.06 | 0.000 | -1.435 | 0.414 |
| SPOC_9P | 6.059 | 1.025 | 5.91 | 0.000 | 6.811 | 2.140 |
| HHIN_15P | -2.436 | 0.316 | -7.70 | 0.000 | -2.322 | 0.677 |
| HESTA_4P | 22.966 | 2.418 | 9.50 | 0.000 | 23.214 | 4.911 |
| HESTA_3P | 17.518 | 1.873 | 9.35 | 0.000 | 17.134 | 3.821 |
| ROOF_5P | 1.896 | 0.279 | 6.80 | 0.000 | 1.275 | 0.589 |
| HESTA_2P | 17.542 | 1.936 | 9.06 | 0.000 | 16.767 | 3.949 |
| SESTA_2P | -15.049 | 1.455 | -10.34 | 0.000 | -14.781 | 2.931 |
| OCCUP_3P | -1.774 | 0.254 | -6.98 | 0.000 | -0.919 | 0.541 |
| HHIN_6P | 4.687 | 0.499 | 9.40 | 0.000 | 2.793 | 1.083 |
| SESTA_4P | -14.830 | 1.431 | -10.36 | 0.000 | -14.341 | 2.877 |
| SPIN_6P | -1.782 | 0.178 | -9.98 | 0.000 | -1.393 | 0.374 |
| SESTA_3P | -14.371 | 1.414 | -10.16 | 0.000 | -13.778 | 2.838 |
| SPIN_11P | -17.538 | 3.699 | -4.74 | 0.000 | -18.192 | 7.701 |
| HEDU_R4P | -2.849 | 0.413 | -6.90 | 0.000 | -2.418 | 0.840 |
| SPACT_2P | -1.454 | 0.229 | -6.34 | 0.000 | -0.951 | 0.458 |
| WATER_CH | 2.550E-08 | 3.260E-09 | 7.81 | 0.000 | 2.384E-08 | 6.771E-09 |
| V81_97 | -1.330E-08 | 1.940E-09 | -6.85 | 0.000 | -1.359E-08 | 4.039E-09 |
| NOV_COV | -1.034 | 0.124 | -8.35 | 0.000 | -1.256 | 0.258 |
| DIST_RIV | -7.900E-06 | 8.820E-07 | -8.95 | 0.000 | -7.222E-06 | 1.952E-06 |

(Continued on the next page)

| Variable | OLS Results | | | | GLS Results | |
|---|---|---|---|---|---|---|
| | Coef. | Std. E | t | P> (t) | Coef. | Std. E |
| MIN_CLD | 0.067 | 0.007 | 9.67 | 0.000 | 0.067 | 0.014 |
| SOILS | 0.054 | 0.008 | 6.90 | 0.000 | 0.056 | 0.017 |
| MIN_PRE | 0.108 | 0.009 | 12.28 | 0.000 | 0.103 | 0.018 |
| MIN_WET | -0.043 | 0.004 | -11.38 | 0.000 | -0.039 | 0.008 |
| HHMRRD | -0.761 | 0.150 | -5.08 | 0.000 | -0.557 | 0.344 |
| HHFEMU_P | 1.121 | 0.266 | 4.22 | 0.000 | 1.090 | 0.570 |
| HHLIT_P | -0.284 | 0.073 | -3.87 | 0.000 | -0.051 | 0.152 |
| WALL_4P | -1.764 | 0.404 | -4.37 | 0.000 | -1.347 | 0.855 |
| MIN_VAP | -0.020 | 0.002 | -9.92 | 0.000 | -0.021 | 0.004 |
| SPOC_1P | 12.672 | 1.663 | 7.62 | 0.000 | 12.183 | 3.494 |
| CROSS261 | 2.54E-09 | 6.26E-10 | 4.06 | 0.000 | 3.145E-09 | 8.545E-10 |
| CROSS661 | 8.48E-08 | 1.41E-08 | 6.01 | 0.000 | 6.952E-08 | 1.738E-08 |
| CROSS873 | 1.91E-08 | 4.88E-09 | 3.92 | 0.000 | 1.810E-08 | 5.741E-09 |
| CROSS1460 | 4.24E-09 | 5.95E-10 | 7.14 | 0.000 | 3.764E-09 | 8.053E-10 |
| CROSS1661 | -2.09E-08 | 3.37E-09 | -6.20 | 0.000 | -1.703E-08 | 4.787E-09 |
| CROSS1686 | 3.18E-08 | 7.13E-09 | 4.45 | 0.000 | 3.092E-08 | 8.354E-09 |
| CROSS2017 | -1.87E-10 | 3.53E-11 | -5.30 | 0.000 | -2.872E-10 | 4.334E-11 |
| CROSS2240 | 3.93E-09 | 8.74E-10 | 4.49 | 0.000 | 3.714E-09 | 1.164E-09 |
| CROSS2780 | -0.003 | 0.001 | -5.14 | 0.000 | -0.003 | 0.001 |
| CROSS4626 | -0.014 | 0.002 | -5.73 | 0.000 | -0.010 | 0.003 |
| CROSS4801 | -0.028 | 0.008 | -3.59 | 0.000 | -0.025 | 0.008 |
| CROSS4807 | 0.040 | 0.007 | 5.54 | 0.000 | 0.037 | 0.010 |
| MIN_WND | -0.043 | 0.016 | -2.65 | 0.008 | -0.024 | 0.031 |
| CONS | 12.463 | 1.806 | 6.90 | 0.000 | 11.513 | 3.663 |

| Variable | Residual Regression Results | | | |
| --- | --- | --- | --- | --- |
| | Coef. | Std. E | t | P> (t) |
| WALL2 | -0.844 | 0.429 | -1.97 | 0.049 |
| SPLIT2 | -1.444 | 0.423 | -3.42 | 0.001 |
| HHINDN1 | -0.449 | 0.196 | -2.29 | 0.022 |
| MAXEDUC4 | -1.668 | 0.644 | -2.59 | 0.010 |
| SPEDUC5 | 2.962 | 0.720 | 4.11 | 0.000 |
| HHEMPSTA4 | 1.993 | 0.571 | 3.49 | 0.000 |
| ROOF4 | 1.490 | 0.558 | 2.67 | 0.008 |
| LOGH | -0.569 | 0.218 | -2.61 | 0.009 |
| DEPRATIO2 | -0.968 | 0.399 | -2.42 | 0.015 |
| LIGHT3*WALL2 | 0.583 | 0.226 | 2.58 | 0.010 |
| LIGHT3*ROOF2 | -0.722 | 0.199 | -3.63 | 0.000 |
| LIGHT3*SPEDUC5 | -2.061 | 0.820 | -2.51 | 0.012 |
| LIGHT3*ROOF4 | -0.654 | 0.284 | -2.30 | 0.021 |
| LIGHT3*WATER6 | -0.985 | 0.440 | -2.24 | 0.025 |
| FLOOR2*WALL2 | -0.438 | 0.192 | -2.28 | 0.023 |
| FLOOR2*HHINDN1 | -0.463 | 0.149 | -3.10 | 0.002 |
| FLOOR2*WATER6 | 1.069 | 0.443 | 2.41 | 0.016 |
| FLOOR2*DEPRATIO2 | 1.098 | 0.431 | 2.55 | 0.011 |
| WALL2*LOGH | 0.465 | 0.221 | 2.10 | 0.036 |
| SPLIT2*HHINDN1 | 0.579 | 0.198 | 2.93 | 0.003 |
| SPLIT2*ROOF2 | 0.358 | 0.170 | 2.10 | 0.035 |
| SPLIT2*ROOF4 | 0.551 | 0.226 | 2.44 | 0.015 |
| SPLIT2*LOGH | 0.622 | 0.237 | 2.63 | 0.009 |
| SPLIT2*MEMSTDNT | -0.228 | 0.068 | -3.37 | 0.001 |
| WALL5*LOGH | 0.403 | 0.084 | 4.77 | 0.000 |
| HHINDN1*MAXEDUC4 | 0.605 | 0.272 | 2.22 | 0.027 |
| HHINDN1*MEMSTDNT | 0.173 | 0.058 | 3.00 | 0.003 |
| MAXEDUC4*ROOF4 | 0.817 | 0.317 | 2.58 | 0.010 |
| MAXEDUC4*LOGH | 0.741 | 0.351 | 2.11 | 0.035 |
| ROOF2*HHEMPSTA4 | -3.772 | 0.691 | -5.46 | 0.000 |
| ROOF2*MEMSTDNT | -0.130 | 0.063 | -2.05 | 0.041 |
| ROOF4*LOGH | -0.807 | 0.279 | -2.90 | 0.004 |
| HHINDN1*HHEMPSTA4 | -1.918 | 0.613 | -3.13 | 0.002 |
| FLOOR2*ROOF2 | 0.762 | 0.199 | 3.84 | 0.000 |
| CONS | -4.000 | 0.399 | -10.03 | 0.000 |

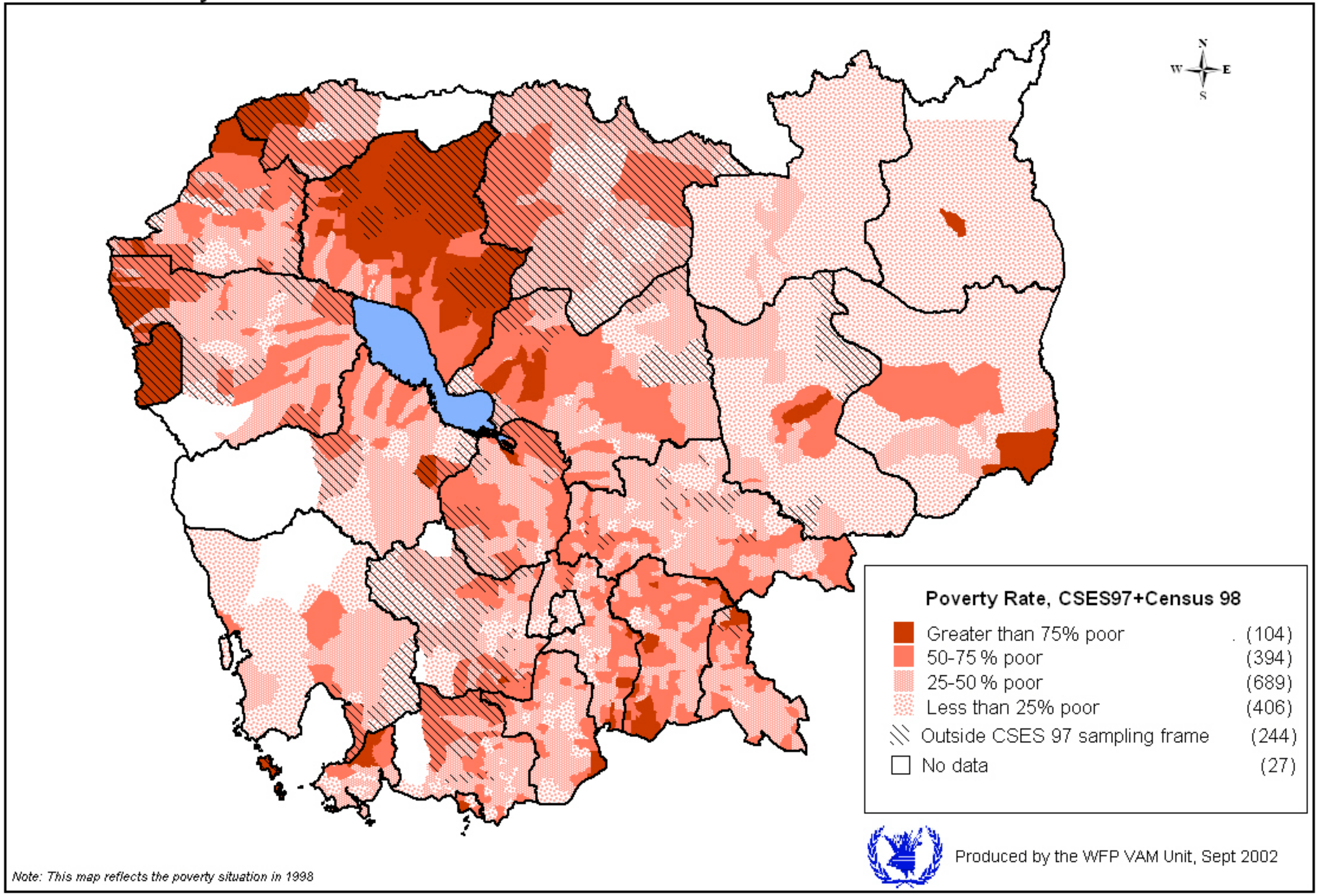
## APPENDIX E First Stage Diagnostics

Items A1 and A2 are the number of observations and the number of clusters respectively. The number of right-hand-side variables in the OLS and GLS regressions by type of variable is given in B1 to B4. These numbers are kept low relative to A1 and A2 to avoid over-fitting. Items C1 to C3 give a general idea about the fit of the model. The Other Urban stratum has the best fit of all three strata. Once a model with a reasonably good fit is obtained, the importance of the location effect is checked. D1 is the overall error and D2 is the error due to the location effect. The magnitude of the location effect is checked by D3. E1 and E2 tell if any location effects not explained by the model are significant. Except for the Other Urban stratum, such effects are significant. F1 to F4 give summary statistics on the residual regressions.

| Item | Description | | Phnom Penh | Other Urban | Rural |
|---|---|---|---|---|---|
| A1 | Descriptive Statistics | #obs | 1200 | 1000 | 3810 |
| A2 | | #cluster | 120 | 100 | 254 |
| B1 | # Right-Hand-Side Variables | #hhlevel | 22 | 10 | 17 |
| B2 | | #census mean | 9 | 13 | 35 |
| B3 | | #gis | 0 | 5 | 11 |
| B4 | | #cross b/w hhlevel and gis | 7 | 12 | 12 |
| C1 | OLS-regression with weight | F-statistics | 28.35 | 55.88 | 57.89 |
| C2 | | R-squared | 0.481 | 0.700 | 0.538 |
| C3 | | Adjusted R-squared | 0.464 | 0.687 | 0.528 |
| D1 | Importance of Location Effect | $\hat{\sigma}_u$ | 0.458 | 0.388 | 0.377 |
| D2 | | $\hat{\sigma}_\eta$ | 0.099 | 0.034 | 0.141 |
| D3 | | $\hat{\sigma}_\eta^2 / \hat{\sigma}_u^2$ | 0.047 | 0.008 | 0.139 |
| E1 | Location effect jointly=0 | F-statistics | 1.474 | 1.024 | 3.108 |
| E2 | | P-value | 0.001 | 0.420 | 0.000 |
| F1 | Residual Regression | # Dep Var for Residual Regression | 29 | 23 | 34 |
| F2 | | F-statistics | 5.54 | 7.71 | 5.19 |
| F3 | | R-squared | 0.121 | 0.066 | 0.045 |
| F4 | | Adjusted R-squared | 0.099 | 0.057 | 0.036 |

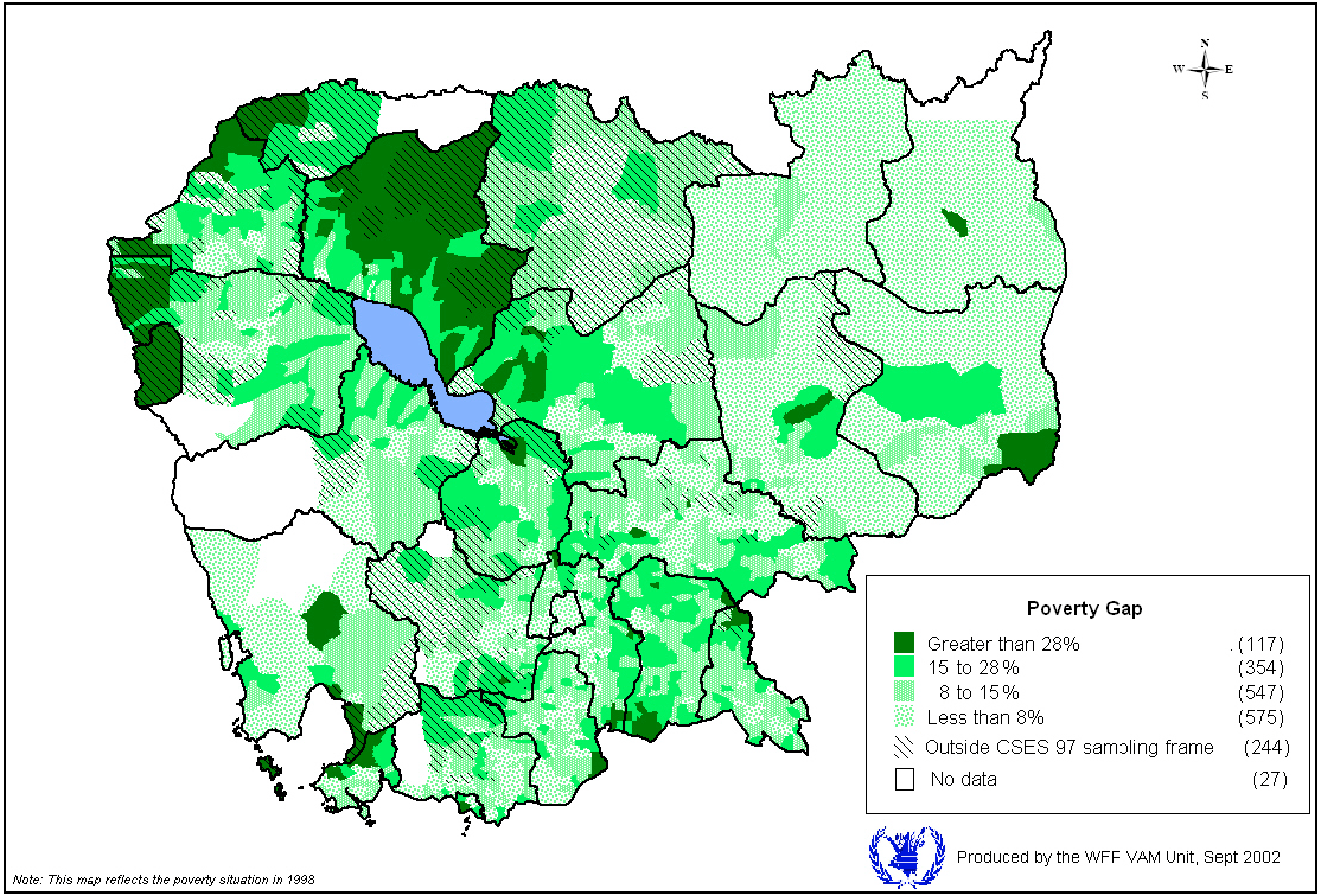
## APPENDIX F Commune-level Poverty Rates[19]

Commune Poverty Rates

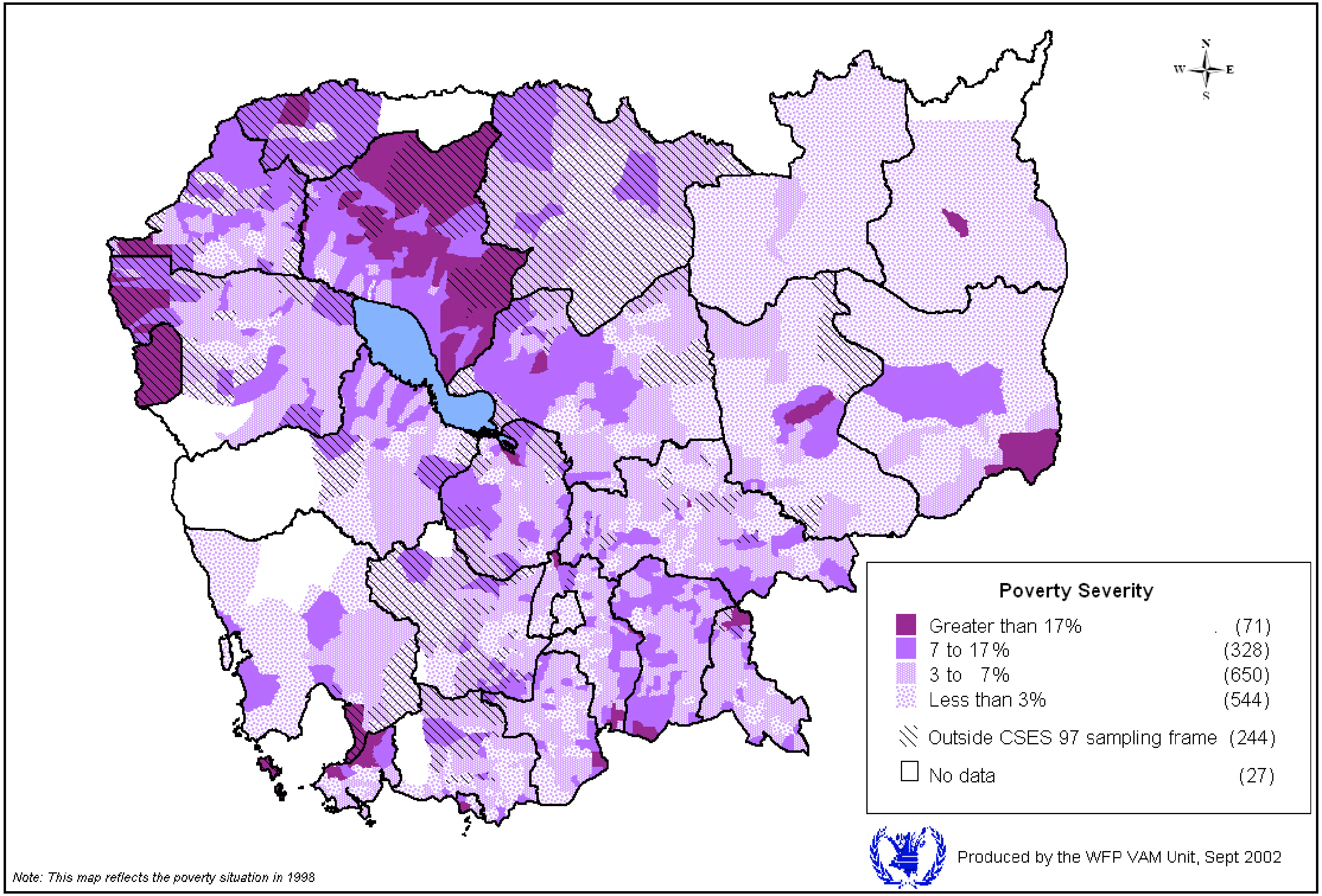



[19] See the last column of Table 2.

## Commune Poverty Gap

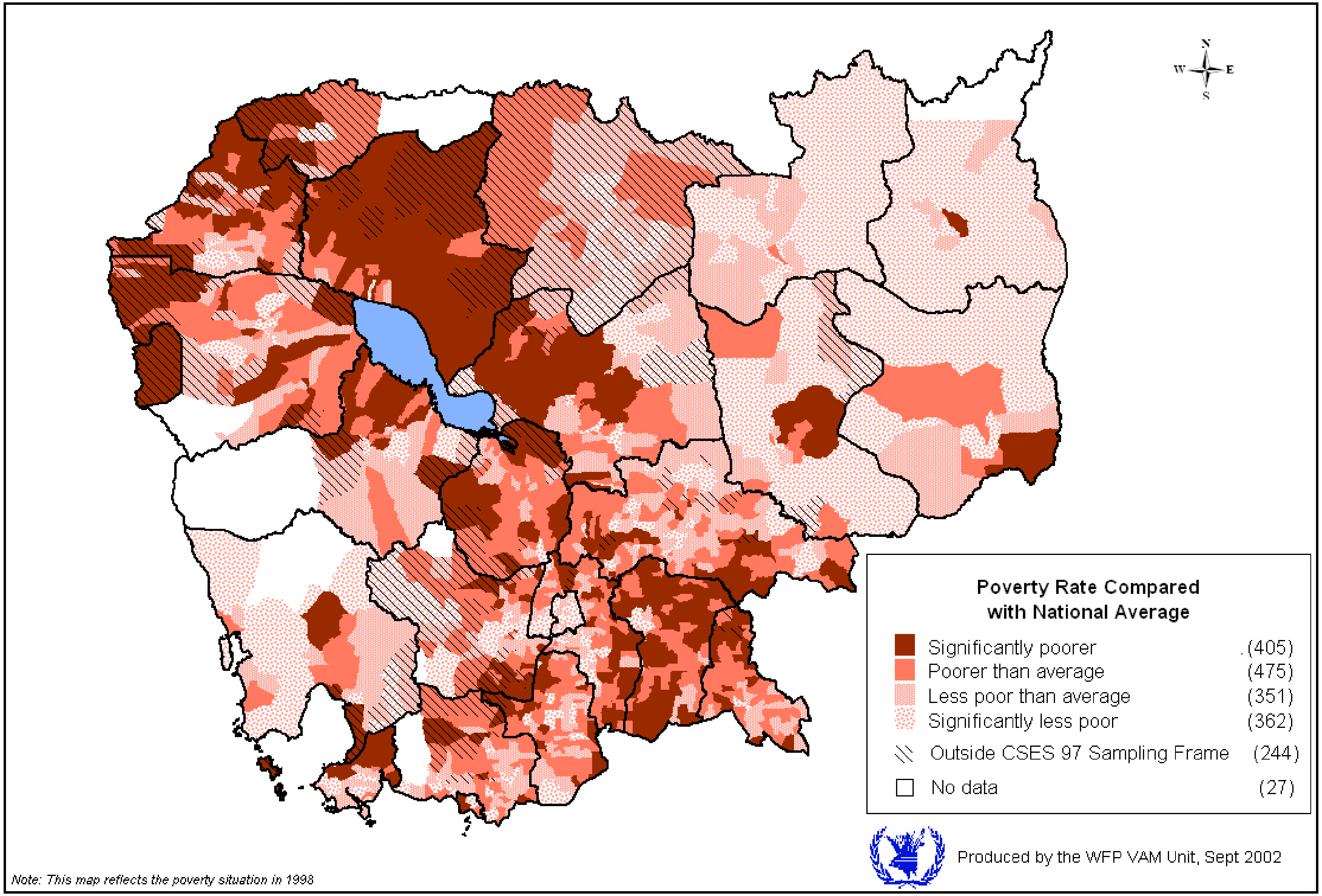


## Commune Poverty Severity

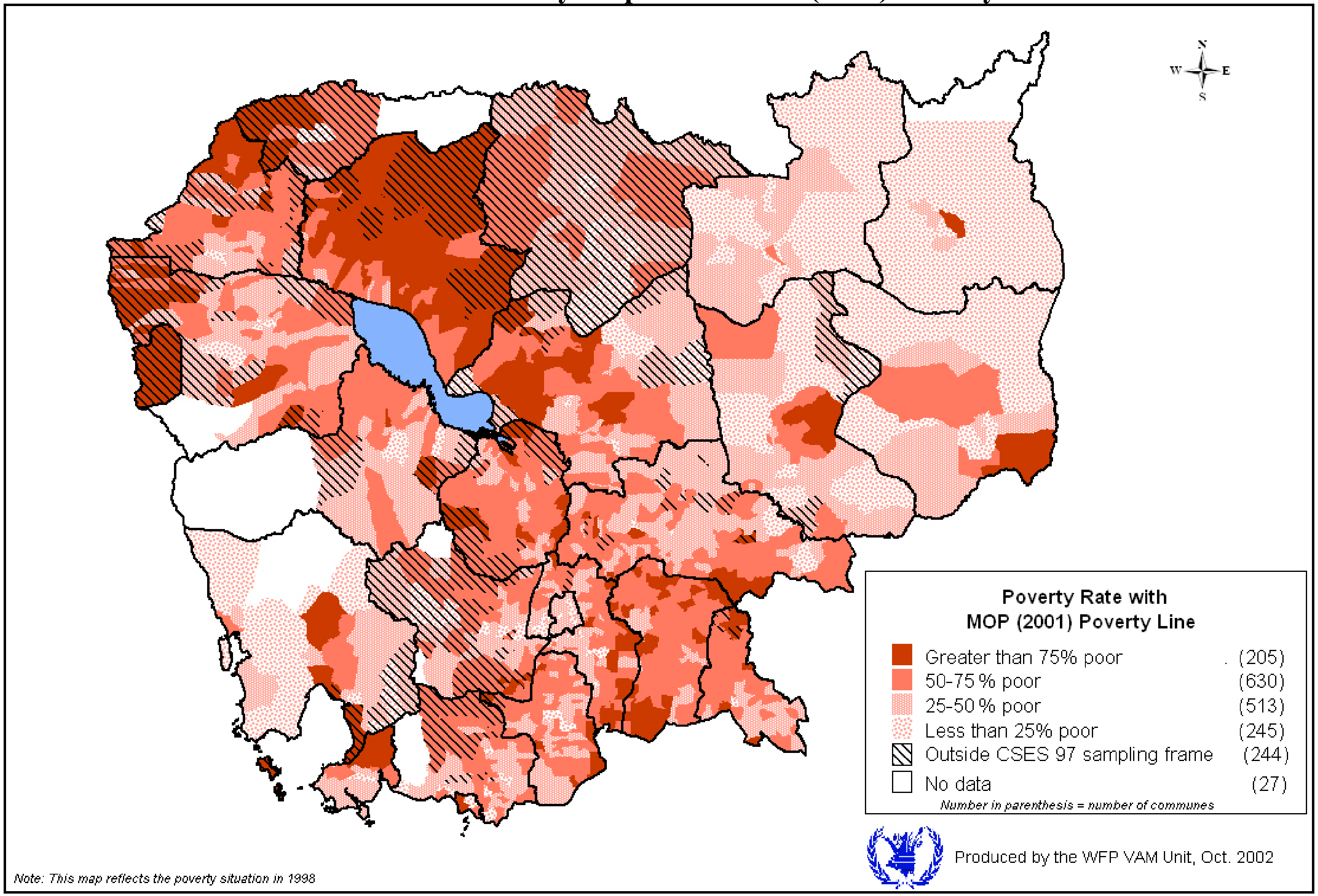


# APPENDIX G Commune-level Poverty Rates in Comparison With the National Poverty Rate

**Commune Poverty Rates Compared with the National Level Poverty Rate**

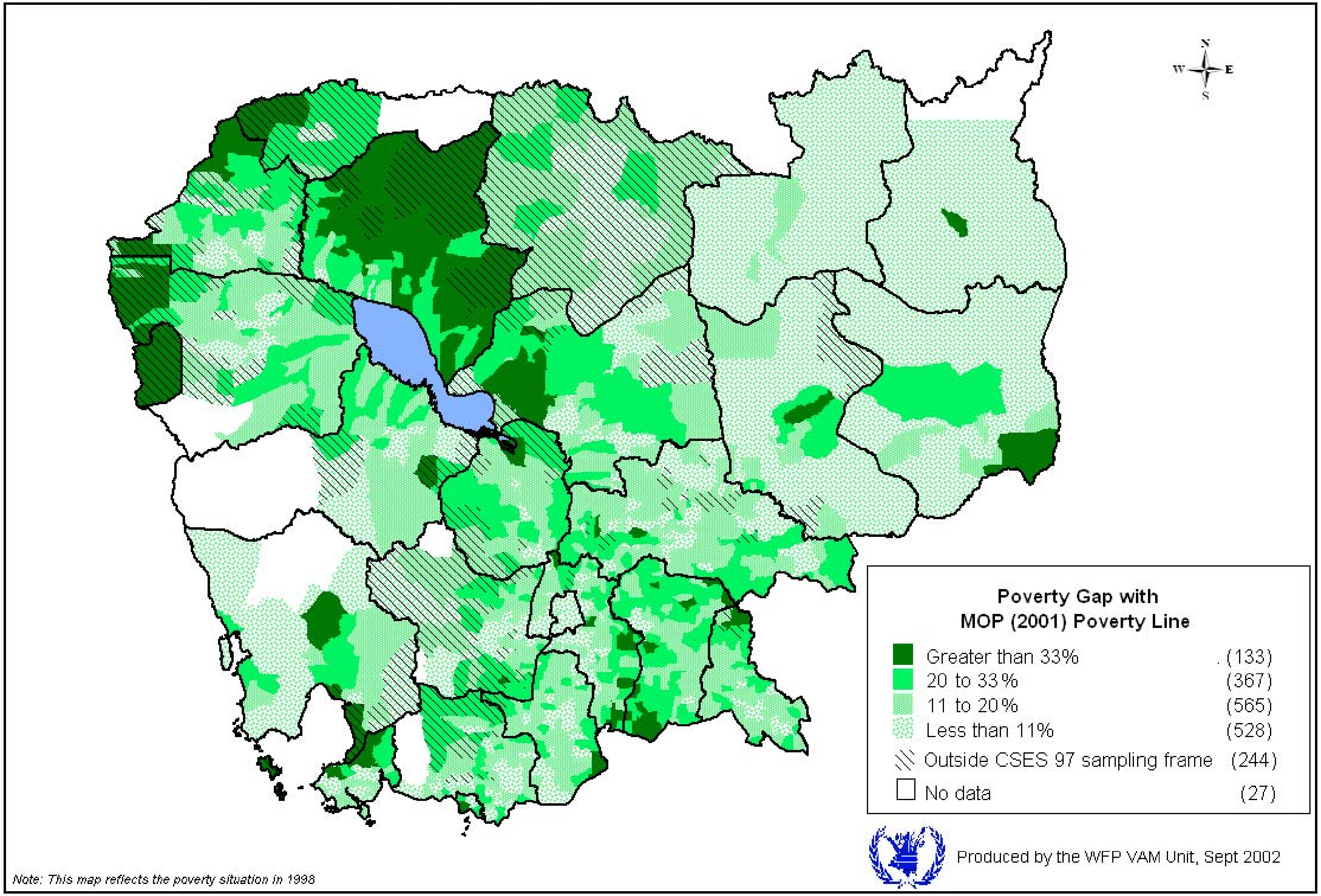


## APPENDIX H Commune-level Poverty Maps With MOP (2001) Poverty Line[20]

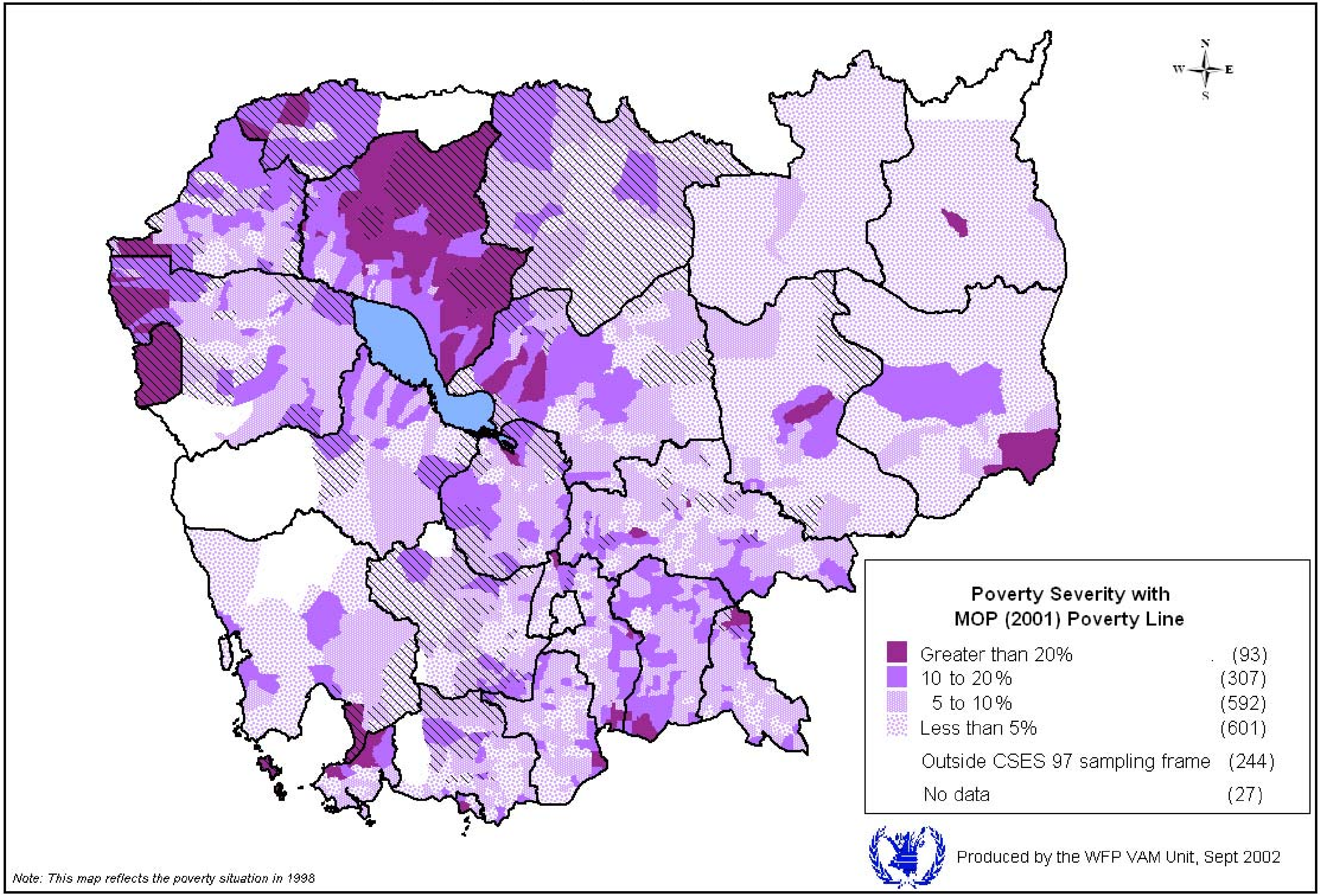



[20] See the CSES 1997 Unadjusted column in Table 2.

## Commune Poverty Gap with MOP (2001) Poverty Line

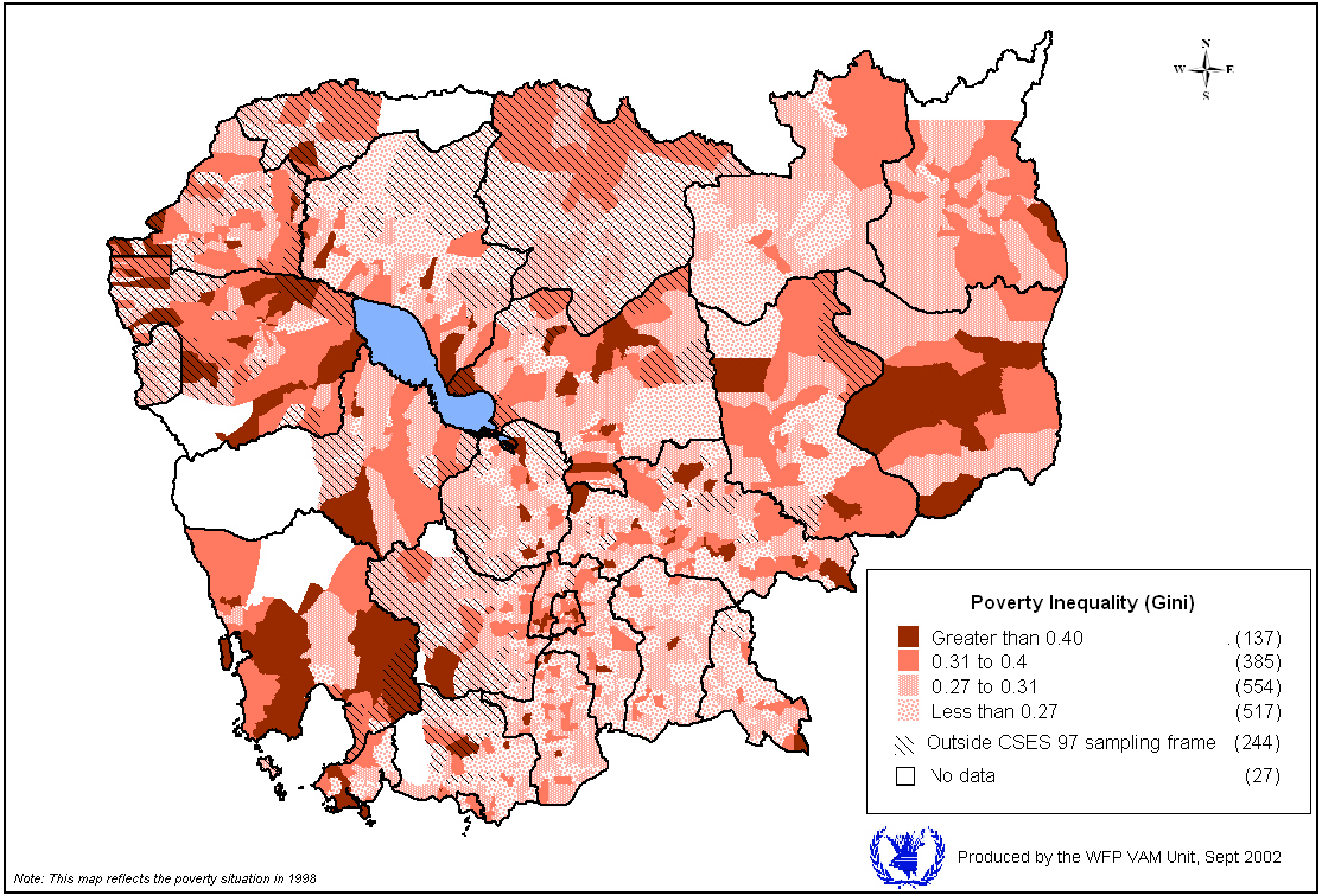


## Commune Poverty Severity with MOP (2001) Poverty Line

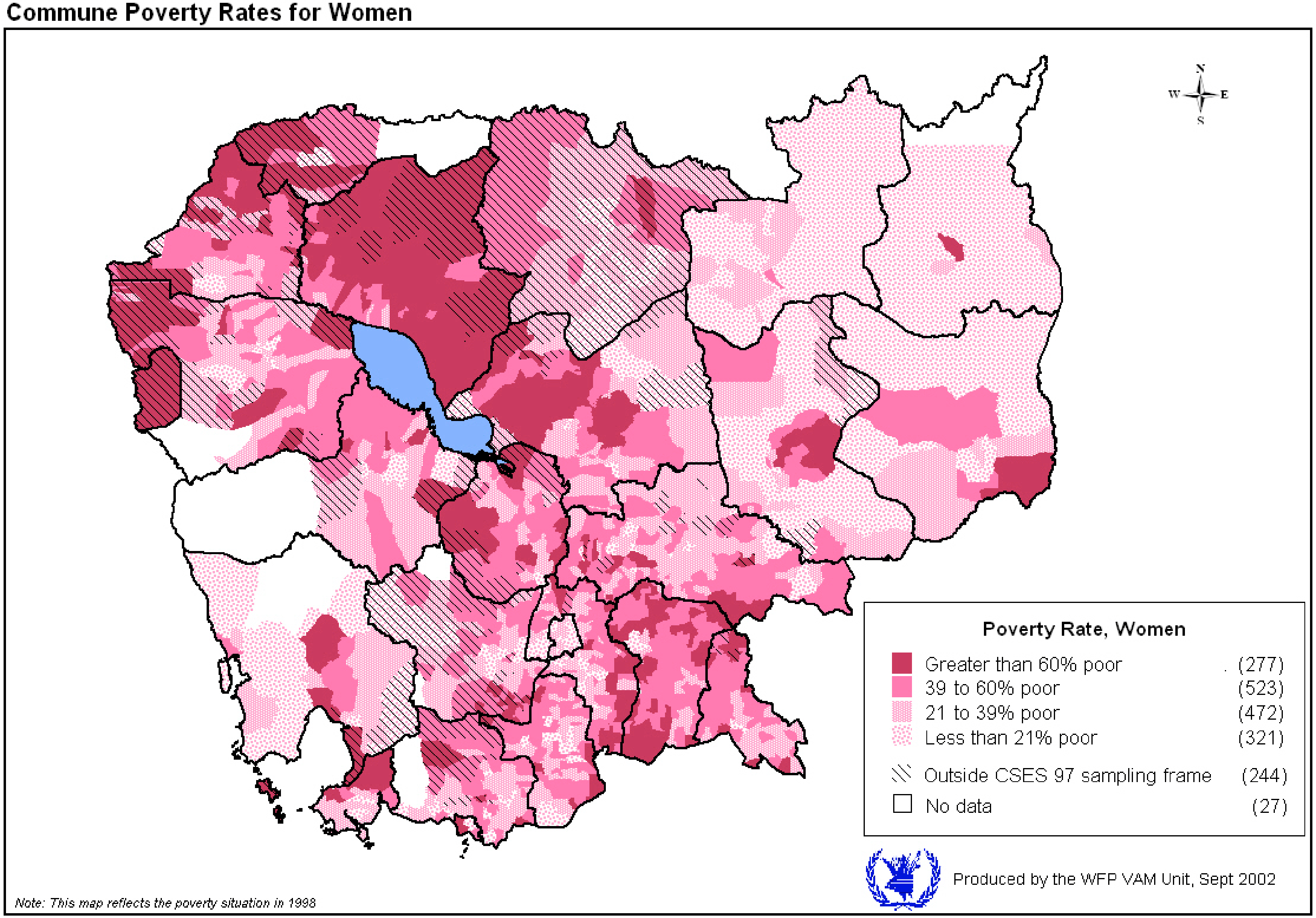


# APPENDIX I Commune-level Inequality (Gini Index)

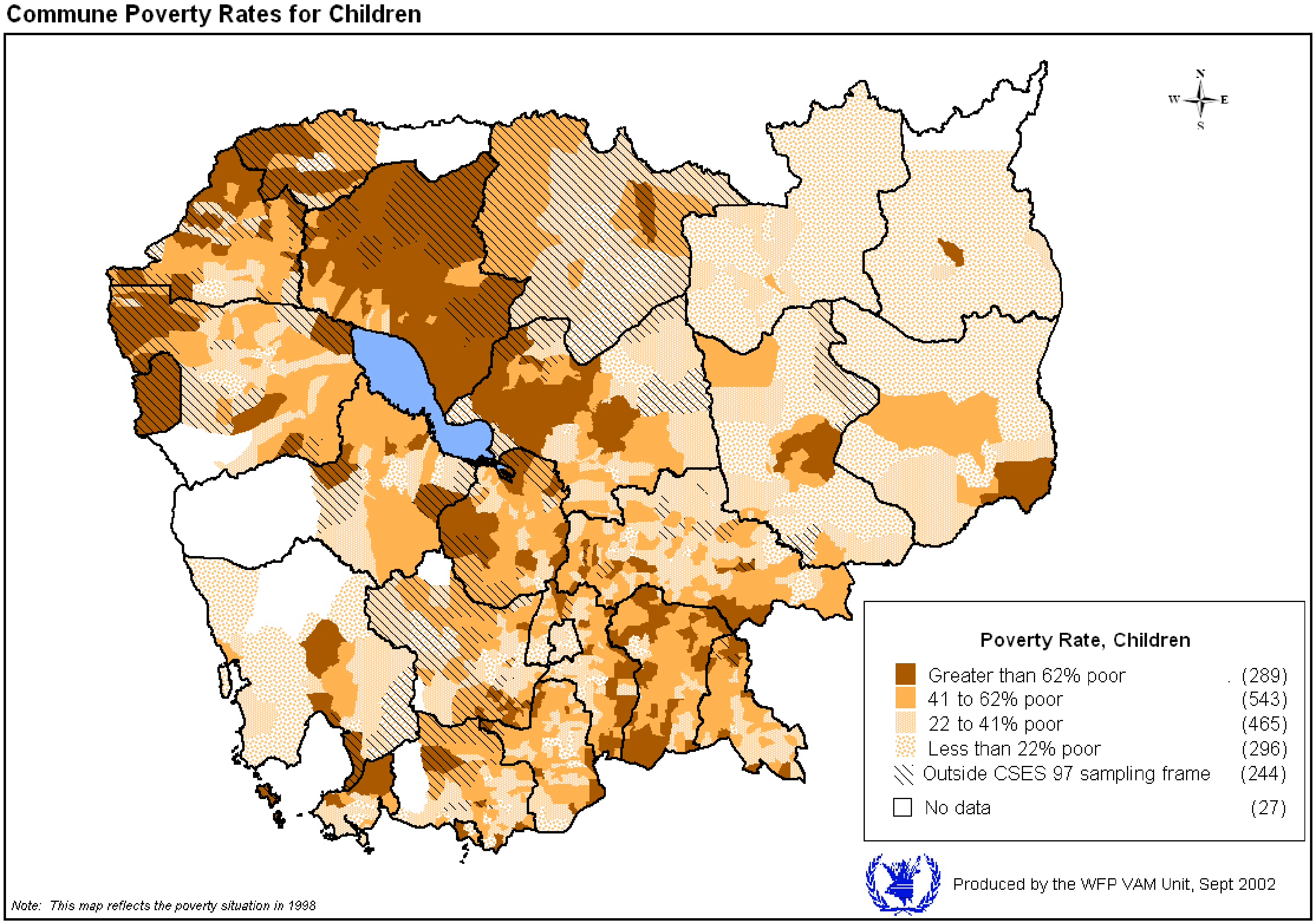
**Commune Poverty Inequality (Gini)**



## APPENDIX J Commune-level Poverty Rates for Women



## APPENDIX K Commune-level Poverty Rates for Children



# APPENDIX L Compiled Lists of Poverty Estimates

Poverty estimates at different levels of aggregation, including the provincial, district and commune levels, are provided in the CD-ROM attached to this report. Description of the contents of the CD-ROM is given in `readme.txt`.

World Food Programme

House No. 250, Street 63
P.O. Box 937
Phnom Penh, CAMBODIA
(855) 23 212 137
(855) 23 212 138
(855) 23 210 943
Fax (855) 23 218 749
Email: wfp.phnompenh@wfp.org