

INTRODUCTION
TO
POVERTY ANALYSIS

Table of Contents

CHAPTER 1: THE CONCEPT OF POVERTY AND WELL-BEING

- 1.1 The concept of well-being and poverty
- 1.2 why measure poverty

CHAPTER 2: MEASURING POVERTY

- 2.1 Steps in measuring poverty
- 2.2 Household surveys
- 2.3 Measuring poverty: choose an indicator of welfare

CHAPTER 3: POVERTY LINE

- 3.1 how to define a poverty line
- 3.2 Issues in choosing poverty line
- 3.3 Solution A: objective poverty lines
- 3.4 Solution B: subjective poverty lines

CHAPTER 4: POVERTY MEASURES

- 4.1 Headcount index
- 4.2 Poverty gap index
- 4.3 Squared poverty gap index
- 4.4 Sen index
- 4.5 Sen-Shorrocks-Thon index
- 4.6 Time taken to exit
- 4.7 Other measures

CHAPTER 5: POVERTY INDEXES--COMPARISON OF MEASURES

- 5.1 Introduction
- 5.2 Measurement errors
- 5.3 Equivalence scales
- 5.4 Robustness of ordinal poverty comparisons
- 5.5 Summary

CHAPTER 6: INEQUALITY MEASURES

- 6.1 Definition of inequality
- 6.2 Commonly used measures of inequality
- 6.3 Inequality comparisons
- 6.4 Decomposition of income inequality

CHAPTER 7: DESCRIBING POVERTY: POVERTY PROFILES

- 7.1 What is poverty profile
- 7.2 Additive poverty measures
- 7.3 Profile presentation
- 7.4 Poverty comparison over time
- 7.5 Excerpts from poverty profiles for Indonesia and Cambodia
- 7.6 Poverty mapping

CHAPTER 8: UNDERSTANDING THE DETERMINANTS OF POVERTY

- 8.1 What causes poverty
- 8.2 regional level characteristics
- 8.3 community level characteristics
- 8.4 Household and individual level characteristics
- 8.5 Understanding the quantitative determinants
- 8.6 Understanding the qualitative determinants
- 8.7 Analyzing the determinants of poverty

CHAPTER 9: POVERTY REDUCTION POLICIES

- 9.1 A framework for action
- 9.2 Practice and good examples

An Overview

This manual presents an introductory course on poverty analysis. The course is designed as part of a broader training program of the World Bank Institute called the Poverty Analysis Initiative (PAI). Its objective is to improve in-country capacity in poverty analysis in poor countries; it focuses especially on developing the skills of statisticians, policy analysts, and researchers in the analysis of data emerging from household surveys, in order to support the policy work related to the PRSP (Poverty Reduction Strategy Paper) process involving poverty identification, measurement, monitoring, and evaluation. This manual covers introductory topics related to poverty measurement and diagnostics, and applies these methods using household survey data. The topics included in this manual are: the concept of well-being, why measure poverty, how measure poverty, setting poverty lines, poverty indices and their comparisons, inequality measures, poverty profiles, the determinants of poverty, and how poverty analysis is linked to poverty reduction policies.

The manual includes discussions of these topics with illustrations from different countries. Many of the materials included in this manual are drawn from the work of Martin Ravallion of the World Bank's Development Research Group. The manual has benefited from work of others from inside and outside the World Bank. These course materials were used for training in a Regional workshop in the Philippines and a national workshop in Cambodia. This manual has been prepared under the general direction of Shahid Khandker of the World Bank Institute with contributions from Jonathan Haughton, Kathleen Beegle, Celia Reyes, and Nidhiya Menon. The STATA exercises were prepared by Hussain Samad and Changqing Sun.

The manual provides the tools for carrying out basic poverty analysis with household data, using STATA software. It includes tailor-made exercises to illustrate the techniques and measures discussed in different chapters. The manual provides a course curriculum with an outline of a full-time 9-day training program of morning lectures and afternoon practice exercises..

The manual will be updated in the future with new exercises based on other software such as SPSS and SAS. This will also provide links to other poverty analysis tools as well as on-line software such as DAD.

We hope that the course materials presented here are useful for self-learning. If you have any questions, please contact Shahid Khandker at skhandker@worldbank.org. Your opinions and suggestions will help improve the presentation of the course materials and make them more useful. The ultimate goal of this course is to enhance local capacity in poverty analysis. We hope that this manual will contribute to this goal.

CHAPTER 1

The Concept of Poverty and Well-being

1.1 The concept of well-being and poverty

There are many different definitions and concepts of well-being. For example, we can think of one's well-being as the command over commodities in general; people are better off if they have a greater command over resources. Or, we can think of the ability to obtain a *specific* type of consumption good (e.g. food, housing). People who have a lack of "capabilities" might have lower well-being (Sen 1987). Lack of capability means inability to achieve certain "functioning" ("being and doings"), lack of well-being, and vulnerability to income and weather shocks. Thus, poverty means either lack of command over commodities in general (i.e., a severe constriction of the choice set (Watts 1968)) or a specific type of consumption (e.g., too little food energy intake) deemed essential to constitute a reasonable standard of living in a society, or lack of "ability" to function in a society.

This course focuses on what is typically referred to as poverty, namely whether households or individuals have enough resources or abilities to meet their needs. This aspect is based on the comparison of individuals' income, consumption, education or other attributes with some defined threshold below which they are considered as being poor in that attribute. Poverty is a deprivation of essential assets and opportunities to which every human being is entitled. Thus, clearly, one can think of poverty from a non-monetary perspective. Although widely used, monetary poverty is not the exclusive paradigm for poverty measurement and non-monetary dimensions of poverty are useful in assessing poverty components, particularly for case study research.

Poverty is also associated with insufficient outcomes with respect to health, nutrition and literacy, to deficient social relations, to insecurity, and to low self-confidence and powerlessness. In some cases, it is feasible to apply the tools that have been developed for monetary poverty measurement to non-monetary indicators of well-being. A few examples of dimensions of well-being for which the techniques could be used include:

- **Health and nutrition poverty:**

The health status of household members can be taken as an important indicator of well-being. One could focus on the nutritional status of children as a measure of outcome, as well as on the incidence of specific diseases (diarrhea, malaria, respiratory diseases), or life expectancy for different groups within the population.

- **Education poverty:**

In the field of education, one could use the level of literacy as the defining characteristic, and some level judged as the threshold for illiteracy as the "poverty line". Another alternative consists in comparing the number of years of education completed to the expected number of years of education that should be in principle completed.

There are certainly other concepts of well-being beyond poverty, both measured through monetary concepts and non-monetary dimensions. Consider inequality. Inequality focuses on the distribution of attributes, such as income or consumption, across the population. This is based on the premise that the relative position of individual of households in society is an important aspect of their welfare. In addition, the overall level of inequality in a country, region or population group, in terms of

monetary and on-monetary dimensions, is also an important summary indicator of the level of welfare in that group. Another notion of well-being is vulnerability. This is defined as the probability or risk today of being in poverty, or falling deeper in to poverty, in the future. Vulnerability is a key dimension of well-being since it affects individuals' behavior (in terms of investment, production patterns, coping strategies) and the perceptions of their own situations.

According to the World Bank (World Bank 2000), "poverty is pronounced deprivation in well-being", where well-being can be measured by an individual's possession of income, health, nutrition, education, assets, housing, and certain rights in a society such as freedom of speech. Also poverty is a lack of opportunities, powerlessness, and vulnerability. Poverty is truly a multi-dimensional phenomenon in such a setting and requires multi-dimensional policy and program interventions in order to improve the well-being of individuals and, hence, make them free from poverty.

For example, economic growth is crucial to the creation of opportunities. However, growth is not enough; the poor and the vulnerable may not be able to benefit from growth, because they lack health, or skills, or access to basic infrastructure. Empowerment is crucial for the poor to take advantage of opportunities created by growth. What do we mean by empowerment? Empowerment means increasing the capacity of poor people to affect the decisions that have a bearing on their lives, by investing in them and by removing barriers that they face to engaging in political, social and economic activities. Again, growth and empowerment may not be enough for a large number of poor in a society. There may be a lot of people who are vulnerable to risks such as illness and injury, economic downturns, and natural disasters, which limit their opportunities as well as capabilities to avail of these opportunities created by growth. Consequently, these people are forced to choose low-risk, low-return activities, and cause them to lose productive assets and be set back further into poverty. Thus, public safety net mechanisms should be there to reduce the impact of such shocks on the poor and help the poor manage the consequences of these shocks.

Although the concepts, measures and analytical tools can be applied to numerous dimensions of well-being, such as income, consumption, health, education, asset ownership, and vulnerability, this course focuses primarily on the consumption dimension and only casually refers to the other dimensions of poverty.

1.2 Why measure poverty

Why measure poverty? Perhaps the strongest justification is that provided by Ravallion (1998), who argued that "a credible measure of poverty can be a powerful instrument for focusing the attention of policy makers on the living conditions of the poor." Poverty data can inform policies intended to reduce poverty. A good measure of poverty would:

- allow one to assess the effects of projects, or crises, or government policies, on poverty,
- permit one to compare poverty over time,
- enable one to make comparisons with other countries, and
- target the poor with a view to improving their position.

Alternatively, we could think of four questions we want to answer with poverty analysis:

- How many are poor or how bad is the poverty problem? *Measuring poverty.*
- Who is poor? *The poverty profile.*
- Why are they poor? *The determinants of poverty.*

- What happens to poverty if...?

Policy implications.

“Our dream is a world free of poverty,” writes the World Bank, and its first mission statement is “to fight poverty with passion and professionalism for lasting results.” The institution’s success in pursuing this goal can only be judged if there are adequate measures of poverty. For policy purposes, the most important reason for measuring poverty is not the need for a descriptive number, but to make poverty comparisons in order to develop antipoverty programs and monitor development progress and growth strategies.

Poverty profiles are typically developed from our measures of poverty. For a straightforward example, see Nicholas Prescott and Menno Pradhan, *A Poverty Profile of Cambodia* (1997). Answering basic policy questions on poverty in Cambodia, for example, requires a systematic information base on the distribution of living standards among the population. Constructing a nationwide poverty profile supports the government’s effort to strengthen poverty reduction policies. A poverty profile sets out the major facts on poverty (and, typically, inequality), and then examines the pattern of poverty, to see how it varies by geography (by region, urban/rural, mountain/plain, etc.), by community characteristics (e.g. in communities with and without a school, etc.), and by household characteristics (e.g. by education of household head, by size of household). A well-presented poverty profile is invaluable. Although it typically uses rather basic techniques such as tables and graphs, it can be immensely informative.

Whatever the justification for measuring poverty, it is commonly done based on data gathered from household surveys. All developed countries, and about two-thirds of developing countries, have undertaken nationally representative household surveys to collect information on consumption and/or income; in many cases, these surveys have been repeated over time.

1.2.1 Why measure poverty: developing a growth strategy

Understanding the characteristics of poverty can help policy makers think about the impact of growth strategies. For example, with measures of poverty over time, we can assess if poverty has increased or decreased, or whether general economic growth helped the poor. With changes in relative prices, we can evaluate how these changes affect the poor. We can use poverty data to inform economy-wide policy reforms and how the poor are affected by such reforms.

Poverty reduction was the central goal of Cambodia’s first Socioeconomic Development Plan (1996-2000). Better and up-to-date information about the poor is essential to assist the Government in designing effective policies for attacking poverty. Who are the poor? How many poor are there? Where do they live? What are their sources of income? Policies intended to help the poor can not succeed unless the Government knows who the poor are and how they are likely to respond different growth strategies.

1.2.2 Why measure poverty: social spending

By collecting information on households and their economic status, we can assess who uses public services and who gains from government subsidies. If programs are cut or there is retrenchment of the public sector, poverty data help inform us of the effects of these plans on the poor. Using information on poverty, we can simulate the impact of different policies.

1.2.3 Why measure poverty: targeted interventions

With data on household poverty status, we can evaluate the impact of programs on the poor and determine whether these programs meet their goals with respect to targeting certain households. Targeting the design and placement of anti-poverty programs is essential to effectively and efficiently reach disadvantaged groups and backward areas. Poverty profiles can help governments identify potential targeting by region, employment, education and gender.

Probably the most important use of the poverty profile is to support efforts to target development resources towards poorer areas, aiming to reduce aggregate poverty through regional targeting. Which regions should command priority in targeting? This question can only be answered at a highly aggregate level by most survey data (like the Cambodian SESC of 1993/94 and the CSES of 1999) because of the limited number of geographic domains that were sampled. So, while this provides a broad sense of the appropriate policy orientation in regional targeting, it is obviously of somewhat limited practical value for choosing the geographic placement of project interventions. However, the poverty rates from these surveys can still be of broad use for targeting. For example, in the CSES 1999, poverty is lowest in Phnom Penh where the headcount poverty rate was 15% compared to the national poverty rate of 51% from the CSES 1999.

Another type of targeting is employment targeting. The ability of the vast majority of households in Cambodia to escape poverty will depend on their earnings from employment. Thus it is important to examine the relationship between poverty and the types of employment of working-age household members. Looking first at the distribution of poverty incidence, the highest poverty rate was found among people living in households headed by farmers (46% in 1993/94 in Cambodia). By contrast, households headed by someone working in the government are least likely to be poor; in these occupations the poverty rate was 20% (1993/94). Clearly policies that aim at reducing poverty through enhancing income-generating capabilities should be targeted towards the agricultural sector.

The relationship between poverty and education is particularly important because of the key role played by education in raising economic growth and reducing poverty. The better educated have higher incomes and thus are much less likely to be poor. Cambodians living in households with an uneducated household head are more likely to be poor, with a poverty rate of 47% in 1993/94. With higher levels of education, the likelihood of being poor falls considerably. Raising education attainment is clearly a high priority in order to improve living standards and reduce poverty.

The relationship between gender and poverty may also indicate another targeting strategy for poverty reduction. One indicator of the gender gap is whether female-headed households are worse off than those headed by males. This might be a concern in Cambodia since about 25% of the population lives in households headed by women. In fact, the CSES 1999 data show that female-headed households in Cambodia are *not* more likely to be poor than male-headed households. There was no difference in poverty rates between the population of female-headed and male-headed households. In fact, the poverty rate was slight *lower* among female-headed households in the CSES 1999 (48% compared to 52% for male-headed households).

1.2.4 Why measure poverty: Poverty Reduction Strategy Paper (PRSP)

Counting the poor is very important to help policymakers to design programs and policies to fight poverty. The World Bank introduced the Poverty Reduction Strategies Paper (PRSP) for the Highly Indebted Poor Countries (HIPC) in 1999, which is supposed to be a country-driven policy paper setting out a strategy for fighting poverty. The PRSP is becoming a central instrument of the Bank's lending programs in many poor countries. The underlying principles of the PRSP are that this policy paper on

fighting poverty should be country-driven, results-oriented and comprehensive, based on the participation of civil societies along with partnerships from donors. The basic thrust of the PRSP process is that this is a long-term perspective of a country's development plan to reduce poverty. The key step for developing the PRSP lies in understanding the characteristics and causes of poverty, once it is known how many poor are there in a country. Thus, measuring poverty is the central theme of policy papers such as the PRSP. But this is only a means to achieving an end. The end is reducing the number of poor. Thus, once poverty is measured and the poor are identified, the next steps in the PRSP are to choose public actions and programs that have the greatest impact on poverty, identify indicators of progress, and monitor change in a systematic manner. Poverty measurement and diagnostics are therefore central to informing policy making for poverty reduction in many countries.

Chapter 2

Measuring Poverty

2.1 Steps in measuring poverty

The goal of this chapter is to set out a method for measuring poverty. There is an enormous literature on the subject, so we just set out the main practical issues, with some suggestions for further reading for those interested in pursuing the subject in more depth.

Three steps need to be taken in measuring poverty (for more discussion see Ravallion, 1998). These include:

- defining an indicator of welfare,
- establishing a minimum acceptable standard of that indicator to separate the poor and the non-poor (often known as a poverty line), and
- generating a summary statistic to aggregate the information from the distribution of this welfare indicator and its position relative to minimum acceptable standards.

This chapter focuses on the indicators of welfare, while chapter 3 discusses the issues involved in setting a poverty line and chapter 4 dwells on measures of aggregate welfare and its distribution.

2.2 Household surveys

2.2.1 Key survey issues

Before discussing the various indicators of welfare, let us look briefly at household surveys, which are the main instruments for collecting data to support poverty analysis. Household surveys are extremely important in making poverty comparisons, but care must be taken in setting up and interpreting the data that are obtained from such surveys. The analyst should be aware of the following issues (see Ravallion (1999) for details):

- The sample frame:* The survey may represent a whole country's population, or some more narrowly defined sub-set, such as workers or residents of one region. The appropriateness of a survey's particular sample frame will depend on the inferences one wants to draw from it.
- The unit of observation:* This can be the household itself or the individuals within the household. A household is usually defined as a group of persons eating and living together.
- The number of observations over time:* A single cross-section, based on one or two interviews, is the most common. Longitudinal surveys in which the same household is re-

surveyed over an extended period (also called panel data sets) have been done in a number of countries.

- iv) *The principal living standard indicator collected:* The most common indicators used in practice are based on household consumption expenditure and household income. The most common survey used in poverty analysis is a single cross-section for a nationally representative sample, with the household as the unit of observation, and it includes either consumption or income data. This form of survey is cheaper per household surveyed than most alternatives, and this allows a larger sample than with longitudinal and or individual based survey. A larger sample of household-level data gives greater accuracy in estimating certain population parameters, such as average consumption per capita, but, of course, it can lose accuracy in estimating other variables, such as the number of under-nourished children in a population (which may require oversampling of the target group). It should not, however, be presumed that the large household consumption survey is more cost-effective for all purposes than alternatives, such as using smaller samples of individual data.

2.2.2 Common survey problems

The following are the main problems for the practitioner to be aware of when interpreting household consumption or income data from a household survey.

2.2.2.1 Survey design

Even very large samples may give biased estimates for poverty measurement if the survey is not random, or if the data extracted from it have not been corrected for possible biases, such as due to sample stratification. A random sample requires that each person in the population or each sub-group in a stratified sample - have an equal chance of being selected. This guarantees statistical independence, the assumption that underlies most of the results used routinely in making statistical inferences about population parameters from sample surveys.

However, the poor may not be properly represented in sample surveys; for example they may be harder to interview because they live in remote areas, or are itinerant, or live illegally in the cities and so do not appear on the rosters of the local authorities. Household surveys often miss one distinct sub-group of the poor: those who are homeless. Also, some of the surveys that have been used to measure poverty were not designed for this purpose, in that their sample frames were not intended to span the entire population; examples include the labor force surveys (widely used for poverty assessments in Latin America) for which the sample frame is typically restricted to the "economically active population," which precludes certain sub-groups of the poor.

Key questions to ask about the survey are:

- a) Does the sample frame (the initial listing of the population from which the sample was drawn) span the entire population?
- b) Is there likely to be a response bias, in that the likelihood of cooperating with the interviewer is not random?

There are various methods of sampling that can help achieve a more cost-effective survey than would be possible with simple random sampling. Stratified random sampling –whereby different sub-groups of the population have different (but known) chances of being selected but all have an equal

chance in any given subgroup - can increase the precision in poverty measurement obtainable with a given number of interviews; for example, one can over-sample certain regions where the poor are thought to be concentrated.

2.2.2.2 Goods coverage and valuation

The coverage of goods and income sources in the survey should be comprehensive, covering both food and non-food goods, and all income sources. Consumption should cover all monetary expenditures on goods and services consumed plus the monetary value of all consumption from income in kind, such as food produced on the family farm and the rental value of owner-occupied housing. Similarly, the income definition should include income in kind. Local market prices often provide a good guide for valuation of own-farm production or owner occupied housing. However, whenever prices are unknown, or are an unreliable guide to reflect opportunity costs, serious valuation problems can arise. The valuation of access to public services is also difficult, though important. For transfers of in-kind goods, prevailing equivalent market prices are generally considered to be satisfactory for valuation. Non-market goods present a more serious problem, and there is no widely preferred method; we return to this problem in more detail below

2.2.2.3 Variability and the time period of measurement

Inter-temporal variability has implications for a number of the choices made in measurement using survey data. One is the choice between income-based and consumption-based measures. One reason for preferring current consumption to current income as the indicator of living standards is because current income usually varies significantly more than current consumption. The incomes of the poor often vary over time in fairly predictable ways, particularly in underdeveloped rural economies depending on rain-fed agriculture. Typically, some kinds of consumption smoothing mechanisms available to the poor, such as through savings and community-based mutual risk-sharing arrangement. This observation has two distinct implications for welfare measurement:

- i) current consumption will almost certainly be a better indicator than current income of current standard of living because current consumption reflects more accurately how much resource households control, and
- ii) current consumption may then also be a good indicator of long-term well-being, as it will reveal information about incomes at other dates, in the past and future.

However, a number of factors can make current consumption a "noisy" welfare indicator. Even with ideal smoothing, consumption will still (as a rule) vary over the life-cycle. This may be less of a problem in traditional societies where resource pooling within an extended family is still the norm, though that is rapidly changing. More importantly, there are other sources of noise in the relationship between current consumption and long-term standard of living. Different households may face different constraints on their opportunities for consumption smoothing. It is generally thought that the poor are far more constrained in their ability to smooth consumption due to lack of borrowing options than the non-poor (also suggesting that life-time wealth is not the only parameter of life-time welfare).

2.2.2.4 Comparisons across households at similar consumption levels

Household size and demographic composition vary across households, as do prices, including wage rates. As a result, it takes different resources to make ends meet for different households. In other words, at a given level of household expenditure, different households may achieve different levels of well-being. There are various welfarist approaches based on demand analysis, including equivalence scales, true cost-of-living indices, and equivalent income measures, which try to deal with this problem. The basic idea of these methods of welfare measurement is to use demand patterns to reveal consumer preferences over market goods. The consumer is assumed to maximize utility, and a utility metric is derived that is consistent with observed demand behavior, relating consumption to prices, incomes, household size, and demographic composition. The resulting measure of household utility will typically vary positively with total household expenditures, and negatively with household size and the prices faced.

The most general formulation of this approach is the concept of "equivalent income", defined as the minimum total expenditure that would be required for a consumer to achieve his or her actual utility level but evaluated at pre-determined (and arbitrary) reference prices and demographics fixed over all households. This gives an exact monetary measure of utility (and, indeed, it is sometimes called "money metric utility"). Quite generally, equivalent income can be thought of as money expenditures (including the value of own production) normalized by two deflators: a suitable price index (if prices vary over the domain of the poverty comparison) and an equivalence scale (since household size and composition varies). The precise form of these deflators will depend on preferences, which (in practice) are usually taken to be revealed by demand behavior. These deflators are discussed further in the next section. There are a number of problems that one should be aware of in all such behavioral welfare measures. A serious problem arises when access to non-market goods (public services, and community characteristics) varies across households. The consumption of market goods only reveal preferences conditional on these non-market goods; they do not, in general, reveal unconditional preferences over both market and non-market goods. A revealed set of conditional preferences over market goods may be consistent with infinitely many utility functions representing preferences over all goods. It is then a big step to assume that a particular utility function that can be found to support observed consumption behavior at an optimum is also the one that should be used in measuring well-being.

Ideally we should not have to rely solely on a household level survey in making interpersonal comparisons of welfare. A separate community survey (done at the same time as the interviews, and possibly by the same interviewers) can provide useful supplementary data on the local prices of a range of goods and local public services. By matching these to the household level data one can greatly improve the accuracy and coverage of household welfare assessments. This has become common practice in the World Bank's LSMS surveys.

2.2.3 Key features of LSMS surveys

The LSMS surveys have two key features: multi-topic questionnaires, and considerable attention to quality control. Let's consider each in more detail.

2.2.3.1 Multi-topic questionnaires

The LSMS surveys ask about a wide variety of topics, and not just demographic characteristics or health experience or some other narrow issue.

- The most important single questionnaire is the *household questionnaire*, which often runs to 100 pages or more. Although there is an LSMS template, each country needs to adapt and test its own version. The questionnaire is designed to ask questions of the best-informed household member. The household questionnaire asks about household composition, consumption patterns including food and non-food, assets including housing, landholding and other durables, income and employment in agriculture/non-agriculture and wage/self-employment, socio-demographic variables including education, health, migration, fertility, and anthropometric information (especially the height and weight of each household member).
- There is also a *community questionnaire*, which asks community leaders (teachers, health workers, village officials) for information about the whole community, such as the number of health clinics, access to schools, tax collections, demographic data, and agricultural patterns. Sometimes there are separate community questionnaires for health and education.
- The third part is the *price questionnaire*, which collects information about a large number of commodity prices in each community where the survey is undertaken. This is useful because it allows analysts to correct for differences in price levels by region, and over time.

2.2.3.2 Quality control:

The LSMS surveys are distinguished by their attention to quality control. Here are some of the key features:

- Most importantly, they devote a lot of attention to obtaining a representative national sample (or regional sample, in a few cases). Thus the results can usually be taken as nationally representative. It is surprising how many surveys are undertaken with less attention to sampling, so one does not know how well they really represent conditions in the country.
- The surveys make extensive use of "screening questions" and associated skip patterns. For instance, a question might ask whether a family member is currently attending school; if yes, one jumps to page x and asks for details; if no, then the interviewer jumps to page y and asks other questions. This cuts down on interviewer errors.
- Numbered response codes are printed on the questionnaire, so the interviewer can write a numerical answer directly on the questionnaire. This makes subsequent computer entry easier, more accurate, and faster.
- The questionnaires are designed to be easy to change (and to translate), which makes it straightforward to modify them in the light of field tests.
- The data are collected by decentralized teams. Typically each team has a supervisor, two interviewers, a driver/cook, an anthropometrist, and someone who does the data entry onto a laptop computer. The household questionnaire is so long that it requires two visits for collecting the data. After the first visit, the data are entered; if errors arise, they can be corrected on the second visit, which is typically two weeks after the first visit. In most cases the data are entered onto printed questionnaires, and then typed into a computer, but some surveys now enter the information directly into computers.
- The data entered are subject to a series of range checks. For instance, if an age variable is greater than 100, then it is likely that there is an error, which needs to be corrected.

This concern with quality has some important implications, notably:

- The LSMS data is usually of high quality, with accurate entries and few missing values;
- Since it is expensive to maintain high quality, the surveys are usually quite small; the median LSMS survey covers just 4,200 households. This is a large enough sample for accurate information at the national level, and at the level of half a dozen regions, but not at a lower level of disaggregation (e.g. province, department, county).
- The LSMS data have a fairly rapid turnaround time, with some producing a statistical abstract (at least in draft form) within 2-6 months of the last interview.

2.3 Measuring poverty: choose an indicator of welfare

There are a number of conceptual approaches to the measurement of well-being. The most common approach is to measure economic welfare based on household consumption expenditure or household income, which is then assigned each resident in the household a share of the total amount. This is a *per capita* measure of consumption/expenditure or income. Of course, there are non-monetary measures of individual welfare, which can include indicators such as infant mortality rates in the region, life expectancy, proportion of spending devoted to food, housing conditions, and child schooling. Well-being is a broader concept than economic welfare, which only measures a person's command over commodities.

In brief, if we choose to assess poverty based on household consumption or expenditure *per capita*, use of an expenditure function will facilitate a lucid analysis. In simple terms, an expenditure function shows the minimum expense required to meet a given level of utility u , which is derived from a vector of goods x , at prices p . It can be derived from an optimization problem in which the objective function (for expenditure) is minimized subject to a set level of utility, in a framework where prices are fixed. The expenditure function thus provides the minimum amount of resources required to attain a set level of well-being, which is essentially what the poverty line is (see chapter 3). The function fits naturally into the framework of poverty assessment, and is the first step needed in most analyses of deprivation.

Let the consumption measure for the household i be denoted by y_i . Then an expenditure measure of welfare may be denoted by:

$$(2.1) \quad y_i = p \cdot q = e(p, x, u)$$

where p is a vector of prices of goods and services, q is a vector of quantities of goods and services consumed, $e(\cdot)$ is an expenditure function, x is a vector of household characteristics (e.g. number of adults, number of young children, etc.) and u is the level of "utility" or well-being achieved by the household. Put another way, given the prices (p) that it faces, and its demographic characteristics (x), y_i measures the spending that is needed to reach utility level u .

Typically, we compute the actual level of y_i from household survey data that include information on consumption. The details of this are discussed below. Once we have computed y_i , we can construct *per capita* household consumption for every individual in the household. Thus, for all people in the household's survey we have one measure of their individual-level well-being, based on consumption. Note that this approach, combined with our poverty line, assumes that all individuals in the household

have the same needs. In reality, different individuals have different needs based on their individual characteristics (age, sex, job, etc).

While estimating per capita consumption might seem straightforward, there are several factors that complicate its estimation. Table 2.1 reports both nominal and real (adjusted for inflation) estimates from three different household surveys in Cambodia. Using the 1997 Cambodia Socio-economic Survey (CSES), for example, nominal and real per capita consumption were 2,223 and 1,887 riels, respectively. Across years, the estimates in real terms for 1993/94 may not be directly comparable with the 1999 estimates because the surveys did not have the same set of questions regarding consumption. For example, consumption per capita was computed as 2,262 riels for 1993/94. Using the data from 1999, the estimate was 1,799. However, as is noted later, since the surveys are not exactly the same, the difference may not be real but rather an artifact of the different ways in which questions were asked. These issues will be discussed in more detail below.

Table 2.1: Summary of per capita consumption from Cambodian Surveys

Surveys	Nominal	Riels
SESC 1993/94	1,833	2,262
CSES 1997 (adjusted)	2,223	2,530
CSES 1997 (unadjusted)	1,887	2,153
CSES 1999 (Round 1)	2,037	1,630
CSES 1999 (Round 2)	2,432	1,964
CSES 1999 (both Rounds)	2,238	1,799

Note: All values are in Riels per person per day. Real values are estimated in 1993/94 Phnom Penh prices, as deflated by the value of the food poverty lines. Adjusted figures from 1997 incorporate corrections for possible underestimation of certain types of consumption (see Knowles 1998, and Gibson 1999 for details). Differences between Rounds 1 & 2 in 1999 are detailed in Gibson (1999).

Source: Gibson (1999)

Traditionally, we need a monetary measure to value household welfare. The two most obvious candidates are income and expenditure.

2.3.1 Candidate 1: Income

It is tempting to measure household welfare by looking at household income. Practical problems arise immediately: what is income? and can it be measured accurately? The most generally accepted measure of income is the one formulated by Haig and Simons:

$$\text{Income} \equiv \text{consumption} + \text{change in net worth.}$$

Example: Suppose I had assets of \$10,000 at the beginning of the year. During the year I spent \$3,000 on consumption. And at the end of the year I had \$11,000 in assets. Then my income was \$4,000, of which \$3,000 was spent, and the remaining \$1,000 added to my assets.

The first problem with this definition is that it is not clear what time period is appropriate. Should we look at someone's income over a year? Five years? A lifetime? Many students are poor now, but have good lifetime prospects, and we may not want to consider them as being truly poor. On the other hand if we wait until we have information about someone's lifetime income, it will be too late to help him or her in moments of poverty.

The second problem is measurement. It is easy enough to measure components of income such as wages and salaries. It may be possible to get adequate (if understated) information on interest, dividends, and income from some types of self-employment. But it is likely to be hard to get an accurate measure of farm income; or of the value of housing services; or of capital gains (e.g. the increase in the value of animals on a farm, or the change in the value of a house that one owns).

For instance, the Vietnam Living Standard Survey (VLSS; undertaken in 1993 and again in 1998) collected information on the value of farm animals at the time of the survey, but not the value a year before. Thus it was not possible to measure the change in the value of animal assets. Many farms that reported negative cash income may in fact have been building up assets, and truly had positive income.

It is typically the case, particularly in societies with large agricultural or self-employed populations, that income is seriously understated. This certainly appears to be the case for Vietnam. Table 2.2 shows income per capita for households in 1993 for each of five expenditure quintiles: a quintile is a fifth of the sample, and quintile 1 contains the poorest fifth of individuals, etc. For every quintile, households on average reported less income than expenditure, which is simply not plausible. This would imply that households must be running down their assets, or taking on much more debt, which was unlikely in a boom year like 1993.

Table 2.2: Income and expenditure by per capita expenditure quintiles, Vietnam

(In dong per capita per year, 1992/93)

	Lowest	Lower-mid	Middle	Mid-upper	Highest	Overall
Income/capita	494	694	956	1,191	2,190	1,105
Expenditure/capita	518	756	984	1,338	2,540	1,227
Memo: food						
... spending/capita	378	526	643	807	1,382	747
... as % of expenditure	73	70	65	60	54	61

Note: In 1993, exchange rate was about 10,000 dong/US\$. *Source:* VLSS93.

There are a number of reasons why income tends to be understated:

- People forget, particularly when asked in a single interview about items they may have purchased up to a year before.
- People may be reluctant to disclose the full extent of their income, lest the tax collector, or neighbors, get wind of the details.
- People may be reluctant to report income earned illegally - for instance from smuggling, or corruption, or poppy cultivation, or prostitution.
- Some parts of income are difficult to observe - e.g. the extent to which the family buffalo has risen in value.

Research based on the 1969-70 socio-economic survey in Sri Lanka estimated that wages were understated by 30%, business income by 39%, and rent, interest and dividends by 78%. It is not clear how much these figures are applicable elsewhere, but they do give a sense of the potential magnitude of the understatement problem.

2.3.2 Candidate 2: Consumption expenditure

Note that consumption includes both goods and services that are purchased, and those that are provided from one's own production ("in-kind").

In developed countries, a strong case can be made that consumption is a better indicator of lifetime welfare than is income. Income fluctuates from year to year, and in addition typically rises and then falls in the course of one's lifetime, whereas consumption remains relatively stable. The smoothing of short-term fluctuations in income is also a hallmark of the permanent income hypothesis, under which transitory income is saved while long-term ("permanent") income is largely consumed.

The life cycle of income and consumption is captured graphically in figure 1. However the available evidence does not provide strong support for this *life-cycle hypothesis* in the context of less-developed countries although households do appear to smooth out the substantial seasonal fluctuations in income that they face during the year (see Alderman and Paxson 1994; Paxson 1993).

A more practical case for using consumption, rather than income, is that households may be more able, or willing, to recall what they have spend rather than what they earned. Even so, consumption is likely to be systematically understated, because:

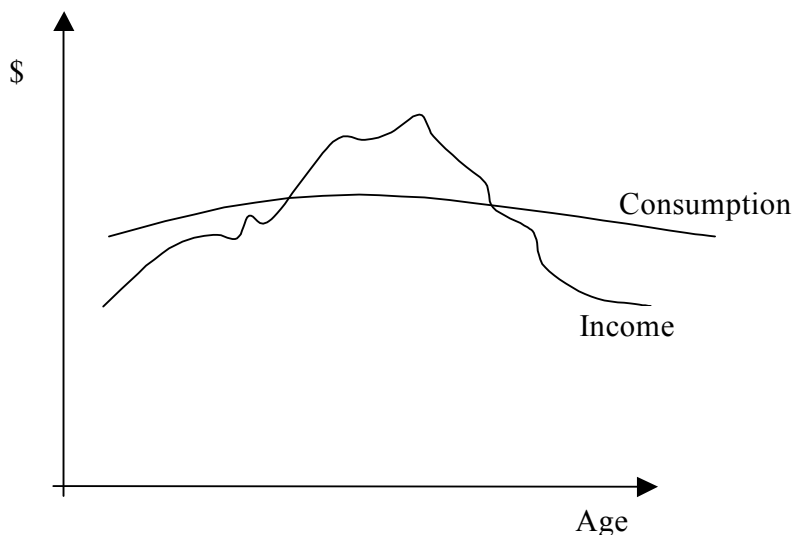


Figure 2.1 Life Cycle Hypothesis: Income and Consumption Profile over Time

- Households tend to under-declare what they spend on luxuries (e.g. alcohol, cakes) or illicit items (drugs, prostitution). For instance, the amount that households said they spent on alcohol, according to the 1972-73 household budget survey in the US, was just half the amount that companies said they sold!
- Questions matter. According to VLSS93, households devoted 1.7% of their expenditure to tobacco; the VLSS98 figures showed that this had risen to 3%. An increase of this magnitude is simply not plausible, and not in line with sales reported by the cigarette and tobacco companies. A more plausible explanation is that VLSS98 had more detailed questions about tobacco use. When the questions are more detailed, respondents are likely to remember in more detail, and report higher spending.

2.3.2.1 Measure durable goods

In measuring poverty it might be argued that only food, the ultimate basic need (which anyway constitutes three quarters of the spending of poor households), should be included. On the other hand even households that cannot afford adequate quantities of food devote some expenditures to other items (clothing, shelter, etc.). It is reasonable to suppose that if these items are getting priority over food purchases, then they must represent very basic needs of the household, and so should be included in the poverty line. This argument also applies to durable goods (housing, pots and pans, etc.).

The problem here is that durable goods, such as bicycles and TVs, are bought at a point in time, and then consumed (i.e. eaten up and destroyed) over a period of several years. Consumption should only include the amount of a durable good that is eaten up during the year, which can be measured by the change in the value of the asset during the year, plus the cost of locking up my money in the asset.

For instance, if my watch was worth \$25 a year ago, and is worth \$19 now, then I used \$6 worth of watch during the year; I also tied up \$25 worth of assets in the watch, money that could have earned me \$2.50 in interest (assuming 10%) during the year. Thus the true cost of the watch during the year was \$8.50.

A comparable calculation needs to be done for each durable good that the household owns. Clearly the margin of potential measurement error is large, since the price of each asset may not be known with much accuracy, and the interest rate used is somewhat arbitrary. The VLSS surveys asked for information about when each good was acquired, and at what price, and the estimated current value of the good. This suffices to compute the current consumption of the durable item, as the illustration in the following box shows.

Box: Calculating the consumption of durable goods - an illustration.

A household is surveyed in April 1998, and says it bought a TV two years earlier for 1.1m dong (about \$100). The TV is now believed to be worth 1m dong. Overall prices rose by 10% over the past two years. How much of the TV was consumed over the year prior to the survey?

- a. Recompute the values in today's prices. Thus the TV, purchased for 1.1m dong in 1996, would have cost 1.21m dong ($=1.1\text{m dong} \times (1+10\%)$) now.
- b. Compute the depreciation. The TV lost 0.21m dong in value in two years, or 0.105m dong per year (i.e. about \$7).
- c. Computer the interest cost. At a real interest rate of 3%, the cost of locking up 1m dong in the TV is now 0.03m dong per annum.

Thus the total consumption cost of the TV was 0.135m dong ($= 0.105 + 0.03$), or about \$10.

One might wonder why attention needs to be paid to calculating the value of durable goods consumption when the focus is on poverty - in practice mainly the ability to acquire enough food. The answer is that when expenditure is used as a yardstick of welfare, it is important to achieve comparability across households. If the value of durable goods were not included, one might have the impression that a household that spends \$100 on food and \$5 on renting a bicycle is better off than a household that spends

\$100 on food and owns a bicycle (that it could rent out for \$5), when in fact both households are equally well off (*ceteris paribus*).

2.3.2.2 Measure the value of housing services

If you own your house (or apartment), it provides housing services, which should be considered as part of consumption. The most satisfactory way to measure the values of these services is to ask how much you would have to pay if, instead of owning your home, you had to rent it.

The standard procedure is to estimate, for those households that rent their dwellings, a function that relates the rental payment to such housing characteristics as the size of the house (in sq. ft. of floor space), the year in which it was built, the type of roof, whether there is running water, etc. This gives

$$Rent = f(\text{area, running water, year built, type of roof, location, number of bathrooms, ...})$$

This equation is then used to impute the value of rent for those households that own, rather than rent, their housing.

In the case of Vietnam there is a problem with this approach: almost nobody rents housing! And of those that do, most pay a nominal rent for a government apartment. Only 13 of the 5,999 households surveyed in VLSS98 paid private-sector rental rates. On the other hand the VLSS surveys did ask each household to put a (capital) value on their house (or apartment). In computing consumption expenditure, the rental value of housing was assumed to be 3 percent of the capital value of the housing. This is a somewhat arbitrary procedure, but the 3 percent is almost certainly too low.

2.3.2.3 How to account for spending on weddings.

Families spend money on weddings. Such spending is often excluded when measuring household consumption expenditure. The logic is that the money spent on weddings mainly gives utility to the guests, not the spender. Of course if one were to be strictly correct, then expenditure should include the value of the food and drink that one enjoys as a guest at other people's weddings, although in practice this is rarely (if ever) included.

2.3.2.4 Accounting for household composition differences

Households differ in size and composition, and so a simple comparison of aggregate household consumption can be quite misleading about the well-being of individuals in a given household. Most researchers recognize this problem and use some form of normalization. The most straightforward method is to convert from household consumption to individual consumption by dividing household expenditures by the number of people in the household. Then, total household expenditure per capita is the measure of welfare assigned to each member of the household. Although this is by far the commonest procedure, it is not very satisfactory, for two reasons:

- First, different individuals have different needs. A young child typically needs less food than an adult, and a manual laborer requires more food than an office workers.

- Second, there are economies of scale in consumption (at least of non-food items). It costs less to house a couple than to house two single individuals.

For example, suppose we have a household with 2 members and monthly expenditure of \$150 total. We would then assign each individual \$75 as their monthly per capita expenditure. If we have another household with 3 members, it would appear that each member is worse off with only \$50 per capita per month. However, suppose we know that the 2-person household contains two adult males age 35 whereas the second household contains 1 adult female and 2 young children. This added information may change our interpretation of the level of well-being in the second household, since we suppose that young children may have much lower costs (at least this is true for food) than adults.

The solution to this problem is to apply a system of weights. For a household of any given size and demographic composition (such as one male adult, one female adult, and two children), an equivalence scale measures the number of adult males (typically) which that household is deemed to be equivalent to. So each member of the household counts as some fraction of an adult male. Effectively, household size is the sum of these fractions and is not measured in numbers of persons but in numbers of adult equivalents. Economies of scale can be allowed for by transforming the number of adult equivalents into “effective” adult equivalents.

In the abstract, the notion of equivalence scale is compelling. It is much less persuasive in practice, because of the problem of picking an appropriate scale. How these weights should be calculated and whether it makes sense to even try is still subject to debate. There is no consensus in the matter. However, equivalence scales are not necessarily unimportant. For example, take the simple argument that typically, in most household surveys, per capita consumption decreases with household size. This is generally taken as evidence that there are economies of scale to expenditure and not necessarily proof that large households are worse off or have a lower standard of living.

There are two possible solutions to this problem: either pick a scale that seems reasonable on the grounds that even a bad equivalence scale is better than none at all, or try to estimate a scale typically based on observed consumption behavior from household surveys. Often the equivalence scales are based on the different calorie needs of individuals of different ages.

OECD scale

Commonly used is the “OECD scale,” which may be written as

$$(2.2) \quad AE = 1 + 0.7(N_{adults} - 1) + 0.5N_{children}$$

where AE refers to “adult equivalent.” A one-adult household would have an adult equivalent of 1, a two-adult household would have an AE of 1.7, and a three-adult household would have an AE of 2.4. Thus the 0.7 reflects economies of scale; the smaller this parameter, the more important economies of scale are considered to be. The 0.5 is the weight given to children, and presumably reflects the lower needs (for food, housing space, etc.) of children. Osberg and Xu (1999) use this scale in their study of poverty in Canada.

Other scales.

Many other scales have been used. For instance, a number of researchers used the following scale in analyzing the results of the living standards measurement surveys that were undertaken in Ghana, Peru and the Côte D'Ivoire:

Age (years)	0-6	7-12	13-17	>17
Weight	0.2	0.3	0.5	1.0

Estimate an equivalence scale.

It is also possible estimate an equivalence scale, by essentially looking at how aggregate household consumption of various goods during some survey period tends to vary with household size and composition. A common method is to construct a demand model in which the budget share devoted to food consumption of each household is regressed on the total consumption per person. Deaton (1997) gives an example using Engel's method with India and Pakistan household expenditure survey data. Specifically, household food share is regressed on per capita expenditure, household size, household composition variables such as ratio of adults and ratios of children at different ages. The equivalence scales or the ratio of costs of a couple with a child to a couple without children can be calculated with the estimated coefficients and displayed in table 2.3:

Table 2.3: Equivalence scales using Engel's method

Age	Maharashtra, India	Pakistan
0-4	1.24	1.28
5-9	1.28	1.36
10-14	1.30	1.38
15-54	1.34	1.42

Note: Reproduced from Deaton (1997) table 4.6.

The numbers show the estimated costs of a family of two adults plus one additional person of various ages calculated relative to the costs of a childless couple. So, for example, a child between 0 and 4 years is equivalent to 0.24 of a couple, or 0.48 of an adult. As the age of the additional member gets older, the extra costs associated with the child rise. We can compare these estimates with the last row, which shows the equivalence scale when an additional adult is added to the household. The third adult is less than half the costs of one adult. That is, an additional adult costs 34% more for the couple, or 68% of one member of the couple. So, by these calculations, these households experience economies of scale to additional adults, plus younger members are not equivalent to adults in terms of costs.

There are a number of problems with this method that will not be detailed here (see Ravallion, 1994 and Deaton 1997). However, consider the example from Ravallion (1994) where there are two hypothetical households as described in table 2.4.

Table 2.4: Consumption within two hypothetical households

	Male adult	Female	First child	Second child	Per person	Per equivalent
--	------------	--------	-------------	--------------	------------	----------------

		adult				male adult
Household A	40	20	10	10	20	40
Household B	30	-	-	-	30	30

Source: Ravallion (1994)

In this example, four persons live in household A while just one in household B. The government can make a transfer to the household that is deemed to be the poorest, but it cannot observe the distribution of consumption within the households. All the government knows is the aggregate expenditure and the household composition. In this case, which of the two households should have priority for assistance? In terms of household consumption per capita, household A looks worse off with a lower average per capita consumption. But using equivalence scales as calculated here, household B would receive assistance priority. This example demonstrates two key points. First, while observable consumption behavior is important data, assumptions about unobservables (e.g. how the aggregate is split between the household) will be required. Second, assumptions in computing consumption for individuals using household data can have considerable bearing on policy choices.

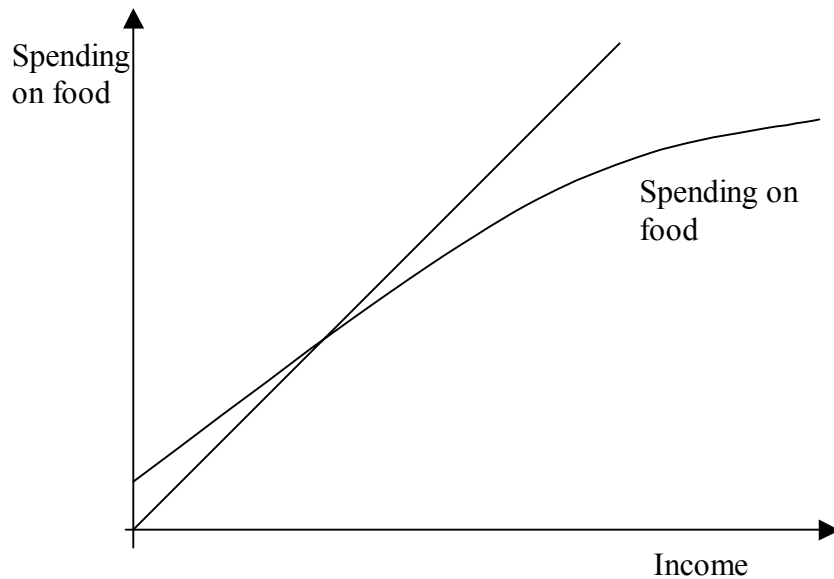
2.3.3 Candidate 3. Other measures of household welfare

Even if they were measured perfectly, neither income nor expenditure would be a perfect measure of household well-being. For instance, neither measure puts a value on the leisure time enjoyed by the household; neither measures the value of publicly-provided goods (such as education, or public health services); and neither values intangibles such as peace and security.

There are other possible measures of well-being. Among the more compelling are:

- *Calories consumed per person per day.* If one accepts the notion that adequate nutrition is a prerequisite for a decent level of well-being, then we could just look at the quantity of calories consumed per person. Anyone consuming less than a reasonable minimum - often set at 2,100 Calories per person per day - would be considered poor. This is an attractive idea, and we will return to it in chapter 3. At this point we just note that it is not always easy to measure calorie intake, particularly if one wants to distinguish between different members of a given household. Not is it easy to establish the appropriate minimum amount of calories per person, as this will depend on the age, gender, and working activities of the individual.

Figure 2.2. Engel curve: food spending rises less quickly than income



- *Food consumption as a fraction of total expenditure.* Over a century ago Ernst Engel observed, in Germany, that as household income (per capita) rises, spending on food rises too, but less quickly; this relationship is shown in figure 2.2. As a result, the proportion of expenditure devoted to food falls as per capita income rises. One could use this finding, which is quite robust and is found everywhere, to come up with a measure of well-being and hence poverty. For instance, households that devote more than (say) 60% of their expenditures to food might be considered as poor. The main problem with this measure is that the share of spending going to food also depends on the proportion of young to old family members (more children, a higher proportion of spending on food), and on the relative price of food (if food is relatively expensive, the proportion of spending going to food will tend to be higher).
- *Measures of outcomes rather than inputs.* Food is an input, but nutritional status (are you underweight?) is an output. So one could measure poverty by looking at malnutrition. Of course, this requires establishing a baseline anthropometric standard against which to judge whether someone is malnourished. Anthropometric indicators have the advantage that they can reveal living conditions within the household (rather than assigning the overall household consumption measure across all members of the household without really knowing how consumption expenditure is divided among household members). However, there is one further point about these measures: by some accounts, the use of child anthropometric measures to indicate nutritional need is questionable when broader concepts of well-being are invoked. For example, it has been found that seemingly satisfactory physical growth rates in children are sometimes maintained at low food-energy intake levels by not playing. That is clearly a serious food-related deprivation for any child.
- *Anthropological method.* Close observation at the household level over an extended period can provide useful supplementary information on living standards in small samples. However, this is unlikely to be a feasible method for national poverty measurement and comparisons. Lanjouw and Stern (1991) used subjective assessments of poverty in a north Indian village, based on classifying households into seven groups (very poor, poor, modest, secure, prosperous, rich and very rich) on the basis of observations and discussion with villagers over that year.

An issue of concern about this method is clearly its objectivity. The investigator may be working on the basis of an overly stylized characterization of poverty. For example, the poor in village India are widely assumed to be landless and underemployed. From the poverty profiles given by Lanjouw and Stern (1991) we find that being a landless agricultural laborer in their surveyed village is virtually a sufficient condition for being deemed poor. By their anthropological method, 99% of such households are deemed poor, though this is only so for 54% when their measurement of permanent income is used. It is clear that the perception of poverty is much more strongly linked to landlessness than income data suggest. But it is far from clear which type of data is telling us the most about the reality of poverty.

When one is looking at a community (e.g. province, region) rather than individual households, it might make sense to judge the poverty of the community by life expectancy, or the infant mortality rate, although these are not always measured very accurately. School enrollments (a measure of investing in the future generation) represent another outcome that might indicate the relative well-being of the population. Certainly, all of these other measures of well-being are not replacements for consumption per capita – and nor does consumption per capita replace these measures. Rather, together, we can get a more complete and multidimensional view of the well-being of a population. Consider the statistics in table 2.5 for 11 different countries. How countries are ranked in terms of living standards clearly depends on which measure or indicator is considered.

Table 2.5: Poverty and quality of life indicators

Countries	GNP per capita (1999 dollars)	% population below poverty line	Female life expectancy at birth, years (1998)	Prevalence of child malnutrition, % children <5 years (1992-1998)	Female adult illiteracy rate, % of people 15+years, (1998)
Algeria	1,550	22.6 (1995)	72	13	46
Bangladesh	370	35.6 (1995/96)	59	56	71
Cambodia	260	36.1 (1997)	55	na	80
Colombia	2,250	17.7 (1992)	73	8	9
Indonesia	580	20.3 (1998)	67	34	20
Jordan	1,500	11.7 (1997)	73	5	17
Morocco	1,200	19.0 (1998/99)	69	10	66
Nigeria	310	34.1 (1992/93)	55	39	48
Peru	2,390	49.0 (1997)	71	8	16
Sri Lanka	820	35.3 (1990/91)	76	38	12
Tunisia	2,100	14.1 (1990)	74	9	42

Source: World Bank (2000)

In sum, there is no perfect measure of well-being. The implication is simple: all measures of poverty are imperfect. That is not an argument for avoiding measuring poverty, but rather for approaching all measures of poverty with a degree of caution, and for asking in some detail about how the measures were constructed.

CHAPTER 3

Poverty Lines

3.1 How to define a poverty line

Let us assume that we have chosen a measure of household well-being - say consumption expenditure. The next step is to choose a poverty line. Households whose consumption expenditure falls below this line are considered to be poor.

Following a common practice, the poor are defined as those who lack command over basic consumption needs, including food and non-food components. The poverty line, thus, is obtained by specifying a consumption bundle considered adequate for basic consumption needs and then by estimating the costs of these basic needs. In other words, the poverty line is conceptualized as a minimum standard required by an individual to fulfill his or her basic food and non-food needs.

Once we have computed a household's consumption, we need to evaluate if that amount places the household "in poverty", or defines the household as "poor". The threshold that we use for this is the poverty line. The poverty line defines the level of consumption (or income) needed for a household to escape poverty. It might appear that the notion of a poverty line implies a distinct "turning point" in the welfare function. That is, by rising from just below to just above the poverty line, households (and individuals therein) move from considerable misery to an adequate minimum amount of well-being. Given that well-being follows a continuum, and given how arbitrary the choice of poverty line is, the notion of such a "turning point" is not very compelling. In fact, it may make sense to define more than one poverty line. For example, one poverty line may mark households that are "poor" and another lower level could indicate those that are "extremely poor." Another approach is to construct a "food poverty line," which is based on some notion of minimum amount of money a household needs to purchase some basic-needs food bundle. If the cost of basic non-food needs is estimated, then the food poverty line added to the non-food needs will equal the overall poverty line.

More formally, the poverty line for a household, z_i , may be defined as the minimum spending/consumption (or income, or other measure) needed to achieve at least the minimum utility level u_z , so:

$$(3.1) \quad z_i = e(p, x, u_z)$$

In practice we cannot measure u_z , or even $e(\cdot)$, and so a more pragmatic approach is needed. There are two approaches. One is to compute a poverty line for each household, adjusting it from household to household to take into account differences in the prices they face and their demographic composition. For example, a small household in a rural area may face low housing costs and relatively inexpensive food prices. Thus, their z_i may be low compared to a large household living in a city where housing is more expensive and food prices are perhaps higher. This gives different poverty lines for each household.

A second approach is to construct one *per capita* poverty line for all individuals, but to adjust *per capita* y_i for differences in prices and household composition. The adjusted *per capita* y_i is then compared with the one poverty line to determine if the individual is living below the poverty line.

The approach taken for Cambodia in 1999 is somewhere between these two extremes. Separate poverty lines were constructed for each of three major "regions", based on the prices prevailing in those areas; whether a household in any given region is poor is then determined by comparing its expenditure per capita with the appropriate regional poverty line. These poverty lines are shown in table 3.1, based on

Gibson's (1999) poverty profile of Cambodia using the CSES 1999 data and Prescott and Pradhan (1997) using the SESC 1993/94 data. We will discuss the construction of these poverty lines in detail below.

Table 3.1: Summary of Cambodia poverty lines

	1993/94 SESC		1999 CSES	
	Food Poverty Line	Poverty Line	Food Poverty Line	Poverty Line
Phnom Penh	1185	1578	1737	2470
Other Urban	995	1264	1583	2093
Rural	881	1117	1379	1777

Note: All values are in Riels per person per day

Source: Prescott and Pradhan (1997); Gibson (1999)

As shown in table 3.1, poverty lines Phnom Penh, the capital of Cambodia, are higher than other areas. This is consistent with other countries. For example, in Vietnam, Duong and Trinh (1999) note that the World Bank concluded that households would need to spend at least 1,071,000 dong (about US\$95) per person per year to be out of poverty. However, for urban areas, the amount was estimated to be 1,342,000 dong; in rural areas it was 1,054,000 dong. This reflects that fact that costs are higher in cities.

Over time, we expect poverty lines to change for a population. This is due to two factors. First, poverty lines reflect the costs of purchasing food and non-food items. As prices change (typically inflation), nominal poverty lines increase. This is what underlies the rising nominal poverty lines in Cambodia, shown in Table 3.1.

The second factor could be a change in the way the poverty line is constructed. So, for Thailand in table 3.2, for example, we see that the poverty line is increasing over time. But without further information we cannot tell whether the rises in the poverty line merely reflect changes in prices, or also represent a revision in the real poverty threshold. This raises the question of whether we should look at relative, or absolute, poverty lines. We now consider each in turn.

Table 3.2: Average poverty line of Thailand

Year	Poverty Line
1988	473
1990	522
1992	600
1994	636
1996	737
1998	878
1999	886

Note: All values are in Baht per person per month.

Source: Thailand Socio-Economic Survey conducted by the National Statistical Office

3.1.1 Relative poverty

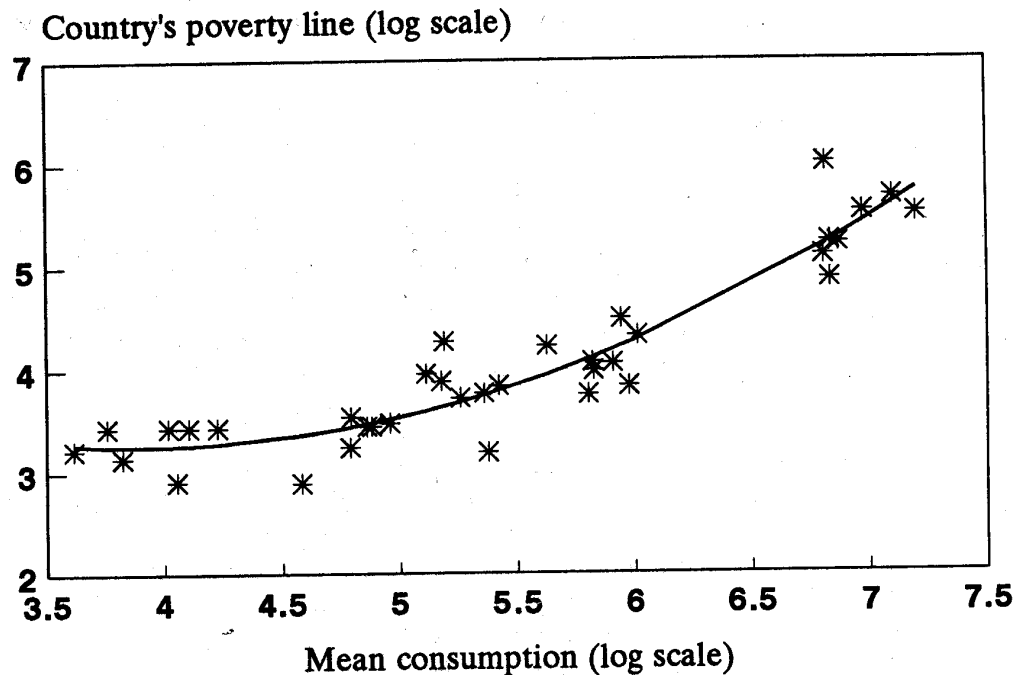
Sometimes we are interested in focusing on the poorest segment (e.g. a fifth, or two-fifths) of the population; these are the relatively poor. When defined in this way, it is a truism

that "the poor are always with us." It is often helpful to have a measure such as this in order to target programs that are geared to helping the poor.

In practice, rich countries have higher poverty lines than do poor countries, as shown clearly in figure 3.1 (from Ravallion 1998, p.26, based on an earlier paper by Ravallion, Datt and van de Walle). This explains why, for instance, the official poverty rate in the early 1990s was close to 15% in the United States and also close to 15% in (much poorer) Indonesia. Many of those counted as poor in the U.S. would be considered to be comfortably well off by Indonesian standards.

As countries become better off, they have a tendency to revise the poverty line upwards – with the notable exception of the United States, where the line has (in principle) remained unchanged for almost four decades. For instance, the European Union typically defines the poor as those whose per capita incomes fall below 50% of the median. As median incomes rise, so does the poverty line.

Figure 5: Poverty lines across countries



Note: Each point is one country. Both poverty line and mean consumption are measured at purchasing power parity.
Source: Ravallion, Datt and van de Walle (1991)

Figure 3.1. Poverty lines across countries.

Based on a sample of 36 countries, Ravallion, Datt and van de Walle estimated the following relationship:

$$(3.2) \quad \ln z_i = 6.704 - 1.773 \ln(C / cap) + 0.228[\ln(C / cap)]^2 + v_i$$

t=5.1 t=-3.6 t=5.1

where $R^2=0.89$. They found that at the mean value of per capita consumption (which they measured in purchasing power parity terms), the elasticity of the poverty line with respect to consumption per capita was 0.71. This means that if per capita consumption were to rise 10%, then the poverty line would rise 7.1% on average. But the non-linear relationship implies that the elasticity of the poverty line with respect to consumption per capita was close to 0 in low-income countries, and was almost 1 in high-income countries.

To the extent that it is to identify and target today's poor, then a relative poverty line is appropriate, and needs to be tailored to the overall level of development of the country. For instance, a \$1 per day poverty line might be useful in Vietnam, where 27% of the population would be considered poor by this standard in 1998 (Haughton 2000), but would be of little relevance in the United States where almost nobody would be poor by this standard.

3.1.2 Absolute poverty

An absolute poverty line is “fixed in terms of the standard of living it commands over the domain of poverty comparisons.” The United States poverty line does not change from year to year, so that the poverty rate today may be compared with the poverty rate of a decade ago, knowing that the definition of what constitutes poverty has not changed.

An absolute poverty line is essential if one is trying to judge the effect of anti-poverty policies over time, or to estimate the impact of a project (e.g. microcredit) on poverty. Legitimate comparisons of poverty rates between one country and another can only be made if the same absolute poverty line is used in both countries. The World Bank needs absolute poverty lines in order to be able to compare poverty rates across countries – useful in determining where to channel resources, and also in assessing progress in the war on poverty. It commonly uses two measures: a) an estimated 1,200 million people worldwide live on less than US \$1 per day, and b) over 2 billion people worldwide live on less than \$2 per day. These are absolute poverty lines.

Box: The \$1 per day standard

Cross-country comparisons of poverty rates are notoriously difficult, but the World Bank has tried to get around this problem by computing the proportion of the population in different countries living on less than \$1 per capita per day (in purchasing power parity terms, and in 1985 US dollars). The numbers shown below suggest that the poverty rate in Vietnam (computed by Haughton 2000) compares favorably with that of India, but lags behind (more affluent) China and Indonesia. One possible lesson that may be drawn from these numbers is that the easy gains in poverty reduction in Vietnam are probably over, and the country will have difficulty reducing its poverty rate substantially in the decade ahead, even if economic growth continues at its current relatively robust rate of between 6% and 7% annually.

	% in poverty			% in poverty	
Vietnam	45	1993	Indonesia	8	1996
Vietnam	27	1998	Nigeria	31	1992-93
China	22	1995	Philippines	27	1994
India	47	4994			

Sources: World Bank. 1999c. *Entering the 21st Century: World Development Report 1999/2000*. Washington DC. Haughton 2000.

3.2 Issues in choosing an absolute poverty line

3.2.1 Decide the standard of living

There is an important conceptual problem that comes up when working with absolute poverty lines, which arises from the issue of what is meant by “the standard of living” (Ravallion, 1998, on which much of this discussion is based).

In practice, almost all absolute poverty lines are set in terms of the cost of buying a basket of goods (the “commodity-based poverty line,” which we denote by z). If we assume that

$$(3.3) \quad u = f(y),$$

which says that utility or “standard of living” (u) depends on income or expenditure (y), then

$$(3.4) \quad y = f^{-1}(u).$$

This says that for any given level of utility, there is some income (or expenditure) level that is needed to achieve it. If u_z is the utility that just suffices to avoid being poor, then

$$(3.5) \quad z = f^{-1}(u_z).$$

In other words, given a poverty line that is absolute in the space of welfare (i.e. gives u_z) there is a corresponding absolute commodity-based poverty line.

But suppose we make another equally plausible assumption, which is that utilities are interdependent. My well-being may depend not just on what I consume, but on how my consumption stacks up against that of the rest of society. A household of four with an income of \$12,000 per year would not be considered poor in Indonesia, but when this household compares its position with average incomes in the U.S., it may feel very poor. We may capture this idea by assuming

$$(3.6) \quad u = g\left(y, \frac{y}{\bar{y}}\right),$$

where \bar{y} is the mean income in the society. In this case

$$(3.7) \quad u_z = g\left(z, \frac{z}{\bar{y}}\right)$$

and so, making the standard assumption of invertibility,

$$(3.8) \quad z = g^{-1}(\bar{y}, u_z).$$

This means that for a poverty line to be absolute in the space of welfare (i.e. to yield u_z), the commodity-based poverty line (i.e. z) may have to rise as \bar{y} rises. The commodity-based poverty line would then look more like a relative poverty line!

3.2.2 Decide u_z and $g(\cdot)$

Even if we assume that the commodity-based poverty line remains constant, we are still left with two problems.

- a) *The Referencing problem.* What is the appropriate value of u_z – i.e. the utility of the poverty line? The choice is of course arbitrary, but “a degree of consensus about the choice of the reference utility level in a specific society may well be crucial to mobilizing resources for fighting poverty” (Ravallion, 1998, p.6).
- b) *The Identification problem.* Given u_z , what is the correct value of z – i.e. of the commodity value of the poverty line. This problem arises both because the size and demographic composition of households vary – an issue we return to in the discussion of equivalence scales below – but also because “the view that we can measure welfare by looking solely at demand behavior is untenable” (Ravallion, 1998, p.7).

The implication is that external information and judgments will be required to answer the referencing and identification problems, and hence to determine the absolute poverty line in practice. But how is this to be done in practice?

Table 3.3 presents absolute and relative poverty rates for different regions in the world. How regions compare with each other depends on which poverty measure is used. For example, by the absolute measure of less than US \$1 a day, Sub-Saharan Africa has the highest portion of the population living in poverty. On the other hand, countries in Latin America and the Caribbean have the highest portion of their population living below one-third the average national consumption.

Table 3.3: Absolute and Relative Poverty rates

	Share of population living on less than \$1 per day (in 1998)	Share of the population living on less than one-third of average national consumption for 1993 (in 1998)
East Asia and Pacific	15.3	19.6
East Asia and Pacific excluding China	11.3	24.6
Europe and Central Asia	5.1	25.6
Latin America and the Caribbean	15.6	51.4
Middle East and North Africa	1.9	10.8
South Asia	40.0	40.2
Sub-Saharan Africa	46.3	50.5
Total	24.0	32.1
Total excluding China	26.2	37.0

Source: World Bank (2000)

In passing we might note that an absolute poverty line is best thought of as one which is fixed in terms of living standards, and *fixed over the entire domain of the poverty comparison* (Ravallion). Thus absolute poverty comparisons will deem two persons at the same standard of living to both be either “poor” or “not poor” irrespective of the time or place being considered, or with or without some policy change, *within the relevant domain*. The poverty comparison is then “consistent” in the specific sense that individuals who are identical in all relevant respects are treated identically. However, depending on the purpose of the comparison, the relevant domain may vary. For example, a global comparison of absolute consumption poverty may entail using a poverty line (e.g. \$1 consumption per capita per day) that is low by the standards of rich countries. If, however, one is trying to form a poverty profile for one country only, the choice of an absolute poverty line should be appropriate to that country (e.g. a poverty

line of \$1 per day might be appropriate in Vietnam, and \$10 per day might be suitable in the United States). Judgments of what constitutes a reasonable absolute poverty line must first specify the domain of comparisons, and recognize that the answer may change if the domain changes.

3.3 Solution A: objective poverty lines.

How then are we to determine poverty lines. One possibility is to pick an “objective” poverty line. The key idea here is that the poverty line should be set at a level that enables individuals to achieve certain capabilities, including a healthy and active life and full participation in society. In practice this almost certainly would imply that the commodity-based poverty line would rise as a country becomes more affluent, because the minimum resources needed to participate fully in society probably rise over time. In Sen’s rather dense prose (Sen 1983, p.168), “an absolute approach in the space of capabilities translates into a relative approach in the space of commodities.”

A common, and fairly satisfactory, way of approaching capabilities is to begin with nutritional requirements. Two common ways of making this operational are the Cost-of-Basic-Needs (CBN) approach, and the Food Energy Intake (FEI) method.

3.3.1 Food Energy Intake method

The goal here is to find the level of consumption expenditure (or income) that allows the household to obtain enough food to meet its energy requirements. Note that consumption will include non-food as well as food items; even underfed households typically consume some clothing and shelter, which means that at the margin these “basic needs” must be as valuable as additional food.

The basic idea is captured in figure 3.2, which shows a *calorie income function*; as income (or expenditure) rises, food energy intake also rises, although typically more slowly. Given some level of just-adequate food energy intake k , one may use this curve to determine the poverty-line level of expenditure, z . Formally, the function shows

$$(3.9) \quad k = f(y)$$

$$\text{So, given monotonicity, (3.10) } \quad y = f^{-1}(k),$$

or, given a minimum adequate level of calorie k_{min} , we have

$$(3.11) \quad z = f^{-1}(k_{min})$$

where z is the poverty line. Notice that this approach is parsimonious in one respect: it does not require any information about the prices of goods consumed.

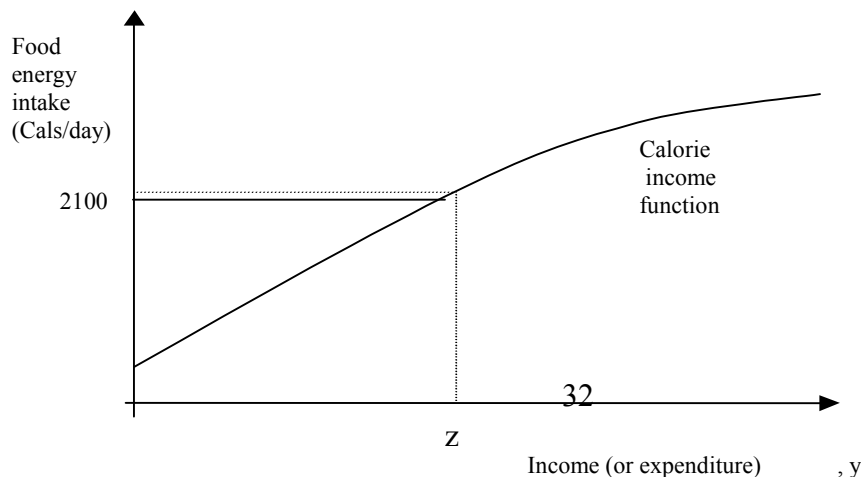


Figure 3.2: Calorie income function

First one needs to determine the amount of food that is adequate. Vietnam pegs this level at 2,100 Calories per person per day, in line with FAO recommendations, but it is recognized that individuals may need more or less food than this – clearly the needs of young children, growing teenagers, manual workers, pregnant women, or sedentary office workers may differ quite markedly; physical stature also plays a role. Not all countries have set the same cut-off point, as table 3.3 shows:

Table 3.3: Per Capita Daily Calorie Intake Used in Poverty Line Construction

Vietnam	2,100
Indonesia	2,100
Philippines	2,000
Thailand	1,978
China	2,150

A variant of this approach was used to measure poverty in Vietnam, using data from the Vietnam Living Standards Survey of 1993. Separate food expenditure lines were estimated for urban and rural areas in each of seven provinces; the cost of obtaining 2,100 Calories of food per person per day was then computed, and the associated poverty lines – one for each rural and urban area in each province. This gave a headcount index of 55% (Dollar et al. 1995).

The Food Energy Intake method is however seriously flawed, and should not be used unless the alternatives are infeasible. The weaknesses of the method were pointed out in an important article by Ravallion and Bidani (1994); in the next few paragraphs we summarize their approach and findings. The method also failed in a recent analysis of data from Vietnam, for slightly different reasons, also summarized below.

3.3.1.1 The urban-rural problem

The problem begins when one recognizes that food energy – typically shown on the Calorie income function – depends on other factors as well as income. The other influences include the tastes of the household (e.g. urban tastes in food may differ from rural tastes); the level of activity of household members; the relative prices of different foods, and of food to non-food items; and the presence of publicly-provided goods.

Figure 3 shows hypothetical (but plausible) calorie income functions for urban and rural households. Rural households can obtain food more cheaply, and tend to be engaged in physically more demanding activities, so their calorie income function will typically be higher than that of urban households. The implication is that for a given level of food energy intake, the poverty line in the rural area will be lower than in the urban area, as figure 3 makes clear. To the extent that this reflects differences in the cost of living, it is not a problem to have two poverty lines of this kind.

The key finding of Ravallion and Bidani (1994), based on 1990 data from the SUSENAS household survey in Indonesia, was that the urban poverty line (20,614 rupiah per person per month) was *much* higher than the rural one (13,295 Rp./person/month). This gap far exceeded the difference in the cost of living between urban and rural areas. Using these poverty lines, Ravallion and Bidani (1994) found that that poverty in Indonesia appeared to be higher in the urban than in the rural areas (Table 3.4), a completely implausible result. The point is also illustrated in figure 3.4, which shows the cumulative

distribution of consumption per capita, for rural and urban areas, and marks the poverty lines and headcount poverty rates.

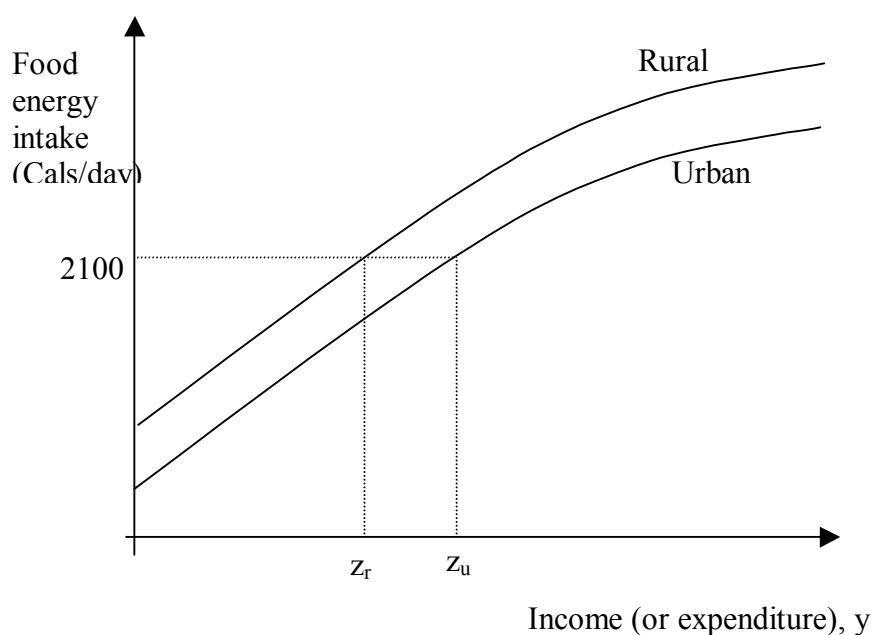


Figure 3.3. Calorie income functions for urban and rural Indonesia

Table 3.4: Poverty Lines in Indonesia using Food Energy Intake Method, 1990

	Indonesia overall	Urban areas	Rural areas
P_0	15.1	16.8	14.3
P_1	2.42	3.23	1.06
P_2	0.66	0.94	0.53

Source: Ravallion and Bidani 1994.

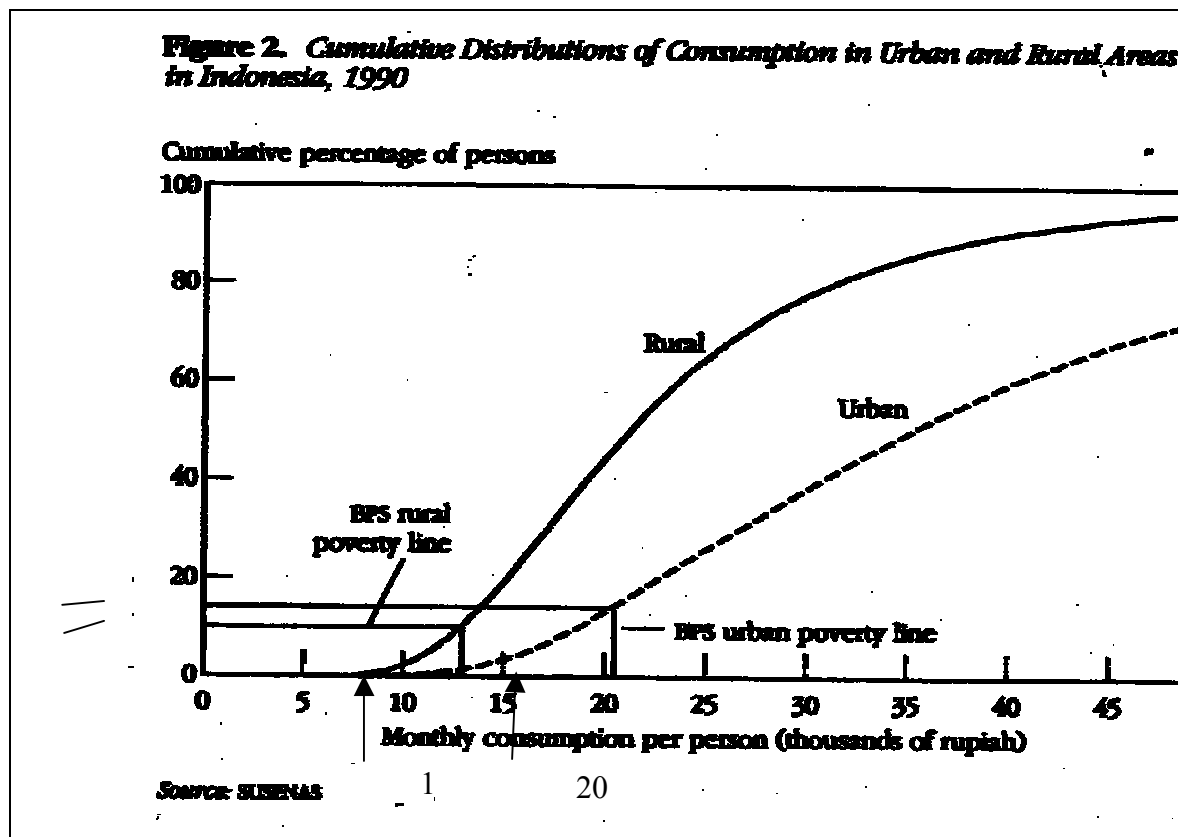


Figure 3.4 Cumulative distribution functions for consumption, Indonesia, 1990
 (Source: Ravallion and Bidani 1994.)

3.3.1.2 The relative price problem

When researchers tried to apply the Food Energy Intake approach to data from the Vietnam Living Standards Survey of 1998, the method failed. As with the 1993 data, the idea was to compute food expenditure functions, find the cost of 2,100 Calories of food, and then find the related level of expenditure per capita, which would then serve as a poverty line. After undertaking this exercise, researchers found a higher level of poverty in 1998 than in 1993, an implausible result in an economy whose real GDP grew by 9% annually between 1993 and 1998, and where there was a widespread sense that the benefits of this growth had spread widely.

What went wrong? Figure 5 shows the situation. The food expenditure function shifted down between 1993 and 1998; in other words, for a given (real) income, households in 1998 would buy less food than in 1993. The main reason was that the price of food rose by 70% between 1993 and 1998, while the price of non-food items rose by just 25%; in other words, food became *relatively* much more expensive. As a result consumers shifted away from food to non-food consumption. This meant that the poverty line rose from z_{93} to z_{98} (see figure 5), a jump that turned out to be implausibly large.

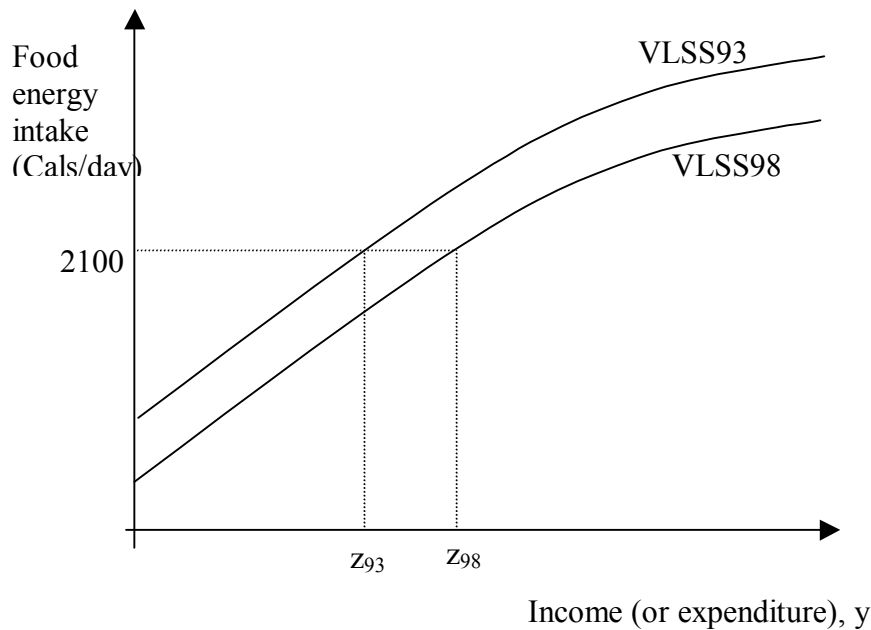


Figure 3.5 The determination of poverty lines for Vietnam, 1993 and 1998

3.3.2 The Cost-of-Basic-Needs method:

A more satisfactory approach to building up a poverty line, while remaining in the spirit of trying to ensure that the line covers basic needs, proceeds as follows:

- Stipulate a consumption bundle that is deemed to be adequate, with both food and non-food components; and
- Estimate the cost of the bundle for each subgroup (urban/rural, each region, etc.).

This is essentially the approach taken by Seebom Rowntree in his seminal study of poverty in York, undertaken in 1936. Note that although we begin with a consumption bundle – so much food, so much housing space, so much electricity, etc. – the poverty line is measured in money. We are therefore not insisting that each basic need be met, only that it *could* be met. Operationally, the steps to follow are these:

- Pick a nutritional requirement for good health, such as 2,100 Calories per person per day.
- Estimate the cost of meeting this food energy requirement, using a diet that reflects the habits of households near the poverty line (e.g. those in the lowest, or second-lowest, quintile of the income distribution; or those consuming between 2,000 and 2,200 Calories). This may not be easy if diets vary widely across the country. Call this food component z^F .
- Add a non-food component (z^{NF}). There is a lot of disagreement about how to do this; we offer some more thoughts on this issue below.
- Then the basic needs poverty line is given by

$$(3.12) \quad z^{BN} = z^F + z^{NF}$$

Box. The US poverty line

In 1963 and 1964, Mollie Orshansky of the Social Security Administration computed the cost of an ‘adequate’ amount of food intake, to get z^F . She then multiplied this number by 3 to get z^{BN} . Why? Because at the time, the average food share for all consumers in the United States was 1/3. This line is still used, updated regularly for price changes.

Source: Dalaker and Naifeh (1998).

To illustrate how this might work, suppose that we use a food energy threshold of 2,100 Calories per day. Let there be three foodstuffs: rice, corn and eggs. Table 5 shows the expenditure on each item, and the amount consumed by a household in the second (from bottom) quintile; since such a household consumes, we suppose, just 2,000 Calories per day, the figures here have to be grossed up to give the cost of purchasing 2,100 Calories. In this example the cost comes to 105 pesos per day.

Table 3.5: Illustration of Construction of Cost of Food Component of Poverty Line

	Calories	Expenditure per day (pesos)	Calories, Adjusted to give 2,100 Calories	Expenditure, adjusted to cover 2,100 Calories
Rice	1,400	60	1,470	63
Corn	400	20	420	21
Eggs	200	20	210	21
Total	2,000	100	2,100	105

The choice of which diet to use when estimating the cost of obtaining 2,100 Calories is not a trivial one, a point emphasized in the context of Indonesia by Pradhan et al. (2000)¹. To illustrate, consider the information in Table 6, drawn from the Vietnam Living Standards Survey of 1992-93. Households in the poorest quintile paid 0.68 dong per Calorie; those in the richest expenditure quintile paid almost twice as much (1.38 dong/Calorie). Depending on which one uses, the poverty line could vary widely.

Table 3.6: Food consumption by expenditure quintile, Vietnam, 1992-93.

Quintile	Expenditure per capita, ‘000 dong/year	% of expenditure devoted to food	Calories per capita per day	Dong per Calorie
Lowest	562	70	1,591	0.68
Low-mid	821	65	1,855	0.79
Middle	1,075	60	2,020	0.87
Mid-upper	1,467	54	2,160	1.00
Upper	2,939	47	2,751	1.38

Source: Vietnam Living Standards Survey 1992-93

In practice, researchers in this case used the price of food for households in the middle quintile, on the grounds that those households were close to the poverty line because they were consuming almost

¹ Pradhan et al. (2000) favor an iterative procedure: pick a reference population that is relatively poor and compute their cost of Calories; now recompute the poverty line; take as the new reference population those households close to this poverty line and re-calculate the cost of Calories; compute the poverty line again; and so on, until the poverty line stabilizes.

2,100 Calories per year. The food expenditure of the middle quintile, grossed up to pay for 2,100 Calories, came to 750,228 dong per capita in 1993; the non-food expenditure of this same group of households was taken to be adequate for those at the poverty line (after a similar grossing up). This gave an overall poverty line of 1,160,842. Individual households lived in regions with different prices, so their expenditure per capita was first deflated, and then compared to this poverty line. The result was an estimated headcount poverty rate of 58% (World Bank 1999).

To compare poverty over time, this poverty line was updated to 1998. The cost of each item in the poverty-line diet of 1993 was recomputed using the prices of 1998 (as taken from the price questionnaire component of the VLSS, mainly); non-food expenditure was inflated using data from the General Statistical Office's price index. This yielded a poverty line of 1,793,903, and an associated poverty rate of 37%. The details are summarized in Table 7.

Table 3.7: Poverty Lines and Headcount Measures of Poverty, Vietnam

	Poverty line (m dong/capita/yr)	Headcount poverty rate
Poverty overall		
1993	1,160 (\$92)	58
1998	1,790 (\$128)	37
Food poverty		
1993	750	25
1998	1,287	15

Note: The food poverty rate excludes any provision for non-food items; it sets the poverty line at z^F .

Sources: Vietnam Living Standards Surveys of 1992-93 and 1997-98.

There is no wholly satisfactory way to measure the non-food component of the poverty line, and the procedures followed tend to be somewhat ad hoc. We saw above that for Vietnam, researchers essentially used the (slightly adjusted) level of non-food spending by households that were in the middle expenditure quintile in 1993. The poverty lines developed for South Korea (KIHASA 2000) measure the cost of food plus the cost of housing that meets the official minimum apartment size plus the cost of non-food items as measured by average spending by households in the poorest two-fifths of the income distribution.

Is there a better way to proceed? Probably not. Even the theory calls for compromise. Consider the food expenditure function shown in figure 3.6. Following Ravallion (1998), let b^F be the cost of buying 2,100 Calories. Then an upper poverty line might be given by

$$(3.13) \quad f^{-1}(b^F) = z^F$$

which measures the income level at which the household would buy 2,100 Calories of food; this is essentially the poverty line used in Vietnam. The non-food component is given by A .

A lower poverty line might be given by

$$(3.14) \quad z_L^F = b^F$$

which measures the income level at which the household could just buy enough food, but would not have any money left over to buy anything else; in Vietnam this is referred to as the food poverty line. But even in this case, households will typically buy non-food items, as shown by C in figure 6. Ravallion suggests one might want to compromise, and measure non-food at the mid-point between these two extremes, giving B . In each case, the poverty line would be given by

$$(3.15) \quad z = z^F + A \text{ (or } +B \text{ or } +C)$$

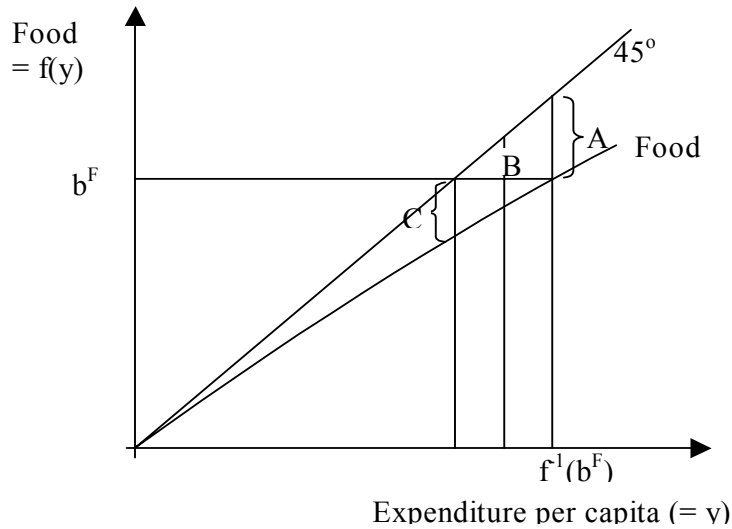


Figure 3.6 **Food Expenditure Function.**

When Ravallion and Bidani apply the Cost of Basic Needs approach to Indonesia, using the SUSENAS data of 1990, they find the poverty rates shown in Table 8. These are radically different from the ones obtained using the Food Energy Intake method. They computed poverty rates using these two measures, for each of the main regions of Indonesia, and found almost no correlation between the two measures.

Table 3.8: Headcount Measures of Poverty in Indonesia, 1990

	Cost of Basic Needs Method		Food Energy Intake Method
	Food	Food + non-food	
Indonesia overall	7.9	19.6	15.1
Urban	2.8	10.7	16.8
Rural	10.2	23.6	14.3

Source: Ravallion and Bidani 1994.

This is a serious indictment of the Food Energy Intake method. But it should also be clear that every measure of poverty can be faulted, because it rests in part on arbitrary assumptions. In measuring poverty, there is no single truth.

3.4 Solution B: Subjective poverty lines

We could measure poverty by asking people to define a poverty line, and using this to measure the extent of poverty. For instance, in a survey one might ask

“What income level do you personally consider to be absolutely minimal? That is to say that with less you could not make ends meet.”

The answers will vary from person to person (and by size of household), but they could be plotted, and a line fitted through them, to get a subjective poverty line such as z^* in figure 3.7. It may also

be possible to get adequate results by asking “do you consider your current consumption to be adequate to make ends meet?”

Mahar Mangahas has amassed extensive information on subjective poverty in the Philippines as part of the social weather stations project. Collected biannually since 1985, and quarterly since 1992, the surveys poll about 1,200 households. Each household is shown a card with a line running across it; below the line is marked “poor” and above the line “non-poor”, and each household is asked to mark on the card where it fits. Separately, households are also asked to define a poverty line.



Figure 3.7 **Estimating a subjective poverty line.**

Gaurav Datt of the World Bank has analyzed the Filipino data in some detail. Here are some of his more interesting findings:

- Self-rated poverty lines are high. In 1997 the median poverty line was about 10,000 pesos per month for a “typical” household; this compares with the government’s “basic needs” poverty line, which at that time stood at 4,495 pesos/month. The implication is that self-rated poverty rates are high – 60% of all households in 1997, compared to 25% using the basic needs line.
- The self-rated poverty line has risen rapidly over time, by about 60-70% between 1985 and 1997. One consequence is that there is no trend in self-rated poverty over time. Another implication is that even when there is an economic slowdown, as occurred in 1997-98, the self-rated poverty rate hardly changes: it rose from 59% in 1996-97 to 61% in 1998.
- Perhaps surprisingly, the self-rated poverty line given by poor households is only slightly lower than that for non-poor households, and in fact the difference is not statistically significant. One might have expected poor households to have a less generous measure of the poverty line.
- There is a clear urban/rural difference in perceptions of the poverty line, with urban households setting a (money) poverty line at about twice the level of rural households, giving:

$$(3.16) \quad z_{self - rated}^u \approx 2 z_{self - rated}^r$$

The cost of living is certainly higher in urban areas, but by a factor of 1.2-1.5 rather than by a factor of 2. Thus the urban self-rated poverty line is, in real terms, higher than its rural counterpart. Why?

- One possibility is that there is more inequality in the urban areas, and that this raises expectations.
- Another plausible explanation is that households in urban areas may have more exposure to the media, and may have been affected more thoroughly by consumerism.
- A third explanation is that urban households may be more attuned to political processes, and their estimates of the poverty line may include an element of strategic behavior – trying to influence policy makers.

Self-rated measures of poverty are rarely collected. If the Filipino experience is at all representative, it is clear that they cannot usefully supplant the more traditional “objective” measures of poverty.

CHAPTER 4

Measures of Poverty

What does Aggregate Poverty Measure? Given information on (say) per capita consumption, and a poverty line, then the only remaining problem is deciding on an appropriate summary measure of aggregate poverty. There are a number of aggregate measures of poverty that can be computed.

4.1 Headcount index

By far the most widely-used measure is the *headcount index*, which simply measures the proportion of the population that is counted as poor, often denoted by P_0 . Formally,

$$(4.1) \quad P_0 = \frac{1}{N} \sum_{i=1}^N I(y_i \leq z) = \frac{N_p}{N},$$

where N is the total population and $I(\cdot)$ is an indicator function that takes on a value of 1 if the bracketed expression is true, and 0 otherwise. So if expenditure (y_i) is less than the poverty line (z), then $I(\cdot)$ equals to 1 and the household would be counted as poor. N_p is the total number of the poor.

The great virtue of the headcount index is that it is simple to construct and easy to understand. These are important qualities. However the measure has at least three weaknesses:

The headcount index does not take the intensity of poverty into account. Consider the following two income distributions:

Headcount Poverty Rates in A and B, assuming poverty line of 125					
	Expenditure for each individual in country				Headcount poverty rate (P_0)
Expenditure in country A	100	100	150	150	50%
Expenditure in country B	124	124	150	150	50%

Clearly there is greater poverty in country A, but the headcount index does not capture this. As a welfare function, the headcount index violates the transfer principle – an idea first formulated by Dalton (1920) that states that transfers from a richer to a poorer person should improve the measure of welfare. Here if a somewhat poor household were to give to a very poor household, the headcount index would be unchanged, even though it is reasonable to suppose that poverty overall has lessened. The headcount index implies that there is a “jump” in welfare, at about the poverty line, so it is meaningful to speak of the poor and the non-poor. In practice, such a jump is not found (Ravallion 1996, p.1330).

- The head-count index does not indicate how poor the poor are, and, hence, does not change if people below the poverty line become poorer. Moreover, the easiest way to reduce the headcount index is to target benefits to people just below the poverty line, because they are the ones who are cheapest to move across the line. But by most normative standards, people just below the poverty line are the least deserving of the poor. Thus, despite its popularity, many problems result from an undue concentration on the head-count statistic.

- It is also important to note that the poverty estimates should be calculated for individuals and not households. What we calculate using the head-count index is the percentage of *individuals* who are poor and not the percentage of households. To be able to do so, we make a critical assumption that all household members enjoy the same level of well-being. This assumption may not hold in many situations. For example, some elderly members of a household may be much poorer than other members of the same household. In reality, not all consumption is evenly shared across household members.

4.2 Poverty gap index

A moderately popular measure of poverty is the *poverty gap index*, which adds up the extent to which individuals fall below the poverty line (if they do), and expresses it as a percentage of the poverty line. More specifically, define the poverty gap (G_n) as the poverty line (z) less actual income (y_i) for poor individuals; the gap is considered to be zero for everyone else. Using the index function, we have

$$(4.2) \quad G_n = (z - y_i)I(y_i \leq z).$$

Then the poverty gap index (P_1) may be written as

$$(4.3) \quad P_1 = \frac{1}{N} \sum_{i=1}^N \frac{G_n}{z}.$$

This measure is the mean proportionate poverty gap in the population (where the non-poor have zero poverty gap). Some people think of this measure as the cost of eliminating poverty (relative to the poverty line), because it shows how much would have to be transferred to the poor to bring their incomes (or expenditure) up to the poverty line. The minimum cost of eliminating poverty using targeted transfers is simply the sum of all the poverty gaps in a population; every gap is filled up to the poverty line. However this interpretation is only reasonable if the transfers could be made perfectly efficiently, for instance with lump sum transfers, which is implausible. Clearly this assumes that the policymaker has a lot of information; one should not be surprised to find that a very “pro-poor” government would need to spend far more than this in the name of poverty reduction.

At the other extreme, one can consider the maximum cost of eliminating poverty, assuming that the policymaker knows nothing about who is poor and who is not. From the form of the index, it can be seen that the ratio of the minimum cost of eliminating poverty with perfect targeting (i.e. G_n) to the maximum cost with no targeting (i.e. z , which would involve providing everyone with enough to ensure they are not below the poverty line) is simply the poverty gap index. Thus this measure is an indicator of the potential saving to the poverty alleviation budget from targeting.

The poverty gap measure has the virtue that it does not imply that there is a discontinuity (“jump”) at the poverty line. Yet a serious shortcoming of this measure is that it may not convincingly capture differences in the severity of poverty amongst the poor. For example, consider two distributions of consumption for four person; the A distribution is (1,2,3,4) and the B is (2,2,2,4). For a poverty line $z=3$ (so that the headcount index is 0.75 in both cases), A and B have the same value for the poverty gap index (i.e. 0.25). However, the poorest person in A has only half the consumption of the poorest in B. One can think of B as being generated from A by a transfer from the least poor of the poor persons (the individual with ‘3’ in A) to the poorest. The poverty gap will be unaffected by such a transfer. Its main drawback is that it ignores inequality *among* the poor. To see this, consider the following example:

Poverty Gap Poverty Rates in A and B, assuming poverty line of 125					
	Expenditure for each individual in country				Poverty gap rate (P_I)
Expenditure in country A	100	100	150	150	0.10
Expenditure in country B	80	120	150	150	0.10

For both of these countries, the poverty gap rate is 0.10, but most people would argue that country B has more serious poverty because it has an extremely poor member.

To summarize, the *Poverty Gap Index* is the average over all people, of the gaps between poor people's standard of living and the poverty line, expressed as a ratio to the poverty line. The aggregate poverty gap shows the cost of eliminating poverty by making perfectly targeted transfers to the poor (i.e., closing all poverty gaps), in the absence of transactions costs and disincentive effects. This is clearly unrealistic but it does convey useful information about the minimum scale of the financial resources needed to tackle the poverty problem. Moreover, the poverty gap index can show the value of using survey information to learn about the characteristics of the poor. A costly way of eliminating poverty would be to make completely untargeted poverty line-sized transfers to everyone in the population. The poverty gap index gives the ratio of the cost of eliminating poverty using perfectly targeted transfers compared with using completely untargeted transfers. Thus, the smaller is the poverty gap index, the greater the potential economies for a poverty alleviation budget from identifying the characteristics of the poor so as to target benefits and programs.

4.3 Squared poverty gap index

To solve the problem of inequality among the poor, some researchers use the squared poverty gap index. This is simply a weighted sum of poverty gaps (as a proportion of the poverty line), where the weights are the proportionate poverty gaps themselves; a poverty gap of (say) 10% of the poverty line is given a weight of 10% while one of 50% is given a weight of 50%; this is in contrast with the poverty gap index, where they are weighted equally. Hence, by squaring the poverty gap index, the measure implicitly puts more weight on observations that fall well below the poverty line. Formally:

$$(4.4) \quad P_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{G_n}{z} \right)^2.$$

The measure lacks intuitive appeal, because it is not easy to interpret and so it is not used very widely. It may be thought of as one of a family of measures proposed by Foster, Greer and Thorbecke (1984), which may be written as

$$(4.5) \quad P_\alpha = \frac{1}{N} \sum_{i=1}^N \left(\frac{G_n}{z} \right)^\alpha, \quad (\alpha \geq 0)$$

where α is a measure of the sensitivity of the index to poverty and the poverty line is z , the value of expenditure per capita for the j -th person's household is x_j , and the poverty gap for individual j is $G_j = z - x_j$ (with $G_j = 0$ when $x_j > z$). When parameter $\alpha = 0$, P_0 is simply the head-count index. When $\alpha = 1$, the index is the poverty gap index P_I , and when α is set equal to 2, P_2 is the poverty severity index. For all $\alpha > 0$, the measure is strictly decreasing in the living standard of the poor (the lower your standard of living, the poorer you are deemed to be). Furthermore, for $\alpha > 1$ it also has the property that the increase in measured poverty due to a fall in one's standard of living will be deemed greater the poorer one is. The

measure is then said to be "strictly convex" in incomes (and "weakly convex" for $\alpha \neq 1$). Another convenient feature of the FGT class of poverty measures is that they can be disaggregated for population sub-groups and the contribution of each sub-group to national poverty can be calculated.

The work by Foster, Greer and Thorbecke provides an elegant unifying framework for measures of poverty. However it begs the question of what the best value of α . Some of these measures also lack emotional appeal.

The measures of depth and severity of poverty provide complementary information on the incidence of poverty. It might be the case that some groups have a high poverty incidence but low poverty gap (when numerous members are just below the poverty line), while other groups have a low poverty incidence but a high poverty gap for those who are poor (when relatively few members are below the poverty line but with extremely low levels of consumption). Table 4.1 provides an example from Madagascar. According to the headcount, unskilled workers show the third highest poverty rate, while the group is in the fifth rank according to the poverty severity. Comparing them with the herders shows that they have a higher risk of being in poverty, but that the poverty tends to be less severe or deep. The types of interventions needed to help the two groups are therefore likely to be different.

Table 4.1: Poverty Indices By sub-groups, Madagascar, 1994

	Head count: %	Rank	Poverty gap: %	Rank	Poverty severity: %	Rank
Small farmers	81.6	1	41.0	1	24.6	1
Large farmers	77.0	2	34.6	2	19.0	2
Unskilled workers	62.7	3	25.5	4	14.0	5
Herders/fishermen	51.4	4	27.9	3	16.1	3
Retirees/handicapped	50.6	5	23.6	5	14.1	4

Source: Coudouel, Hentschel and Wodon (2001)

4.4 Sen Index.

Sen (1976) proposed an index that sought to combine the effects of the number of poor, the depth of their poverty, and the distribution of poverty within the group. The index is given by

$$(4.6) \quad P_s = P_0(1 - (1 - G^P) \frac{\mu^P}{z}),$$

where P_0 is the headcount index, μ^P is the mean income (or expenditure) of the poor, and G^P is the Gini coefficient of inequality among the poor. The Gini coefficient ranges from 0 (perfect equality) to 1 (perfect inequality), and is discussed further below in the context of measuring inequality. The Sen Index can also be written as the average of the headcount and poverty gap measures weighted by the Gini coefficient of the poor, giving:

$$(4.7) \quad P_s = P_0 G^P + P_1(1 - G^P)$$

The Sen index has been widely discussed, and has the virtue of taking the income distribution among the poor into account. However the index lacks intuitive appeal, and cannot be decomposed satisfactorily into its constituent components, which explains why it is rarely used in practice.

4.5 The Sen-Shorrocks-Thon index.

The Sen index has been modified by others, and perhaps the most compelling version is the Sen-Shorrocks-Thon (SST) index, defined as

$$(4.8) \quad P_{SST} = P_0 P_1^P (1 + \hat{G}^P),$$

which is the product of the headcount index, the poverty gap index (applied to the poor only), and a term with the Gini coefficient of the poverty gap ratios (i.e. of the G_n 's). This Gini coefficient typically is close to 1, indicating great inequality in the incidence of poverty gaps.

Example

In 1996, 12.4% of the population of Quebec province (Canada) was in poverty. The poverty gap index, *applied to the poor only*, stood at 0.272. And the Gini coefficient of the poverty gap ratios was 0.924. Thus the Sen-Shorrocks-Thon index was 0.065 ($=0.124 \times 0.272 \times (1+0.924)$).

Osberg and Xu (1999) use the SST index to compare poverty across the 10 Canadian provinces for 1984, 1989, 1994, 1995 and 1995, as well as to put the degree of Canadian provincial poverty into an international context. A number of graphs from their study are reproduced below. Figure 1 provides an international comparison, using the SST index, and shows that the US is an outlier with its relatively high poverty rate (as measured by the SST). A comparison of the US and Canada over time (figure 2) shows that while poverty was similar in the two countries a generation ago, it is now higher in the US. Figure 3 provides information on some Canadian provinces: Newfoundland was the poorest in 1984, but by 1996 had become much less of an outlier.

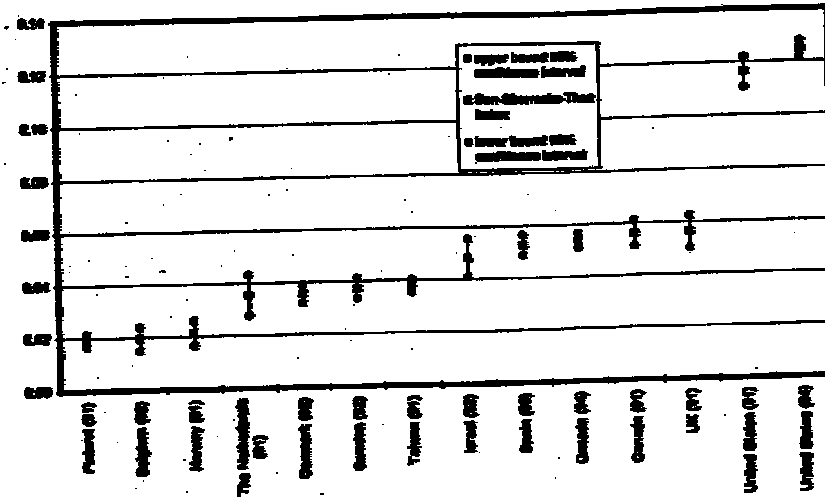
One strength of the SST index is that it can help give a good sense of the sources of change in poverty over time. This is because the index may be decomposed into

$$(4.9) \quad \Delta \ln P_{SST} = \Delta \ln P_0 + \Delta \ln P_1^P + \Delta \ln(1 + \hat{G}^P),$$

which may be interpreted as, % change in SST index = % change in headcount index + % change in poverty gap index (among poor) + % change in (1+Gini coefficient of poverty gaps).

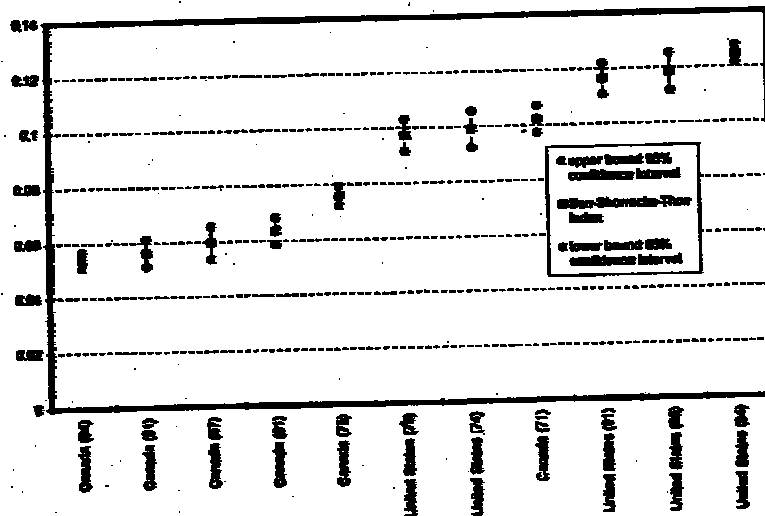
In plain English, this allows us to decompose poverty into three aspects: are there more poor? are the poor poorer? and is there higher inequality among the poor?

FIGURE 1
1990s – Country Rankings by SST Poverty Index – Poverty Line Based on Half the Median Equivalent Income



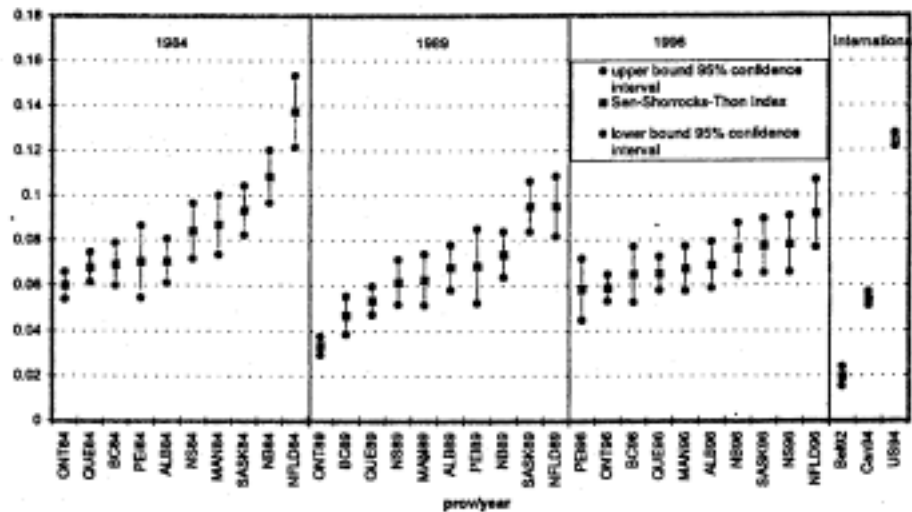
Note: [95% confidence interval = mean \pm 2 standard deviations] of 300 bootstraps.
 Source: Luxembourg Income Study, Osberg and Xu (1997).

FIGURE 2
Canada-United States Comparison – The SST Index from 1971 to 1994 – Poverty Line Based on Half the Median Equivalent Income



Note: [95% confidence interval = mean \pm 2 standard deviations] of 300 bootstraps.
 Source: Luxembourg Income Study, Osberg and Xu (1997).

FIGURE 3
Sen-Shorrocks-Thon Index of Poverty Intensity – 1/2 Median After-Tax Income, 1984, 1989, 1996



Note: [95% confidence interval = mean +/-2 standard deviations] of 300 bootstraps.
Source: Authors' calculations, Survey of Consumer Finances, various years, Luxembourg Income Study, Osberg and Xu

Example:

The information in table 4.2 comes from Osberg and Xu, and traces the evolution of poverty in the Canadian province of Newfoundland between 1984 and 1996. It is clear that most of the change in the poverty rate over time was due to variations in the number of people in poverty (P_t), rather than in the size of the poverty gap per poor person (P_t^P) or the distribution of poverty among the poor (G^P).

Table 4.2: Decomposition of poverty, and changes in poverty, in Newfoundland

	SST index	P_0	P_1^P	$1+G^P$	$\Delta \ln \text{SST index}$	$\Delta \ln P_0$	$\Delta \ln P_1^P$	$\Delta \ln(1+G^P)$
1984	.137	.245	.304	1.844				
1989	.095	.169	.296	1.897	-.370*	-.372*	-.027	.028
1994	.105	.184	.304	1.884	.104	.086	.026	-.007
1995	.125	.212	.316	1.864	.168	.141	.038	-.010
1996	.092	.164	.294	1.897	-.307	-.254	-.071	.018

Notes: * denotes statistically significant at the 95% level. Poverty line is half of median equivalent income, using the “OECD scale” (i.e. equivalent income = $1 + 0.7(N_{\text{adults}} - 1) + 0.5(N_{\text{children}})$).

Source: Osberg and Xu, 1999.

Note that the values of the Sen-Shorrocks-Thon index provided by Osberg and Xu do not give just a single point estimate for each province; they also provide a confidence interval. Because the SST index is complex, it is not possible to compute these confidence intervals analytically. Instead, they are

computed artificially using *bootstrapping*. The basic idea behind the bootstrap is straightforward and clever. Suppose we have a survey sample of 2,000 households. Now pick a sample of 2,000 from this sample *with replacement* – i.e. pick a household, then put it back into the sample, pick another household, put it back into the sample, and so on, until you have picked 2,000 households. Compute the SST index using this artificial sample. Then repeat the process many times; Osberg and Xu use 300 repetitions. The result is a distribution of values of the SST, from which it is easy to find (say) the 95% confidence interval. A sample code to generate confidence intervals for the SST index is given in the exercise for Chapter Five.

4.6 Time taken to exit

Previous poverty profiles for Cambodia, and indeed for most countries, rely on the three basic classes of Foster Greer Thorbecke poverty statistics discussed above. But when thinking about poverty reduction strategies, it may be useful to show how long it would take, at different potential economic growth rates, for the average poor person to exit poverty. A poverty statistic with this property is derived by Morduch (1998); the statistic is decomposable by population sub-groups and is also distributionally sensitive. For the j th person below the poverty line, the expected time to exit poverty (i.e., to reach the poverty line), if consumption per capita grows at positive rate g per year is:

$$(4.10) \quad t_g^j \approx \frac{\ln(z) - \ln(x_j)}{g}.$$

What effect can economic growth have on the elimination of poverty? Figure 4 shows the average time it would take to raise the consumption level of a poor person in Cambodia to the poverty line, for various hypothetical growth rates. It is assumed that this growth rate is continuous, is in real terms, and is distributionally neutral *among* the poor. If the economic growth rate enjoyed by the poor is only one percent per year, it would take over 20 years for the average poor person to exit poverty. But at a growth rate of four percent per year it would take less than six years for the average poor person to exit poverty. Hence, economic growth that acts to raise the real consumption levels of the poor can have a powerful effect on the elimination of poverty.

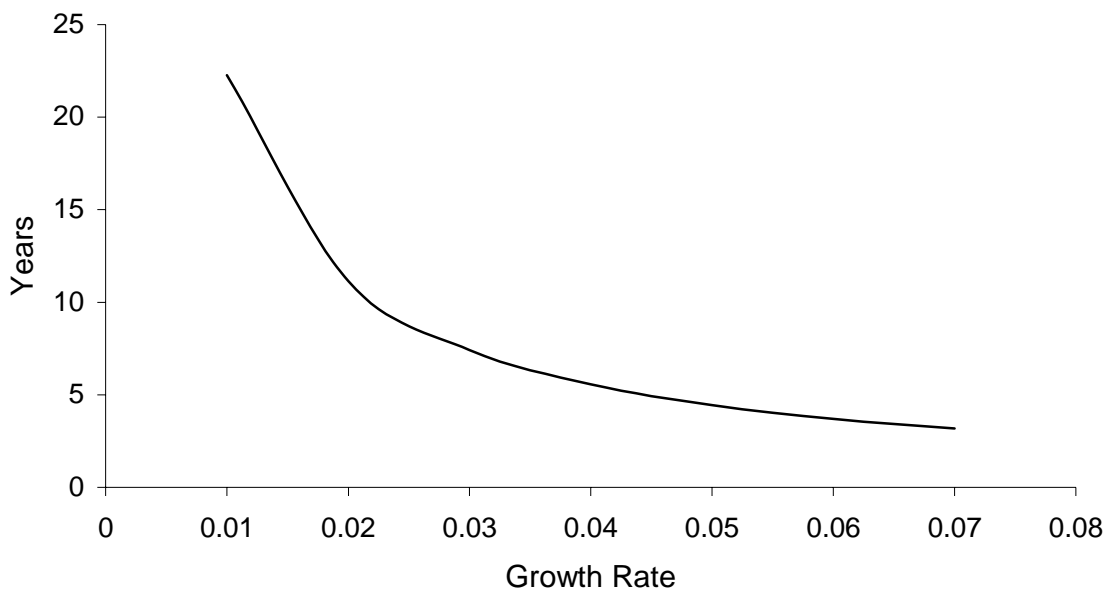


Figure 4.4: Average exit time from poverty

Despite the potency of economic growth, it will take more than just growth to rapidly improve the lives of the very poor. The expected time to exit poverty for those people who are so poor that they are below the food poverty line in Cambodia – i.e. they cannot even afford enough food, even if they were to devote all their consumption spending to food – is more than 15 years, even at a three percent continuous annual growth rate. Thus, targeted programs are needed to deliver benefits to the poor, in the form of improvements in their human and physical assets and through interventions (e.g., infrastructure, markets) that improve the returns they get from those assets.

4.7 Other Measures

There are other additive poverty measures that are distribution sensitive. For example, the first distribution sensitive poverty measure was proposed by Watts (1968), and takes the form:

$$(4.11) \quad W = \frac{1}{N} \sum_{i=1}^N \log\left(\frac{z}{y_i}\right)$$

Following Atkinson (1987), one can characterize a general class of additive measures, encompassing W , the FGT (Foster, Greer and Thorbecke) class of measures, and some other measures (such as the second measure proposed by Clark, Hemming and Ulph, 1981), as taking the following form:

$$(4.12) \quad P = \frac{1}{N} \sum_{i=1}^N p(z, y_i)$$

where $p(z, y_i)$ is the individual poverty measure, taking the value zero for the non-poor ($y_i > z$) and some positive number for the poor, the value of which is a function of both the poverty line and the individual living standard, non-decreasing in the former and non-increasing in the latter. The class of measures defined by (4.12) has strong appeal in constructing poverty profiles, as discussed in Chapter Seven.

Chapter 5

Poverty Indices: Comparison of Measures

5.1 Introduction

Given the intellectual energy that has gone into the theory of poverty measurement over the last 15 years, and the virtual plethora of poverty measures from which one can now choose, it is of interest to ask: Does it really matter in poverty comparisons which of these measures one uses? The answer depends on whether, and how, relative inequalities in society have changed across the situations being compared. If all consumption levels (poor and non-poor) have changed by the same proportion - a "distributionally neutral" growth or contraction - then all of these poverty measures will yield the same ranking in the poverty comparison, and the ranking in terms of absolute poverty will depend solely on the direction of change in the mean of the distribution.

However, the differences between these measures can become quite pronounced. Consider, for example, two policies: Policy A entails a small redistribution from people around the mode of the distribution, which is also where the poverty line happens to be located, to the poorest households. (This is actually a fair characterization of how a reduction in the prices of domestically produced food-staples would affect the distribution of welfare in some Asian countries). Policy B entails the opposite change - the poorest lose while those at the mode gain. (An increase in food-staple prices in the above example). A moment's reflection will confirm that the head-count index (H) will prefer policy B; since changes in H depend solely on which direction people are crossing the poverty line (under policy A, people who were just on the poverty line would now fall below it since they transfer resources to the very poor, thus the total number of poor increases). However, a measure such as the Squared Poverty Gap Index will indicate the opposite ranking, since it will respond relatively more to the gains among the poorest than among the not-so-poor.

The need to examine higher order poverty measures, such as the Poverty Gap Index and the Squared Poverty Gap Index, also depends on whether or not the poverty comparison in terms of the head-count index has considered more than one poverty line. If only one poverty line is used then it is especially important to check the higher-order measures. If, for a given poverty line, the higher order measure gives a different result to the head-count index, then this will also hold for a head-count index of "ultra-poverty" based on a sufficiently low poverty line.

5.2 Measurement errors

Poverty measures can be quite sensitive to certain sorts of measurement error in the underlying parameters, and quite robust to others. To give a stark example, suppose that the observed welfare indicator contains an additive random error with zero mean; one gets the indicator's value right on average, but there is an error in any one observation. Suppose also (for the purpose of this example) that the poverty line is at the mode of the distribution. Then it can be shown that the expected value of the observed head-count index will be locally un-affected by changes in the degree of imprecision in the welfare measurements; on average one will predict the same index with a noisy indicator as a precise one, and the mean will be unbiased. However, this need not hold for other poverty lines or for higher order poverty measures; Ravallion (1988) gives a general statement of the necessary and sufficient conditions

for greater variability in the welfare indicator to increase the expected value of a poverty measure defined on that indicator. Under plausible assumptions, greater imprecision in the welfare indicator will yield higher estimates of any smooth distribution-sensitive poverty measure, such as P_2 .

Now consider, instead, errors in the mean of the distribution on which the poverty measure is being estimated; suppose, for instance, that consumption is systematically underestimated. It may be demonstrated that the elasticity of the head-count index to errors in the mean, holding the Lorenz curve constant, is simply the elasticity of the cumulative distribution function evaluated at the poverty line. This may be estimated, and values around two are quite common, at least for developing countries. Thus a five percent under-estimation of mean consumption at all consumption levels could easily translate into a 10 percent over-estimation in the head-count index and, hence, the number of poor. Furthermore, amongst the FGT class of measures, the elasticities tend to be higher, the higher the value of alpha.

Note that this assumes that there is no corresponding error in the Lorenz curve – i.e. where errors occur, they are not systematically higher for one income group than for another. However, it is generally thought to be more likely that survey measurement errors will tend to lead to an under-estimation of the extent of inequality, because the rich are proportionally more likely to under-declare their income (or expenditure). Indeed, it may well be that surveys overestimate the consumption of the poor (through social stigma facing respondents or because the sample frame excludes sub-groups of the poor, such as the homeless, who are amongst the poorest), and underestimate that of the rich (such as through their fear of revealing black market transactions or tax evasion). But, it is less clear how this will translate into poverty measurement errors. For example, if the under-estimation of income or consumption were entirely above the poverty line then it would make no difference to measured poverty. More than this, there is little that one can say in general about the effects on the head-count index of measurement error in the Lorenz curve when the estimated consumption of the is are affected. However, for the poverty gap index and the squared poverty gap index, survey underestimation of consumption by the poor will lead to overestimation of poverty. Furthermore, the higher the value of alpha, the more sensitive the measure will be to consumption measurement errors for the poorest.

Inflation raises the nominal value both of expenditure and of the poverty line, and so has no net effect on the measures of poverty.

5.3 Equivalence scales

Poverty studies usually measure living standards using expenditure (or income) per capita. As we have discussed in Chapter Two, because needs vary among household members, and because there are economies of scale in consumption, poverty measures based on per capita welfare indicators may not be good estimates. An alternative is to base our poverty measures on expenditure (or income) *per adult equivalent*. If poverty estimates are not affected by the adult equivalence weights that we choose, it is safe to say that those poverty estimates are not biased because of weighting procedure.

Kathleen Short et al. have recently investigated the sensitivity of the US 1997 poverty rate to various variations on the OECD equivalence scales (*Experimental Poverty Measures*, June 1999). The results are summarized in the table below. While different definitions of adult equivalent do change the measured poverty rate, the most striking feature of this table is how small these differences are.

Table 5.1: Sensitivity of Adult Equivalence Scale

Measuring the sensitivity of the US poverty rate (P_0) to different definitions of adult equivalence.

	Headcount poverty rate, %
$N_a + N_c$ (i.e. use income per capita)	13.3
$(N_a + 0.7 N_c)^{0.65}$	13.1
$(N_a + 0.7 N_c)^{0.70}$	12.3
$(N_a + 0.7 N_c)^{0.75}$	12.7
$(N_a + 0.7 N_c)^{0.5}$	13.4
$(N_a + 0.7 N_c)^{0.6}$	12.7
$(N_a + 0.85 N_c)^{0.65}$	12.7
$1 + 0.4(N_a - 1) + 0.4(\text{child } 1) + 0.3(N_c - 1)$ (Canadian scale)	13.8
One adult: 1. Two adults: 1.41. Single parents: $(N_a + 0.8 + 0.5 N_c - 1)$. All other families: $(N_a + 0.5 N_c)^{0.7}$	13.1

Source: Short et al., 1999.

Notes: N_a = number of adults. N_c = number of children.

Table 5.2 displays adult equivalence weights that have been in India and in Taiwan. A study of the importance of weights computed the correlation coefficient between (unweighted) expenditure per capita and expenditure per adult equivalent using these weights, using data from Sri Lanka (1969-70), Taiwan (1974) and Peninsular Malaysia (1973). All the correlation coefficients were 0.96 or higher; they are shown in table 5.2 below.. The implication? Equivalence scales give similar results to expenditure per capita.

Table 5.2: Adult equivalents, India and Taiwan

India: adult equivalents			Taiwan: adult equivalents		
Age	Male	Female	Age	Male	Female
0	.43	.43	0-1	.3	.3
1-3	.54	.54	2-4	.4	.4
4-6	.72	.72	5-7	.5	.5
7-9	.87	.87	8-10	.7	.7
10-12	1.03	.93	11-14	.8	.8
13-15	.97	.80	15-20	.9	.9
16-19	1.02	.75	21+	1.0	.9
20-39	1.00	.71			
40-49	.95	.68			
20-59	.90	.64			
60-69	.80	.51			
70+	.70	.50			

Table 5.3: Correlation coefficients, expenditure per capita with expenditure per adult equivalent

Survey	Years	India weights	Taiwan weights
Sri Lanka	1969-70	0.99	0.96
Taiwan	1974	0.98	0.96
Peninsular Malaysia	1973	0.99	0.97

5.4 Robustness of ordinal poverty comparisons

At a number of points in the discussion so far we have seen that there is pervasive uncertainty about possibly crucial aspects of a poverty comparison. There are likely to be errors in our living

standards data, unknown differences in needs between households at similar consumption levels, uncertainty and arbitrariness about both the poverty line and precise poverty measures. Given these problems, how robust are poverty comparisons? Would they alter if we made alternative assumptions? A recent strand of research in poverty analysis has shown how we can answer such questions, drawing on and developing results from the theory of stochastic dominance. Ravallion (1998) gives an elementary exposition of the approach, again oriented to the needs of the analyst trying to make a reasonably robust poverty comparison. The analysis is easier for a single dimension of well-being, but Ravallion (1998) also gives an introduction to multi-dimensional dominance.

5.4.1 A single measure of standard of living

As mentioned before, it is always a good idea to consider one or more extra poverty lines. Let us take this a step further, and imagine the curve that is traced out as one plots the head-count index H on the vertical axis and the poverty line on the horizontal axis, allowing the latter to vary from zero to the maximum consumption. This is simply the cumulative distribution function, which can be thought of as the "poverty incidence curve" $F(z)$ – each point on the curve gives the proportion of the population consuming less than the amount given on the horizontal axis (figure 5.1). If one calculates the area under this curve up to each point and then plots it against the poverty line, one traces out the "poverty deficit curve" $D(z)$. In fact, each point on this curve is simply the value of the poverty gap (or, equivalently, the poverty gap index times the poverty line z).

If one again calculates the area under the poverty deficit curve at each point and plots it against the poverty line, then one obtains a new curve, which can be termed the "poverty severity curve" $S(z)$; each point on this curve is directly proportional to the FGT measure P_2 .

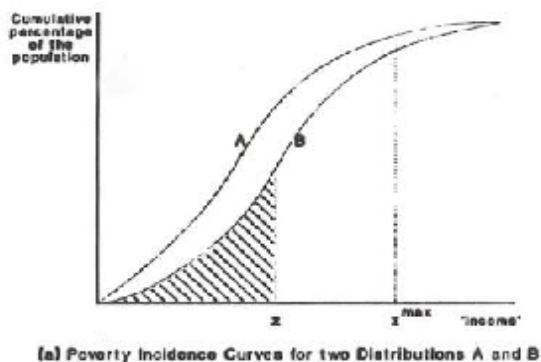


Figure 5.1: Poverty incidence curve for two distributions A and B

Although the measures discussed above generally meet the set of desirable axioms it is possible that they will rank the same set of distributions in different ways, simply because of their differing sensitivity to incomes in different parts of the distributions. When rankings are ambiguous, the alternative method of stochastic dominance can be applied. We discuss first- and second- order stochastic dominance below, before moving on to consider the one dominance category that is associated with unambiguous comparisons of inequality across distributions: mean-normalised second-order dominance, or Lorenz dominance.

5.4.1.1 First order stochastic dominance

Consider two income distributions y_1 and y_2 with cumulative distribution functions (CDFs) $F(y_1)$ and $F(y_2)$. If $F(y_1)$ lies nowhere above and at least somewhere below $F(y_2)$ then distribution y_1 displays first order stochastic dominance over distribution y_2 : $F(y_1) \leq F(y_2)$ for all y . Hence in distribution y_1 there are no more individuals with income less than a given income level than in distribution y_2 , for all levels of income. We can express this in an alternative way using the inverse function $y=F^{-1}(p)$ where p is the share of the population with income less than a given income level: first order dominance is attained if $F_1^{-1}(p) \geq F_2^{-1}(p)$ for all p . The inverse function $F^{-1}(p)$ simply plots incomes against cumulative population, usually using ranked income quintiles. First order stochastic dominance of distribution y_1 over y_2 implies that any social welfare function that is increasing in income will record higher levels of welfare in distribution y_1 than in distribution y_2 .

5.4.1.2 Second order stochastic dominance

Consider now the deficit functions (the integral of the CDF) of distributions y_1 and y_2 :

$$(5.1) \quad G(y_{i,k}) = \int_0^{y_i} F(y_i) dy$$

where $i=1,2$. If the deficit function of distribution y_1 lies nowhere above and somewhere below that of distribution y_2 , then distribution y_1 displays second order stochastic dominance over distribution y_2 : $G(y_{1,k}) \leq G(y_{2,k})$ for all y_k . The dual of the deficit curve is the Generalized Lorenz curve defined as

$$(5.2) \quad GL(p) = \int_0^{y_1} y dF(y),$$

which plots the cumulative income shares, scaled by the mean of the distribution, against cumulative population, where the height of the curve at p is given by the mean of the distribution below p . Second order dominance established by comparisons of the deficit curves for complete, uncensored distributions implies and is implied by Generalized Lorenz dominance: $GL_1(p) \geq GL_2(p)$ for all p . Second order dominance of distribution y_1 over distribution y_2 implies that any social welfare function that is increasing and concave in income will record higher levels of welfare in y_1 than in y_2 . It should now be apparent that second order stochastic dominance is therefore implied by first order stochastic dominance, although the reverse is not true.

Let us now apply these ideas to the study of poverty analysis. Suppose we do not know the poverty line z , but we can be sure that it does not exceed z^{max} . Nor do we know the poverty measure, but we can identify some desirable properties for such a measure, including the additivity property. Then it can be shown that poverty will unambiguously fall, between year 1 and year 2, if the poverty incidence curve (the cumulative distribution) for the year 2 lies nowhere above that for year 1, up to z^{max} . Poverty falls unambiguously since the overall probability reflected by the cumulative distribution is smaller if the poverty incidence curve lies nowhere above. This is called the *First-Order Dominance Condition* (FOD). Figure 5.1 illustrates first-order dominance. When we plot the cumulative frequency distributions (cumulative percentages of the population below various consumption levels) in states A and B, we find that the curve for A is everywhere above that for B. Poverty is higher in state A than in B, *no matter what the poverty line or measure*. If the curves intersect (and they may intersect more than once), then the ranking is ambiguous. Some poverty lines and some poverty measures will rank the distributions

differently to others. One can restrict the range of poverty lines, or one can impose more structure on the poverty measure.

If one excludes the head-count index, and restricts attention to additive measures that do reflect the depth of poverty such as the poverty gap index (P_1) and the squared poverty gap index (P_2) (i.e., measures that are strictly decreasing and at least weakly convex in incomes of the poor) then we can use a *Second-Order Dominance Condition* (SOD). A fall in poverty then requires that the poverty deficit curve, given by the area under the cumulative distribution, is nowhere lower for year 1 at all points up to the maximum poverty line, and at least somewhere higher. What happens above the z^{max} is not relevant, and in fact the overall probability need not be smaller under the lower poverty deficit curve, and could in even be higher; this case is illustrated in figure 5.2.

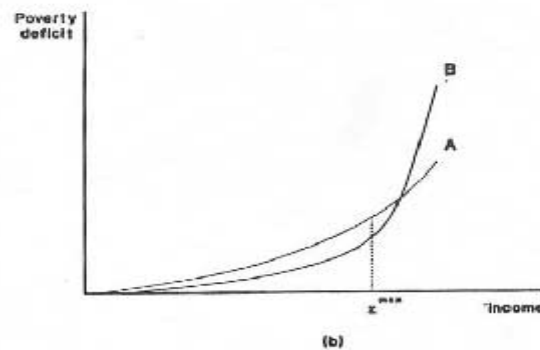


Figure 5.2: Second-Order Dominance Condition

Second-order dominance over the entire distribution is equivalent to generalized Lorenz curve dominance. The generalized Lorenz curve (GLC) is simply the ordinary Lorenz curve *scaled up by the mean*; thus it plots (on the vertical axis) the cumulative value of the welfare indicator (normalized by the population size) for the poorest p percent of the population ranked by that indicator (on the horizontal axis). If the GLC of distribution A is everywhere above that of B, then the area under A's cumulative frequency distribution must be everywhere lower than B's. Notice that the highest point on the GLC is the mean; thus a necessary condition for poverty to fall for all possible poverty lines and all measures reflecting the depth of poverty is that the mean has not fallen. By similar reasoning, another necessary condition is that the lowest level of living has not fallen (the lowest point on the GLC - just before it hits zero - is the lowest level of living). If z^{max} is the highest income then it is immaterial whether one tests SOD using the poverty deficit curve or the generalized Lorenz curve, though for lower values of z^{max} it is better to use the poverty deficit curve.

5.4.1.3 Third order stochastic dominance

When second-order dominance is inconclusive, one can further restrict the range of admissible poverty measures. If one is content to rely solely on distributionsensitive measures such as P_2 (but now excluding P_0 and P_1) then a Third-Order Dominance Condition (TOD) can be tested; an unambiguous poverty comparison for all poverty lines then requires that the poverty severity curve is everywhere higher in one of the two situations being compared. If necessary, one can go on to test higher order dominance, though the interpretation of the (increasingly) restricted class of measures becomes less clear.

To illustrate the three dominance tests, consider an initial state in which three persons have consumptions in amounts (1, 2, 3). Any final state in which one or more of these persons has a higher consumption, and none have a lower consumption, will imply a lower poverty incidence curve (strictly lower and no higher anywhere), and hence no higher poverty for any poverty line or poverty measure; examples of such final states are (2, 2, 3) or (1, 2, 4). Now suppose that the final state is in fact (2, 2, 2).

The poverty incidence curves now cross each other: some poverty lines and some poverty measures will judge this final state to be an improvement, while others will judge it to be worse than the initial state. (Compare the head-count indices for $z=1.9$ and $z=2.1$.) However, the poverty deficit curves do not cross each other; for the initial state the poverty deficit curve is (1, 3, 6) (corresponding to consumptions 1, 2, and 3), while it is (0, 3, 6) for the final state (2, 2, 2). Thus poverty will have fallen (or at least not increased) for all poverty lines and all measures that are decreasing in consumptions of the poor, such as P_1 and P_2 . But what if the final state has consumption vector (1.5, 1.5, 2)? Table 5.3 gives the poverty incidence, deficit and severity curves. Even if we confine attention to distribution-sensitive poverty measures, some poverty lines will rank the states differently to others. But note that the intersection point of the poverty severity curves is above 2; any poverty line less than this point will indicate that poverty has fallen for all distribution sensitive measures, such as P_2 .

Table 5.3: Poverty Incidence, Depth and Severity Curves

Consumption (z)	Poverty Incidence Curve $F(z)$		Poverty Depth Curve $D(z)$		Poverty Severity Curve $S(z)$	
	Initial	Final	Initial	Final	Initial	Final
1	1/3	0	1/3	0	1/3	0
1.5	1/3	2/3	2/3	2/3	1	2/3
2	2/3	1	4/3	5/3	7/3	7/3
3	1	1	7/3	8/3	14/3	15/3

Note: three people with initial consumption (1,2,3) and final consumption (1.5,1.5,2)

When two frequency distributions are quite close, we may also want to assess whether the difference between them is statistically significant. For first-order dominance, this can be done quite easily using the Kolmogorov-Smirnov test, based on the largest vertical distance between the two cumulative frequency curves; expositions on this simple test, and tabulations of critical values are readily available. Statistical inference is more difficult for higher order dominance, and more advanced methods are needed.

5.4.2 Robustness: More than One Dimension

Similar ideas can be applied in circumstances in which poverty lines vary across households or individuals in an unknown way. For example, errors in measuring the standard of living may mean that we are not identifying the true poverty lines for different individuals. Unknown differences in "needs" at given consumption levels could also mean that the true poverty lines vary. There may be considerable, unknown, inter-individual variation in nutritional requirements. Errors in accounting for differences between households in their demographic composition or the prices they face may also entail some underlying variation in the appropriate poverty lines.

Poverty comparisons are clearly more difficult when the poverty line has an unknown distribution, but even then unambiguous conclusions may be possible if one is willing to make some assumptions. Provided that the distribution of poverty lines is the same for the two (or more) situations being compared and is independent of the distribution of living standards, first-order dominance of one distribution over another implies an unambiguous poverty ranking. This holds no matter what the underlying distribution of poverty lines may be.

Another case of interest is when one knows the distribution of "needs" (such as family size) as well as consumption, but one does not know precisely how these two variables interact to determine welfare. For two dimensions of welfare, such as aggregate consumption and family size, one can derive "bi-variate dominance tests," which are more or less stringent depending on

the assumptions one is willing to make about the way in which differences in "needs" interact with consumption in determining well-being; the precise tests depend on (amongst other things) whether the marginal social valuation of consumption is higher or lower in larger families. In the special case in which the marginal valuation of consumption is independent of family size, and the marginal distribution of size is fixed, the problem collapses back to the standard dominance tests above.

Let us suppose first that we know nothing about how needs interact with consumption in determining poverty. For additive poverty measures and a fixed distribution of the population across different needs, all of the above dominance tests can be applied separately to each of the groups identified as having different needs. Thus one can test for first-order dominance amongst (say) rural households, separately for urban households, or large families separately from small families. If we find that FOD holds for each group separately then we can conclude that FOD also holds for the aggregate, no matter what the difference in needs is between the groups. If FOD fails then, by restricting attention to measures of the depth and severity of poverty, one can then test for second-order dominance for each "needs" group separately, or third-order dominance if necessary.

These will often be quite stringent tests. Weaker tests can be invoked if one is willing to rank needs groups in terms of the marginal welfare attached to an increment of consumption. Suppose one can, and let group 1 have the highest marginal social valuation of consumption (i.e., the steepest individual poverty measure, typically the poorest group). Let us also assume that this ranking is the same at all possible consumption levels, so group 1 always has the highest marginal valuation of consumption. When ranking distributions in terms of poverty measures we will also need to assume that the poverty measure, as a function of consumption, is not discontinuous at the poverty line. This precludes the head-count index, but few other measures; the condition holds for P_1 and P_2 . Under these conditions one can apply simple partial dominance tests, where the test is done cumulatively by the ranked needs of groups starting from group 1, rather than separately for each group. Thus dominance is tested on the cumulative frequency distributions for group 1 in the two situations compared, then for the population weighted sum of groups 1 and 2, then for 1,2, and 3, etc. This makes dominance more likely. For example, although poverty may increase in some needs groups, aggregate poverty may be found to have fallen as a result of some policy change.

However, these tests have to be modified further when the distribution of needs also changes, such as when the proportion of the population living in urban areas has increased over the period of the poverty comparison, as is typically the case in inter-temporal poverty comparisons for developing countries. It is theoretically possible that first-order dominance may hold separately for each of urban and rural areas, and yet not hold in the aggregate for all possible distributions of needs between the two sectors and all possible ways in which consumption and needs interact to determine well-being. More general tests can be devised for such situations, though they are difficult to explain non-mathematically and unlikely to be very convincing to policy makers.

5.5 Summary

Dominance testing can be a useful tool for making poverty comparisons. The tests can be robust to many of the measurement problems that routinely confound poverty assessments and they are easy to apply. For the first-order test (which assumes very little about how poverty is measured, but is thus the hardest test to pass) it is simply a matter of plotting the cumulative frequencies of consumption in each of the situations being compared up to the maximum admissible poverty line. The ordinal poverty comparison is unambiguous as long as the two curves do not cross over. When the test is inconclusive, a second-order test may prove useful, based on the area under the plotted cumulative distribution. However this restricts attention to measures that reflect the depth of poverty. Higher order tests can then be applied, if needed. There are also tests that can handle multiple dimensions of well-being that cannot be precisely aggregated. On-going research on this topic will probably further enhance the advantages of this approach in situations where data are imperfect and measurement is inherently controversial.

CHAPTER 6

Inequality Measures

6.1 Definition of inequality

The poverty measures we have discussed depend on the average level of income or consumption in a country, and the distribution of income or consumption. Based on these two elements, poverty measures then focus on the situation of those individuals or households at the bottom of the distribution. Inequality is a broader concept than poverty in that it is defined over the entire population, not only for the population below a certain poverty line. Most inequality measures do not depend on the mean of the distribution, and it is this property of mean independence that is considered to be a desirable property of an inequality measure. Instead, inequality is concerned with the distribution.

Inequality indicators can be harder to develop than income/consumption poverty indicators because they essentially summarize one dimension of a two-dimensional variable. Note that inequality measures can be calculated for any distribution, not just for consumption, income or other monetary variables, but also for land and other continuous and cardinal variables.

Sometimes we are more interested in measuring inequality than poverty per se. The commonest way to begin is by dividing the population into fifths (*quintiles*) from poorest to richest, and reporting the levels or proportions of income (or expenditure) that accrue to each level. Table 6.1 shows the level of expenditure per capita, in '000 dong per year, for Vietnam in 1993, based on data from the Vietnam Living Standards Survey. A fifth of the individuals (not households) included in the survey were allocated to each expenditure quintile. The figures show that 8.4% of all expenditures were made by the poorest fifth of households, and 41.4% by the top fifth. Quintile information is easy to understand, but sometimes one wants a summary measure rather than a whole table of figures.

Table 6.1: Breakdown of expenditure per capita by quintile, Vietnam 1993

	Expenditure quintiles					Overall
	Lowest	Low-mid	Middle	Mid-upper	Upper	
Per capita expenditure ('000 dong/year)	518	756	984	1,338	2,540	1,227
% of expenditure	8.4	12.3	16.0	21.8	41.4	100.0

Source: Vietnam Living Standards Survey 1993.

6.2 Commonly used measures of inequality

6.2.1 Gini coefficient of inequality

The most widely used single measure of inequality is the Gini coefficient. The Gini coefficient is based on the Lorenz curve, a cumulative frequency curve that compares the distribution of a specific variable (eg. income) with the uniform distribution that represents equality. To construct the Gini coefficient, graph the cumulative percentage of households (from poor to rich) on the horizontal axis and the cumulative percentage of expenditure (or income) on the vertical axis. This gives the Lorenz curve as shown in figure 1. The diagonal line represents perfect equality. The Gini coefficient is defined as

$A/(A+B)$, where A and B are as shown on the graph. If $A=0$ the Gini coefficient becomes 0 which means perfect equality, whereas if $B=0$ the Gini coefficient becomes 1 which means complete inequality.

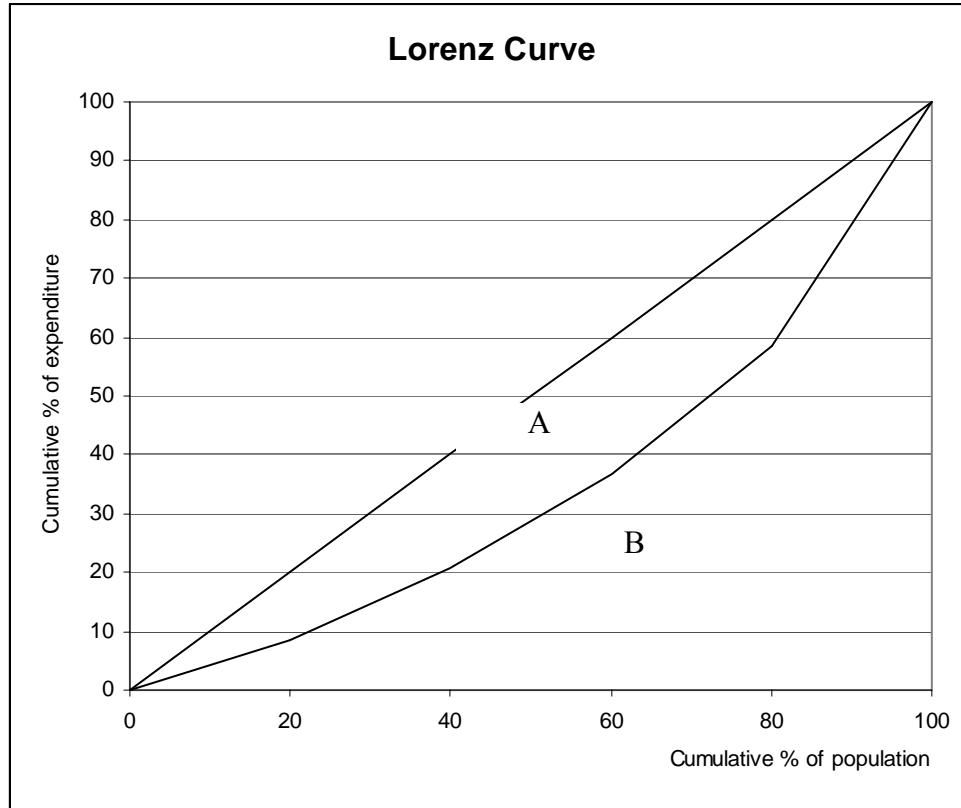


Figure 6.1. Lorenz Curve

Let x_i be a point on the X-axis, and y_i a point on the Y-axis. Then

$$(6.1) \quad Gini = 1 - \sum_{i=1}^N (x_i - x_{i-1})(y_i + y_{i-1}).$$

When there are N equal intervals on the X-axis this simplifies to

$$(6.2) \quad Gini = 1 - \frac{1}{N} \sum_{i=1}^N (y_i + y_{i-1}).$$

The Gini coefficient is not entirely satisfactory. To see this, consider the criteria that make a good measure of income inequality, namely:

- *Mean independence.* This means that if all incomes were doubled, the measure would not change. The Gini satisfies this.
- *Population size independence.* If the population were to change, the measure of inequality should not change, ceteris paribus. The Gini satisfies this too.
- *Symmetry.* If you and I swap incomes, there should be no change in the measure of inequality. The Gini satisfies this.

- *Pigou-Dalton Transfer sensitivity.* Under this criterion, the transfer of income from rich to poor reduces measured inequality. The Gini satisfies this too.

It is also desirable to have

- *Decomposability.* This means that inequality may be broken down by population groups or income sources or in other dimensions. The Gini index is not decomposable or additive across groups. That is, the total Gini of society is not equal to the sum of the Ginis for its subgroups.
- *Statistical testability.* One should be able to test for the significance of changes in the index over time. This is less of a problem than it used to be because confidence intervals can typically be generated using bootstrap techniques.

6.2.2 Generalized Entropy measures

There are a number of measures of inequality that satisfy all six criteria. Among the most widely used are the Theil indexes and the mean log deviation measure. Both belong to the family of generalized entropy inequality measures. The general formula is given by:

$$(6.3) \quad GE(\alpha) = \frac{1}{\alpha(\alpha-1)} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \right)^\alpha - 1 \right]$$

where \bar{y} is the mean income. The values of GE measures vary between 0 and ∞ , with zero representing an equal distribution and higher value representing a higher level of inequality. The parameter α in the GE class represents the weight given to distances between incomes at different parts of the income distribution, and can take any real value. For lower values of α , GE is more sensitive to changes in the lower tail of the distribution, and for higher values GE is more sensitive to changes that affect the upper tail. The commonest values of α used are 0,1 and 2. GE(1) is Theil's T index, which may be written as

$$(6.4) \quad GE(1) = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\bar{y}} \ln \left(\frac{y_i}{\bar{y}} \right)$$

GE(0), also known as Theil's L, is called mean log deviation measure because it gives the standard deviation of $\log(y)$:

$$(6.5) \quad GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \left(\frac{\bar{y}}{y_i} \right)$$

6.2.3 Atkinson's inequality measures

Another class of inequality measures was proposed by Atkinson. This class also has a weighting parameter ε (which measures aversion to inequality) and some of its theoretical properties are similar to those of the extended Gini Index. The Atkinson class is defined as:

$$(6.6) \quad A_\varepsilon = 1 - \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \right)^{1-\varepsilon} \right]^{1/(1-\varepsilon)}$$

6.2.4 Decile dispersion ratio

A simple, and widely-used, measure is the decile dispersion ratio, which presents the ratio of the average consumption of income of the richest 10 percent of the population divided by the average income of the bottom 10 percent. This ratio can also be calculated for other percentiles (for instance, dividing the average consumption of the richest 5 percent – the 95th percentile – by that of the poorest 5 percent – the 5th percentile). This ratio is readily interpretable, by expressing the income of the rich as multiples of that of the poor.

6.3 Inequality comparisons

Many of the tools used in the analysis of poverty can be similarly used for the analysis of inequality. One could draw a profile of inequality, which would look at the extent of inequality among certain groups of households. This provides information on the homogeneity of the various groups, an important element to take into account when designing policy interventions.

One may also analyze the nature of changes in inequality over time. One could focus on changes for different groups of the population to show whether inequality changes have been similar for all or have taken place, say, in a particular sector of the economy. In rural Tanzania, although rural incomes increased substantially between 1983 and 1991, inequality increased (with the Gini coefficient increasing from 0.52 to 0.72), especially among the poor. This can be linked to important reforms that took place in agricultural price policy, which intensified inequalities, with the poor and less-efficient farmers unable to participate in the growth experienced by wealthier, more efficient farmers (Ferreira, 1996).

It is often instructive to analyze other dimensions of inequality. For instance, in a country where public health provision is well developed and reaches all strata of the population, one could expect to see lower levels of inequality in health outcomes than in income levels. This can be done using tabulations along the lines of table 6.2, which presents measures of inequality for rural Egypt, applied to a number of different measures.

The analysis could also focus on the inequality of different consumption categories or income sources. In Egypt, it was found that agricultural income represented the most important inequality-increasing source of income, while non-farm income has the greatest inequality-reducing potential. Table 6.2 sets out this decomposition and shows that while agricultural income only represents 25% of total income in rural areas, it accounts for 40% of the inequality.

Table 6.2: Decomposition of income inequality in rural Egypt, 1997

Income Source	Percentage of households receiving the income source	Share in total income (%)	Gini coefficient for the income source	Percentage contribution to overall income inequality
non farm	61	42	0.63	30
agricultural	67	25	1.16	40
Transfer	51	15	0.85	12
livestock	70	9	0.94	6
Rental	32	8	0.92	12
Total	100	100		100

Source:

6.4 Decomposition of income inequality

The common inequality indicators mentioned above can be used to assess the major contributors to inequality, by different subgroups of the population and regions as well as by income source. In static decompositions, household and personal characteristics, such as education, gender, occupation, urban and rural, and regional location, are determinants of household income. If that is the case, then at least part of the value of any given inequality measure must reflect the fact that people have different educational levels, occupations, genders, and so on. This inequality is referred to as the “between-group” component.

But for any such partition of the population, whether by region, occupation, sector or any other attribute, some inequality will also exist among those people within the same subgroup; this is the “within-group” component. The Theil indexes, and those of the Generalized Entropy class, can be decomposed across these partitions in an additive way, but the Gini index cannot.

To decompose Theil’s T index (i.e. GE(1)), let Y be the total income of the population, Y_j the income of a subgroup, N the total population, and N_j the population in the subgroup. Using T to represent GE(1)

$$(6.7) \quad T = \sum_{i=1}^N \frac{y_i}{N y} \ln \left(\frac{y_i N}{y N} \right) = \sum_{i=1}^N \frac{y_i}{Y} \ln \left(\frac{y_i N}{Y} \right)$$

$$= \sum_j \left(\frac{Y_j}{Y} \right) T_j + \sum_j \left(\frac{Y_j}{Y} \right) \ln \left(\frac{Y_j / Y}{N_j / N} \right)$$

So the inequality measure can be decomposed into two components. The first term represents the within-group inequality and the second term represents the between-group inequality. This decomposability is very useful as it allows one to consider what is driving changes in inequality at the national level.

Similarly, GE(0) can also be decomposed. Using L to represent GE(0):

$$(6.8) \quad L = \sum_{i=1}^N \frac{I}{N} \ln \left(\frac{Y}{Y_i N} \right) = \sum_j \left(\frac{N_j}{N} \right) L_j + \sum_j \frac{N_j}{N} \ln \left(\frac{N_j / N}{Y_j / Y} \right)$$

Here are some examples of different measures of inequality (Dollar and Glewwe 1999, p.40):

Table 6.3: Expenditure Inequality in Selected Less Developed Countries

Country	Gini coefficient	Theil T	Theil L
Côte d'Ivoire, 1985-86	0.435	0.353	0.325
Ghana, 1987-88	0.347	0.214	0.205
Jamaica, 1989	n/a	0.349	0.320
Peru, 1985-86	0.430	0.353	0.319
Vietnam, 1992-93	0.344	0.200	0.169

Source: Reported in Dollar and Glewwe (1999), p.40.

For a typical decomposition of inequality, consider the following simple example, again from Dollar and Glewwe (p.41) and for Vietnam in 1993. Using Theil's T, 22% of the total inequality is attributable to between-group inequality - i.e. to the difference in expenditure levels between urban and rural areas. The remaining 78% of all inequality is due to the inequality in expenditure per capita that occurs within each region.

Table 6.4: Decomposition of expenditure inequality by area, Vietnam

	Theil T	Between-group inequality	Population share (%)
All Vietnam	0.200		100
Urban	0.196	0.044 (22% of total)	20
Rural	0.136		80

Source: Dollar and Glewwe (1999), p.41.

Similar results were found for Zimbabwe in 1995-96. There a decomposition of Theil's T coefficient showed that the within-area (within rural areas and within urban areas) contribution to inequality was 72 percent, while the between-area (between urban and rural areas) component was 28 percent. In many Latin American countries, the between-area component of inequality explains a much higher share of total inequality.

Of equal interest is which of the different income sources, or components of a measure of well-being, are primarily responsible for the observed level of inequality. For example, if total income can be divided into self-employment income, wages, transfers, and property income, one can examine the distribution of each income source. Raising one of the income sources by 1 percent, what would happen to overall inequality? Table 6.5 shows the results for the Gini coefficient for both income and wealth sources in Peru (1997). As the table shows, self-employment income is the most equalizing income source, while agricultural property is the most equalizing wealth asset. An increase in some income sources would actually lead to increased inequality, essentially when these sources are less equally distributed than overall income). The results depend on two factors: (1) the importance of the income source in total income (for larger ones, one percent increase is larger in absolute terms) and (2) the distribution of that income source (if it is more unequal than overall income, it will lead to a reduction; if it is less unequal, it will result in an increase in overall inequality). The size of the impact will be greater, the greater the distance from the overall level of inequality.

Table 6.5: Peru: Expected change in income inequality resulting from a one percent change in income source, 1997 (as percentage of Gini change)

Income source	Expected change	Wealth sources	Expected change
Self-employment income	-4.9	Housing	1.9
Wages	0.6	Durable goods	-1.5
Transfers	2.2	Urban property	1.3
Property income	2.1	Agricultural property	-1.6
		Enterprises	0

CHAPTER 7

Describing Poverty: Poverty Profiles

7.1 What is poverty profile

A poverty profile sets out the major facts on poverty (and typically, inequality), and then examines the pattern of poverty, to see how it varies by geography (by region, urban/rural, mountain/plain, etc.), by community characteristics (e.g. in communities with and without a school, etc.), and by household characteristics (e.g. by education of household head, by size of household). Hence, a "poverty profile" is simply a comprehensive poverty comparison, showing how poverty varies across sub-groups of society, such as region of residence or sector of employment. A well-presented poverty profile can be immensely informative and extremely useful in assessing how the sectoral or regional pattern of economic change is likely to affect aggregate poverty, although it typically uses rather basic techniques (tables, graphs).

For example, regional poverty comparisons are important for targeting development programs to poorer areas. A recent poverty study for Cambodia showed that headcount poverty rates were highest in the rural sector and lowest in Phnom Penh in 1999 (Figure 7.1). Approximately 40% of the rural population, 10% of the Phnom Penh population and 25% of other urban residents live in households that are below the poverty line. Figure 1 also shows the 95% confidence interval that surround the estimates of the headcount index for each area. These standard error bands can be helpful when the sub-populations include only a small number of observations, because the bar charts may otherwise give a misleading sense of confidence in the precision of the poverty comparison that is illustrated. The sampling errors are sufficiently small to have full confidence in the conclusion that headcount poverty rates are lower in Phnom Penh than in other urban areas, which are lower in turn than in rural areas. In terms of contribution to the total amount of poverty, 90.5 % of people living below the poverty line live in rural areas, 7.2% live in other urban areas and 2.3% live in Phnom Penh.

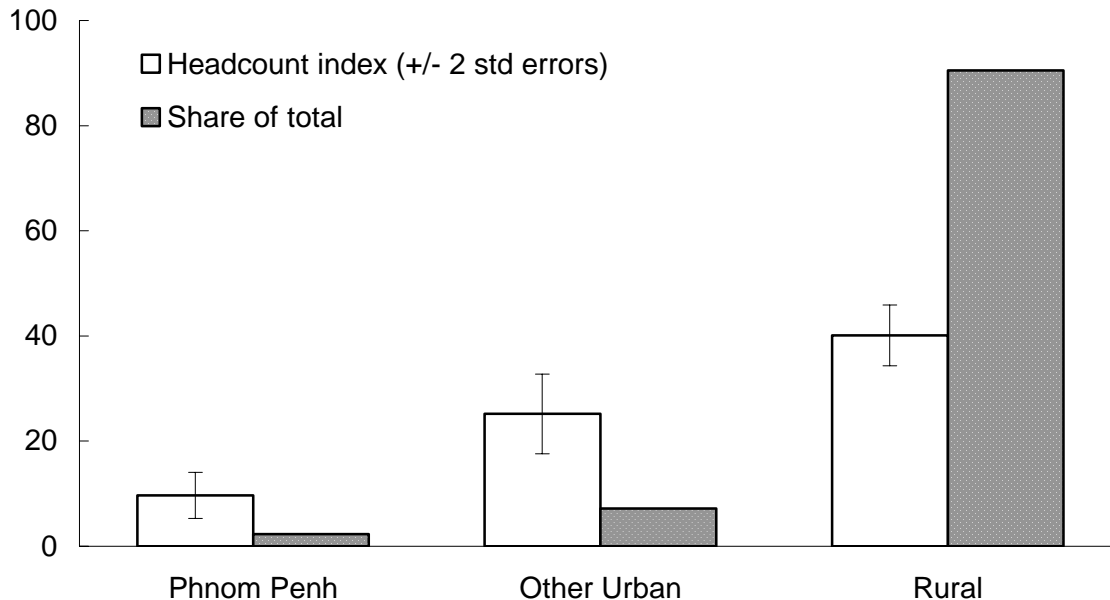


Figure 7.1: Headcount poverty by region

Similarly, table 7.1 presents information on Ecuadorian households' access to services, which mainly differ at the community level. The table shows that the poor have, on average, lower access to services. However, when looking at urban and rural areas separately, it appears that access to services such as electricity is very similar for the poor and non-poor in urban areas.

Table 7.1: Some Characteristics of the Poor in Ecuador, 1994

% with access to basic services	Urban		Rural		Total	
	poor	Non-poor	poor	Non-poor	poor	Non-poor
Sewerage connection	57.3	83.4	12.4	28.2	29.6	63.8
Electricity supply	97.8	99.5	62	75.8	75.8	91.1
Water from public net	61.2	78.8	18.3	23	34.8	59.3
Waste collection	59.7	76.7	1.1	5.6	23.5	51.5

Source: World Bank (1996)

By way of a further illustration, Table 7.2 shows poverty measures by household characteristics — gender and education levels of household head — for Malawi in 1997-98. Clearly, the higher the education level that household heads achieve, the less likely that the household is poor.

Table 7.2: Poverty among various household groups in Malawi, 1997/98

Household Characteristics	Headcount	Poverty gap	Squared Poverty Gap
Gender of Head			
Male	57.9	0.22	0.11
Female	65.6	0.28	0.15

Education levels of Head			
no education	70.6	0.31	0.17
less than standard IV	63.2	0.25	0.13
standard IV	58.1	0.22	0.11
Primary school	47.2	0.15	0.06
Secondary school	29.8	0.08	0.03
University	15.5	0.07	0.04

Source: National Economic Council (2000)

The World Bank's *Poverty Reduction Handbook* (1992) sets out some key questions that one may ask when preparing a poverty profile, as follows:

1. Does poverty vary widely between different areas in the country?
2. Are the most populated areas also the areas where most of the poor live?
3. How is income poverty correlated with gender, age, urban and rural, racial, or ethnic characteristics?
4. What are the main sources of income for the poor?
5. On what sectors do the poor depend for their livelihood?
6. What products or services—tradables and non-tradables—do the poor sell?
7. To what extent are the rural poor engaged in agriculture? In off-farm employment?
8. How large a factor is unemployment? Underemployment?
9. Which are the important goods in the consumption basket of the poor? How high is the share of tradables and non-tradables?
10. How is income poverty linked with malnutrition or educational outcomes?
11. What are fertility characteristics of the poor?
12. To what public services do the poor have access? What is the quality of the service?
13. How important are private costs of education and health for the poor?
14. Can the poor access formal or informal credit markets?
15. What assets—land, housing, and financial—do the poor own? Do property rights over such assets exist?
16. How secure is their access to, and tenure over, natural resources?
17. Is environmental degradation linked to poverty?
18. How variable are the incomes of the poor? What risks do they face?
19. Are certain population groups in society at a higher risk of being poor than others are?

The extent to which a detailed poverty profile can be constructed depends on what data are available. While certain variables such as educational, health indicators, and access to essential services are the most basic components of a poverty profile, the relevance of many other variables depends on country circumstances. The general rule is that all variables that correlate with poverty and are relevant for policies under consideration should be included. By this rule, income generating activities, asset positions, access to social and infrastructure services, and the composition of consumption are all of interest. Cross-links to non-income measures of poverty may also be useful.

7.2 Additive poverty measures

The use of additive poverty measures can greatly facilitate poverty comparisons. Consider the general class of Foster, Greer and Thorbecke poverty measures described by equation (5) of chapter 4. Suppose the population can be divided into m mutually exclusive sub-groups. The poverty profile is

simply the list of poverty measures P_j for $j=1, \dots, m$. Aggregate poverty can then be written as the average of the sub-group poverty measures, weighted by their population:

$$(7.1) \quad P = \frac{1}{N} \sum_{j=1}^m P_j N_j$$

where

$$(7.2) \quad P_j = \frac{1}{N_j} \sum_{i=1}^{N_j} p(z_j, y_i^j)$$

is the poverty measure for the sub-group j with population N_j . Here y_i^j is the welfare indicator of individual i who belongs to the sub-group j , where $i=1, \dots, N_j$. The total population N is equal to $\sum_{j=1}^m N_j$

Analogously, one can also define "clusters" of sub-groups; as one disaggregates further and further, the poverty profile at each step adds up to that of the previous step, using population weights.

In addition to the computational convenience in forming poverty profiles, additive poverty measures guarantee "subgroup consistency" in the sense that when poverty increases in any sub-group of the population, aggregate poverty will also increase. Indeed, researchers have shown that (given certain assumptions of technical nature) subgroup consistency implies, and is implied by, the class of measures defined by equation (5) of chapter 4. This property is intuitively appealing for any poverty profile. Indeed, an evaluation of the effects on aggregate poverty of targeted poverty alleviation schemes – whereby the benefits are concentrated in certain sub-groups – may be quite misleading unless the poverty measure used has this property; some measures of aggregate poverty may show an increase even if poverty fell in the target group and there were no changes elsewhere. Subgroup consistency can thus be viewed as a desirable property in evaluating anti-poverty policies.

One possible objection to additivity is that it attaches no weight to one aspect of a poverty profile: the differences between sub-groups in the extent of poverty. Consider two equal-sized groups - "rural" and "urban" sectors - with initial poverty indices of 0.70 and 0.20 respectively. Aggregate poverty is 0.45 according to any (population weighted) additive measure. One is to choose between two policies X and Y. Under policy X, the poverty profile changes to 0.70 and 0.10, while under policy Y the profile becomes 0.60 and 0.20. By any additive poverty measure one should be indifferent between X and Y; both yield an aggregate poverty index of 0.40. Yet, in contrast to X, the gains under policy Y have gone to the poorer rural sector. Should we prefer policy Y? The answer is only "yes" if one is concerned about inequalities between groups independently of absolute living standards, and it is by no means clear why we should be.

7.3 Profile Presentation

There are two main ways of presenting a poverty profile. The first ("type A") gives the incidence of poverty or other poverty measure(s) for each sub-group defined in terms of some characteristic, such as place of residence; this is done in Table 7.2. The second ("type B") gives the incidence of characteristics amongst sub-groups defined in terms of their poverty status, such as "poor" and "non-poor", as is done in the dark columns in Figure 7.1.

Which is more useful will depend on the purpose of the poverty profile. Suppose that one is using the poverty profile to select a target region for a poverty alleviation scheme. The scheme will allocate a small sum of money to all residents in the chosen target region (“indicator targeting”). It is an imperfect form of targeting because (as is invariably the case) the policy maker does not know who has which standard of living, even when a distribution of living standards can be constructed from a household sample survey. The policy maker therefore has to make do with an imperfect indicator of living standards, in this case region of residence. In this case a type A profile will be more useful, because it identifies the areas (or groups) where poverty rates are highest, and hence where universal cash transfers are least likely to be handed out to the non-poor.

7.4 Poverty Comparisons over time

If several consecutive rounds of household surveys are available, changes in income poverty over time can be assessed. This requires poverty measures that are comparable and that reflect differences over time in the cost of living across regions. The commonest method for preparing comparisons over time consists of converting nominal income or consumption data from different surveys and regions into real income and consumption by deflating the indicators in space and time. A constant poverty line can then be applied to these real values to infer poverty measures. Ideally, to obtain robust poverty comparisons over time, one would want to use surveys with similar sampling frame and methods, with corrections for prices differences, and with similar definitions of consumption or income. In practice, however, differences exist in some of these dimensions. This does not imply that no comparison can be made. It simply means that the analyst will need to:

- correct for major differences in the sampling frame and sampling method for the different surveys or the different rounds of a panel survey;
- use regional and temporal price indices to ensure a similar definition of the poverty line over time and across regions;
- adjust the definition of consumption or income aggregates over time to ensure a similar definition is used. Changes in definitions, and in particular in the degree to which home production is included in the definition, can lead to important distortions of poverty measurement.

When several rounds of survey data are available, the analyst can investigate changes in the regional distribution of poverty or in the major characteristics of the poor, such as ethnicity, gender, age, urban and rural location, employment, access to social programs and basic services, etc. Although the various population groups identified in the first period of time should clearly form the basis of the analysis over time, it is also important to investigate whether or not ‘new’ groups of poor people have appeared. This is particularly relevant for countries that undergo rapid changes linked to such factors as economic reforms, conflicts, natural disasters, and epidemics such as HIV/AIDS. For example, comparing the baseline poverty profile for Cambodia derived from the 1993/94 data with that of the CSES 1997 suggests that the food poverty line (consisting of the food poverty line plus a non-food allowance equal to the level of non-food consumption of persons whose per capita consumption just equals the food poverty line) increased from 1,578 to 1,819 riels per day in Phnom Penh (15%), from 1,264 to 1,407 riels per day in other urban areas (11%), and from 1,117 to 1,210 riels (8%) in rural areas.

Table 7.3: Cambodia: Poverty Measures, 1993-94 and June 1997

	Head count index (%)		Poverty gap index (%)		Poverty severity index (%)	
	1993/94	1997	1993/94	1997	1993/94	1997
Food Poverty Line						
Phnom Penh	6.2	3.4	1.3	0.5	0.4	0.1
Other Urban	19.6	15.4	4.4	3.3	1.4	1.1
Rural	21.9	20.0	4.0	3.9	1.1	1.2
Total	20.0	17.9	3.7	3.5	1.1	1.1
Poverty Line						
Phnom Penh	11.4	11.1	3.1	2.2	1.2	0.6
Other Urban	36.6	29.9	9.6	7.5	3.6	2.7
Rural	43.1	40.1	10.0	9.7	3.3	3.4
Total	39.0	36.1	9.2	8.7	3.1	3.1

The estimates in table 7.3 indicate that the incidence of poverty declined modestly in Cambodia as a whole (from 39% to 36%) during the period 1993/94 to June 1997. On a regional basis, poverty declined significantly in other urban areas (from 37% to 30%), modestly in rural areas (from 43% to 40%) and not at all in Phnom Penh (where it remained at 11%). During the same period, the estimates indicate that two other measures of poverty (i.e. the poverty gap and poverty severity index) declined significantly, both in Phnom Penh and in other urban areas but not in rural areas.

One can also look at changes in the characteristics of different poverty groups. For example, the distribution of access to services in the base year can be compared with the distribution of services in the second year. This will then show whether changes made in the supply of the services have been pro-poor.

Poverty measures sometimes are translated into the *relative risks* of being poor for different household groups. These risks indicate whether the members of a given group are poor in relation to the corresponding probability for all other households of society. This concept may be applied to examine whether, over time, the relative poverty risk of specific population groups decreases or increases. Table 7.4 compares the relative poverty risk of various groups in Peru in 1994 and 1997. It shows, for example, that the poverty risks of households of seven persons or more increased over time (from 71% to 106%), while that of households where the spouse of the head is working diminished (from -11% to -21%).

Table 7.4. Peru: Poverty Risks for Selected Groups of Households, (%)

Household characteristic	1994	1997
Households using house for business purposes	-28	-29
Rural households with at least one member in off-farm empl.	-24	-23
Households with head's spouse working	-11	-21
Households without water or sanitation	54	50
Households without electricity	63	69
Households with head less than secondary education	73	72
Households of 7 persons or more	71	106
Source:		

7.5 Excerpts from poverty profiles for Indonesia and Cambodia

7.5.1 Indonesia

Table 7.5 gives an example of a poverty profile in which the sampled households in Indonesia's 1987 SUSENAS have been classified into 10 groups according to their principal income source. Results are given for the three main poverty measures discussed above. The following points should be noted:

- An urban-rural cost-of-living differential of 10% has been assumed. From the only empirical work that has been done, this appears to be a reasonable assumption, though it is lower than what has been assumed in most other poverty profiles for Indonesia.
- The poverty measures are based on the estimated population distributions of persons ranked by household consumption per person, where each person in a given household is assumed to have the same consumption. Household specific sampling rates have been used in estimating the distributions.

Table 7.5: A Sectoral Poverty Profile for Indonesia in 1987

Principal sector of employment	Population share (1987)	Head-count index (%)	Poverty gap index (%)	FGT P ₂ measure (x100)
Farming				
Self-employed	41.1	31.1	6.42	1.97
Labourer	8.6	38.1	7.62	2.21
Industry				
Urban	3.0	8.1	1.26	0.32
Rural	3.4	19.4	3.00	0.76
Construction	4.3	17.4	2.92	0.80
Trade				
Urban	6.3	5.0	0.71	0.17
Rural	7.6	14.7	2.42	0.61
Transport	4.1	10.7	1.53	0.34
Services				
Urban	7.6	4.2	0.61	0.14
Rural	7.3	11.6	1.84	0.49
Other	6.7	17.1	3.55	1.03
Total	100.0	21.7	4.22	1.24

- In forming the poverty profile, households have been grouped by their stated "principal income source". Many households will have more than one income source. In principle one could form sub-groups according to the various interactions of primary and secondary income sources, but this would rapidly generate an unwieldy poverty profile. An alternative is to calculate average incomes from various sources for various consumption groups (for example, ultra-poor, poor, near-poor, etc.).
- The three measures are in close agreement on the ranking of sectors in terms of poverty, with very few re-rankings. For example, the two farming sub-groups are the poorest for all three measures.

Changes in the poverty profile may arise from the contributions of different sub-groups in the poverty profile to changes over time in aggregate poverty. Table 7.6 provides information on the relative contribution of various sectors to aggregate poverty alleviation in Indonesia between 1984 and 1987. These are the "intra-sectoral effects", expressed as a percentage of the reduction in aggregate poverty for each poverty measure. The table also gives the aggregate contribution of shifts in population and the interaction effects between sectoral gains and population shifts.

Table 7.6: Sectoral Decomposition of the Change in Poverty in Indonesia, 1984-1987

Principle sector of employment	Population share (1984)	Contribution of sectoral change in:		
		Head-count index (%)	Poverty gap index (%)	FGT P ₂ measure
Farming				
Self-employed	45.0	49.8	54.6	57.4
Labourer	9.0	11.2	14.8	16.5
Industry				
Urban	2.6	0.8	0.4	0.3
Rural	3.3	2.8	3.1	2.7
Construction	4.1	3.2	2.6	2.2
Trade				
Urban	5.4	2.2	1.6	1.4
Rural	6.6	7.2	5.6	4.7
Transport	3.8	3.6	2.7	2.2
Services				
Urban	6.5	1.0	1.0	0.9
Rural	5.8	2.9	2.4	2.0
Total sector effects (incl. omitted sectors)		89.3	93.8	95.1
Contribution of population shifts		13.2	10.4	9.4
Interaction effects		-2.6	-4.3	-4.5
Total	100.0	100.0	100.0	100.0

Note: minor sectors omitted.

Source: Adapted from Huppi and Ravallion (1991).

The drop in poverty among self-employed farmers had the largest influence on aggregate poverty reduction, and most particularly on the reduction in the severity of poverty as measured by P₂. About 50% of the reduction in the national head-count index was due to gains in this sector, while it accounted for 57% of the gain in P₂. The second most important contribution came from gains to farm workers, whose reduction in poverty as measured by the head-count index contributed 11% to the reduction in the aggregate index, while the decline in this sector's P₂ measure contributed almost 17% of the aggregate decline. These two groups jointly accounted for 61% of the reduction of the aggregate head-count index and 74% of the reduction of the aggregate value of P₂. Note that the rural farm sector's impressive participation in the reduction of aggregate poverty is due to both significant declines in their poverty measures, and the large share of national poverty accounted for by this sector.. Over 13% of the decline in the national head-count index was due to population shifts between various sectors of employment, and over 9% of the decline in the P₂ measure can be traced back to these shifts. The sectors that gained in population share were almost all urban (Huppi and Ravallion 1991), and had initially lower poverty measures. This is the main factor underlying the contribution of population shifts to poverty alleviation. The fact that population was moving out of the rural sector, where poverty was falling faster, accounts for the negative interaction effects in Table 7.6.

7.5.2 Cambodia

A basic breakdown of Cambodian poverty rates, by region, is given above in Table 7.3. The numbers show that at least 85% of the poor are concentrated in rural areas.

Some more recent figures are shown in table 7.7, using data from the CSES 1999. Data in 1999 were collected in two rounds, and table 7.7 contains estimates for each round (and the pooled sample) of the three main poverty statistics, and also reports the results from the previous surveys for comparison.

Table 7.7: Comparisons of Poverty Estimates from Cambodian Surveys

	Headcount	Poverty gap	Poverty severity
SESC 1993/94	39.0	9.2	3.1
1997 CSES (as adjusted by Knowles)	36.1	8.7	3.1
1997 CSES (unadjusted)	47.8 (1.5)	13.7 (0.7)	5.3 (0.3)
CSES 1999 (Round 1)	64.4 (2.3)	23.9 (1.3)	11.3 (0.8)
CSES 1999 (Round 2)	35.9 (2.4)	6.5 (0.7)	2.0 (0.4)
CSES 1999 (both Rounds combined)	51.1 (1.8)	15.4 (0.9)	6.7 (0.5)

Note: No sampling errors (reported in parentheses for the other years) are reported by the two previous poverty profiles but the relative errors for SESC 1993/94 and the adjusted 1997 CSES would likely be higher than the relative error in 1999 because the sampling scheme used previously was not as efficient (fewer clusters and broader stratification). The poverty line used for the unadjusted 1997 CSES results takes values of 1923 Riel per person per day in Phnom Penh, 1398 in other urban and 1195 in rural.

An interesting feature of these results is that there is a substantial discrepancy in the poverty estimates from the two survey rounds in 1999. The headcount index is almost 30 percentage points higher in Round 1 than in Round 2, while the poverty gap and poverty severity indexes are between four and six times higher. These discrepancies are also large relative to the variation across previous survey estimates of poverty in Cambodia. If the discrepancies between the two survey rounds are ignored, and the data are pooled, the resulting poverty estimates are fairly similar to the unadjusted 1997 estimates, showing a slight *increase* in all three poverty measures (table 7.7). This comparison can be considered the “do nothing” one because it uses all of the survey data in both years.

The pattern of poverty with respect to the age group of the household head is reported in table 7.8. It is apparent that poverty rates rise with age, reaching a maximum for the 36-40 year old group of household heads, and then declining. A similar pattern was reported in the 1997 poverty profile. Once again, the definition of headship and its economic interpretation may confound the results so more detailed examination would be needed before any interventions might be designed on the basis of these age patterns. For example, the household head need not be the major economic contributor to the household, respondents may simply have nominated the oldest or most senior member. Thus, the relatively low poverty rate for people living in households whose head is aged 61 years and above may reflect the wealth accumulation that this elderly head has achieved or it could be that there is a younger generation within the household whose economic success is sufficient to allow them to support their elders within the same household.

Table 7.8: Distribution of Poverty by Age and Gender of Household Head

	Headcount		Poverty gap		Poverty severity		Share of total pop (%)
	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	
<i>Poverty Line</i>	35.9	100.0	6.5	100.0	2.0	100.0	100.0
18-30 years	36.7	10.7	5.6	9.1	1.4	7.5	10.5

31-35 years	35.4	10.9	5.4	9.2	1.6	8.8	11.1
36-40 years	43.6	21.2	8.0	21.6	2.7	23.3	17.5
41-45 years	40.3	15.7	7.3	15.8	2.2	15.3	14.0
46-50 years	36.5	14.4	7.7	16.9	2.4	16.9	14.2
51-60 years	28.3	15.8	5.3	16.3	1.7	16.8	20.0
61 and Above	32.0	11.3	5.6	11.1	1.8	11.3	12.7
Male	36.4	84.4	6.6	84.2	2.1	85.1	83.3
Female	33.6	15.7	6.1	15.8	1.8	14.9	16.7

Source: Gibson(1999).

In addition to the gender of the household head, marital status and widowhood (including male widowers) is sometimes used as a sub-population characteristic for forming poverty profiles. There are two reasons why widow-headed households, and households where there has been a dissolution (i.e., separation or divorce), could be at greater risk of poverty. The loss of an economically active household member, as would occur with the death of a husband in war for example, is likely to cause a large income shock that could push a household into poverty. The second factor, and the one that links marital status with household size, is that widow-headed households tend to be smaller than average, which will constrain the effective living standards of their residents if there are economies of household size.

In Cambodia, the headcount poverty rate in 1999 increased smoothly with household size to a maximum rate for households with eight members (figure 7.2). In the round 1 data the highest headcount poverty rate is for households with nine members. A relationship like that shown in figure 7.2 might normally be doubted because uncontrolled for household size economies may be the cause of the pattern: large households may have lower expenditures not because their members are poor but because they don't need to spend as much per person to reach the same standard of living. But the available evidence suggests that size economies may not be important for Cambodian households, so the pattern shown by figure 7.2 may be a useful basis for identifying the poor.

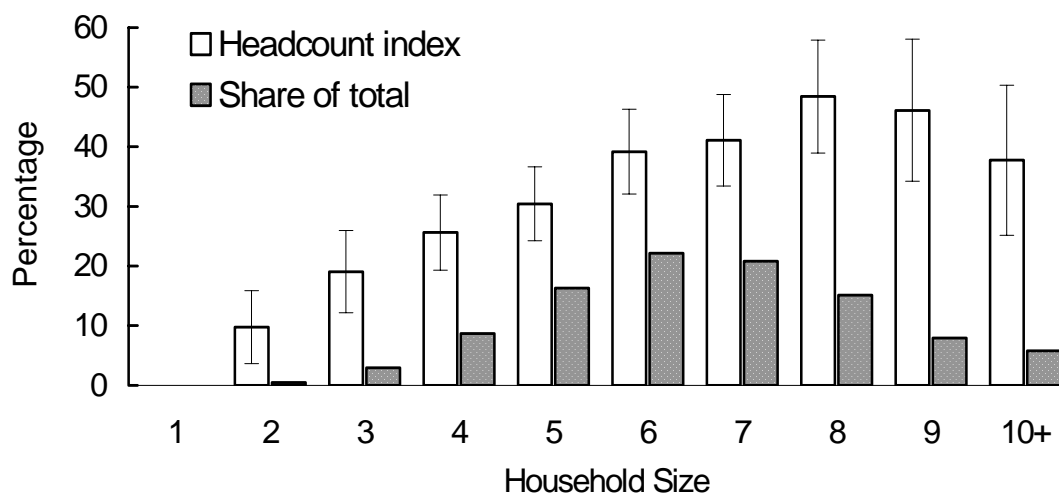


Figure 7.2: Poverty by household size

Previous poverty profiles showed that poverty rates were relatively high among those whose household head either has no schooling or has only primary schooling. Poverty rates then fall with the attainment of lower secondary education, fall farther with upper secondary and are almost zero if the household head is a university graduate. But those whose household head had a technical/vocational or

other form of education had a higher poverty rate than those with primary schooling (at least in the 1997 poverty profile).

According to the survey estimates, there was no difference, in 1999, in poverty rates between those whose household head has no schooling and those whose head has some primary education (figure 7.3). Although the survey estimate of the headcount poverty rate is slightly higher for the primary schooled group, the estimates for both groups are

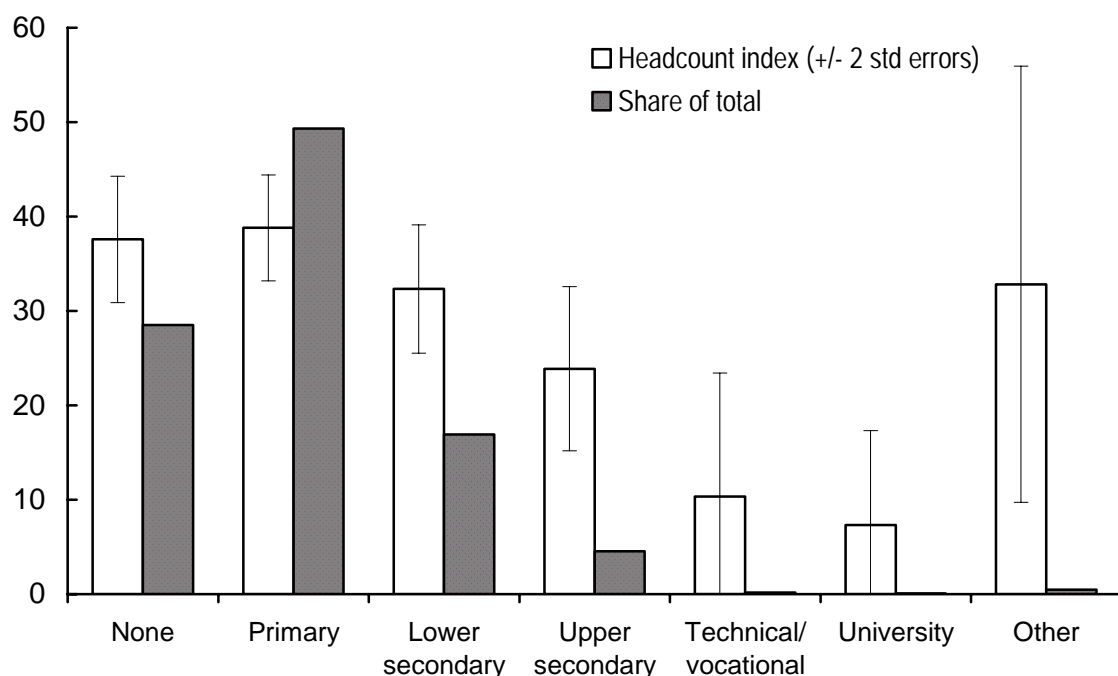


Figure 7.3: Poverty by education level of household head

surrounded by wide and overlapping confidence intervals, and the ranking reverses with the poverty gap and poverty severity indexes and at the food poverty line (table 7.9).

Table 7.9: Cambodia: Distribution of Poverty by Education Level of Head

	Headcount		Poverty gap		Poverty severity		Share of total pop (%)
	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	
<i>Poverty Line</i>	35.9	100.0	6.5	100.0	2.0	100.0	100.0
No schooling	37.6	28.5	7.0	29.3	2.2	30.1	27.3
Primary	38.8	49.3	6.8	47.9	2.1	48.2	45.7
Lower secondary	32.3	16.9	6.0	17.4	1.8	16.4	18.8
Upper secondary	23.9	4.5	4.5	4.7	1.3	4.5	6.8
Technical/vocational	10.3	0.2	0.8	0.1	0.1	0.0	0.6
University	7.3	0.1	2.2	0.1	0.8	0.1	0.3
Other	32.8	0.5	6.7	0.5	2.4	0.6	0.5

Literate	35.2	73.1	6.3	72.7	2.0	72.1	74.7
Illiterate	38.2	26.9	7.0	27.3	2.2	27.9	25.3

Source: Gibson(1999).

Table 7.10 reports poverty rates and contributions to the total according to the employment status of the household head. Among the major groups, it is those whose household head is self-employed that have the highest headcount poverty rate, although this ranking changes when considering the other poverty measures and using the food poverty line. The depth and severity of poverty appear to be greatest for those people whose household head is an unpaid family worker and this is also reflected in their relatively large contribution to total poverty (i.e., this group comprises 1.3% of the population and 1.2% of the headcount poor but contribute 12.4% of the total poverty severity measure at the food poverty line).

Table 7.10: Distribution of Poverty by Employment Status of Head

	Headcount		Poverty gap		Poverty severity		Share of total pop (%)
	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	Index (%)	Contribution to total (%)	
Poverty Line	35.9	100.0	6.5	100.0	2.0	100.0	100.0
Employee – private sect	34.8	6.6	7.5	7.9	2.4	8.0	6.8
Employee – public sect	19.4	5.1	3.7	5.4	1.1	5.2	9.5
Self-employed	38.8	80.5	6.6	75.8	1.9	71.6	74.5
Unpaid family worker	33.2	1.2	16.0	3.1	10.4	6.4	1.3
Other (incl. Employer)	33.3	0.2	7.5	0.3	1.8	0.2	0.2
Unemployed	53.1	0.2	6.8	0.1	0.9	0.0	0.1
Not in the labour force	29.7	6.2	6.4	7.4	2.3	8.7	7.5

Source: Gibson(1999).

7.6 Poverty mapping

Poverty analysis is often based on national level indicators that are compared over time or across countries. The broad trends that can be identified using aggregate information are useful for evaluating and monitoring the overall performance of a country. For many policy and research applications, however, the information that can be extracted from aggregate indicators may not be sufficient, since they hide significant local variation in living conditions within countries. For example, poverty within a region can vary across districts. This makes small-area estimates of poverty very appealing. However, often we are unable directly to compute poverty estimates for small areas like districts. Instead, we usually have poverty estimates for regions or entire countries only.

The main reason that poverty measures are computed for large areas and not usually available for small areas is data availability. There are two main types of welfare-related information sources available to policy-makers. *Household surveys* often include a detailed income and/or consumption expenditure module (such as the Cambodian CSES 1999). However, due to relatively small sample size, the collected information is usually only representative for broad regions of the country. For example, with the CSES 1999 we can compute poverty estimates for Phnom Penh, other urban areas and rural areas but not for every district. *Census data* (and sometimes large household sample surveys) are available for all households (or very large samples of households) and can provide reliable estimates at highly

disaggregated levels such as small municipalities, towns, and villages. But censuses do not contain the income or consumption information necessary to yield reliable indicators of the level and distribution of welfare such as poverty rates or inequality measures.

However, it is possible to merge information from these two types of data sources (detailed household surveys like the CSES 1999, and census data) so that “poverty maps” can be constructed. These detailed poverty maps capture the heterogeneity of poverty within a country. That is, areas that are better off and those that are worse off will be more clearly defined. Sometimes regions that have less poverty overall may have substantial pockets of poverty that are lost in the aggregate poverty statistics.

Poverty maps can improve the targeting of interventions. In designing poverty alleviation projects and allocating subsidies, resources will be used more effectively if the most needy groups can be better targeted. This reduces the leakage of transfer payments to non-poor persons, and it reduces the risk that poor persons will be missed by a program. Many countries use geographic targeting schemes based on relatively unreliable welfare indicators such as the “basic needs indicators”, which are *ad hoc* combinations of various indicators into one index.

Poverty maps can also help governments to articulate their policy objectives. Basing allocation decisions on observed geographic poverty data rather than subjective rankings of regions increases the transparency of government decision making. Such data can thus help limit the influence of special interests in allocation decisions. There is a related role for well-defined poverty maps to lend credibility to government and donor decision making.

To create a poverty map, first use the household survey data to estimate a model of per capita consumption expenditure (or any other household or individual-level indicator of well-being) as a function of variables that are common to both the household survey and the census. Then use the resulting estimated equation to predict per capita expenditures for each household in the census. The estimated household-level measures of poverty and inequality may then be aggregated for small areas, such as districts, villages, or even neighborhoods. For a recent application to Vietnam, see Baulch and Minot (2001); discussions of the methodology occur in Hentschel et al. (2000), Elbers, Lanjouw and Lanjouw (2000), and Alderman et al. (2000).

CHAPTER 8

Understanding the Determinants of Poverty

8.1 What causes poverty?

The previous chapters have already touched on some of the determinants of poverty. The factors could be macro, sector-specific, community, household and individual characteristics. This chapter summarizes some of the characteristics of the poor by region, community, household and individual characteristics followed by quantitative and qualitative description of the major determinants of poverty. It then discusses how regression techniques can be used to determine the factors causing poverty, or at least the factors correlated to poverty. Note that correlation means association but does not necessarily mean causality.

8.2 Regional level characteristics

At the regional level, there are numerous characteristics that might be associated with poverty. The relationship of these characteristics with poverty is country specific. In general, however, poverty is high in areas characterized by geographical isolation, a low resource base, low rainfall, and other inhospitable climatic conditions. For example, many argue that economic development in Bangladesh is severely retarded due to its susceptibility to annual floods. In the case of Cambodia, low purchasing power and the remoteness of rural areas is responsible for generating food insecurity among the poor. Inadequate public services, weak communications and infrastructure, as well as underdeveloped markets are dominant features of life in rural Cambodia that contribute to poverty.

Other important regional and national characteristics that affect poverty include good governance, a sound environmental policy, economic, political and market stability, mass participation, global and regional security, intellectual expression and a fair, functional, and effective judiciary. Regional-level market reforms can boost growth and help poor people, but it is important to note that they can also be a source of dislocation. The effect of market reforms are complex, deeply linked to institutions and to political and social structures. The experience of transition, especially in countries of the former Soviet Union, vividly illustrates that market reforms in the absence of effective domestic institutions can fail to deliver growth and poverty reduction. There is also a case for bringing vulnerability and its management to center stage. Participatory poverty work underlines the importance of vulnerability to economic, health, and personal shocks. So do the financial crises of the 1990s – not least in East Asia, the shining example of success in development and poverty reduction – and the sequence of devastating natural disasters.

Inequality is also back on the agenda. New work shows the importance of gender, ethnic, and racial inequality as a dimension – and a cause – of poverty. Social, economic, and ethnic divisions in regions are often sources of weak or failed development. In the extreme, vicious cycles of social division and failed development erupt into internal conflict (within or across regions), as in Bosnia and Sierra Leone, with devastating consequences for people.

8.3 Community level characteristics

As with regional characteristics, there are a variety of community-level characteristics that may be associated with poverty for households in that community. At the community level, infrastructure is a major determinant of poverty. Indicators of infrastructure development that have often been used in econometric exercises include proximity to paved roads, whether or not the community has electricity, proximity to large markets, availability of schools and medical clinics in the area, and distance to local administrative centers. Other indicators of community level characteristics include human resource development, equal access to employment, social mobility and representation, and land distribution.

Recent research has also stressed the importance of social networks and institutions, and social capital in the community. In addition to removing social barriers, effective efforts to reduce poverty require complementary initiatives to build up and extend the social institutions of the poor. Social institutions refer to the kinship systems, local organizations, and networks of the poor and can be thought of as different dimensions of social capital. Research on the roles of different types of social networks in poor communities confirms their importance. An analysis of poor villages in North India, for example, shows that social groups play an important role in protecting the basic needs of poor people and in mediating risk. Studies of agricultural traders in Madagascar show that social relationships are more important to traders than input prices. Close relationships with other traders are used to lower the transactions costs of exchange, while ties to creditors and others who can help out during times of financial hardship are vital sources of security and insurance.

How does social capital affect development? The narrowest view holds social capital to be the social skills of an individual – one’s propensity for cooperative behavior, conflict resolution, tolerance and the like. A more expansive meso view associates social capital with families and local community associations and the underlying norms (trust, reciprocity) that facilitate coordination and cooperation for mutual benefit. A macro view of social capital focuses on the social and political environment that shapes social structures and enables norms to develop. This environment includes formalized institutional relationships and structures, such as government, the political regime, the rule of law, the court system, and civil and political liberties. Institutions have an important effect on the rate and pattern of economic development. An integrating view of social capital recognizes that micro, meso, and macro institutions coexist and have the potential to complement one another. Macro institutions can provide an enabling environment in which micro institutions develop and flourish. In turn, local associations help sustain regional and national institutions by giving them a measure of stability and legitimacy – and by holding them accountable for their actions. Social capital is clearly a complicated characteristic and often researchers find it difficult to identify appropriate variables that measure social capital quantitatively.

8.4 Household and individual level characteristics

Some of the important characteristics in this category would include the age structure of household members, education, gender of the household head, and the extent of participation in the labor force. In recent times, other components that fall under this category have included domestic violence prevention, and gender-based, anti-discrimination policies. The following discussion organizes these characteristics into groups and discusses them in greater detail. These groups are demographic, economic and social characteristics.

8.4.1 Demographic characteristics

Demographic characteristics of the household can be broadly classified into three categories, as follows:

8.4.1.1 Household size and structure

This indicator is an important one as it shows a possible correlation between the level of poverty and household composition. Household composition, in terms of the size of the household and characteristics of its members (such as age), is often quite different for poor and non-poor households. The Cambodian CSES of 1993/94 shows that the poor tend to live in larger households, with an average family size of 6.6 persons in the poorest quintile compared to 4.9 in the richest quintile; similar patterns are found in most countries. The poor also tend to live in younger households – with the bottom quintile having twice as many children under 15 per family as the top quintile – and slightly fewer elderly people over age 60. Better-off households also tend to have heads that are somewhat older.

8.4.1.2 Dependency ratio

The dependency ratio is calculated as ratio of the number of family members not in the labor force (whether young or old) to those in labor force in the household. This ratio allows one to measure the burden weighing on members of the labor force within the household. One might expect that a high dependency ratio will be correlated positively with the level of household poverty.

8.4.1.3 Gender of the household head

It is widely believed that the gender of the household head significantly influences household poverty, and more specifically that households headed by women are poorer than those headed by men. This is of particular importance to Cambodia. Due to male casualties in past wars, women are often the heads of households. Women play an important role in the labor force, both in the financial management of the household and in the labor market, but appear to face large degree of discrimination. They are severely affected by both monetary and non-monetary poverty, for example, they have low levels of literacy, are paid lower wages, and have less access to land or equal employment. According to a report based on a joint conference between the Cambodian Institute for Cooperation and Peace and the World Bank Institute, 43 percent of women are illiterate and 90 percent of these women are poor. According to the Cambodian Ministry of Women's and Veterans' Affairs, the following contribute to poverty—lack of opportunities for employment and education, lack of access to finance, mass illiteracy, food insecurity, malnutrition, human trafficking, powerlessness, no resources, overwork in wage employment and in households, discrimination in the labor market and in work places, and domestic violence. So many observers are surprised to learn that poverty rates are *not* higher among female-headed than male-headed households in Cambodia. Likewise, female-headed households in neighboring Vietnam are no more likely to be in poverty than their male-headed counterparts.

8.4.2 Economic characteristics

Economic characteristics include employment, income, consumption spending and household property:

8.4.2.1 Household employment

There are several indicators for determining household employment. Within this array of indicators, economists focus on the rate of participation in the labor force, the real rate of unemployment, the rate of underemployment and job changes.

8.4.2.2 Household incomes

Income represents a very important area to consider when characterizing the poor. Of interest is the level of income as well as its distribution among the household members and among the various socio-economic groups. However, in practice income indicators present us with certain problems. As noted in chapter 2, income is difficult to define, as it includes several components, of which only some are monetary (for example, farm households consume most of their production on-site). Second, individuals tend to make false declarations about their income level, which is generally under-estimated. It is possible, in part, to correct these declarations but only at the cost of carrying out a large-scale data-gathering operation on economic activities, the costs of production factors and inputs and the prices of products. Given these limitations and the fact that savings are low, there is often a tendency to use the household's total spending as an approximation to its disposable income.

8.4.2.3 Structure of household consumption expenditures

The structure of household consumption expenditures can be used to characterize households by describing the make-up of food and non-food spending. What is of interest is to measure the relative weight of the goods and services consumed by the household according to its poverty level. This measurement gives some indication as to the probable impact of price variation on household purchasing power. We can expect basic products, especially food, to represent a more significant part of total spending by the poor.

8.4.2.4 Household property

The property of a household includes its tangible goods (land, cultivated areas, livestock, agricultural equipment, machinery, buildings, household appliances and other durable goods) and its financial assets (liquid assets, savings and other financial assets). These indicators are of interest as they represent the household's inventory of wealth and therefore affect its income flow. Furthermore, certain households, especially in rural areas, can be poor in terms of income, but wealthy when their property is taken into consideration. Despite its importance, property is difficult to evaluate in practice in any reliable way. First, one encounters the same problem of under-declaration. Second, it is very difficult to measure certain elements of property such as livestock. Finally, the depreciation of assets may be difficult to determine for at least two reasons: 1) the life span of any given asset is variable; 2) the acquisition of these assets occurs at different moments in each household. Therefore, property is more difficult to use than certain other elements in the characterization of poverty.

8.4.3 Social characteristics

Aside from the demographic and economic indicators, we also have recourse to social indicators to characterize poverty and household living standards. The social indicators generally selected are health, education and shelter.

8.4.3.1 Health within the Household

Four types of indicators are normally used to characterize health in analyzing a household's living standards. These indicators include nutritional status (for example, anthropometric indicators such as weight for age, height for age, and weight for height), disease status (for example, infant and juvenile mortality and morbidity rates as related to certain diseases such as malaria, respiratory infections, diarrhea and sometimes poliomyelitis), the availability of health care services (primary health-care centers, maternity facilities, hospitals and pharmacies, basic health care workers, nurses, midwives, doctors and traditional healers, and medical service such as vaccinations, access to medicines and medical information) and the use of these services by poor and non-poor households.

The data from the Cambodian Socio-Economic Survey of 1993/94 show that water and sanitation are especially important influences on health and nutritional status. The CSES shows that the poor are extremely disadvantaged in access to safe sources of water supply and sanitation. Only 4 percent of the poorest quintile have access to piped water, while more than 17% of the richest quintile have the same. Similar differences are apparent in access to sanitation. Few of the poor (9 percent) have access to a toilet in the home, while around half of the richest 20 percent do. Another indicator of housing standards is access to electricity. Here again the access of the poor lags far behind. Access to electricity from a generator or line connection rises sharply with income, from a mere 1 percent among people in the bottom quintile to 37 percent of Cambodians in the richest quintile. Other indicators of household wealth include ownership of transportation. Access to bicycles is quite evenly distributed, with at least one half of households owning a bicycle in every quintile, even the poorest. However access to cars, jeeps or motorbikes is very rare among the poor and rises sharply with income. Overall, the shift from bicycles to motorized transport is a strong indicator of wealth of families with access to a wider variety of services and amenities.

8.4.3.2 Education

Three types of indicators are normally used to characterize education in an analysis of household living standards. These include the household members' level of education (literacy rate – with poor households having lower literacy), the availability of educational services (primary and secondary schools) and the use of these services by the members of poor and non-poor households (children's registration in school, drop out rate of children by age and gender and reasons for dropping out, percentage of children who are older than the normal age for their level of education and average spending on education per child registered).

Literacy and schooling are important indicators of the quality of life in their own right, as well as being key determinants of poor people's ability to take advantage of income-earning opportunities. Cambodia (based on 1993/94 CSES data) has achieved a (self-reported) basic literacy rate of 67 percent among adults (older than age 15), implying a high degree of literacy among the poor. However, the literacy gap remains quite large, with literacy ranging from just over half of adults (58 percent) among the poorest 20 percent of the population to 77 percent among the richest. Much larger differentials appear in the distribution of schooling attainment. Years of schooling among adults aged over 15 averages only 3.1 years in the bottom 20 percent of the population, and increases to 5.3 years of schooling among the richest 20 percent. There is a large gender gap, with mean grade attainment being 5.1 years among men compared with 3.2 years among women.

8.4.3.3 Shelter

Shelter refers to the overall framework of personal life of the household. It is evaluated, by poor and non-poor household groups, according to three components (some of which overlap with the indicators mentioned above): housing, services, and the environment. The housing indicators include the type of building (size and type of materials), the means through which one has access to housing (renting or ownership), and household equipment. The service indicators focus on the availability and the use of drinking water, communications services, electricity, and other energy sources. Finally, the environmental indicators concern the level of sanitation (touched on before), the degree of isolation (availability of roads and paths which are usable at all times, length of time and availability of transportation to get to work) and the degree of personal safety.

It is generally established that poor households live in more precarious, less sanitary environments, which contribute to the poorer health and lower productivity of household members.

8.5 Understanding the quantitative determinants of poverty: An example from Cambodia

Some of the characteristics of the poor in Cambodia are apparent from the poverty estimates for various sub-populations reported above (e.g., they are more likely to live in large households, in the countryside, with the household head having minimal education). But the degree to which relevant characteristics vary, on average, between the poor and the non-poor is not always apparent from the sub-population poverty rates, and there is also a wider set of characteristics that may potentially be used as indicators of poverty. In what follows we focus on one characteristic of the poor that has not been examined by most poverty profiles, which is the structure of their consumption.

One of the characteristics of the poor that is most important to understand is the structure of their consumption and how it differs from the non-poor. This characteristic can provide a link between general economic policies and poverty reduction strategies. For example, by identifying the items consumed mainly by poor households, it may be possible to design indirect tax reforms that avoid undue increases in the cost of living for the poor. Such preventative design is important where countries lack comprehensive income tax and transfer systems that can be used to compensate for the burden of indirect taxes and government charges. Another long-term benefit of identifying the items consumed mainly by the poor is that it can help to give research and investment strategies a pro-poor bias. For example, investments in production or processing technology for foods whose consumption is most heavily concentrated amongst the poor will tend to reduce disparities in living standards.

One simple way of finding out which items are relatively more important to the consumption of the poor is to compare the average budget shares of poor and non-poor households; this is done for 1999 for Cambodia in Table 8.1. One may also develop a measure that is both scale neutral and allows for equity weights (for instance, zero weights to consumption by households above the poverty line). The distributional characteristic of a good is defined by Newbery (1995) as:

$$(8.1) \quad d_i \equiv \frac{\sum_h \beta^h q_i^h}{\bar{\beta} Q_i} ,$$

which is comprised of four components: a) the social weight β^h is the social marginal utility of transferring one riel to household h , b) q_i^h is consumption of the i -th good by household h , c) $\bar{\beta}$ is the average of the social weights over all of the households, and d) Q_i is the aggregate consumption of good.

The more heavily the consumption of good i is concentrated on the socially deserving (i.e., those with high social marginal values of consumption, β^h) the higher will be the distributional characteristic. Commonly used social weights are based on a utilitarian framework using a constant elasticity of substitution social welfare function:

$$(8.2) \quad \begin{aligned} W &= \frac{(c^h)^{1-\nu}}{1-\nu} & \nu \geq 0, \nu \neq 1 \\ W &= \ln(c^h) & \nu = 1, \end{aligned}$$

where W is the value that is placed on the income distribution, c^h is the consumption level for household h , and ν is the coefficient of inequality aversion. With this social welfare function, the social weight applied to consumption by household h is given by the social marginal utility of consumption, $\beta^h = (c^h)^{-\nu}$. Different values of the coefficient ν reflect different judgments about the desirability of making transfers to reduce income inequality. Because of this element of value judgement, a range of values for ν are commonly used to see whether conclusions are robust to particular ethical judgements.

One way to view the value for ν is to consider judgments about the effect of taking 100 riels from a richer household, giving some of this to a poorer household and destroying the rest (e.g., due to efficiency losses). If one household has twice the income of the other, then when $\nu=0$ the judge would approve of this transfer only if the poorer household received all 100 riels. But it can be shown that when ν takes the values of 0.5, 1, and 2, then the amount that the poorer household receives needs to be only 71, 50, and 25 riels in order for the resulting distribution to give the same level of social welfare as before the transfer. Hence, in the current context, when the inequality aversion parameter is, say, $\nu=2$, taking one extra riel in indirect tax from a poor household is judged to cause four times as much social cost as does taking one extra riel in tax from a household with twice the consumption level.

The difference in average budget shares between poor and non-poor households, and the distributional characteristic, were calculated for each of the 36 items with consumption data available from round 2 of CSES 1999 (see Gibson 1999). This level of commodity detail is clearly more aggregated than the fine detail with which taxes (including those on trade such as tariffs) are usually defined, but it is still a useful starting point in the search for items for which targeting of effective tax rates in order to alleviate poverty is appropriate. The social weights for the distributional characteristic were based on the CES utility function (defined over the real value of per capita consumption), with two values of the inequality aversion parameter used ($\nu=0.5$ and $\nu=2$). These values cover the usual range used in similar studies in other countries.

The ranking of items according to the difference in average budget shares between poor and non-poor households is reported in table 8.1. The item with the largest difference is rice, which makes up one-quarter of the total budget for poor households (including imputed values for consumption out of gifts and self-production), but only 16 percent of the budget for non-poor households. Other items which take a larger share in the budgets of the poor than the rich include fish, fresh vegetables and wood fuel. At the other extreme, items that are more important in the budgets of the non-poor are house rent and repairs (including imputed rent for owner-occupiers), medical care, and transport and communication.

The distributional characteristic, at both low and high levels of inequality aversion, shows that rice is the item whose consumption is most heavily skewed towards poor households (columns 4-7 in table 8.1). Other items where consumption is concentrated on the poor include tubers, sugar and salt,

spices and seasonings, fresh vegetables, wood fuel, and nuts and seeds (only at low inequality aversion). Several of these items (e.g., tubers) have only a small overall share of household consumption, and so in the aggregate; variation in the tax rates for these items may not have large revenue implications.

It is notable that both the distributional characteristic and the difference in budget shares between poor and non-poor show that items that are commonly considered as ‘merit goods’ (i.e., items whose consumption should be promoted on social grounds) have consumption that is skewed away from the poor. In terms of having consumption concentrated on the poor, medical care ranks only 31st and education spending only 33rd out of 36 goods, according to the distributional characteristic calculated when inequality aversion is high. Similarly, medical care is the item where the average budget share for the poor is second lowest, relative to the budget share for the non-poor, while for education it is seventh lowest. These results tend to suggest that any subsidies in health and education will tend to be captured predominantly by the non-poor. Thus it would be very useful for any public expenditure review carried out in the context of poverty reduction strategies to include a detailed incidence analysis to see exactly what type of households (i.e., poor or non-poor) are benefiting from any subsidies.

Table 8.1: Disaggregated Consumption Characteristics of the Poor (selected items)

Commodity	Average value of household consumption ^a	Budget share of poor versus non-poor	Distributional characteristic			
			Low inequality aversion ($\nu=0.5$)		High inequality aversion ($\nu=2.0$)	
			d_i	Rank ^b	d_i	Rank ^b
	Riel/day	% points				
Rice (all varieties)	2492	9.17	1.010	1	1.015	1
Fish (fresh, fermented, salted, canned etc)	1411	1.14	0.965	9	0.830	9
Fresh vegetables (trakun, cabbage etc)	680	1.06	0.975	5	0.870	5
Wood fuel (firewood, charcoal)	459	0.61	0.971	7	0.847	6
Poultry (chicken, duck and other)	507	0.47	0.954	14	0.819	11
Tobacco products	348	0.37	0.956	13	0.841	8
Sugar and salt	222	0.30	0.982	3	0.895	2
Tubers (sweet potato, cassava, traov, etc)	224	0.25	0.990	2	0.876	3
Spices and seasonings	277	0.25	0.975	6	0.874	4
Fruits (banana, orange, mango etc)	438	0.24	0.945	15	0.795	13
Eggs (duck, chicken, fermented, etc)	222	0.18	0.968	8	0.841	7
Nuts and seeds (coconut, peanut, etc)	193	0.10	0.977	4	0.826	10
Oils and fats	200	0.09	0.958	11	0.817	12
Alcoholic beverages	240	-0.04	0.907	22	0.709	20
Other cereals (bread, noodles, biscuit etc)	413	-0.07	0.929	16	0.754	17
Furniture & household equipment	115	-0.12	0.896	25	0.684	23
Pulses and legumes (gram, cowpea etc)	162	-0.12	0.963	10	0.789	14
Recreation and entertainment	46	-0.17	0.706	35	0.349	36
Personal care (soap, toothpaste etc)	237	-0.27	0.923	20	0.754	16
Fuel and power (kerosene etc, excl. wood)	189	-0.29	0.826	30	0.573	30
Tea, coffee, cocoa	116	-0.32	0.897	24	0.644	26
Clothing and footwear	451	-0.33	0.929	17	0.751	18
Education (school fees, textbooks etc)	323	-0.65	0.770	33	0.461	33
Personal effects (jewelry, handbags etc)	131	-0.66	0.807	31	0.472	32
Miscellaneous non-food items	640	-0.69	0.891	26	0.687	22
Food taken away from home	449	-0.89	0.904	23	0.671	24
Transport & communication	455	-1.57	0.743	34	0.421	34

Medical care (doctors fees, drugs etc)	818	-2.82	0.792	32	0.498	31
House rent (incl. imputed) and repairs	1784	-2.85	0.698	36	0.417	35

Notes: Results are calculated from the 3000 households in Round 2 of CSES 1999

^a In 1999 Phnom Penh prices, with the food poverty line used as the spatial deflator.

^b The item ranked “1” has consumption most heavily biased towards poor households, the item ranked “36” has consumption least heavily concentrated on poor households.

Source: Gibson (1999).

8.6 Understanding the qualitative determinants of poverty: An example from Cambodia

Qualitative indicators can only have values of “yes” or “no” at the level of an individual household. Because there is no need to measure quantities, qualitative indicators may be especially valuable for rapid appraisal methods that attempt to assess poverty status on the basis of a few simple questions. The percentages of three groups – below the food poverty line, those between the food poverty line and poverty line, and the non-poor – having the characteristics examined are reported in table 8.2 for Cambodia in 1999.

Only one of the two characteristics of household heads examined – whether the head can speak a foreign language fluently – is able to discriminate amongst the three poverty states. Despite the statistically significant differences in percentages across the three groups in foreign language fluency, this indicator is probably not a useful one for identifying the poor because it occurs with such low frequency in the population.

Most of the qualitative employment and income indicators also do not look promising as key questions to ask in rapid appraisals of poverty status. Although there are statistically significant differences amongst the three poverty states in the percentage of households who are not primarily dependant on agricultural incomes, this indicator gives no way of distinguishing the food poor from the non-poor, because of the non-linear relationship. Knowing that the household’s consumption in the past 12 months was the same as normal or higher than normal is also no help in determining poverty status. However, over one-fifth of households who report that their past year’s consumption is lower than normal are below the food poverty line and this percentage is significantly higher than that for the other two poverty states (the percentage is indistinguishable between other poor and the non-poor).

The selected housing characteristics all show statistically significant differences between the poor and non-poor but, with the exception of the use of fuels other than wood as the primary cooking fuel, are not able to distinguish between the different severities of poverty. Despite this discriminating ability of questions about non-wood cooking fuels, their practical importance is likely to be limited by the rarity of the indicator: almost nine out of every ten non-poor households use wood for cooking so many errors would be made if this was the basis made for allocating households into various poverty categories. Perhaps the most useful of the housing characteristics considered are the ones related to the composition of roofs and walls because the indicated conditions are externally verifiable and occur commonly but with a large difference in frequency between poor and non-poor.

Table 8.2: Qualitative Poverty Indicators

Description of indicator	Percent with the indicated characteristic			
	Poor		Non-Poor	Total
	below food poverty line	between food poverty line and poverty line		

Head of household characteristics				
belongs to ethnic minority	3.6 (2.2)	5.7 (2.3)	4.4 (1.3)	3.7 (1.2)
speaks a foreign language fluently	0.3 (0.2)	1.4 (0.6)	5.3 (0.7)	3.8 (0.5)
Employment and Income				
main income other than agriculture	37.3 (6.4)	22.6 (3.2)	39.0 (2.2)	33.1 (2.0)
Consumption expenditure				
Lower than normal	21.2 (4.3)	12.9 (3.0)	10.7 (1.3)	12.2 (1.4)
Same as normal	56.8 (6.3)	62.7 (3.5)	64.7 (2.1)	63.5 (2.1)
Higher than normal	22.0 (4.3)	24.4 (3.0)	24.6 (1.8)	24.3 (1.7)
Housing Characteristics				
main source of lighting is electricity	11.3 (3.5)	4.9 (0.9)	22.4 (1.8)	17.3 (1.5)
main drinking water source is piped, public tap, tanker	6.5 (2.4)	3.7 (1.0)	15.4 (1.5)	11.8 (1.3)
Toilet is located in dwelling	10.2 (3.2)	12.6 (2.8)	23.7 (1.7)	19.8 (1.6)
primary cooking fuel is not wood	0.8 (0.6)	3.6 (1.0)	11.4 (0.9)	8.6 (0.8)
concrete or plywood wall of dwelling	30.9 (6.6)	28.4 (3.0)	49.7 (2.1)	43.0 (2.0)
Not use thatching or mixed thatch for roof	49.1 (5.3)	49.8 (3.4)	65.2 (2.1)	60.1 (1.9)
.....not earth or wood floor	1.5 (0.8)	1.7 (0.6)	14.2 (1.1)	10.1 (0.9)
Assets and Liabilities				
owns the following livestock (rural only):				
Pig	47.0 (5.5)	43.8 (4.1)	46.6 (2.8)	46.0 (2.4)
Buffalo	15.1 (5.0)	14.0 (2.6)	12.5 (2.0)	13.1 (1.8)
Cow	37.8 (5.3)	57.5 (3.9)	46.9 (2.9)	48.6 (2.5)
Poultry	57.6 (6.6)	65.2 (4.1)	63.8 (2.2)	63.5 (2.2)
has a fish pond (rural only)	2.0 (1.2)	3.4 (1.0)	3.4 (0.8)	3.2 (0.6)
has at least one outstanding loan	22.8 (4.7)	18.2 (2.3)	16.5 (1.7)	17.5 (1.6)
owns a radio or cassette recorder	33.3 (3.8)	40.4 (3.4)	51.2 (2.0)	46.9 (1.9)
owns a TV	14.1 (3.0)	14.7 (1.9)	31.7 (1.5)	26.0 (1.3)
owns a bicycle	55.4 (5.0)	62.9 (3.0)	57.8 (1.9)	58.7 (1.7)
owns a boat	1.7 (0.7)	7.8 (1.8)	10.4 (1.8)	8.9 (1.4)
owns a motorbike or scooter	14.8 (3.0)	13.5 (1.9)	28.6 (1.4)	23.8 (1.1)

owns a car or truck	0.0 (0.0)	0.1 (0.0)	2.4 (0.3)	1.6 (0.2)
owns a cart (rural only)	33.0 (5.4)	34.5 (3.6)	35.5 (2.8)	35.0 (2.3)
owns a tractor (rural only)	2.1 (1.7)	2.9 (1.3)	4.0 (1.1)	3.5 (1.0)
Village Characteristics				
has a road	91.5 (3.8)	87.4 (4.0)	91.3 (2.1)	90.5 (2.1)
has a primary school	62.6 (7.5)	52.7 (5.5)	46.1 (3.8)	49.3 (3.5)
has a development project	29.4 (6.5)	44.8 (5.5)	41.2 (3.8)	40.8 (3.4)

Note: Results are from the 3000 households surveyed in round 2 of CSES 1999, with household sampling weights used in the calculation of the percentages. Standard errors in () are corrected for the effect of clustering, stratification and weights.

^a Includes French, English, Chinese and Thai.

Source: Gibson (1999).

Most of the indicators of household assets and durable goods ownership show little power in discriminating between the various poverty states. The ownership rates for pigs, buffalo and poultry (rural households only) show no significant differences among any of the three poverty states, while the proportion of households owning cows above the food poverty line but below the poverty line is significantly higher the cow-owning proportion below the food poverty line.

The ownership rates of bicycles, carts and tractors (the last two for rural households only) show no statistically significant differences amongst any of the three poverty states. The ownership rates of car, motorbikes and scooters, televisions and radio cassette players show statistically significant differences between the poor and non-poor, but no differences between the food poor and the other poor. It is only the ownership rate for boats that shows a statistically significant difference between the food poor and the other poor (but no significant difference between the non-poor and the poor who are above the food poverty line) and this characteristic may be limited by its environmental specificity.

The qualitative village characteristics considered do not seem to be good predictors of poverty status, with no significant differences amongst the three groups in terms of road access and an unexpected inverse relationship between poverty status and access to primary schools (i.e., primary schools are more widely available in the villages of the food poor than of the other poor and the non-poor).

The important implication of this discussion is that there is no quick and easy way, in Cambodia at least, to identify the poor. Indeed this is a relatively common finding; there are no satisfactory short cuts to the measurement of poverty.

8.7 Analyzing the determinants of poverty: Regression techniques

An analysis of the correlates of poverty can be carried if a multi-topic household survey is available, using income and consumption regressions. In these regressions, the logarithm of consumption or income (possibly divided by the poverty line) is typically used as the left-hand variable. Right-hand “explanatory” variables span a large array of possible poverty correlates along the lines of those discussed above. .

The regression estimates show how closely each independent variable is related to the dependent variable (e.g. consumption per capita), holding all other influences constant. There is scope for a wide variety of regressions; for instance the dependent variable could measure child nutrition, or morbidity, or schooling, or other measures of capabilities; the regressions could be used to examine the determinants of employment or labor income; and they can be used to estimate agricultural production functions (which relate production with information on type of crops grown per area, harvest, inputs into agricultural production, and input and output prices). For an accessible discussion and lots of examples, in the context of Vietnam, see Haughton et al., *Health and Wealth in Vietnam* (1999).

A regression makes fuller use of information than do the tabulations of a poverty profile. More formally, let w_i be the normalized indicator divided by the poverty line, so that $w_i = y_i/z$, where z is the poverty line and y_i is (per capita) income or consumption. A unitary value for w_i signifies that the household has its level of income or consumption exactly at the level of the poverty line. Denote by X_i the vector of independent variables. Then a popular form of regression is

$$(8.3) \quad \text{Log } w_i = \gamma X_i + \varepsilon_i$$

Note that w_i is in log form, which is a common way of allowing for the log normality of the variable.

Some researchers prefer to use, on the left hand side, a binary variable that is set equal to 1 if the household is poor, and to 0 otherwise. Some of the information is lost by doing this, and the resulting logit or probit regression is relatively sensitive to specification errors. However it may be more helpful for targeting analysis, to assess the predictive power of various explanatory variables used for means-testing. A fuller treatment of regression analysis is given in the Advanced Poverty Analysis course.

Table 8.3 shows a simple example of a regression of per capita household expenditure in Côte d'Ivoire. It indicates that education plays a different role in urban and rural areas (where it does not seem to significantly influence consumption), as do different types of assets. In rural areas, infrastructure has substantial predictive power—households located in villages that are nearer to both paved roads and public markets are better off, as are households living in areas with higher wage levels. The results present further questions about the quality of education in rural areas and to the importance of rural infrastructure in helping families grow out of poverty, that could be addressed in putting together a poverty reduction strategy.

Table 8.4. **Determinants of Household Spending Levels in Côte d'Ivoire in 19xx.**

Education level of most educated male	Urban	Rural
Elementary	0.38 (5.3)	0.04 (0.6)
Junior secondary	0.62 (8.6)	0.08 (0.9)
Senior secondary	0.80 (9.6)	0.05 (0.4)
University	0.93 (9.4)	-
Elementary	0.11 (1.7)	0.07 (1.0)
Junior secondary	0.24 (3.1)	0.27 (2.2)

Senior secondary	0.34 (4.1)	-
University	0.52 (4.1)	-
Value of selected household assets		
Home	0.06 (5.3)	-
Business assets	0.04 (3.3)	0.16 (4.9)
Savings	0.08 (4.7)	-
Hectares of agriculture land		
Cocoa trees	-	0.17 (4.3)
Coffee trees	-	0.04 (1.3)
Distance to nearest paved road	-	-0.04 (-2.9)
Distance to nearest market	-	-0.09 (-3.3)
Unskilled wage	-	0.37 (6.4)
<i>Note:</i> t-statistics in brackets.		

Note: Figures in parentheses are t-statistics.

Source: Adapted from Grosh and Munoz 1996, p.169, based on Glewwe 1990.

Several variations of these multiple income regressions can be used to examine the correlates of the income of the poor. Poverty analysis focuses on correlates of income and expenditure at the lower end of the distribution rather than the correlates at the top end. One can then perform different regressions for each quintile, or quartile, of the population. Whether these regressions can be conducted will in part depend on the sample size of the survey. Alternatively, the regression can examine structural differences in parameter estimates for different income or expenditure groups.

When multiple cross-sectional surveys are available, the same regression can be repeated for different years to see how the association of certain correlates with income or consumption varies over time. Variations over time will be reflected in changes in coefficients or parameters. The results of repeated cross-section regressions can also be used to decompose changes in poverty between changes in household characteristics, and changes in the returns to (or impact of) these characteristics (e.g., Baulch et al. 2001; van de Walle 2000; Wodon 2000).

CHAPTER 9

Poverty Reduction Policies

So far we have looked at the concept of poverty and well-being, the various indicators used to measure poverty, the idea of poverty profiles, and the factors that determine poverty. In this section, we consider poverty reduction strategies, as well as touch upon certain policy implications.

9.1 A framework for action

The approach to reducing poverty has evolved over the past 50 years in response to deepening understanding of the complexity of development. Large investment in physical capital and infrastructure were once viewed as the primary means of development. Later the development community realized that improvements in health and education were important not only in their own right but also to promote growth in the incomes of poor people. Following the development experience of East Asia miracle, emphasis was placed on improving economic management and allowing greater play for market forces. Labor-intensive growth, economic openness, investment in infrastructure and providing basic services to the poor in health and education become the main strategies. In the 1990s, governance and institutions moved toward center stage.

The World Bank (2000) now separates its anti-poverty activities into three groups: promoting opportunity, facilitating empowerment, and enhancing security.

9.1.1 Opportunity

The lack of material opportunities such as jobs, credit, and public services, schools and health services is a direct cause of poverty.

There is a strong link between economic growth and the income and non-income dimensions of poverty. The stark cross-country difference in poverty outcomes generally reflect cross-country differences in economic growth over the long run. Poverty trends have tracked growth trends very closely in 1980s and 1990s. According to Chen and Ravallion (2000), on average, growth in the consumption of poorest fifth of the population tracked economic growth one-for-one over this period. In the vast majority of countries that they study, growth led to rising consumption in the poorest fifth of the population, while economic decline led to falling consumption. So pro-growth policies in most cases actually are also pro-poor.

Well-functioning markets are important in generating sustainable growth and expanding opportunity for poor people because poor people rely on formal and informal markets to sell their labor and products, to finance their investments, and to insure against risks. Market-friendly reforms can not only deliver growth, but also deliver benefits to poor people. For example, recent studies have examined the impact of market-friendly policies—such as openness to international trade, low inflation, a moderate-size government, and strong rule of law—on the incomes of poor people in a large cross-country sample. The findings: these policies on average benefit poor people more than others. Case studies of Chile, China, Ghana, Uganda and Vietnam show that

agricultural reforms have helped raise producer prices for small farmers by eliminating marketing boards, changing real exchange rates through broader economic reforms, lowering tariffs and eliminating quotas.

The human, physical, natural, financial and social assets that poor people possess—or have access to—affect their prospects for escaping poverty because these assets can enable poor people to take advantage of opportunities. Expanding the assets of poor people can strengthen their position and their control over their lives. A recent study of irrigation in Vietnam (van de Walle 2000a) has uncovered important complementarities between education and gains from irrigation. Households with higher education levels received higher returns to irrigation.

The core policies and institutions

Creating more opportunities involves complementary actions to stimulate overall growth, make markets work for poor people, and build their assets including addressing inequalities in the distribution of endowments such as education:

- Encouraging effective private investment is essential because investment and technological innovation are the main drivers of growth in jobs and labor incomes. Fostering private investment requires reducing risk for private investors—through stable fiscal and monetary policy, stable investment regimes, sound financial systems, and a transparent business environment. Certainly, the rule of law and anti-corruption measures are also important. Special measures are frequently required to ensure that micro enterprises and small businesses can participate effectively in markets that are more vulnerable, yet employ a large number of poor people. For example, ensuring access to credit, lowering transaction costs of reaching export markets, and reducing restrictions on the informal sector will all help creating a sound business environment for poor households and small firms. Public investment in expanding infrastructure and communications, and upgrading the skills of the labor force have to complement private investment to enhance competitiveness and create new market opportunities.
- Opening to international markets offers a huge opportunity for job and income growth as long as countries have the infrastructure and institutions to stimulate a strong supply response. Therefore, the opening needs to be well designed with special attention to bottlenecks.
- Building human, physical, natural and financial assets that poor people own or can use requires actions on three fronts. First, increase public spending on basic social and economic services. Second, reform public delivery or privatize those services in order to ensure good quality service delivery. Third, have the poor communities and households participate in planning and monitoring to keep providers accountable.
- Addressing asset inequalities across gender, ethnic, racial and social divides. Special action is required to tackle socially-based inequality such as concentrated farm land ownership in rural communities, under-schooling of girls relative to boys, and the limited independence of women due to lack of access to productive means.
- Getting infrastructure and knowledge to poor areas. Special action is also needed in order to improve the social and economic infrastructure in poor and remote areas, which to a great extent also contribute the poverty problem. Similarly, basic urban services should be provided to city slums so that urban poor people may have chance to participate more actively in over growth.

9.1.2 Empowerment

Lack of representation in the process of policy-making due to social and institutional barriers has impeded poor people's access to market opportunities and to public sector services.

State institutions must be responsive and accountable to poor people. In nearly every country the public sector often pursues activities that are biased against poor people and poor people have trouble getting prompt, efficient service from the public administration. Meanwhile, due to weak popular monitoring institutions, public spending may not reach its targeted poor people, but poor people may bear the burden of corruption disproportionately. Poor people's participation in policy making and performance monitoring at both state and community levels is crucial.

Some social norms and practices help generate and perpetuate poverty. Discrimination on the basis of gender, ethnicity, race, religion, or social status can lead to social exclusion and creates barriers to upward mobility, constraining people's ability to participate in economic opportunities and to benefit from and contribute to economic growth. For example, one cross-country study (Klasen, 1999) indicates that countries that invest in girls' education have higher rate of economic growth. The second half of the 20th century also witnessed devastating ethnic conflicts due to ethnic cleavages and lack of institutions of conflict mediation. On the other hand, good social institutions—kinship, community organizations, and informal networks—can play an important role in poverty reduction. For example, many development programs succeed because they mobilize local groups of project beneficiaries in program implementation.

The key policies

To empower poor people, policies need to facilitate active collaboration among poor and other groups in society include strengthening the participation of poor people in political processes and local decision-making and making changes in governance that make public administration, legal institutions, and public service delivery more efficient and accountable to all citizens, and removing the social barriers that result from distinctions of gender, ethnicity, race and social status:

- Laying political and legal basis for inclusive development including institutionalizing transparent, democratic and participatory mechanisms for making decision and monitoring implementation and promoting legal assistance to poor people who usually have little resource and information to access legal system.
- Strengthening monitoring mechanisms of public administrations' performance by providing access to budgetary information and participatory mechanisms.
- Promoting decentralization and community development to enhance the control that poor people and their communities have over the services to which they are entitled. Decentralization needs to be combined with effective participation and monitoring mechanisms.
- Promoting gender equality by promoting women's representation in decision-making and providing special assistance for women's productive activities.
- Tackling social structures and institutions that are obstacles to the upward mobility of poor people by fostering debate over exclusionary practices and supporting the socially excluded participating into political processes.
- Supporting poor people's social capital by assisting networks of poor people to engage with market and non-market institutions so as to strengthen their influence over policy.

9.1.3 Security

Poor people are exposed to a wide array of risks that make them vulnerable to income shocks and losses of well-being. Reducing poor people's vulnerability to ill health, economic shocks, natural disasters, and violence will enhance well-being on its own and encourage investment in human capital and in higher-risk, higher-return activities as well.

Households and communities respond to their risk exposure through diversification of assets and sources of income and various types of self-insurance and networks of mutual insurance mechanisms. These means help to reduce the risk or soften the impact, but the effect is limited. To counter the incentive and information problems that exclude poor people from many market-based insurance mechanisms, the state has a special role in providing or regulating insurance and setting up safety nets. Health, environmental, labor market, and macroeconomic policies can all reduce and mitigate risk.

Large adverse shock—economic crises and natural disasters—cause poor people to suffer not only in the short run. They undercut the ability of the poor to move out of poverty in the long run as well, by depleting their human and physical assets, which may be irreversible. So it is crucial to prevent economic crises and natural disasters, as well to protect poor people when they occur.

The main policy tools

National programs to manage economy-wide shocks and effective mechanisms to reduce the risks faced by poor people, as well as helping them cope with adverse shocks when they occur, are necessary:

- Formulating programs to helping poor people managing risk. Microinsurance programs, public work programs and food transfer programs may be mixed with other mechanisms to deliver more effective risk management.
- Developing national programs to prevent and respond to macro shocks—financial or natural.
- Design national systems of social risk management that are also pro-growth
- Supporting minority rights and providing the institutional basis for peaceful conflict resolution can help prevent civil conflict and mobilize more resource into productive activities.
- Tackling health problems including widespread illnesses such as malaria and tuberculosis, as well as moderately common but serious conditions such as HIV/AIDS.

There is no simple, universal blueprint for implementing this strategy. Each developing country needs to prepare its own mix of policies to reduce poverty, reflecting national priorities and local realities. Given the important complementarities among these three dimensions, an effective poverty reduction strategy will require action on all three fronts, by all stakeholders in society—government, civil society, the private sector and poor people themselves.

9.2 Practice and good examples

Any good poverty reduction plan has to be based on a comprehensive poverty analysis that identifies the nature and evolution of poverty, the profile of poor people, and all contributing factors of poverty. Building on an accurate understanding of poverty, the strategy for poverty reduction has to prioritize the poverty reduction goals and take into account complementarities and compatibilities of various policy tools. Then specific implementation modules including resource allocation and monitoring mechanisms have to be designed. The participation of poor people at the last two stages is essential since overall they are the main actors in the fight against poverty.

9.2.1 Poverty Analysis: nature and evolution, profile, and causes

Poverty has many dimensions. Material deprivation in term of low consumption or income is certainly the most important welfare indicator, other dimensions such as low achievements in education and health, vulnerability to risk, and powerlessness should be considered as well.

Within each country, poverty can vary and move in different directions for different sub-population defined by region, area, income level, gender, ethnicity, and caste. Households may move in

and out of poverty very frequently due to their exposure to risk and their capability in managing and coping with risk. Identifying who is poor or vulnerable is required for efficient policy targeting and provides vital information for the analysis of the causes of poverty .

Besides a poverty profile and a quantitative analysis of poverty determinants, another route for investigating the causes of poverty is to examine the factors highlighted by poor people – a participatory poverty assessment. For a good example, applied to Vietnam, see World Bank (2000).

9.2.3 Monitoring and evaluation of poverty reduction strategy

A poverty reduction strategy has to make specific goals for actions within a certain time period and with a budget allocation plan. Usually it involves building capacities in data collection and data analysis in order to keep track of progress and problems. A more important aspect of this capacity is poor people's access to this information and mechanisms for them to voice their concerns. The participatory process needs to be institutionalized and community organizations need to be included for better representation.

9.2.4 Examples of poverty reduction strategies

9.2.4.1 Brazil

Brazil saw significant improvements in its social indicators in the 1990s. Net enrollment in primary education increased from 88.2 percent in 1992 to 97.1 percent in 1997. Infant mortality fell from 62 per 1,000 live births in the mid-1980s to 38 in the mid-1990s.. New programs guarantee minimum per capita spending for basic health care and minimum per student spending in primary schools. Innovative actions to get children into school includes the Bolsa Escola, which gives poor families grants if their children go to school.

Despite the advances, the inequalities in health and education remain great, with the poorest fifth of the population having three years of education, and the top fifth more than nine years. The income-poor still leave school with skills inadequate for a middle-income country integrated into the global economy. And reducing income poverty has proved difficult. Indeed, in the unstable macroeconomic environment of the 1980s and early 1990s, poverty rose. Two recent events confirm that the groups most vulnerable to economic insecurity are those with the highest incidence of poverty. Drought in the Northeast hit poor rural workers severely, and the ripple effects of the East Asian crisis, though more benign than expected, reduced the income of workers with the least education.

Some illustrative priorities for action: job growth through productive investment and prudent macroeconomic management is clearly central to increasing income opportunities. But unless structural inequalities are tackled effectively, the gains for poor people will be modest. To reduce structural inequalities, a large land reform program is under way, and there have been promising experiments in negotiated land reform in the Northeast. In the ongoing education effort the next steps will probably require even broader, deeper, and more participatory reforms—many of these are now under discussion.

9.2.4.2 China

China stands out for the extraordinarily rapid decline in income poverty in the 1980s and its high levels of education and health. But it has also had significant increases in inequality—between town and country, and between the coastal areas and inland China, with the poor, semiarid areas of inland China participating little in growth.

The formal structures of security are in transition, and there are deep concerns about the less dynamic parts of urban China, which are experiencing the beginnings of a major shakeout in state enterprise and government employment. The formal provision of security was always weaker in rural areas, but micro evidence suggests that village mechanisms continue to provide high levels of food security through land allocation processes—confirmed as politically popular in villages by democratic votes. Ensuring the voice of the new poor in urban areas and those left behind in inland China will be important in guiding action.

Three areas for action are illustrative. First, maintaining rapid growth through high non-state investment is crucial if there is to be a smooth process of job destruction in inefficient state activities and smooth reform of the social protection arrangements for state employees. If there is a sustained slowdown, insecurity in areas dependent on now inefficient state production could be acute. Second, the smooth integration of China into the global trading system will be key in locking in reforms and ensuring economic stability and steady job growth. But if new opportunities are not to widen disparities, this will have to be accompanied by greater emphasis on building the assets of people living in the poorer areas. Third, continuing area-based integrated rural development activities in poor areas of inland China and, more generally, balancing investment across geographic areas need to be important parts of any overall strategy.

9.2.4.3 Uganda

Having emerged from a period of destructive conflict a decade and a half ago, Uganda still suffers deep poverty in many dimensions. But it also shows what an immensely poor Sub-Saharan African country can achieve.

The first country to receive enhanced debt relief on the basis of its poverty reduction strategy, Uganda stands out for its steady growth in the 1990s. It also stands out for significant reductions in income poverty, impressive efforts toward universal primary education, and a major effort to ensure transparent, poverty-focused budgets, both centrally and locally. One of its main vulnerabilities is health. HIV/AIDS hit Uganda early: a tenth of adults are now infected, and AIDS orphans are straining traditional systems of fostering children.

Three areas are priorities for future action. Consolidating and deepening the accountability and participation in resource allocation and strengthening central and local state institutions, to provide a basis for sound local investment programs in social and physical capital. Tackling the perceived risks in the business environment so that job-creating growth can take off. And furthering current efforts to stop the spread of HIV/AIDS and such diseases as tuberculosis. These efforts would be easier if Uganda were to pull back from its costly and ill-conceived military adventurism in nearby countries.

9.2.4.4 India

India suffers severe deprivations in education and health—especially in the North, where caste, class, and gender inequities are particularly strong. In studies in Bihar and Uttar Pradesh, poor women and men emphasized their extreme vulnerability and the ineffectiveness of state institutions, from schools to police.

In the past, poverty reduction in India lagged behind that of East Asia because of slower growth and significantly less progress in promoting mass education and basic health. More recently, however, growth has accelerated and there is a growing consensus that poverty has fallen substantially.

There are also marked differences within India—with the South, particularly the state of Kerala, having sharply better education and health. Kerala has life expectancies greater than those in Washington,

D.C., despite vastly lower income levels. The effectiveness of public action in Kerala has been attributed to its strong tradition of political and social mobilization.

What are the priorities for action in India? Accelerated poverty reduction requires continued solid economic growth, which in turn demands further liberalization, especially in agriculture, and better provision of infrastructure, sorely lacking in most of India. In areas with deep deprivation in health and education, the development of social infrastructure is critical. Expanding education and health services requires that state governments reverse the deterioration in their fiscal positions, as subsidies to the loss-making power sector crowd out spending in the social sectors. The higher spending needs to be matched by better service provision. This requires deep improvements in governance, often weakest in India's poorest regions, and in combating teacher absenteeism. Also needed is more equitable service provision, which requires empowering women and members of lower castes.

9.2.4.5 Tanzania

Since independence, the government has been focused on three development problems: ignorance, disease and poverty. National efforts to tackle these problems were initially channeled through centrally directed, medium-term and long-term development plans; despite high levels of foreign aid, these efforts were a complete failure, and poverty was higher in 1990 than at the time of independence. Despite sustained effort over the past decade, and an acceleration in economic growth, half of the population today is considered basically poor, and approximately one third lives in extreme poverty.

The current poverty reduction strategy takes a decentralized approach, based on broad consultation with all stakeholders. The Zonal Workshops, which included a large number of representatives of the poor at the village level, identified priority concerns and requirements for poverty reduction. Through an assessment of the poverty profiles and trends using 6 household surveys of various purposes during the past two decades, the key findings include: a very high incidence of poverty (48% in 1991/92); poverty is more widespread in rural areas than in urban centers; the poor are concentrated in subsistence agriculture; urban poverty is also acute in urban areas other than Dar es Salaam, and poor people are mostly in the informal sector; youth, the old and large households are more likely to be poor; the incidence of poverty declined between 1983-1991 and 1991-1993, but rose during 1993-1998; infant mortality rate remains to be high and AIDS is the leading killer of youths in Dar es Salaam and several other sampled districts; malnutrition continues to be severe; access to clean drinking water is still limited and the majority of poor people have no access to piped water and, as a result, have much higher exposure to cholera and waterborne diseases; the judiciary system has corruption problems.

Two participatory poverty assessment studies have suggested the following causes of poverty and obstacles to development: insecure land tenure, lack of access to agricultural inputs, credit, technology, transportation, markets, and quality health services, vulnerability to shocks, gender inequality. These factors were reiterated in the Zonal Workshops. The information then was summarized for each dimension of poverty and further discussed with other stakeholders.

The poverty reduction strategy aims at (1) reducing income poverty; (2) improving human capabilities, survival and social well-being; and (3) containing the extreme vulnerability among the poor. The key instrument is high and equitable growth. Specifically, over the next three years, the annual GDP growth rate is targeted to accelerate from 5.2% to 6% - a somewhat unrealistic goal. Agriculture should increase at 5% on average, compared to 3.6% during 1990-1998. The growth of the industrial and service sectors is expected to be higher than that of agriculture. At these growth rates, the incidence of poverty is expected to drop to 42% by 2003 and 24% by 2010. In pursuit of the above poverty reduction objectives, reforms aimed at bolstering market efficiency and factor productivity will be implemented. Besides

maintaining low inflation and a small fiscal imbalance, special budgetary attention is to be given to rural, export, and private sectors development programs.

Specific goals are also identified for other dimensions of poverty such as education and health, and especially HIV/AIDS related issues. Twenty seven indicators for the monitoring and evaluation of the poverty reduction strategy have been selected and at least two observations will be made during the next three years. These key welfare indicators will be at the heart of further discussion and assessment of the poverty reduction strategy.

Appendix I: Data Introduction

In this course, we will be working extensively with STATA using a subset of information from the Household Survey 1998/99, conducted jointly by Bangladesh Institute of Development Studies (BIDS) and the World Bank. The information was collected at individual, household, and community levels. A description of the data sets and file structure used for these exercises is given below.

File Structure

We will use and generate a lot of files. There are mainly three types of STATA files. Some contain data sets (identified by the suffix `.dta`), others contain STATA programs (identified by the suffix `.do`), and yet others contain a record and output of the work we do in STATA (identified by the suffix `.log`). To keep these files organized, it is useful to create a well-structured set of directories. The following directory structure has been created on each computer:

```
c:\intropov
c:\intropov\data
c:\intropov\dofiles
c:\intropov\logfiles
```

Currently under `c:\intropov\data`, there are four data files:

1. `hh.dta` includes 20 household level variables such as household location, household business type, asset ownership, access to service, etc.
2. `ind.dta` includes 10 individual level variables such as age, gender, education or schooling, marital status, main activity, working status, occupation, sector, relation to household head, etc.
3. `consume.dta` includes 30 categories of expenditure such as various food and non-food items, rent or rental value of housing, etc.
4. `vprice.dta` includes village level price information on the main food items.

Data Description

hh.dta

hhcode: household identification number

thana: thana code (a *thana* is an administrative center comprising a number of villages). It ranges from 1 to 32 as there are 32 thanas.

vill:	village code (when combined with thana it uniquely identifies a village). It ranges from 1 to 4 as a maximum four villages are selected from a thana.
region:	region code 1 Dhaka (the capital) 2 Chittagong 3 Khulna 4 Rajshahi
weight	sampling weight for household
distance	distance to nearest paved road (km)
d_bank	distance to nearest commercial/agricultural bank (km)
toilet	type of latrine used in the household 1 sanitary 0 non-sanitary
hhelec	if household has electricity 1 yes 0 no
hassetg	household total assets (in <i>taka</i>)
famsize	household size
sexhead	gender of household head 1 male 0 female
agehead	age of household head (years)
educhead	years of schooling of household head
hhlandd	land (in <i>decimals</i> which is one-hundredth of an acre) owned by household
<u>ind.dta</u>	
pid	household member identification number (unique for a household member, so becomes unique in the sample after being combined with household id)
indsave	individual savings (in <i>taka</i>)
snaghr	non-farm self employment working hours per month

sagrhr	farm self employment working hours per month
wnagrhr	non-farm wage job working hours per month
wagrhr	farm wage job working hours per month
iempgrhr	total working hours per month
rel_hh	code for relation to household head
	1 Head himself/herself
	2 Wife/husband
	3 Son/daughter
	4 Grandson/granddaughter
	5 Father/mother
	6 Sister/brother
	7 Niece/nephew
	8 Son-in-law/daughter-in-law
	9 Spouse of brother or sister
	10 Brother or sister of spouse
	11 Father-in-law/mother-in-law
	12 Other relatives of head or spouse
	13 Servant/maid servant
	14 Other _____ (specify)

educ	years of schooling completed
sex	gender
age	age (in years)

consume.dta

Ten items have been selected from the survey: rice, wheat, pulses, milk, oil, meat, fish, vegetables, fruits, sugar. Let X denote any items, so:

qX	quantity (kg) of item X consumed last week
eX	value of item X consumed last week (in <i>taka</i>)
expfd	household total food consumption per month (in <i>taka</i>)
expnfd	household total expenditure on regular non-food items per month (in <i>taka</i>)
hhexp	household total expenditure per month (in <i>taka</i>)

vprice.dta

Eleven price items (vegetables in the above now has two entries: potatoes and other vegetables) were selected from the survey. Again, denote an item by X :

pX	village price per kg
------	----------------------

Appendix II: STATA Preliminary

STATA is a statistical software package that offers a large number of statistical and econometric estimation procedures. With STATA we can easily manage data and apply standard statistical and econometric methods such as regression analysis and limited dependent variable analysis to cross-sectional or longitudinal data.

1. Getting Started

1.1. Starting STATA

Start a STATA session by double-clicking on the STATA icon in your desktop. The STATA computing environment comprises four main windows. The size and shape of these windows may be moved about on the screen. Their general look and description are shown below:

Review	Stata Results
Lists previous commands entered in the STATA Command window during current session	Displays each command and results or feedback generated by the command
Variables	
Lists all variables of the currently open dataset	
	Stata Command
	STATA commands are entered here. It shows a blinking command prompt

It is useful to have the **Stata Results** window be the largest so you can see a lot of information about your commands and output on the screen. In addition to these windows STATA environment has a menu and a toolbar at the top (to perform STATA operations) and a directory status bar at the bottom (that shows the current directory). You can use menu and

toolbar to issue different STATA commands (like opening and saving data files), although most of the time it is more convenient to use the **Stata Command** window to perform those tasks. If you are creating a log file (see below for more details), the contents can also be displayed on the screen; this is sometimes useful if one needs to back up to see earlier results from the current session.

1.2. Opening a Dataset

You open a STATA dataset by entering following command in the **Stata Command** window:

```
use hh
```

(or possibly use `c:\intropov\data\hh`, depending on where STATA has opened up; see below for more details). STATA responds by displaying the following in the **Stata Results** window:

```
. use hh
.
```

The first line repeats the command you enter and the second line with no error message implies that the command has been executed successfully. From now on, we will show only the **Stata Results** window to demonstrate STATA commands. The following points should be noted:

- STATA assumes the file to be in STATA format with an extension `.dta`. So, typing `hh` is the same as typing `hh.dta`.
- We can only have one data set open at a time in STATA. So if we open another data set `ind.dta`, we will be replacing `hh.dta` with `ind.dta`.
- The above command assumes that the file `hh.dta` is in the current directory (shown by the directory status bar at the bottom). If that is not the case then you can do one of the following two things (assuming current directory is `c:\stata\data` and the file `hh.dta` is in `c:\intropov`):

- Type the full path of the data file:

```
. use c:\intropov\hh
.
```


- Make `c:\intropov` as the current directory and then open the file as before:

```
. cd c:\intropov
. use hh
.
```

- If the memory allocated to STATA (which is 1,000K or 1M by default) is too little for the data file to be opened we will see an error message like the following:

```
. use hh
no room to add more observations
r(901);
.
```

The third line displays the code associated with the error message. All error messages in STATA have associated codes like this; further explanations are available in the STATA Reference Manuals. In this case we have to allocate more memory to STATA. The following commands allocate 10M to STATA and then tries again to open the file:

```
. set memory 10m
(10240k)
. use hh
.
```

Since the file opens successfully, allocated memory is sufficient. If you continue to get an error message, you can try 25m or 30m. But be careful not to specify too much memory for STATA. If you specify too much, the computer will use virtual memory that will actually slow down your computer. Another important thing to remember is that the memory allocation command (`set memory 10m` for example) works only if no data set is open. Otherwise you will get following error message:

```
. use hh
. set memory 10m
no; data in memory would be lost
r(4);
```

You can clear the memory using one of the two commands: `clear` or

drop _all. The following demonstration shows the first command:

```
. use hh
. set memory 10m
no; data in memory would be lost
r(4);
. clear
. set memory 10m
.
```

1.3 . Saving a Data set

If you make changes in an open STATA data file and want to save those changes, you can do that by using the STATA `save` command. For example, the following command saves the `hh.dta` file:

```
. save hh, replace
file hh.dta saved
```

You can optionally omit the filename here (just `save, replace` is good enough). If you do not use the `replace` option STATA does not save the data but issues the following error message:

```
. save hh
file hh.dta already exists
r(602);
```

The `replace` option unambiguously tells STATA to overwrite the pre-existing original version with the new version. If you do NOT want to lose the original version, you have to specify a different filename in the `save` command:

```
. save hh1
file hh1.dta saved
```

Notice that there is no `replace` option here. However, if a file named `hh1.dta` already exists then you have to either use the `replace` option or use a new filename.

1.4. Exiting STATA

An easy way to exit STATA is to issue the command: `exit`. However, if you have an unsaved data set open, STATA will issue the following error message:

```
. exit
no; data in memory would be lost
r(4);
```

To remedy this problem you can save the data file and then issue the `exit` command. If you really want to exit STATA without saving the data file, you can first clear the memory (using `clear` or `drop _all` command as shown before) and issue the `exit` command. You can also simplify the process by combining two commands:

```
. exit, clear
```

1.5. STATA Help

You can access STATA help facilities following ways:

- The reference manuals (the index therein).
- If you know which command or keyword you want the help information about, you can issue the command `help` followed by the command name or keyword. This command only works if you type the full command name or keyword unabbreviated. For example, following will not help:

```
. help mem
help for mem not found
try help contents or search mem
```

But following will work:

```
. help memory

(output omitted)
```

- If you can not recall the full command name or keyword, or you are not sure about which command you want you can use the command `lookup` or `search` followed by the command name or keyword. So following will work:

```
. search mem

(output omitted)
```

This command will list all commands associated with this keyword and display a brief description of each of those commands. Then you can pick what you think is

- relevant and use “help” to obtain the specific reference.
- The STATA website (<http://www.stata.com>) has excellent help facilities, for example, Online Tutorial, Frequently Asked Questions (FAQ), etc.

1.6. Notes on STATA Commands

Here are some general comments about STATA commands:

- STATA commands are typed in lower case.
- All names, including commands or variable names, can be abbreviated as long as there is no ambiguity. So `describe`, `des` or simply `d` do the same job as there is no confusion.
- In addition to typing, some keystrokes can be used to represent a few STATA commands or sequences. The most important of them are the **Page-Up** and **Page-Down** keys. To display the previous command in **Stata Command** window, you can press **Page-Up** key. You can keep doing that until the first command of the session appears. Similarly, the **Page-Down** key displays the command that follows the currently displayed command in the **Stata Command** window.

2. Working with data files: looking at the content

From now on, we will mostly list the command(s) and not the results, to save space. To go through this exercise open the `hh.dta` file, as we will use examples extensively from this data file.

2.1. Listing the variables

To see all variables in the data set, use the `describe` command (fully or abbreviated):

```
. describe
```

This command provides you with information about the data set (name, size, number of observations) and lists all variables (name, storage format, display format, label).

To see just one variable or list of variables use the `describe` command followed by the variable name(s):

```
. d hhcode vill

      storage   display    value
variable name  type    format    label    variable label
-----
```

hhcode	double %7.0f	COMPLETE HOUSEHOLD CODE
vill	float %2.0f	VILLAGE CODE

As you can see, the `describe` command shows also the variable type, its length and a short description of the variable (if available). The following points should be noted:

- You can abbreviate a list of variables by typing only the first and last variable names, appended by a hyphen (-), provided the two variables appearing in the list are in the same order as they are in the data set. For example, to see all variables from `hhcode` up to `famsize`, you could type:

```
. describe hhcode-famsize
```

- Also you can use the wildcard symbol (*) to save yourself some typing. For example, to see all variable that start with "hh", you could type:

```
. describe hh*
```

You can abbreviate a variable or variable list this way in any STATA command (where it makes sense), not just in `describe`.

2.2. Listing data

To see actual data stored in the variables we use the `list` command (abbreviated as `l`). If you type the command `list` all by itself STATA will display values for all variables and all observations, which is not desirable for any practical purpose. We usually want to see the data for certain variables and for certain observations. We achieve this by typing a `list` command with a variable list and with conditions, as shown in the following examples:

The following command lists all variables of the first three observations:

```
. list in 1/3
```

Here STATA displays all observations starting from observation 1 and ending with observation 3. STATA can also display data as a spreadsheet. There are two icons in the toolbar called **Data Editor** and **Data Browser** (fourth and third from right). By clicking one, a new window will pop up and the data will be displayed as a table, with observations as rows and variables as columns. **Data Browser** will only display the data, whereas you can edit data

with **Data Editor**. The commands `edit` and `browse` will also open the spreadsheet window.

The following command lists household size and head's education for households headed by a female who is younger than 45:

```
. list hhid hhsiz headedu if headsex==2 & headage<45
```

The above statement uses two relational operators (`==` and `<`) and one logical operator (`&`). Relational operators impose a condition on one variable, while logical operators combine two or more relational operators. The following list shows the relational and logical operators that are used in STATA:

Relational operators	Logical operators
<code>></code> (greater than)	<code>~</code> (not)
<code><</code> (less than)	<code> </code> (or)
<code>==</code> (equal)	<code>&</code> (and)
<code>>=</code> (greater than or equal)	
<code><=</code> (less than or equal)	
<code>!=</code> or <code>~=</code> (not equal)	

You can use relational and logical operators in any STATA command (where it makes sense), not just in the `list` command.

2.3. Summarizing data

The command `summarize` (abbreviated `sum`) calculates and displays a variety of summary statistics, for example means and standard deviations. If no variable is specified, summary statistics are calculated for all variables in the data set. The following command summarizes the household size and education of the household head:

```
. sum famsize educhead
```

Any observation that has a missing value for the variable(s) being summarized is excluded from this calculation by STATA (missing values are discussed later). With no option, `summarize` provides the mean, standard deviation, minimum and maximum for each variable.

If we want to know the median and percentiles of a variable, we need to add the `detail` option (abbreviated `d`):

```
. sum famsize educhead, d
```

STATA allows the use of weights. The `weight` option is useful if the sampling probability of one observation is different from another. In most household survey, the sampling frame is stratified, where the first primary sampling units (often villages) are sampled, and conditional on the selection of primary sampling unit, secondary sampling units (often households) are drawn. Household surveys generally provide weights to correct for the sampling design differences and sometimes data collection problems. The implementation in STATA is straightforward:

```
. sum famsize educhead [pw=weight]
```

Here the variable `weight` has the information on the weight to be given to each observation and `pw` is a STATA option to incorporate the weight into the calculation. We will discuss the use of weights further in chapter exercises that follow this exercise.

For variables that are strings, `summarize` will not be able to give any descriptive statistics except that the number of observations is zero. Also, for variables that are categorical, it can be difficult to interpret the output of the `summarize` command. In both cases, a full tabulation may be more meaningful, which we will discuss next.

Many times we want to see summary statistics by group of certain variables, not just for the whole data set. Say in the above example, we want to see mean of family size and education of household head by the gender of household head. Of course, we can use a condition in the `sum` command (for example, `sum famsize educhead if sexhead==1 [pw=weight]`), but this is not convenient if the variability of the group variable increases (for example, the variable `region`). In those cases, we use by option of STATA which makes life much simpler. For example, we want to see summary statistics of family size and education of household head for each region. Before we use STATA by option we must make sure that data is sorted by the group variable. You can check this by issuing `describe` command after opening each file. `describe` command, after listing all the variables, informs if the data set is sorted by any variable(s). If there is no sorting information listed or the data set is sorted by variable(s) that is (are) different from what you want it to be, you can issue a `sort` command and then save the data set in this form. Following commands sort the data set by `region` and show summary statistics of family size and education of household head by region:

```
. sort region
```

```
. by region: sum famsize educhead [pw=weight]
```

2.4. Frequency distributions (tabulations)

We often need frequency distributions and cross tabulations. We use the `tabulate` (abbreviated `tab`) command to do this. The following command gives the regional distribution of the households:

```
. tab region
```

The following command gives the gender distribution of household heads in region 1:

```
. tab sexhead if region==1
```

In passing, note the use of the `==` sign here. It indicates that if the regional variable is identically equal to 1, then do the tabulation.

We can use the `tabulate` command to make a two-way distribution. For example, we would like to check whether there is any gender bias in the education of household heads. We use the following command:

```
. tab educhead sexhead
```

To see percentages by row or columns we can add options to the `tabulate` command:

```
. tab region sexhead, col row
```

2.5. Distributions of descriptive statistics (table command)

Another convenient command is `table` which combines features of `summ` and `tab` commands and `by` option. In addition, it displays the results in a more presentable form. Following table command show mean of family size and education of household head by region:

```
. table region, c(mean famsize mean educhead)
```

```
-----
```


region	mean(famsize)	mean(educhead)
Dhaka	5.23	2
Chittagon	5.82	3
Khulna	5.03	3
Ragfhahi	5.03	2

Results are as we expected. But why is mean of educhead displayed as integer, not fraction? This is because educhead variable is stored as an integer number and STATA simply truncated numbers after the decimal. Look at the description of this variable.

```
. d educhead

      storage   display   value
variable name  type    format   label      variable label
educhead      float   %2.0f    label      Education (years)
of HH Head
```

You see that educhead is a float variable its format (%2.0f) shows that its digits occupy 2 places and it has no digit after decimal. You can force STATA to reformat the display. Say we want it display two place after decimal for three digit display. Following command shows that and subsequent table command:

```
. format educhead %3.2f

. table region, c(mean famsize mean educhead)

-----
      region | mean(famsize) mean(educhead)
-----+-----
      Dhaka |          5.23          2.09
Chittagon |          5.82          3.14
      Khulna |          5.03          2.91
      Ragfhahi |          5.03          2.15
-----
```

That is much better. Please consult STATA manuals for formatting options available for all types of variables. Remember that formatting changes only the display of the variable, not the internal representation of the variable in the memory. `table` command can display up to five statistics and not just mean. Also, what we showed here is a one-way table. But it is possible to display two-way, three-way or even higher dimensional tables. Let us look at an example of a two-way table. Suppose we want to see the education of household head and not just by region but also by sex of household head. Here is how we do that:

```
. table region sexhead, c(mean famsize mean educhead)
```

region	sex head	
	0	1
Dhaka	3.36	5.37
	0.18	2.24
Chittagon	4.17	6.08
	1.50	3.39
Khulna	4.18	5.11
	1.36	3.05
Ragfhahi	3.70	5.13
	0.00	2.31

2.5. Missing Values in STATA

In STATA, a missing value is represented by a dot (.). A missing value is considered larger than any number. The `summarize` command ignores the observations with missing values and the `tabulate` command does the same, unless forced include missing values.

2.6. Counting observations

We use `count` command to count the number of observations in the data set:

```
. count
      519
.
```

The `count` command can be used with conditions. The following command gives the number of households whose head is older than 50:

```
. count if agehead>50
      161
.
```

3. Working with data files: changing data set

3.1. Generating new variables

In STATA the command `generate` (abbreviated `gen`) creates new variables, while the command `replace` changes the values of an existing variable. The following commands create a new variable called `oldhead`, then set its value to 1 if the household head is older than 32 years and to 0 otherwise:

```
. gen oldhead=1 if agehead>32
      (90 missing values generated)
. replace oldhead=0 if agehead<=32
      (90 real changes made)
```

What happens here is, for each observation, the `generate` command checks the condition (whether household head is older than 32) and sets the value of the variable `oldhead` to 1 for that observation if the condition is true, and to missing value otherwise. The `replace` command works in a similar fashion (however, the condition is reversed here and `oldhead` gets 0 if the condition is met). After the `generate` command STATA informs us that 90 observations failed to meet the condition and after the `replace` command STATA informs us that those 90 observations have got new values (0 in this case). The following points should be made:

- STATA variable names cannot be longer than 8 letters. So `generate/replace` commands must conform to this.
- If a `generate` or `replace` command is issued without any conditions, that command applies to all observations in the data file.
- While using the `generate` command, care should be taken to handle missing values properly.
- The right hand side of the `=` sign in the `generate` or `replace` commands can be any expression involving variable names, not just a value.

- The command `replace` does not have to always follow the `generate` command. The `replace` command can be used to change the values of any existing variable, independently of `generate` command.

STATA provides many useful functions to be used in `generate` and `replace` commands, for example `mean(.)` or `max(.)`. For example, in the `ind.dta` file. the following command calculates the maximum share of employment among four sectors for each household:

```
. gen maxhr=max(snaghr , saghr ,wnaghr ,waghr)
```

An extension of the `generate` command is `egen`. Like the `gen` command, the `egen` command can create variables to store descriptive statistics like the mean, sum, maximum and minimum or other statistics. For example, an alternative way to create `maxhr` variable is:

```
. egen maxhr=rmax(snaghr saghr wnaghr waghr)
```

Note the difference in syntax. The more powerful feature of `egen` command is its ability to create statistics involving multiple observations. For example, the following command creates average individual employment hours for the data set:

```
. egen avgempshr=mean(iempshr)
```

All observations in the data set get the same value for `avgempshr`. The following command creates the same statistics, this time for males and females:

```
. egen avghrmf=mean(iempshr) , by(sex)
```

Here observations for males get the value that is average of male employment hours, while observations for females get the equivalent for female employment hours.

3.2. Labeling

3.2.1. Labeling variables

You can attach labels to variables to give a description to them. For example, the variable `oldhead` does not have any label now. You can attach a label to this variable by typing:

```
. label variable oldhead "HH Head is over 32"
```

In the `label` command, `variable` can be shortened to `var`. Now to see the new label, type:

```
. des oldhead
```

3.2.2. Labeling Data

There are other types of labels we can create. To attach a label to the entire data set, which appears at the top of our describe list:

```
. label data "Bangladesh HH Survey 1998/99"
```

To see this label, type:

```
. des
```

3.2.3. Labeling Values of variables

Variables that are categorical, like those in `sexhead` (1=male, 0=female), can have labels that help one to remember what the categories are. For example, if we tabulate the variable `sexhead` we see only 0 and 1 values:

```
. tab sexhead
```

To attach labels to the *values* of a variable, we have to do two things. We have to define a value label. Then we have to assign this label to our variable(s). Using the new categories for `sexhead`:

```
. label define sexlabel 0 "Female" 1 "Male"  
. label values sexhead sexlabel
```

Now, to see the labels, type:

```
. tab sexhead
```

If you want to see the actual values of the variable `sexhead`, which is still 0s and 1s, you can add an option to not display the labels we have assigned the values of the variable:

```
. tab sexhead, nolabel
```

3.3. Keeping and Dropping Variables and Observations

We can select variables and observations of a data set by using the `keep` or `drop` commands. Suppose we have a data set with 6 variables: `var1`, `var2`, ... , `var6`. We would like to keep a file with only three of them, say `var1`, `var2`, and `var3`. You can use either of the following two commands:

```
keep var1 var2 var3 (or keep var1-var3)
drop var4 var5 var6 (or drop var4-var6)
```

Note the use of a hyphen (-) in both commands. For practical purpose you should use the command that involves fewer variables or less typing (and hence less risk of error!). We can also use relational or logical operators. For example, the following command drops those observations where the head of the household is 80 or older :

```
. drop if agehead>=80
```

The following command keeps those observations where household size is 6 or less:

```
. keep if famsize<=6
```

The above two commands drop or keep all variables based on the conditions. You cannot include a variable list in a `drop` or `keep` command that also uses conditions. For example, the following command will fail:

```
. keep hhcode famsize if famsize<=6
invalid syntax
r(198);
```

You have use two commands to do the job:

```
. keep if famsize<=6
. keep hhcode famsize
```

You can also use the keyword `in` in a `drop` or `keep` command. For example, to drop the first 20 observations:

```
. drop in 1/20
```

3.4. Producing Graphs

STATA is quite good at producing basic graphs, although considerable experimentation may be needed to produce really beautiful graphs. The following command shows the distribution of the age of the household head in a bar graph:

```
. graph agehead
```

Whether you use a categorical variable or a continuous variable, say income, does not make a difference to STATA: the default is 5 bars (“bins”). The number of bars may be increased, up to a maximum of 50, by adding the option `bin(#)`.

```
. graph agehead, bin(12)
```

In order to save the graph we would add the option `saving(filename)`. You can also provide titles. For a scatter plot of two variables, type the following command:

```
graph headedu headage, t1(education by age) saving(eduage,  
replace) ylab xlab s(.)
```

3.5. Combining Data sets

3.5.1. Merging data sets

STATA can only have one data set in memory at a time. However, in many occasions you need variables that are spread in two or more files and you would like to combine those files for the purpose of analysis. For example, we want to see how individual's education varies by household head's gender. Since household head's gender (variable `sexhead`) and individual's education (variable `educ`) come from two different files (`hh.dta` and `ind.dta`) we have to merge these two files to do the analysis. We want to combine these two files at household level, so the variable that is used for merging is `hhcode` (let's call it merge variable). Before merging is done, both files must be sorted by merge variable. Following command opens, sorts and saves the `ind.dta` file:

```
. use ind  
. sort hhcode  
. save, replace
```

Once both data sets are sorted you can go ahead with merging. Following steps show how we do this:

```
. use hh  
. merge hhcode using ind
```

In the context `hh.dta` is called the *master file* (this is the one that remains in the memory before merging) and `ind.dta` is called *using file*. To see how the merge operation went, we issue following command:

```
. tab _merge
```

_merge	Freq.	Percent	Cum.
3	2767	100.00	100.00
Total	2767	100.00	

The variable `_merge` is created by STATA after each merge operation and it can have three possible values:

- 1 shows those observations from master file that could not be merged
- 2 shows those observations from using file that could not be merged
- 3 shows those observations that were successfully merged

Total number of observations in the resulting data set is the sum of these three `_merge` frequencies. Possible candidate for `_merge=1` is an observation in `hh.dta` file that has a `hhcode` values which cannot be found in `ind.dta` file. Similarly if the `ind.dta` file has a `hhcode` that is not found in the `hh.dta` file that observation will appear with `_merge=2`. In the above example, however, each household in `hh.dta` file has an exact match in the `ind.dta` file and that is why we got `_merge=3` and not 1's or 2's. If you keep only the matched observations, you can do that by this command: `keep if _merge==3`. Once we have merged the data sets we can go ahead with our analysis:

```
. sort sexhead
. by sexhead: sum educ
```

-> sexhead = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	179	2.329609	3.424591	0	14

-> sexhead = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	2588	2.340417	3.324081	0	16

The result shows that there is not much difference in education by the gender of household head.

Notice that in order to show the results by `sexhead` variable, we have to sort the data by that variable, otherwise we will get an error message. The preference of master and using data is not important and a matter of convenience.

3.5.2. Appending data sets

Consider what would happen in the above merging scenario if we have `_merge=1` and `2` only but no `3`s. This can happen if individual data (`ind.dta`) come from households that are completely different from the households in `hh.dta`. In this case resulting number of observations after merging is the sum of observations in the two files (observations with `_merge=1` + observations with `_merge=2`). STATA in this case would actually append the two data sets; however variables that are only included in one file will have missing values for all observation from the other file.. Although this is not what we intend in the above example, appending is necessary when we need to combine two data sets that have same (or almost same) variables but are mutually exclusive. For example, assume we have four regional version of `hh.dta` file: `hhdhak.dta` (has households only from Dhaka region), `hkhul.dta` (has households only from Khulna region), `hhraj.dta` (has households only from Rajshahi region) and `hhchit.dta` (has households only from Chittagong region). These data files have same variables as does `hh.dta` but represent four distinct sets of households. How do we combine them to get a overall data set of the whole country? We append them. STATA has a command (`append`) to that. Following commands show how we do that.

```
. use hhdhak
. append using hkhul
. append using hhraj
. append using hhchit
```

At this stage we have a data set in the memory that has household information from all four regions. If we need this data set for subsequent use we should save this data set after arranging in a defined order (say, sorting by `hhcode`).

4. Working with .log and .do files

STATA can work interactively. You type one command line each time and STATA will process that command, display the result if any and wait for the next command. It is very flexible. But what if you want to save these results, or perhaps have a printed version of the results? You can do that in STATA by creating `.log` files. A `.log` file is created by issuing a `log using` command and closed by a `log close` command; and all commands issued in between plus corresponding outputs are saved in the `.log` file. Let us go back to the example in Section 3.5.1. Assume that we want to save only the education summary by household gender, not the merging outcomes. Here are commands that we issue:

```
. log using educm.log

. by sexhead:sum educ

-----
-> sexhead = 0

      Variable |      Obs      Mean   Std. Dev.      Min      Max
-----+-----
```

```

      educ |      179      2.329609      3.424591      0      14
-----+-----
-> sexhead = 1

Variable |      Obs      Mean      Std. Dev.      Min      Max
-----+-----
      educ |     2588      2.340417      3.324081      0      16
.log close

```

What happens here is STATA creates a text file named `educm.log` in the current directory and saves the summary output in that file. If you want the `.log` file to be saved in directory other than current directory you can specify full path of the directory in the `.log` creation command.

The `.log` creation command shown above fails if the `.log` file already exists. There are two ways you can handle this in above scenario:

- You can replace the existing `.log` file by issuing the command: `log using educm.log, replace.`
- You can append any new output to the existing file by issuing command: `log using educm.log, append.`

If you really want to keep the existing `.log` file unchanged, then you can rename either this file or the file in the `.log` creation command. In a `.log` file if you want to suppress portion of it, you can issue `log off` before that portion and `log on` command after that. You have to close a log file before opening a new one, otherwise you will get an error message.

If you find yourself using same set of commands repeatedly, you can save those commands in a file and run them together whenever you need it. These command files are called `.do` files and they are extremely useful. Again going back to the example of Section 3.5.1., you can type all commands, starting from "use ind" to "by sexhead: sum educ", in a text file called `educm.do`, save it in the current directory and then issue the command "do educm" from STATA prompt to run it. In order to save the output of all these commands you can also include `.log` creation commands in the file. So the `educm.do` file may look like this:

```

log using educm.log
use ind
sort hhcode
save, replace
use hh
merge hhcode using ind
tab _merge
sort sexhead
by sexhead:sum educ
log close

```

The main advantage of using .do file instead of typing commands line by line is repeatability. Usually if it takes quite some steps to obtain the desired output, you should edit a .do file because you may need to do it tens of times.

There are certain commands that are useful in a .do file. We will discuss them from the following sample .do file:

```
*This is a STATA comment which is not executed
/*****This is a do file that shows some very useful
  commands used in do files. In addition, it creates a
  log file and uses some basic STATA commands      ***/

#delimit ;
set more 1;
drop _all;
cap log close;
log using c:\intropov\logfiles\try1.log, replace;

use "c:\intropov\data\hh.dta";
describe;
list in 1/3;
list hhid hhsz headedu if headsex==2 & headage<45;
summarize hhsz;
summarize hhsz, detail;
sum hhsz headedu [pw=weight], d;
tabulate headsex;
tabulate headedu headsex, col row chi;
tabulate headedu, summarize(headage);
label define sexlabel 1 "MALE" 2 "FEMALE";
label values headsex sexlabel;
tabulate headsex;
label variable headsex "Head Gender";
use "c:\intropov\data\hh.dta";
```

```

sort hhid;
save temp, replace;
use "c:\intropov\data\consume.dta", clear;
sort hhid ;
merge hhid using temp;
tabulate _merge;
keep if _merge==3;

log close;
exit;

```

The first line in the file is a comment. STATA treats any line that starts with an asterisk (*) as a comment and ignores it. You can write multi-line comment by using forward slash and asterisk (/*) as the start of the comment and end the comment by asterisk and forward slash (*). Comments are very useful for documentation purpose and you should include at least following information in the comment of a do file: general purpose of the do file, and last modification time and date. You can include comment anywhere in the do file, not just at the beginning.

<code>#delimit ;</code>	By default STATA assumes that each command is ended by the carriage return (ENTER key press). If, however, a command is too long to be fit in one line you can spread it over more than one line. You do that by letting STATA know what the command delimiter would be. The command in the example says that a command is ended by a semicolon (;). Every command following the <code>delimit</code> command has to end with a ; until the file ends or another <code>delimit</code> command appears which makes carriage return again the command delimiter. Although for this particular .do file we don't need to use the <code>delimit</code> command it is done to explain the command.
<code>set more 1</code>	STATA usually displays results one screenful at a time, and waits for the user to press any key. But this would soon become a pain if after starting a .do file to run you have to press a key every time this happens until the program ends. This command displays the whole output skipping page after page automatically.
<code>drop _all</code>	This command clears the memory.
<code>cap log close</code>	This command closes any open .log file. If no log file is open STATA just ignores this command.

You can edit .do files either using the STATA do-file editor or using a text editor such as Word and WordPad. If you are editing a .do file with the do-file editor, you can run your program directly within the do-file editor window by clicking the “Do current file” icon. Similarly, you can read .log files in any text editor. Now open the do-file editor and type the above code into it, then save it as `c:\intropov\dofiles\try.do`. Click the “Do current file” icon and switch to the “STATA Results” window. You should indicate command lines be processed without pause (using the `set more off` command). When you see “end of do file”, open `c:\intropov\logfiles\try11.log` in Word and check the results.

Follow-up practice

Now let’s do some practice using all three datasets. Remember, do not overwrite these three data files.

1. Generate a new variable, `agegroup`, that categorize individuals according to their age. For example, assign 1 to `agegroup` if the person is less older than 30. You can make your own rule that you think is appropriate. Label this variable and its categorical values, and tabulate it.
2. Calculate the sex ratio of sampled population, and the labor participation rates for both genders.
3. Count the number of children younger than 15 and the number of elders older than 65. Compare the mean per capita staple food consumption (in kg) for households with no children, one child, and two or more children.
4. Calculate the mean per capita food consumption of those households whose heads are aged between 30 and 39. Compare it to the mean per capita food consumption for those whose heads are aged between 50 and 59.
5. Report the mean and median per capita food consumption for each educational level of household head. (Hint: “`by`” or “`table xxx, c(mean xxx)`”).
6. Calculate the food share in total household expenditure and compare the mean food share for households headed by men with that of households headed by women.
7. Tabulate mean household size and mean educational level by region and area.

Exercises

Exercise 1 (Chapter Two):

Let's look at the data sets that we are going to use for our poverty analysis. It is quite important to know the general population structure and household characteristics. We also need some steps to construct the data set that we will use in the later exercises.

1.1 Household Characteristics

Open `c:\intropov\data\hh.dta` which consists of household level variables. Answer the following questions:

- i. How many variables are there?
- ii. How many observations (households) are there?
- iii. There are four regions. Household characteristics may vary by regions. Fill in the following table (hint: use `table` command).

	Dhaka	Chittagong	Khulna	Rajshahi
Total number of households	_____	_____	_____	_____
Total number of population	_____	_____	_____	_____
Average distance to paved road	_____	_____	_____	_____
Average distance to nearest bank	_____	_____	_____	_____
% Household has electricity	_____	_____	_____	_____
% Household has sanitary toilet	_____	_____	_____	_____
Average household assets	_____	_____	_____	_____
Average household land holding	_____	_____	_____	_____
Average household size	_____	_____	_____	_____

Are the sampled households very different across regions?

- iv. Gender of households heads may also affect household characteristics:

Male-headed households	Female-headed households
---------------------------	-----------------------------

Average household size	_____	_____
Average head schooling	_____	_____
Average head age	_____	_____
Average household assets	_____	_____
Average household land holding	_____	_____

Are the sampled households headed by males very different from those headed by females?

1.2 Individual Characteristics

Open `c:\intropov\data\ind.dta`. This file consists of household member information. Merge this data with the household level data (`hh.dta`) and answer the following questions for individuals who are 15 years old or older:

i. Regional variation	Dhaka	Chittagong	Khulna	Rajshahi
Average schooling years	_____	_____	_____	_____
Gender ratio (% of female)	_____	_____	_____	_____
% Working population (with positive working hours)	_____	_____	_____	_____
% Working population working	_____	_____	_____	_____ in farm

Are the sampled individuals very different across regions?

ii. gender difference	Among males	Among females
Average schooling years(age \geq 5)	_____	_____
Average schooling years(age $<$ 15)	_____	_____

Average age	_____	_____
% Working population (with positive working hours)	_____	_____
% Working population working in farm	_____	_____
Average working hours	_____	_____
Average working hours in farm	_____	_____
Average working hours in non-farm	_____	_____

Are the sampled women very different from men?

1.3 Expenditure

Open `c:\intropov\data\consume.dta`. It has household level consumption expenditure information. Merge it with `hh.dta`.

- i. Create three variables: per capita food expenditure (let's call it `pcfood`), per capita nonfood expenditure (call it `pcnfood`) and per capita total expenditure (call it `pcexp`). Now let's look at the consumption patterns.

Average per capita expenditure:	<code>pcfood</code>	<code>pcexp</code>
by region:		
Whole	_____	_____
Dhaka region	_____	_____
Chittagong region	_____	_____
Khulna region	_____	_____
Rajshahi region	_____	_____
by head gender:		
male-headed households	_____	_____
female-headed households	_____	_____
by head education level:		

head has some education	_____	_____
head has no education	_____	_____
by household size:		
Large household (>5)	_____	_____
Small household (<=5)	_____	_____
by land ownership:		
Large land ownership (>0.5/person)	_____	_____
Small land ownership or landless	_____	_____

Summarize your findings on per capita expenditure comparison

ii. Now add another measure of household size which take into account that children actually consume less than adults. Let's assume children (age <15) will be weighted as .75 of an adult. For instance, a household of a couple with one child aged at 7 has 2.75 adult-equivalence scale instead of 3. Go back to the `ind.dta` and create this variable (let's call it `famsize2`), then merge with the household data and the consumption data). Create per adult equivalent expenditure variables (let's call them `pafood` and `paexp`) and repeat the above exercise.

Average per capita expenditure:	pcfood	pcexp
by region:		
Whole	_____	_____
Dhaka region	_____	_____
Chittagong region	_____	_____
Khulna region	_____	_____
Rajshahi region	_____	_____
by head gender:		
male-headed households	_____	_____
female-headed households	_____	_____

by head education level:

head has some education _____

head has no education _____

by household size:

Large household (>5) _____

Small household (<=5) _____

by land ownership:

Large land ownership (>0.5/person) _____

Small land ownership or landless _____

Compare your new results with those of per capita expenditure. In analyzing poverty, is it better to use adult-equivalents?

iii. Besides looking at the mean or the median value of consumption, we can also easily look at the whole distribution of consumption using `graph`. The following is to plot the cumulative distribution function curve of per capita total expenditure.

```
. cumul pccexp, gen(pccexpcdf)
. label variable pccexpcdf "Cumulative Distribution of pccexp"
. label variable pccexp "Per Capita total expenditure"
. graph pccexpcdf pccexp, ylabel(0,.2,.4,.6,.8,1) xlabel
saving(cdf, replace)
```

The `cumul` command creates a variable called `pccexpcdf` which is defined as the empirical cumulative distribution function (cdf) of `pccexp`. After labeling the two variables, `pccexpcdf` (on the vertical axis) is plotted against `pccexp` (on the horizontal axis). Note the options `ylabel` and `xlabel` are used to have each axis properly labeled. Without arguments, `graph` will choose “round” values to be labeled; with arguments, `graph` will only label those values that you choose. The graph is also saved in a file called `cdf.gph`. When you want to look the graph later, you just need to type “`graph using cdf`”.

The cumulative distribution function curve of a welfare indicator can reveal quite a lot information about the poverty and inequality. For example, if we know the value of a poverty line, we can easily find the corresponding percentage value of people below the line.

- iv. Keep `pcfood pccexp pafood paexp famsize2 hhcode`, and merge with `hh.dta` and save as `pce.dta` in the `c:\intropov\data` directory.

1.4 Household Weights

In most household surveys, observations are selected through a random process, but different observations may have different probabilities of selection. Therefore, we need use weights which are equal to the inverse of the probability of being sampled. A weight of w_j for the j th observation means, roughly speaking, that the j th observation represents w_j elements in the population from which the sample was drawn. Omitting sampling weights in the analysis usually gives biased estimates, which may be far from the true values.

Various post-sampling adjustments to the weights are sometimes necessary. A household sampling weight is provided in the `hh.dta`. In estimating individual level parameters such as per capita expenditure, we need to transform the household sample weights into individual sample weights and use as an option in STATA commands:

```
. gen weighti = weight*famsize
. table region [pweight=weighti], c(mean pccexp)
```

STATA has four types of weights: `fweight`, `pweight`, `aweight`, and `iweight`. `fweight` or frequency weights, indicate duplicate observation and only allow integers; `pweight` or sampling weights, denote the inverse of probability of selection due to sample design; `aweight`, or analytic weights, are inversely proportional to the variance of an observation; and `iweight`, or importance weights indicate the relative “importance” of the observation and are rarely used. The most commonly used are `pweight` and `aweight`. Not every command supports every kind of weight, but each do have a default “natural” weights. So if instead `[weight=weighti]` is specified in the above command, then `fweight`—the default treatment of weights for `table` is assumed.

Now let’s repeat some previous estimation with weights:

	Dhaka	Chittagong	Khulna	Rajshahi
Average household size	_____	_____	_____	_____
Average per capita food expenditure:	_____	_____	_____	_____
Average per capita total expenditure:	_____	_____	_____	_____

Are weighted averages very different from unweighted ones?

1.5 Challenge: Effects of clustering and stratification

If the survey under consideration has a complex sampling design, then the standard errors of estimates may well be biased if ignoring clustering and stratification. Both have effects on the estimates of standard errors.

Household surveys can have multistage sampling, i.e., sample groups of households first (villages or city blocks) and then sample households within sampled primary sampling units. This technique is called clustering. Because observations in the same cluster are not independent as what we always assume, the standard errors may be underestimated.

Different groups of clusters are often sampled separately. These groups are called strata. For example, villages might be divided into strata according to their regions. Then 100 villages can be sampled from each region. Because sampling is implemented independently across strata, strata are statistically independent and can be analyzed as such. In many cases, this produces smaller estimates of standard errors.

STATA has a set of command specially dealing with survey estimation. For instance, one has to start with `svyset` to set variables: sampling weight, primary sample unit, strata. Then `svymean` will give estimates of population means and their correct standard errors. `svyreg` will perform linear regression, taking into account of survey design. Repeat the above exercise (1.4) and compare the results.

Exercise 2 (Chapter Three):

The degree of consistency of the poverty line depends on the extent to which the chosen poverty line represents similar levels of well-being over time and across groups so that the poverty measures can be compared over time and across groups. Three methods have been used for deriving the poverty line in Bangladesh: direct caloric intake, food energy intake and cost of basic needs. The cost of basic needs method is both representative and consistent for comparisons over time and across groups.

The following table gives a nutritional basket considered minimal for the healthy survival of a typical family in rural Bangladesh.

Food items	Per capita normative daily requirements		Average rural consumer price (taka/kg)
	Calorie	Quantity (gm)	
Rice	1386	397	15.19
Wheat	139	40	12.81
Pulse	153	40	30.84
Milk (cow)	39	58	15.90
Oil (mustard)	180	20	58.24
Meat (beef)	14	12	66.39
Fish	51	48	46.02
Potato	26	27	8.18
Other vegetables	26	150	38.30
Sugar	82	20	30.49
Fruits	6	20	28.86
Total	2112	832	

2.1 Direct Calorie Intake

The direct caloric intake method will consider any household not meeting the nutritional requirement of 2112 calories per day per person as poor. For this method, we need to know the quantity of every food item consumed by households and its calorie content. Then we calculate the total calorie content of those consumed food and derive an equivalent daily caloric intake per capita for each household. `c:\intropov\data\consume.dta` includes quantity of 10 food items consumed (“potato” and “other vegetables” listed in the above are combined into one item “vegetables” in the survey, let’s assume that the total per capita daily caloric requirement of this combined item is 52 and the quantity is 177 gm). Use the quantity information (hint: the unit is kg per week and needs to convert into gm per day) and the calorie content information of the above table to calculate each household's per capita caloric intake (calorie per day). Create a new variable `cpcap` to store this caloric intake. Now identify the households for which `cpcap` is less than 2112. These households are considered "poor" based on direct caloric intake method.

Create a variable `directp` which equals 1 if the household is poor and 0 otherwise. What percentage of people are poor by this method?

	Whole	Dhaka region	Other regions
% of being poor	_____	_____	_____

2.2 Food-Energy Intake

The food-energy intake method finds the value of per capita total consumption at which a household can be expected to fulfill its caloric requirement determines poverty based on that expenditure. Note this expenditure automatically includes allowance for both food and non-food thus avoiding the tricky problem of determining the basic needs for those goods. It does not need price data either. A simple method is to rank households by their per capita caloric intakes and calculate the mean expenditure for the group of households who consume approximately the stipulated per capita caloric intake requirement.

- i. Merge `cpcap` with `hh.dta` and calculate the average `pcexp` for the households whose per capita calorie intake is within 10% minus/plus range of 2112.
- ii. Call the average value `feipline` and identify the households for which `pcexp` is less than `feipline`. These households are considered "poor" based on food-energy intake method. Create a variable `feip` which equals 1 if the household is poor and 0 otherwise.

```
. sum pcexp [aw=weighti] if cpcap<2112*1.1 & cpcap>2112*.9
. gen feipline = r(mean)
. gen feip = 0
. replace feip = 1 if pcexp <= feipline
```

Note that STATA commands that report results also save the results so that other commands can subsequently use those results. `r-class` commands such as `summarize` save results in `r()` in version 6.0 or version 7.0. After any `r-class` commands, if you type “`return list`”, STATA will list what was saved. Another group—`e-class` commands such as `regress` save results in `e()` and “`estimates list`” will list saved results. For example, `e(b)` and `e(V)` store the estimates of coefficients and the variance-covariance matrix, respectively. To access coefficients and standard errors, there is an easier way. `_b(varname)` or `_coef(varname)` contains the coefficient on `varname` and `_se(varname)` refers to the standard error on the coefficient.

What percentage of people are poor by this method?

	Whole	Dhaka region	Other regions
--	-------	--------------	---------------

% of being poor _____

iii.Challenge: a more sophisticated method is to regress per capita total expenditure on per capita calorie intake and then predict the expected per capita expenditure at 2112 kcal level. Do this (hints):

```
. regress pcexp cpcap [aw=weighti]
. gen feipline=_b[_cons] + _b[cpcap]*2112
```

Should there be separate regression for each region? _____

2.3 Cost of Basic Needs

The cost of basic needs method is to find the value of consumption necessary to meet minimum subsistence needs. Usually it involves a basket of food items based on nutritional requirements and consumption patterns, and a reasonable allowance for non-food consumption.

i.According to the above basket and the average rural consumer prices, how much money does a household of four need each day to meet calorie requirement?

ii.One way to derive the non-food allowance is simply to assume a certain percentage of the value of minimum food consumption. How much total expenditure does the family need at least per year for being non-poor if assuming that nonfood expense is about 30 percent of food expense?

iii.vprice.dta gives village level price information on all 11 food items. Therefore, we can actually calculate a food poverty line (call it foodline) and a total poverty line (call it cbnpline) for each village using the cost of basic needs method and merge this variable with pce.dta (hint: need to sort both data sets and merge by thana vill). Do this and create a variable cbnp which equals 1 for the poor and 0 for the non-poor.

What percentage of people are poor by this method?

	Whole	Dhaka region	Other regions
% of being poor	_____	_____	_____

Do percentages of people being poor agree using the three methods? _____

Keep all imputed poverty lines and poverty indicators, merge with `pce.dta`, save it as `final.dta`.

Exercise 3 (Chapter Four):

3.1 A Simple Example

In STATA, open the data file `example.dta` and browse the data using STATA “Data Browser”. You should see a spreadsheet listing information exactly as presented in the following table.

Country A	Country B	Country C
110	118	119
115	120	120
119	120	122
120	125	123
125	126	123
127	127	125
138	133	135
141	141	140
178	173	173
222	212	215

It consists of the individual consumption information of three countries. All three countries have 10 people.

1. Summarize consumption for three countries. What do you find?
2. Assume a poverty line of 125, calculate the poverty rates for each country using the formulas provided in the Chapter Four notes:

	Country	A	B	C
a.	Using the headcount index:	_____	_____	_____
b.	Using the poverty gap index:	_____	_____	_____
c.	Using the squared poverty gap index:	_____	_____	_____
3. Which country has the highest incidence of poverty? Justify your answer.

3.2 Poverty Measures for Rural Bangladesh 1999

Now let’s work with the per capita food expenditure and the per capita total expenditure (`pcfood` and `pcexp` in `c:\intropov\data\final.dta`) that we have created in Exercise 1, and use `cbnpline` (the cost of basic needs poverty line that we derived in Exercise 2). To facilitate and generalize the calculation, three STATA `.ado` programs (under `c:\ado\personal`) are provided so that you do not have to program those formulas yourself. All you need to do is to prepare a welfare indicator and a poverty line, and invoke those programs in STATA just as what you do with STATA commands. Therefore, these programs can

be readily applied in your own poverty research. Please see section appendix for the documentation. Note that as a general rule in STATA, commands are case-sensitive.

FGT.ado can calculate the head count index (or FGT(0)), the poverty gap index (or FGT(1)), and the squared poverty gap index (or FGT(2)). For example,

```
. FGT y, line(1000) fgt0 fgt1 fgt2
. FGT y, vline(varname) fgt0 fgt1 fgt2
```

will calculate the headcount ratio, the poverty gap ratio, and squared poverty gap index using a poverty line 1000 on welfare indicator *y*. Instead of typing all three measures you can specify *all* option. However if only one measure is needed, you can specify just corresponding option. *line(#)* sets the poverty line. However if poverty line varies among observations you can use *vline(varname)* where *varname* contains poverty line for each observation. If neither option for poverty line is provided, the poverty line is assumed to be the value of half the median of welfare indicator *y*.

Sen.ado and SST.ado calculate the Sen index and SST index respectively. The syntax follows the same format:

```
. Sen y, line(1000)
. SST y, line(1000)
```

You can specify conditions, range and weights with these commands. For example, the following command calculates headcount ratio for Dhaka region based on a poverty line of 3000.

```
. FGT pexp [aw=weighti] if region==1, line(3000) fgt0
```

Now we are ready to turn to the measurement of poverty using the data from the BHES 1991/2.

1. Compute the five main measures of poverty (headcount, poverty gap, squared poverty gap, Sen index and Sen-Shorrocks-Thon index) for per capital expenditure, using both the food poverty line and the total poverty line derived by the cost of basic needs method in the previous exercise.

	Food poverty line	Total poverty line
i. Headcount index:	_____	_____
ii. Poverty gap index:	_____	_____
iii. Squared poverty gap index:	_____	_____
iv. Sen index:	_____	_____
v. Sen-Shorrocks-Thon index	_____	_____

2. Compute the headcount index and poverty gap index for specific subgroups using the food poverty line.

- | | Headcount index | Poverty gap index |
|---------------------------------|-----------------|-------------------|
| i. Dhaka region: | _____ | _____ |
| ii. Other three regions: | _____ | _____ |
| iii. Households headed by men: | _____ | _____ |
| iv. Households headed by women: | _____ | _____ |
| v. Large households (>5): | _____ | _____ |
| vi. Small households (<=5): | _____ | _____ |
3. Repeat the above exercise using the total poverty line.

- | | Headcount index | Poverty gap index |
|---------------------------------|-----------------|-------------------|
| i. Dhaka region: | _____ | _____ |
| ii. Other three regions: | _____ | _____ |
| iii. Households headed by men: | _____ | _____ |
| iv. Households headed by women: | _____ | _____ |
| v. Large households (>5): | _____ | _____ |
| vi. Small households (<=5): | _____ | _____ |

3.3 Appendix

Description----Poverty measure programs: FGT . ado, Sen . ado, and SST . ado

The general syntax is:

```
program varname [weight] [if exp] [in range] [, options]
```

where `program` is one of `FGT`, `Sen` and `SST`. `FGT` computes the three most common poverty measures (headcount ratio, poverty gap ratio, squared poverty gap ratio) based on the (income) distribution described by `varname` and their standard errors while `Sen` and `SST` compute the Sen index and the Sen-Shorrocks-Thon index, respectively.

`aweight` and `fweight` are allowed.

The poverty line is either specified by a value using `Line(#)`, or specified by a variable using `vline(varname)`. User can select any positive value in the first option. The second option should be used when there are multiple poverty lines and one variable with a name “`varname`” contains poverty line values for each observation. If neither option is provided, the default poverty line is then directly computed as half the median of welfare indicator.

Options only for `FGT`:

All requests that all three measures are computed and reported. By default, no measure is produced if nothing is specified. The following options specify the measures to be computed and reported:

```
fgt0: headcount ratio [FGT(0)]
fgt1: poverty gap ratio [FGT(1)]
fgt2: poverty gap ratio [FGT(2)]
```

When `sd` option is specified, the standard errors of specified measure will be calculated and reported next to the measure.

The results will be saved in the following global macros:

S_1 = total number of observations in the data
S_2 = number of observations used to compute the indices
S_3 = weighted number of observations
S_4 = value of the poverty line if a single value poverty line is used
S_5 = weighted number of observations identified as poor
S_6 = headcount ratio in FGT, or Sen index in Sen, or Sen-Shorrocks-Thon index
in SST
S_7 = poverty gap ratio [FGT(1)]
S_8 = squared poverty gap ratio [FGT(2)]

Note: `FGT.ado` file is based on `poverty.ado` written by Philippe VAN KERM and the standard error calculation follows Deaton (1997)

Exercise 4 (Chapter Five):

The robustness of poverty measures is very important because if poverty measures are not as accurate as we wish they could be then many conclusions that we draw from poverty comparisons between groups and over time may not be right.

4.1 Sampling Error

For example, the fact that poverty calculations are based on a sample of households rather than the population implies that calculated measures carry a margin of error. When the standard errors of poverty measures are large, small changes in poverty may well be statistically insignificant and should not be interpreted for policy purpose.

FGT can also compute the standard errors of its poverty measures if option *sd* is specified:

```
. FGT y, Line(1000) fgt0 fgt1 sd
```

Now let's re-compute the headcount index and poverty gap index for Dhaka region and other three regions using the total poverty line, and compute the standard errors of the two measures as well. Does the factor of standard errors change any conclusion of poverty comparison between Dhaka region and other regions?

	Headcount index	Poverty gap index
a) Dhaka region:	_____	_____
(Standard Errors)	_____	_____
b) Other three regions:	_____	_____
(Standard Errors)	_____	_____

4.2 Measurement Error

Another reason that we need to be very careful in poverty comparison is because data collected are measured incorrectly. This could be due to recall error on the part of respondents while answering survey questions, or because of enumerator error when the data were entered into specific formats. Let us simulate measurement error in per capita expenditure, and then investigate what the effect of this error is on basic poverty measures.

```
. sum pcexp [aw=weighti]
. gen mu = r(sd)*invnorm(uniform())/10
. gen pcexp_n1 = pcexp + mu
```

So this measure error is assumed to be a normal random variable with a standard error as big as a tenth of the standard error of observed per capita expenditure. Let us assume that the measurement error “mu” is additive to observed per capita expenditure. Note that this error by design is independent of observed per capita expenditure and of any other household or community characteristics.

Now re-compute the headcount ratio and poverty gap ratio using this new per capita expenditure. Then copy the values of those two measures using observed per capita expenditure from Exercise 3.

		pcexp	pcexp_n1
i.	Headcount index:	_____	_____
ii.	Poverty gap index:	_____	_____

Are those measures different? Now let's brainstorm the following situation. If the measurement error is correlated with household characteristic, for example, subsistence farmers usually underreport their consumption of own production, then is the measurement error problem going to be more severe? _____

4.3 Sensitivity Analysis

Apart from taking standard errors into account, it is also important to test the sensitivity of poverty measures for alternative definitions of consumption aggregates and alternative ways of setting the poverty line. For example, some non-food items are excluded from the expenditure aggregate on the basis that those items are irregular and do not reflect a household's command over resources on average. Also a 30% allowance for non-food expenditure is quite ad hoc.

1) Create a new total expenditure which includes the previously excluded irregular non-food expenditure (`expnfd2`) and re-compute the three FGT poverty measures on per capita expenditure and compare the results with those of original definition.

		pcexp	pcexp_n2
i.	the headcount index:	_____	_____
ii.	the poverty gap index:	_____	_____

2) the non-food allowance can be estimated from data. Two methods have been considered. The first is to find the average non-food expenditure for households whose total expenditure is equal to the food poverty line. The non-food expenditure for this group of households must be necessities since the households are giving up part of minimum food consumption in order to buy non-food items. The second is to find the non-food expenditure for households whose food expenditure is equal to food poverty line. Since the second is more generous than the first, the two are usually referred as the "lower" and the "upper" allowances and the poverty lines constructed using them are called "lower" and "upper" poverty lines, respectively.

```

. sum pcnfood [aw=weighti] if pcfood<foodline*1.1 &
  pcfood>foodline*.9
. gen line_u = foodline + r(mean)
. sum pcnfood [aw=weighti] if pcexp<foodline*1.1 &
  pcexp>foodline*.9
```

```
. gen line_1 = foodline + r(mean)
```

Now let's compare the results using these two poverty lines:

	Poverty line	30%	lower	upper
i.	the headcount index:	_____	_____	_____
ii.	the poverty gap index:	_____	_____	_____

Are those measures different? Can you explain why they are different? _____

3) Challenge: compare poverty measures when using per adult equivalence scale expenditure (paexp), with those of using per capita expenditure. _____

4.4 Stochastic Dominance

An ultimate tool of testing the robustness of poverty comparison is stochastic dominance test. The first-order stochastic dominance test compares the cumulative distribution functions of per capita expenditure. Let's compare the cumulative distributions between Dhaka region and other three regions. First, generate the cumulative distribution function of Dhaka region:

```
. keep if region == 1
. sort pcexp
. gen cump1 = sum(weighti)
. replace cump1 = cump1/cump1[_N]
. keep cump1 pcexp
. save temp, replace
```

Then, similarly generate the cumulative distribution cump2 for the other three regions. Keep cump2 and pcexp, and append with the "temp" by:

```
. append using temp
. label variable cump1 "Dhaka"
. label variable cump2 "other regions"
. graph cump1 cump2 pcexp xla() ylab(0,.2,.4,.6,.8,1)
```

Does one distribution dominate another? If the two lines cross at least once, then we may need to test for the second-order stochastic dominance. The poverty “deficit” curve is just the integral of the cumulative distribution up to every per capita expenditure value. After creating `cump1`,

```
. gen intcump1 = sum(cump1)
. keep intcump1 pcexp
. save temp, replace
```

Create `intcump2` for the other three regions. After combining variables and labeling them properly,

```
. graph intcump1 intcump2 pcexp xla() ylab()
```

Does one distribution dominate another here?

4.5 Challenge: Bootstrapping standard error for SST index

The bootstrapping technique can be applied in calculating of standard errors of poverty measures, especially those whose standard errors are impossible to solve analytically. The idea is quite simple. Repeat the calculation of the poverty measure a large number of times and each time use a new random sample drawn from the original one with replacement. For this purpose, you need to learn to use macros and loops in STATA. The following code is an example:

```
local i = 1
while `i' <= 100 {
    use " c:\intropov\data\final.dta ", clear;
    keep pcexp weighti cbnpline
    bsample _N
    SST pcexp [aw=weighti], vline(cbnpline)

    drop _all
    set obs 1
    gen sst = $S_1
    if `i' == 1 {
        save temp, replace
    }
    else {
        append using temp
        save temp, replace
    }

    local i = i + 1
}
```



```
sum sst
```

The above code is to repeat calculation of SST index 100 times.

Exercise 5 (Chapter Six):

5.1 Lorenz Curve

The Lorenz curve can give a clear graphic interpretation of the Gini coefficient. Let's make the Lorenz curve of per capita expenditure distribution of rural Bangladesh. First, we need to calculate the cumulative shares of per capita expenditure and population:

```
. sort pcexp
    . gen cumy = sum(pcexp*weight)
    . gen cump = sum(weight)
    . qui replace cumy = cumy/cumy[_N]
    . qui replace cump = cump/cump[_N]
```

Second, we need to plot the cumulative share of expenditure against the cumulative share of population. We may also want to add a 45° line so that we can present the perfect equality case.

```
. sort pcexp
    . graph cumy cump cump, c(11) s(..) xlabel
(0,.2,.4,.6,.8,1)
    ylabel(0,.2,.4,.6,.8,1) border
```

Now repeat this exercise for Dhaka region and compare its Lorenz curve with the Lorenz curve for the whole rural area. What conclusion can you draw? _____

5.2 Inequality Measures for Rural Bangladesh

Three programs are provided to compute the Gini coefficient, generalized entropy family and Atkinson family of inequality measures, respectively. As in Exercise 3, you can use these programs just like other STATA commands. The syntax is:

```
. gini y [if...] [in...] [weight]
. GE y [if...] [in...] [weight], alpha(#)
. Atkinson [if...] [in...] [weight], averse(#)
```

Alpha(#) and *Averse(#)* set the parameter values for the generalized entropy measure and Atkinson measure, respectively. Details see the appendix.

Let's continue using the per capita total expenditure to calculate inequality measures:

- i. Compute the Gini coefficient, the Theil index and the Atkinson index with inequality aversion parameter equal to 1 for the four regions.

	Gini	Theil	Atkinson
All regions	_____	_____	_____
Dhaka region:	_____	_____	_____
Other three regions:	_____	_____	_____

- ii. Now repeat the above exercise using two decile dispersion ratios and the share of consumption of poorest 25%. STATA command `xtile` is good for dividing the sample by ranking. For example, to calculate the consumption expenditure ratio between richest 20% and poorest 20%, you need to identify those two groups.

```
. xtile group = y, nq(5)
```

`xtile` will generate a new variable `group` which splits the sample into 5 groups according to the ranking of `y` (from smallest to largest, i.e., the poorest 20% will have `group==1`, while the richest 20% will have `group==5`). Similarly, to identify the poorest 25%, you need to split the sample into 4 groups.

	top 20 %	top 10%	
	_____	_____	
	bottom 20%	bottom 10%	Percentage of consumption of poorest 25%
whole	_____	_____	_____
for Dhaka region:	_____	_____	_____
for other three regions:	_____	_____	_____

- iii. Challenge: many inequality indexes can be decomposed by subgroups. Try to decompose Theil index by region. _____

Appendix

Description: inequality measure programs: `gini.ado`, `GE.ado`, `Atkinson.ado`

The general syntax is as the above one except where `program` is one of `gini`, `GE` and `Atkinson`. `gini` computes the most commonly used inequality measure—gini coefficient, while `GE` and `Atkinson` compute the Generalized Entropy family and the Atkinson family of inequality measures, respectively.

`aweight` and `fweight` are allowed.

Option for GE :

Alpha(#) sets the parameter value which determines the sensitivity of the inequality measure to changes of the distribution. The measure is sensitive to changes at the lower end of the distribution with a parameter value close to zero, equally sensitive to changes across the distribution for the parameter equal to one (which is the Theil index), and sensitive to changes at the higher end of the distribution for higher values.

Option for Atkinson:

Averse(#) sets the parameter value which measures aversion to inequality.

The result will be saved in the following global macro:

`S_1` = the inequality measure

Note: `gini.ado` file is based on Deaton (1997)

Exercise 6 (Chapter Seven):

In the previous exercises, we have computed poverty measures for various subgroups, such as regions, head gender, household size, etc. Another way is to present poverty profile by comparing characteristics of “poor” and “non-poor”.

6.1 Characteristics of the poor

“poor” and “non-poor” are defined by `cbnp` in Exercise 2.

	poor	non-poor
% Households	_____	_____
% Population	_____	_____
Average distance to paved road	_____	_____
Average distance to nearest bank	_____	_____
% Household has electricity	_____	_____
% Household has sanitary toilet	_____	_____
Average household assets	_____	_____
Average household land holding	_____	_____
Average household size	_____	_____
% Household headed by men	_____	_____
% Household headed by women	_____	_____
Average schooling of head	_____	_____
Average age of head	_____	_____
Average household total working	_____	_____
Hours worked on non-farm	_____	_____

6.2 More Poverty Comparison across subgroups

Calculate headcount and poverty gap index using `cbnpline` for the following subgroups:

	Headcount index	Poverty gap index
c) Head with no education:	_____	_____
d) Head with primary education	_____	_____
e) Head with secondary or higher education	_____	_____
f) Large land ownership (>0.5/person)	_____	_____
g) Small land ownership or landless	_____	_____
h) Large asset ownership (>50000)	_____	_____
i) Small asset ownership (<=50000)	_____	_____

Combined with the poverty measures computed in Exercise Three, describe the most significant poverty patterns in Bangladesh?

Exercise 7 (Chapter Eight):

Develop and estimate a model that explains $\log(\text{pcexp}/\text{cbtnpline})$ using available data. The regressors may include demographic characteristics such as gender of head and family structure; access to public services such as distance to a paved road; household members' employment such as working hours on farm and off farm; human capital such as average education of working members; asset positions such as land holding; etc. You need to identify potentially relevant variables and the direction of their effect. Then put all those variables together, and run the regression. Report the result and discuss whether it matches your hypothesis. If not, give possible reasons.

```
. gen y = log(pcexp/cbtnpline)
. reg y age age2 workhour x1-x3 [aw=weighti]
```

where $x1-x3$ are other explanatory variables that you want to include.

Note that if you want to include categorical variables, you need to convert them into dummy variables if the ranking of categorical values do not have any meaning. For example, you need to generate three (NOT four) dummy variables from `region`:

```
. gen dhaka = 0
. replace Dhaka = 1 if region==1
```

After the regression, you can plot the residuals against the fitted values to see whether it is a good fit.

```
. predict yhat, xb
. predict e, residuals
. graph e yhat
```

References

- Alderman, Harold, Miriam Babita, Jean Lanjouw, Peter Lanjouw, Nthabiseng Makhatha, Amina Mohamed, Berk Özler, and Olivia Qaba. 2000. *Is Census Income an Adequate Measure of Welfare? Combining Census and Survey Data to Construct a Poverty Map of South Africa*. Statistics South Africa Working Paper. Forthcoming.
- Atkinson, Anthony B. 1987. "On the Measurement of Poverty", *Econometrica*. Vol. 55. pp.749-64.
- Bidani, Benu, and Martin Ravallion. 1992. "Constructing a Consistent Poverty Profile". Mimeo. Poverty Analysis and Policy Division. World Bank, Washington DC.
- Clark, S., R. Hemming and D. Ulph. 1981. "On Indices for the Measurement of Poverty". *Economic Journal*. Vol. 91. pp. 515-526.
- Dalaker, Joseph and Mary Naifeh. 1998. *Poverty in the United States: 1997*, U.S. Department of Commerce, Washington DC.
- Dalton, Hugh. 1920. *Some aspects of the inequality of incomes in modern communities*. London, G. Routledge; New York, E.P. Dutton.
- Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, Johns Hopkins University Press, Baltimore.
- Duong, Nguyen Binh and Dinh Hien Trinh. 1999. "Identification of Poverty in Vietnam," in Dominique and Jonathan Haughton et al. (eds.), *Health and Wealth in Vietnam: An Analysis of Household Living Standards*, Institute of Southeast Asian Studies Press, Singapore.
- Elbers, Chris, Jean Lanjouw, Peter Lanjouw. 2000. *Welfare in Towns and Villages. Micro-Level Estimation of Poverty and Inequality*. Tinbergen Institute Working Paper. Forthcoming.
- Ferreira, M. L. 1996. "Poverty and Inequality during Structural Adjustment in Rural Tanzania". World Bank Policy Research Working Paper 1641, Washington D.C.
- Gibson, Jonathan. 1999. *A Poverty Profile of Cambodia, 1999*. Department of Economics, University of Waikato, New Zealand.
- Grosh, Margaret and Juan Munoz. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey." LSMS Working Paper 126, The World Bank, Washington D.C.
- Haughton, Jonathan. 2001. *The Impact of the East Asian Crisis: Poverty Analysis Using Panel Data*. Lecture notes prepared for the World Bank. Suffolk University and Beacon Hill Institute, Boston, MA

- Hentschel, Jesko, Jean Lanjouw, Peter Lanjouw and Javier Poggi. 2000. "Combining Survey Data with Census Data to Construct Spatially Disaggregated Poverty Maps: A Case Study of Ecuador," *World Bank Economic Review*, Volume 14 No. 1.
- Huppi, Monika and Martin Ravallion. 1991. "The Sectoral Structure of Poverty During an Adjustment Period. Evidence for Indonesia in the Mid-1980s". *World Development*.
- Knowles, James. 1998. *Estimation of Updated Poverty Rates for Cambodia – 1997 Technical Report*. The World Bank.
- Morduch, Jonathan. 1998. "Politics, growth, and inequality in rural China: does it pay to join the party?". Boston: Harvard Institute for Economic Research.
- Newbery, David M.G. 1995. *Tax and benefit reform in Central and Eastern Europe*. London: Center for Economic Policy Research.
- Newbery, David, and Nicholas H. Stern, (eds.). 1987. *The Theory of Taxation for Developing Countries*. Oxford: Oxford University Press.
- Osberg, Lars and Kuan Xu. 1999. *Poverty Intensity: How Well Do Canadian Provinces Compare?* *Canadian Public Policy* (XXV:2).
- Pradhan, Menno Asep Suryahadi, Sudarno Sumarto and Lant Pritchett. 2000. *Measurements of Poverty in Indonesia: 1996, 1999 and Beyond*, SMERU, Jakarta, June.
- Prescott, Nicholas and Menno Pradhan. 1997. *A Poverty Profile for Cambodia*, World Bank, Washington DC.
- Ravallion, Martin. 1988. "Expected Poverty under Risk-Induced Welfare Variability". *The Economic Journal*. Vol 98. pp. 1171-1182.
- Ravallion, Martin and Benu Bidani. 1994. "How Robust Is a Poverty Profile?", *The World Bank Economic Review*, 8(1): 75-102.
- Ravallion, Martin. 1996. "How well can method substitute for data? Five experiments in poverty analysis," *The World Bank Research Observer*, 11(2), 199-221, August.
- Ravallion, Martin. 1998. *Poverty Lines in Theory and Practice*, Living Standards Measurement Study Working Paper No. 133, World Bank.
- Ravallion, Martin. 1999. *Poverty Comparisons*. World Bank document.
- Sen, Amartya. 1976. "Poverty: An Ordinal Approach to Measurement" *Econometrica*, Vol. 46. pp.437-446.

Sen, Amartya. 1983. "Social theory, social understanding, and political action" In *Social Theory and Political Practice*. Edited by Christopher Lloyd. Oxford: Clarendon Press; New York:: Oxford University Press.

Watts, Harold W. 1968. *The measurement of poverty – an exploratory exercise*. University of Wisconsin.; Institute for Research on Poverty.; Discussion papers; 12-68.

World Bank. 1999. *Vietnam: Attacking Poverty*, Hanoi. Various questionnaires related to the Vietnam Living Standards Surveys.

World Bank. 2000. *World Development Report 2000-2001*. Washington DC.