

**A GUIDE TO THE**

**SEDLAC**

**SOCIO-ECONOMIC DATABASE**

**FOR LATIN AMERICA AND THE CARIBBEAN**

CEDLAS \*  
and  
The World Bank \*\*

---

**INDEX**

0. Introduction .....	2
1. Household surveys .....	4
2. Income .....	13
3. Poverty .....	25
4. Inequality .....	35
5. Demographics .....	40
6. Education .....	42
7. Employment .....	47
8. Housing .....	55
9. Infrastructure .....	57
10. Durable goods and services .....	59
11. Poverty-alleviation programs.....	60
12. Aggregate welfare .....	62
13. Pro-poor growth .....	63
References .....	65

---

\*CEDLAS is the Centro de Estudios Distributivos, Laborales y Sociales at Universidad Nacional de La Plata. E-mail: cedlas@depeco.econo.unlp.edu.ar

\*\* The World Bank's LAC poverty group (LCSPP).

## **0. INTRODUCTION**

The SEDLAC is a database of socio-economic statistics constructed from microdata of the Latin American and Caribbean (LAC) household surveys, developed by CEDLAS (Universidad Nacional de La Plata) and The World Bank's LAC poverty group (LCSPP). This guide describes the contents of each table in the database, and briefly explains the main methodological decisions taken to construct the statistics.

So far, the SEDLAC is divided into 13 sections: household surveys, incomes, poverty, inequality, demographics, education, employment, housing, infrastructure, durable goods and services, poverty-alleviation programs, aggregate welfare and pro-poor growth. As the project develops we are going to include more sections.<sup>1</sup> Each section contains at least one Excel file with several worksheets. Each sheet contains a table with statistics on a specific issue for 25 LAC countries (data permitting). All statistics are computed from microdata of household surveys by routines documented in Stata do files.

A large-scale dataset like the SEDLAC implies a daily work of revising and controlling the construction of variables and statistics, identifying mistakes, and searching for more efficient ways of processing the data and presenting the statistics. In this sense, the SEDLAC is expected to be changing over time. The suggestions of researchers and users are important inputs to improve the database. The SEDLAC is intended to be a space of interactions among researchers on social issues in LAC.

### **On the comparability of statistics**

Household surveys are not uniform across LAC countries. They significantly differ in geographical coverage and questionnaires. Surveys are also different within countries over time. LAC governments have been improving their household surveys over the past decade, changing coverage and questionnaires. The issue of comparability is, hence, of a great concern. In particular, how comparable (across countries and over time) are the statistics shown in the SEDLAC?

We make all possible efforts to make statistics comparable across countries and over time by using similar definitions of variables in each country/year, and by applying consistent methods of processing the data. However, perfect comparability is not assured, as the coverage and questionnaires of household surveys differ among countries, and frequently also within countries over time. Hence, a trade-off arises between accuracy and coverage. If we want to be ambitious in the analysis, we have to pay the cost of losing accuracy and

---

<sup>1</sup> See Next Steps in the web site.

getting into comparability problems. Sometimes these problems are too severe and it is convenient to restrict the analysis. This guide and other documents in our web page provide the user with relevant information to decide on that trade-off. The final decision whether making a comparison or not depends on the preferences and specific needs of each user.

Household surveys have several problems. However, they are still the best source of information for national socio-economic statistics. We think we should avoid the two extreme positions toward household surveys: to discard them or to use them without qualifications. With all their limitations household surveys still provide valuable information, being the best available source to generate representative statistics of the population. However, it is important to be aware of their drawbacks. Despite LAC governments and international organizations have taken important steps in the last decade (*e.g.* the MECOVI program), they still have a long way to go in order to have a more reliable, richer and more homogeneous set of national household surveys. Latin America is lagging behind some other regions in that area.

This guide is organized in 13 chapters. In each chapter we describe the tables available in each section of the database, provide basic information to read the tables, and discuss the methodological decisions taken to construct them.

### **Citation**

Information taken from this database should be cited as “Source: SEDLAC (CEDLAS and The World Bank)” or “Source: Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank)”. We advise making reference to the date when the database was consulted, as statistics may change.

## 1. HOUSEHOLD SURVEYS

This section has one file (*surveys\_LAC.xls*) containing information on the main characteristics of the household surveys:

- name of the survey
- acronym
- year when the survey was conducted
- date of the field work
- geographic coverage
- number of households in the dataset
- number of individuals in the dataset
- contents of each survey.

In particular, the file records whether a specific survey has at least some basic information on

- demographics
- housing
- education
- employment
- income
- expenditures/consumption
- health
- social programs
- anthropometrics
- migration

It also reports whether the survey has other special modules, and whether parallel surveys were carried out (*e.g.* to record local prices).

### *Methodological issues*

#### **Countries**

So far, the SEDLAC includes information from the following 25 countries: Argentina, Bahamas, Belize, Bolivia, Brazil, Colombia, Costa Rica, Chile, Dominica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Uruguay and Venezuela. The sample covers all countries in mainland Latin America and eight countries in the Caribbean. In each period the sample of countries represents more than 97% of LAC total population.

The database mainly covers the 1990s and 2000s, although we also present information for previous decades in a few countries.

Most household surveys included in the sample are nationally representative. The main exception is Argentina, where surveys cover only urban population, which nonetheless represents more than 85% of the total population in that country. The household surveys of Suriname and Bahamas cover only urban areas. We also work with some surveys that cover only urban areas in Bolivia, Colombia and Paraguay for the early 90s, and Uruguay until 2005.

In the rest of the section we briefly summarize the information for each country.

### *Argentina*

All estimates are computed from the Encuesta Permanente de Hogares (EPH). This survey is carried out by the Instituto Nacional de Estadística y Censos (INDEC) since the early 1970s in the Greater Buenos Aires area, and since the 1980s in most large cities (more than 100,000 inhabitants). Before 2003 it was carried out in two rounds: May and October. During 2003 a major methodological change was implemented by INDEC, including changes in the questionnaires and in the timing of the survey visits. The new survey (known as EPH Continua or EPH-C) is now conducted over the whole year. INDEC also started to make imputations to income non-response observations. To assess the impact of these methodological changes we present four sets of statistics for 2003 in most tables: one computed from the EPH carried out in May, and three computed from the EPH-C of the second semester of 2003: the first two ignoring the income imputations - one with the old weights and the other using weights that consider income non-response - and the last one including the observations with imputations. The EPH-C covers 31 large urban areas which are home to around 70% of the Argentine urban population. Since the share of urban areas in Argentina is 87%, the sample of the EPH represents around 60% of the total population of the country.

### *Bolivia*

Bolivia's current household survey is the Encuesta Continua de Hogares-Mecovi (ECH). This survey is nationally representative. It has been carried out since 1999. In 2003-04 the survey was modified to record incomes over the whole year. That change implies serious comparability problems with previous years. Bolivia had other two household surveys: the Encuesta Integrada de Hogares (until 1995) that covered only regional capital cities and the city of El Alto, and the Encuesta Nacional de Empleo (ENE), conducted three times between 1996 and 1997.

### *Brazil*

The Brazilian household survey is called Pesquisa Nacional por Amostra de Domicílios (PNAD) and is carried out by the Instituto Brasileiro de Geografia y Estadísticas (IBGE) on a yearly basis. The PNAD was first conducted in the state of Rio de Janeiro in 1967. Now, it is a national survey that covers urban and rural areas. The exceptions are the rural areas of Rondonia, Acre, Amazonas, Roraima, Pará and Amapá. The survey was not carried out in 1970, 1980, 1991 and 2000 (census years) and 1994. Some rural areas from the North of the country were added in the sample frame since 2004.

### *Chile*

The Encuesta de Caracterización Socioeconómica Nacional (CASEN) is a nationally and regionally representative household survey carried out by the Ministry of Planning MIDEPLAN through the Department of Economics at the Universidad de Chile, that is responsible for the data collection, digitalization and consistency checking of the database. ECLAC is responsible of making adjustments for non-response, missing income values, and the under (or over) reporting of different income categories before the databases are officially available for public use. The survey has been regularly implemented every two years since 1985 during November and in some cases, up to mid December. The last available CASEN was carried out in 2006.

### *Colombia*

The Encuesta Nacional de Hogares-Fuerza de Trabajo (ENH) was the main household survey in Colombia. It was carried out by the Departamento Administrativo Nacional de Estadística (DANE) until 2000. Initially, the survey covered only the main four urban areas of the country. New urban areas were included progressively, and in 1992 rural areas were included in the September round of the survey. We present information from this survey for 1992, 1996, 1999 and 2000. In 2000 the ENH was changed for the Encuesta Continua de Hogares (ECH). The new ECH has significant differences with the ENH in terms of frequency, coverage and questionnaires. In 2006 we present information drawn from the new Gran Encuesta Integrada de Hogares.

### *Costa Rica*

Costa Rica's main household survey is the Encuesta de Hogares de Propósitos Múltiples (EHPM), conducted since 1987 by the Instituto Nacional de Estadísticas y Censos (INEC).<sup>2</sup> The EHPM is carried out once a year (on July) and is nationally representative, covering both urban and rural areas of the country. The EHPM, which surveys around 40,000 individuals, does not have a rich questionnaire. Although the survey has not significantly changed over time, changes in the sample weights after the Census generate some comparability problems.

### *Dominican Republic*

The main household survey in Dominican Republic is called Encuesta Nacional de Fuerza de Trabajo (ENFT). The ENFT is conducted since 1991 by the Banco Central de la República Dominicana. From this year to 1992 the survey was carried out in four rounds: January, April, July and October. Since 1993 the ENFT was carried out in two waves: April and October. The survey is nationally representative with a sample size of more than 20,000 inhabitants. Significant changes in the survey starting in 2000 generate serious comparison problems with previous surveys.

### *Ecuador*

The Encuesta de Condiciones de Vida (ECV) is the Ecuador's version of the Living Standards Measurement Survey. Five national surveys have been conducted so far: 1994, 1995, 1998, 1999 and 2006. Ecuador has also the Encuesta de Empleo, Desempleo y Subempleo (ENEMDU), which became national in 2003. The various differences between this survey and the ECV make the comparison difficult. We also report information from the urban Encuesta Periódica de Empleo y Desempleo (EPED) between 1995 and 2000.

### *El Salvador*

The Encuesta de Hogares de Propósitos Múltiples (EHPM) is the household survey of El Salvador. Prior to the incorporation in the MECOVI program, household surveys in El Salvador were narrow in terms of coverage and questionnaires. The EHPM is conducted

---

<sup>2</sup> Household surveys in Costa Rica were implemented in 1966. The Encuesta Centroamericana de Hogares (ECH) covered the period 1966-1971 and was carried out in March, July and November. From 1976 to 1986 the INEC carried out the Encuesta Nacional de Hogares, Empleo y Desempleo (ENHED), for the same three waves: March, July and November.

under the MECOVI program since 1997. Major contributions of this program have been the expansion and deepening of thematic coverage of the questionnaire.

### *Guatemala*

The MECOVI program has promoted household surveys in Guatemala. The main household survey is the ENCOVI (Encuesta Nacional de Condiciones de Vida) carried out in 2000 and 2006. The ENEI (Encuesta Nacional de Empleo e Ingresos) was carried out between 2002 and 2004. We include statistics computed from both surveys.

### *Haiti*

Haiti has conducted the Enquete sur les Conditions de Vie en Haiti (ECVH) during 2001. It is a national survey with more than 33,000 observations. The survey has information on demographic, education, and employment variables. It reports data on income and consumption, as well as the access to social programs and health services.

### *Honduras*

The main household survey in Honduras is called Encuesta Permanente de Hogares de Propósitos Múltiples. During 1990 and 1991 the EPH was conducted once a year. Two rounds were included since 1992 (April-May and September-October) with the exception of 2000. Non-monetary and non-labor incomes were not included in the EPH until 1996. In many tables we show the statistics in two groups: the first one, labeled *EPHPM 1*, includes statistics computed with only labor monetary incomes (1992 and 1997), while the second group, labeled *EPHPM 2*, includes statistics with the complete definition of income.

### *Jamaica*

Jamaica has two main household surveys: the Jamaica Survey of Living Conditions (JSLC) and the Labour Force Survey (LFS). The JSLC is linked to the ongoing quarterly LFS. The households are visited once for the standard LFS. Then a subset of households is revisited about a month later for the JSLC. When the data sets are merged, the LFS serves as the employment module of the combined LFS/SLC. Zero income report is a particularly relevant problem in Jamaica.

### *Mexico*

The main household survey in Mexico is the Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). This survey contains information on socioeconomic characteristics, incomes and detailed information on different categories of expenditures. It is a nationally representative survey, covering urban and rural areas. The ENIGH is collected by the Instituto Nacional de Estadística y Geografía (INEGI) and is available for 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005 and 2006.

### *Nicaragua*

Nicaragua's household survey is the Encuesta Nacional de Hogares sobre Medición del Nivel de Vida (EMNV), a survey with national coverage that has been conducted four times: 1993, 1998, 2001 and 2005.<sup>3</sup> It has been developed by the Instituto Nacional de Estadística y Censos de Nicaragua (INEC) in the framework of the MECOVI Program.

### *Panama*

The main household survey in Panama is the Encuesta de Hogares (EH), carried out in August of each year. It is a nationally representative household survey. Panama carried out Living Standards Measurement Surveys in 1997 and 2003. Our analysis is based only on the EH. Although this is a well-established survey with more than 50,000 observations, the information is limited to demographic, education, employment, migration and income variables.

### *Paraguay*

Social variables in Paraguay can be traced at a national level only since 1995 with the microdata of the Encuesta de Hogares-Mano de Obra (EH-MO, 1995), the Encuesta Integrada de Hogares (EIH, 1997-1998 and 2000-2001) and the Encuesta Permanente de Hogares (EPH, 1999 and yearly since 2002). A historical perspective of the evolution of social indicators in Paraguay can be computed only for the Asunción Metropolitan area since 1983 with data from the Encuesta de Hogares-Mano de Obra.

### *Peru*

Peru has two household surveys. The ENNIV, corresponding to the LSMS group, was carried out in 1985, 1991, 1994, 1997 and 2000. The ENAHO is carried out in four waves since 1997, and continues until today. The fourth wave of the survey is nationally representative, and it is officially used to estimate poverty rates. The SEDLAC presents

---

<sup>3</sup> A specific round to analyze the consequences of the Mitch Hurricane was introduced in 1999.

statistics computed from the ENAHO for the period 1997-2007. After year 2000 the survey was enlarged and a new sample frame was used, implying comparability problems with previous surveys. In most tables we show statistics for 1997-1999 and for 2001-2007, separately.

### *Suriname*

The SEDLAC includes data from the Expenditure Household Survey (EHS), 1999. This survey, carried out by UNDP with the National Statistical Office of Suriname, covers only the urban area of the capital city (Paramaribo), which represents more than half of the population of the country. The EHS is a small survey of 1,700 observations (individuals).

### *Uruguay*

The Encuesta Continua de Hogares (ECH) is the main household survey in Uruguay. The ECH is conducted by the Instituto Nacional de Estadística (INE). It was first carried out in 1968 in Montevideo, and for a long time covered only urban areas. Anyway, the sample of the ECH represented around 80% of the total population of the country. In 2006 the survey became national, covering also rural areas.

### *Venezuela*

The Encuesta de Hogares por Muestreo (EHM) is the household survey in Venezuela. The EHM has a national coverage of the non-indigenous country's resident population. It is conducted since 1967, two times per year. The period of data collection is 6 months. The main objective of this survey is to provide general information on socioeconomic and sociodemographics characteristics of the labor force.

## **Changes in geographical coverage**

Changes in the geographical coverage of a given household survey can severely affect the comparability of results within a country over time. Since regions differ in their economic and social situations, adding a new region into the survey usually significantly affect the national statistics.

In most countries where changes in geographical coverage of the survey occurred during the last decade we use all the data available in the most recent survey (which include areas not covered in previous surveys) and provide ways of assessing the impact of the change in coverage. For instance, in Bolivia the household survey was urban in 1993 and

nationally representative in 1997. We present two sets of statistics for Bolivia 1997: one for the whole sample and one for those urban areas also surveyed in 1993. In Argentina, for instance, the National Statistical Office (INDEC) introduced additional cities into the sample of the survey (EPH) during the 1990s. We compute statistics for the larger sample available in the last rounds of the survey, and also statistics for a smaller sample of cities that are present since the 1980s. To avoid multiplying the information in the tables, we report statistics for both the smaller and the larger sample only for one year.

### **National and urban surveys**

Household surveys in Argentina, Bahamas and Suriname cover only urban areas. They are then *urban* surveys. However, Argentina is one of the most urban countries in the world, with more than 85% of the population living in cities. The survey in Suriname covers only Paramaribo, which is home of more than half of the population of the country. When we report *national* statistics we include Argentina, Bahamas and Suriname, although the statistics come from an *urban* household survey.

### **Changes in surveys**

Most Latin American countries have experienced significant improvements in their household surveys in the last decades. In particular, major changes have been implemented in several LAC countries since mid 1990s after the MECOVI Program. Although these changes are very welcome, they pose significant problems for comparison purposes within countries over time. Since this is a project intended to be continued in the future we try to exploit all the information in the latest surveys, although in some cases this decision implies difficulties when comparing with previous surveys in the country. For instance, Honduras' household survey includes questions on non-labor income since 1996. We can either ignore non-labor income to compare the old surveys with the new ones, or include non-labor income to have a most precise picture of incomes, and pay the cost of losing comparability with the older surveys. We take the second option, although in some cases (e.g. poverty and inequality indicators in Honduras) we include estimates using the two alternatives.

## Weights

All NSOs provide the individual and household sample weights in the data sets. We use these official weights to compute all the statistics in the database. All statistics in the SEDLAC are population-weighted statistics.

## 2. INCOMES

This section has two files: *construction\_incomes\_LAC.xls* contains information on the items included in the construction of income variables, while *incomes\_LAC.xls* reports per capita income by deciles, areas and regions.

The file *construction\_incomes\_LAC.xls* includes three tables aimed at providing information on the items included into (or excluded from) the individual and household income variables constructed in the SEDLAC.

### *total*

- General information on the construction of individual and household income.

### *labor*

- Information on the construction of labor income.

### *non\_labor*

- Information on the construction of non-labor income

### Codes

Each cell in the tables of this file admits several possible answers:

- (1) *Yes* = The item is in the survey, it has a monetary value and it is added up in the definition of income.

It could be the case that there is not a specific separate question to report the item, but we are sure that the surveyed person includes it into the answer to a more general question. Suppose a survey asks: Have you received rents, interest payments, benefits or pensions during the last month? In the table *non\_labor* we enter a YES in the row for Pensions (and also for Profits and benefits, Rents, Interests), although we cannot single out the amount.

- (2) *W.S.Q* (without specific question) = There is not a specific question on the item, but we believe it is included in a more general variable.

Example: The Argentina's EPH asks a general question on labor income. This variable in principle includes items like per diem or commissions, but the survey does not ask for

them explicitly. We believe they are included in the labor income report of the surveyed person but we are not sure. Hence, in the file we write *W.S.Q.*

Notice the difference with the previous point. If the EPH asked, What is your labor income including wages, per diem and commissions?, then we would write in the row for per diem and commissions YES, instead of *W.S.Q.*

(3) *Yes, but ignored* =there is a specific question in the survey with a monetary value, but we ignore it in our definition of income.

Suppose the survey records the non-ordinary income from selling an asset. We ignore this kind of income source in our definition of household income, so we write *Yes, but ignored.*

(4) *Only binary*= there is a question on the item, but the questionnaire allows only a binary (yes or no), or maybe categorical answer. There is not a monetary value for the item, so it is not included in the income variables.

The Argentina's EPH asks whether the worker has received non-monetary payments, but the possible answers are just yes or no. Then, we write *Only binary* in the corresponding rows in the sheets *total* and *labor*.

(5) *No* = The item is not asked in the survey, and it is not included into a more general variable. Our household income variable does not include that concept.

(6) *No, but estimated* = The item is not asked in the survey, and it is not included into a more general variable. However, we estimate the item and include it in our definition of household income.

Example: The Argentina's EPH does not include questions on the implicit rent for own-housing. However, for comparison purposes with other countries we include a rough estimation of this concept in our definition of household income (see below).

The file *incomes\_LAC.xls* reports mean household per capita income in local currency units. It contains two tables:

#### *deciles\_pci*<sup>4</sup>

- Mean household per capita income in local currency units (LCU) by deciles of the household per capita income distribution. National, urban and rural statistics.

#### *regions\_pci*

- Mean household per capita income in local currency units (LCU) by area (rural and urban) and region.

### ***Methodological issues***

#### **The household**

During a household survey round National Statistical Office (NSO) agents visit a dwelling. The agent then identifies each household in a dwelling, in case that more than one family lives in the house. Each household has a household head and a variable number of members. NSOs across countries differ in the treatment of the “external” members of a household: renters, domestic servants and their families. For this study we assume that these persons do not share the same budget of the primary members of the household, and then they are not considered part of the household. Of course, in practice there may be cases where, for instance, domestic servants are really part of the household. However, as most household surveys do not have information on intra-household behavior, we make the decision of excluding external members from the household. Exploratory analysis for some countries suggests that for the most part results are not significantly affected by this decision.

#### **Why income and not consumption?**

It is well known that household consumption is a better proxy for well-being than household income.<sup>5</sup> Three main reasons justify this preference: (i) if people can lend and borrow current consumption is closer to permanent income (or consumption) than current income, (ii) differential under-reporting by strata is usually a more severe problem for

---

<sup>4</sup> Due to lack of space this table is divided into two sheets in the Excel file.

<sup>5</sup> See for instance Deaton and Zaidi (2003).

income than for consumption, and (iii) incomes are frequently reported before taxes, while consumption is an after-tax concept.

Despite this dominance, nearly all distributional studies in LAC use income as the well-being indicator. A simple reason justifies this practice: few countries in the region routinely conduct national household surveys with consumption/expenditures-based questionnaires, while all of them include questions on individual and household income. Most countries have expenditure surveys that are mainly devoted to the construction of weights for the Consumer Price Index. However, these surveys are usually carried out every long periods of time (e.g. 10 years), so they are not appropriate for closely monitoring poverty, inequality and other social indicators.

At this stage of the project we follow the literature on distributional issues in LAC and use income as the well-being indicator. We expect to include expenditures in the following stages of the project, as countries start to collect information on household spending more frequently.

### **Construction of income variables**

We construct individual income by adding all income sources. Whenever possible we distinguish among income from salaried work, self-employment and salaries assigned to owners. Whenever possible we compute labor income from the main activity. Individual non labor-income is divided into three groups: (i) pensions, (ii) capital and benefits, and (iii) transfers. Countries differ in the questions devoted to capturing capital income, interests, profits, rents, and dividends. For comparison purposes we prefer to gather all these questions in a single category. The same criterion applies to transfers, although we also construct a variable that, whenever possible, identifies those transfers made by the government, and other that captures transfers clearly associated to poverty-alleviation programs (e.g. transfers from Programa Jefes de Hogar in Argentina). Once we have individual incomes, we construct household income by adding incomes for all members of the household. Finally, we compute household income adjusted for several equivalence scales (see below for a discussion on equivalence scales).

### **Recall period**

Most household surveys report incomes obtained during the month previous to the survey. Some surveys also include information on incomes earned in the last 6 months (e.g. ENIGH in Mexico). In those cases, and for comparative purposes, we only include income in the last month.

Incomes are transformed into monthly incomes if values are not reported on a monthly basis. However, some ambiguities arise as information on income frequency is sometimes scarce. Suppose for instance that an individual reports income for \$100 earned on a weekly basis. If no other information is available two situations are consistent with that report: (i) the individual earns \$100 a week, so her income is \$400 a month, or (ii) the individual worked only that week in the month, and her income is then \$100 a month. When no other information is available to figure out the correct alternative, we assume the first option.

The use of *current income* as the welfare indicator warrants some important caveats. If income is volatile and people can save or have some access to the credit markets, individual welfare might not be closely associated to current income. When using current instead of permanent income, inequality is usually overestimated and the estimations of poverty are also biased. In that case, the sign of the bias depends on the location of the poverty line in the income distribution. Suppose an economy with two individuals: *A* earns 100 in  $t_1$  and 0 in  $t_2$ , while *B* earns 0 in  $t_1$  and 100 in  $t_2$ . Assuming no discount rate, “permanent” income is 50 for both individuals: there is complete equality. However, if the survey only captures current income, inequality statistics will picture a very unequal society in  $t_1$  and  $t_2$ . If the poverty line is 40 and people have some access to the credit markets (e.g. can use savings or borrow from relatives or neighbors during bad months), then real poverty in this economy is 0. However, statistics based on current income will report a headcount ratio of 50%, *i.e.* overestimation of poverty. Instead, if in the same economy the poverty line is 60, then real poverty is 100% but the statistics will show only a headcount ratio of 50%, *i.e.* underestimation of poverty.

Two ways of alleviating these problems are using current consumption, and computing average incomes from panels. We will explore these two alternatives in following stages of this project. So far, most LAC countries either do not have budget surveys and panels, or when available, budgets surveys are conducted every long periods of time, and panels are short and often the microdata is not available to the public.

### **Non-current income**

Since we are interested in capturing current income, non-current items are not included in our definition of income. Sometimes extraordinary items are explicitly marked in the survey. For instance, the Argentina’s EPH asks for “bonificaciones *no habituales*” and severance payments. We exclude these concept from our definition of income, and hence depart from the official definition.

The same criterion leads to the exclusion of income from the sale of some goods and assets like vehicles, houses, or stocks. We also exclude income from gifts, life insurance, gambling and inheritances.

### **Implicit rent from own-housing**

The concept of income refers to the flow of resources obtained as remuneration to the use of all the assets owned by an individual or household. According to this definition, income should include not only the returns for the use of labor and capital, but also any other rents produced by the possession of durable goods, such as houses or cars. Families living in their own dwellings implicitly receive a flow of income equivalent to the market value of the service that the use of this property represents for them. This remuneration should be computed as part of household income, even though it is never recorded in a formal market.

In some surveys owners are asked to estimate the rent they would have to pay if they had to rent the houses they occupy. The answers to this question are used to impute rents to own-housing, although issues of reliability in the answers are usually raised, in particular in areas where housing markets are not well developed. In those surveys where this information is not available or is clearly unreliable we increase household income of housing owners by 10%, a value that is consistent with estimates of implicit rents in the region.

### **Non-response and missing incomes**

Not all the individuals selected to respond the household survey answer all the questions. Non-response is more usual for the income questions. The existence of missing incomes can bias the statistics if non-response is correlated with income. Even when that occurs, trends may not be affected if people who do not answer the income questions share similar characteristics.

The typical way to alleviate this problem is imputing earnings to non-respondents by matching techniques or by applying the coefficients of a Mincer equation. This equation can be estimated by OLS, Heckman ML, quantile regression or other methods, and its specification depends on the information available in the survey. The practice of imputing incomes faces two kind of problems. On the one hand the  $R^2$  of the Mincer equations are typically low: many determinants of earnings are non-observable (*i.e.*, individual talent) or often not measured (*i.e.*, education quality), and then imputations may imply large errors. This problem is particularly severe for capital income, which is the source with higher rates of non-response. The matching techniques face the same drawback. The

second problem is related to the reliability of the imputation procedure. As stated above, several decisions should be taken to implement an adjustment. The researcher should choose an estimation procedure, pick the dependent and independent variables, select a method for imputing error terms, and so on. Even when all the steps are clearly documented, people may be suspicious on the way the data is treated and on the choice of a particular imputation strategy. Working with the raw data has the advantage of more transparency.

In this version of the SEDLAC we compute the statistics with the official datasets, as it has been done in most academic and official studies.<sup>6</sup>

Suppose income from source  $s$  is missing for individual  $i$ . Should we record as missing that individual's total income? If we take that alternative, should we in turn record as missing the total income of individual  $i$ 's household? We make the following (necessarily arbitrary) decisions. If  $s$  is not the main source of income for  $i$ , then we compute the individual total income ignoring source  $s$ .<sup>7</sup> If instead  $s$  is the main source, we record total income as missing. This alternative has the advantage of not dropping from the datasets individuals who do not respond questions on income sources of secondary importance. The cost to be paid is the income under-estimation for these individuals. Regarding household income, we record it as missing if the household head's total income is missing. Otherwise, we compute household income assigning zero income to non-heads with missing income.

### Zero income

In many surveys a non-negligible fraction of the working population reports zero income. This answer can be the consequence of different situations: (i) the individual in fact did not earn any income during the period covered by the survey, (ii) she earned only non-monetary income, which is not recorded in the survey, or (iii) she misreported income.

Household monthly income is used as a proxy for well-being. As argued above, one of the main caveats arises from the volatility of monthly income compared to well-being. The case of zero income is probably the most illustrative one. While people may have zero household income in one particular month, that cannot be a good proxy for their well-being, insofar her family can get (monetary or non-monetary) income from some

---

<sup>6</sup> In some countries these datasets include some basic imputations for income non-response (e.g. Chile and Paraguay).

<sup>7</sup> Most household surveys identify the main individual source of income, even when the individual does not answer the income questions. In those cases where also that information is missing we assume that labor is the main income source for adults.

sources (charity, transfers, savings, etc.). For that reason zero income is a particularly important case of either misreport or failure in the proxy for well-being.

At this stage we compute the tables accepting zero income for poverty statistics and ignoring them for inequality indicators, as it is mostly done in academic papers.<sup>8</sup> The differential treatment of zero income for poverty and inequality arises from the assumption that zero household income mostly comes from households with all members unemployed, and/or from misreport by low-income people who forget or are not asked to report some income sources (*e.g.* charity, in-kind payments). Under this assumption zero income respondents should be considered as poor. However, some inequality measures collapse when considering zero income. Inequality indicators are scale invariant and then rely on proportional income differences. Accepting zero income implies dividing by zero, which generates computational problems. Given this fact, and the likely unreliability of zero household income, families who report zero income are usually ignored when computing inequality indicators.

Suppose income from source  $s$  is reported to be zero for individual  $i$ . If  $s$  is not identified as the main source of income for  $i$ , then we compute the individual total income ignoring source  $s$ , even when there is information that lead us to believe that the individual should have had some income from source  $s$  (*e.g.* positive hours of work). If instead  $s$  is the main source, we record total income as missing. There is one exception: if the individual identifies herself as a “worker with no earnings”, we keep income as zero. Regarding household income, we record it as missing if the household head’s total income is reported as zero.

The above discussion applies to the case of negative income, which can be the consequence of negative profits in an economic activity. Although very infrequent, some farmers in the household surveys report negative income. So far we have treated negative incomes as missing incomes.

### **Unreliable income**

Some income responses are clearly unreliable. Unreliable income may be the consequence of measurement errors or the deliberate misreport of income. Some NSOs identify inconsistent answers in the dataset based on their expertise. When that occurs we accept these decisions. As in the case of missing income, in this stage of the project we also accept the NSOs imputations replacing unreliable answers, but will implement our own imputation procedure in a later stage.

---

<sup>8</sup> See for instance the discussion in Székely and Hilgert (1999).

In some special cases we also exclude observations that seem clearly unreliable, even when the NSO does not do so. For instance, in Paraguay 1995 the highest income reported by one individual is deemed to be implausible by experts who have worked with Paraguay's data (see Székely and Hilgert, 1999). In these cases we mark the observations as inconsistent in the do file, and ignore them in the analysis. These observations should be clear outliers, and its number should be very small.

### **Under-reporting**

Under-reporting can be the consequence of the deliberate decision of the respondent to misreport, or to the absence of questions to capture some income sources, or to the difficulties in recalling or estimating income from certain sources (earnings from informal activities, in-kind payments, home production, capital income). This problem likely implies a downward bias on the measured living standards of poor people, who rely on a combination of informal activities and/or production for own consumption, and of rich people who derive a larger proportion of income from non-labor sources and are probably more prone to under-report. Differential misreporting behavior among respondents and differential efforts in the survey design can distort comparisons across countries. If these behaviors and efforts change over time they can also distort trends.

Researchers apply three kinds of strategies to alleviate these problems. The first one is restricting the analysis to more homogeneous variables less affected by problems of misreporting. Typically people look at the distribution of labor income, or even more restricted, at the distribution of monetary wages from salaried work in urban areas. Of course, the cost of doing that is ignoring a sometimes sizeable part of the overall income distribution.

The second strategy is applying some grossing-up procedure. Income from a given source in the household survey is adjusted to match the corresponding value in the National Accounts. This adjustment usually leads to inflating capital income relatively more than the other income sources. It relies on the dubious assumptions that data from national accounts is error-free (Deaton, 2003).

Finally, a third strategy is estimating underreported incomes from other pieces of information in the survey. Mincer regressions can be run to estimate wages for workers who clearly misreport wages but reliably report individual characteristics.

Researchers who have done different types of adjustments have generally found that most results for inequality trends are robust. Gasparini and Sosa Escudero (2001) for instance

found that the measured increasing trend in income inequality in Argentina is robust to the three types of adjustments mentioned above. Cross-country results are somewhat less robust to methodological changes. Székely and Hilgert (1999) find that some inequality rankings among LAC countries vary as they perform a wide range of adjustments. The results for poverty vary across countries, in some cases underreporting adjustments significantly lower poverty estimates.

As in the case of non-response, so far we have computed statistics with the raw data as in most academic and official studies. In the near future we will produce a report on the robustness of some poverty and inequality measures to adjustments for non-response.

It is important to notice that we compute some statistics for a wide range of variables, some of which presumably have fewer problems of under-reporting (*e.g.* earnings for salaried formal urban workers). Users may restrict the comparisons to these variables if they are particularly worried about under-reporting issues.

### **The case of Chile**

In Chile in order to alleviate under-reporting problems incomes from the household survey (CASEN) are adjusted to match some National Accounts figures. Unfortunately, for this study we could not completely undo these adjustments to make Chile comparable to the rest of the countries. Pizzolitto (2005) reports that income growth, poverty and inequality patterns are robust to these adjustments.

### **Absence of “very rich” people**

A common observation among users of household surveys is that they do not typically include “very rich” individuals: millionaires, rich landlords, powerful entrepreneurs and capitalists do not usually show up in the surveys. The highest individual incomes in LAC surveys mostly correspond to urban professionals. This fact can be the natural consequence of random sampling (there are so few millionaires that it is unlikely that they are chosen by a random sample selection procedure to answer the survey), non-response, or large under-reporting. The fact is that rich people in the surveys are “highly educated professionals obtaining labor incomes, rather than capitalist owners living on profits” (Székely and Hilgert, 1999). The omission of this group does not affect poverty estimates but surely implies an underestimation of inequality of a size difficult to predict. Studies for other regions have used tax information to estimate income for rich individuals (Piketty and Saez, 2003). At least in this early stage we do not plan to perform any adjustment to alleviate this problem.

## Real and nominal income

Real rather than nominal incomes should be used in any distributional analysis. However, if prices faced by all households were the same, the distinction would be irrelevant. But prices usually differ by location: if two households located in different regions have the same nominal income but face different prices they will have different living standards. Despite many authors have highlighted the importance of considering spatial variations of prices in a distributional study (*e.g.* Deaton (1997), Ravallion and Chen (1997)), price adjustments are rarely performed in countries that do not routinely collect information on local prices as part of the household survey. Unfortunately, most LAC countries fall into this category. All countries have some regional price study, which does not completely solve the problem, since price dispersion may be high within a region, especially between urban and rural areas. More important, these studies substantially differ in methodology and results across countries.

In this database all rural incomes are increased by a factor of 15% to capture differences in rural-urban prices. That value is an average of some available detailed studies of regional prices in the region. Although certainly arbitrary, we believe this alternative is better than (i) ignoring the problem of regional prices altogether, or (ii) using for each country the available price information, despite the enormous differences in methodology, scope, and results.

Another problem arises in those countries where the survey is carried out during several months. If there is inflation, nominal incomes reported in different months should be deflated to make them comparable. In all countries where this happens, we use the official consumer price index to adjust nominal incomes.

## Equivalence scales

Individuals usually live in households and share a common budget. This fact implies that an individual's well-being depends on the resources available in the household and on the size, structure and sharing rule within the household. Probably, the most common indicator of individual well-being is household per capita income: household total income divided just by the number of persons in the household. Although widely used, this variable ignores three relevant factors: (i) consumption economies of scale within the household, that for instance allow a couple to live with less than double the budget of a person living alone, (ii) differences in needs among individuals, basically as a function of age and gender (these differences are behind the adjustments for *adult equivalents*), and

(iii) unequal allocations of resources within the household.<sup>9</sup> While points (i) and (ii) are considered in this stage of the project, we leave point (iii) for potential exploration in future stages of the project, given the scarce data on this issue.

Define an individual's *equivalized* household income as total household income divided by  $(A + \alpha_1.K_1 + \alpha_2.K_2)^\theta$ , where  $A$  is the number of adults,  $K_1$  the number of children under 5 years old, and  $K_2$  the number of children between 6 and 14. Parameters  $\alpha$  allow for different weights for adults and kids, while  $\theta$  regulates the degree of household economies of scale. Deaton and Zaidi (2002) suggest intermediate values of the  $\alpha$ s ( $\alpha_1=0.5$  and  $\alpha_2=0.75$ ), and a rather high value of  $\theta$  (0.9) for countries like those in the region. We take that as the benchmark case, but also experiment with alternative vectors of parameters. Although it would probably be more correct to assign different parameters to LAC countries in different states of development, we prefer to use the same scale across countries in each simulation, for transparency in the comparisons.

---

<sup>9</sup> See Buhmann, Rainwater, Schmaus and Smeeding (1988) and Deaton (1997) for discussions on these points.

### 3. POVERTY

The first part of this section reproduces poverty figures from official country sources, while the second part reports our own poverty estimates.

#### 3.1. Poverty from official sources

##### *Country estimates*

The file *poverty\_official\_LAC.xls* reports poverty estimates from official sources (mostly National Statistical Offices) in each country. The file has 4 sheets

##### *Source*

- Source of information for official poverty estimates.

##### *National*

- Poverty headcount ratio. National estimates. Extreme and moderate poverty. Individuals and households (wherever available)

##### *urb-rur*

- Poverty headcount ratio. Urban and rural estimates. Extreme and moderate poverty. Individuals and households (wherever available)

##### *region*

- Poverty headcount ratio. Individuals. Moderate poverty. Regional estimates.

#### 3.2. Poverty using CEDLAS methodology

This section includes poverty estimates calculated with CEDLAS methodology; *i.e* using the income constructed with the methodology described in section 2 as the welfare indicator. The file *poverty\_LAC.xls* shows own poverty estimates at the country level, while the file *poverty\_LA\_cedlas.xls* presents poverty headcount ratios for Latin America and its regions, using the USD-a-day international lines.

##### *Country estimates*

The file *poverty\_LAC.xls* has several sheets:

##### *poverty USD 2.5*

- Headcount ratio, poverty gap and FGT(2) using the USD-2.5-a-day poverty line. National, rural and urban statistics.

#### *intervals USD 2.5*

- Standard error, coefficient of variation and 95% confidence interval for the headcount ratio using the USD-2.5-a-day poverty line. National, rural and urban statistics. Computed by bootstrap with 100 replications.

#### *poverty USD 4*

- Headcount ratio, poverty gap and FGT(2) using the USD-4-a-day poverty line. National, rural and urban statistics.

#### *intervals USD 4*

- Standard error, coefficient of variation and 95% confidence interval for the headcount ratio using the USD-4-a-day poverty line. National, rural and urban statistics. Computed by bootstrap with 100 replications.

#### *poverty median*

- Headcount ratio, poverty gap and FGT(2) using the 50% of the median poverty line. National, rural and urban statistics.

#### *intervals median*

- Standard error, coefficient of variation and 95% confidence interval for the headcount ratio using the 50% of the median poverty line. National, rural and urban statistics. Computed by bootstrap with 100 replications.

#### *regions\_2.5*

- Poverty headcount ratio. USD-2.5-a-day poverty line. By region.

#### *regions\_4*

- Poverty headcount ratio. USD-4-a-day poverty line. By region.

#### *NBI\_1*

- Poverty headcount ratio computed using an indicator of basic needs (NBI) (see below). Urban areas. Share of people meeting each of the criteria used to define poverty as NBI. Proportion of all NBI criteria met by the poor.

#### *NBI\_2*

- Poverty headcount ratio computed using an indicator of basic needs (NBI) by equalized income quintiles. Proportion of all NBI criteria met by quintiles. Urban areas.

#### *NBI\_3*

- Poverty headcount ratio computed using a combined indicator of basic needs and income (see below). Urban areas.

#### *Regional estimates*

The file *poverty\_LA\_cedlas.xls* reports poverty estimates for Latin America and its regions. That file has the following sheets

#### *table\_PL2.5*

- Poverty headcount ratio for Latin America and its regions. USD-2.5-a-day poverty line.

#### *table\_PL4*

- Poverty headcount ratio for Latin America and its regions. USD-4-a-day poverty line.

#### *figure\_1*

- Poverty headcount ratio. Weighted and unweighted means for LA.

#### *figure\_2*

- Poverty headcount ratio by regions. Weighted and unweighted means.

#### *Methodological issues*

Most of the poverty statistics shown in the SEDLAC are on *income poverty*, defined as the inability of achieving a certain minimum income level, known as the poverty line (PL). Even within this limited framework, the measurement of poverty faces two main problems: the identification and the aggregation of the poor. The identification problem is certainly the most difficult to solve. There are neither normative nor objective clear

arguments to set a line below which everybody is poor and above which everyone is non-poor (Deaton, 1997). Despite this central conceptual ambiguity, reducing poverty is a deliberate policy objective for governments around the world. The international community has embraced this goal as reflected in the first Millennium Development Goal (halving poverty by 2015), the Poverty Reduction Strategy Papers, and the World Bank Poverty Assessments.

Since there is a fundamental arbitrariness in defining poverty, different authors and agencies use different poverty lines. There is a wide range of poverty estimates across studies for each country. The SEDLAC includes a set of poverty estimates based on (i) international poverty lines (USD 2.5 and 4 a day at PPP), (ii) national poverty lines (extreme and moderate), and (iii) relative poverty lines (50% of median income). Using a range of lines is especially relevant given the arbitrariness in the definition mentioned above. While the measurement of poverty with national lines takes into consideration that societies differ in the criteria used to identify the poor, the international lines are unavoidable instruments to compare absolute poverty levels and trends across countries, and provide regional and world poverty counts.

#### *International lines*

The USD 1 a day at PPP prices is an international poverty line meant to define an international norm to gauge at the inability to pay for food needs. The USD1-a-day line was proposed in Ravallion *et al.* (1991) and used in World Bank (1990). Originally, it was a value measured in 1985 international prices and adjusted to local currency using purchasing power parities (PPP) to take into account local prices. The USD 1 standard was chosen as being representative of the national poverty lines found among low-income countries. The line was recalculated in 1993 PPP terms at USD1.0763 a day (Chen and Ravallion, 2001). Recently, this basic line was set at USD 1.25 a day at 2005 PPP (Ravallion, Chen and Sangraula, 2008).

The USD-1.25 line seems too low for LAC, a region comprised of mostly middle-income countries. For that reason we compute poverty with the USD-2.5-a-day line, which coincides with the median value of the extreme poverty lines officially set by the LAC governments. We also include poverty estimates using a USD-4-a-day line, which is close to the median value of the moderate poverty lines officially set by governments in the region. These values are multiplied by 30.42 to get monthly poverty lines.

Since 2009 we use the new PPP computed in the 2005 round of the *International Comparison Program* (World Bank, 2008). Due to the use of this set of prices, poverty in some countries is significantly different from previous estimations. The main case is

Colombia, where poverty looks very high with the new PPPs. We are currently studying this issue.

Although the USD lines have been criticized, their simplicity and the lack of reasonable and easy-to-implement alternatives have made them the standard for international poverty comparisons.<sup>10</sup> For instance, the United Nations' Millennium Development Goal 1 – eradicate extreme poverty and hunger – is stated in terms of USD-1-a-day poverty – halving between 1990 and 2015 the proportion of people whose income is less than USD 1 a day.

Following Deaton (2003) and WDI (2004) we compute the poverty line for 2005 in local currency units using the PPP adjustment, and then take that value to the month(s) of a given survey using the national consumer price index of the country. We apply this line to a homogeneous definition of household per capita income across countries/years that includes all the ordinary sources of income and estimates of the implicit rent from own-housing (see section 2). Of course, even when we follow the same procedure, since household surveys differ across countries, we may end up with non-strictly comparable variables.

It is well known that household consumption is a better proxy for well-being than household income.<sup>11</sup> Despite this dominance, nearly all comparative poverty studies in LAC use income as the well-being indicator. A simple reason justifies this practice: few countries in the region routinely conduct national household surveys with consumption/expenditures-based questionnaires, while all of them include questions on individual and household income. Some authors and agencies adjust average income to accord with consumption data from national accounts to estimate poverty (ECLAC, 2003; Wodon, 2000; WDI, 2002). However, it is not clear that the adjustment for consumption increases comparability, since the reliability of national accounts need not be greater than the reliability of household surveys. Deaton (2003) strongly argues for the use of only survey data to compute poverty, as adjustments to match national accounts “tend to overstate the reduction of poverty over time, and to exaggerate poverty differences across countries”. In this database we do not perform any adjustment to compute poverty from household surveys. WDI (2003) reports that poverty measures based on consumption and those based on income without adjustment do not significantly differ, due to two effects that roughly cancel each other out: mean income is higher than mean consumption, but income inequality is higher than consumption inequality. For this reason, since 2003 the

---

<sup>10</sup> See Srinivasan (2004), Kakwani (2004) and Ravallion (2004) for a discussion on the merits and demerits of the USD-1-a-day line.

<sup>11</sup> See for instance Deaton and Zaidi (2003).

WDI reports poverty statistics computed without adjusting income to match consumption from national accounts.

### *National lines*

Most LAC countries have national extreme poverty lines which are mostly based on the cost of a basic food bundle, and moderate poverty lines computed from the extreme lines using the Engel/Orchansky ratio of food expenditures. This methodology is also used by ECLAC, which in some cases help governments to implement the national poverty lines. Despite some similarities, methodologies for national poverty estimates substantially differ across nations. Some countries use expenditures (*e.g.* Mexico), others use incomes (*e.g.* Argentina) and others a mix of income and expenditures (*e.g.* Bolivia).

In the file [poverty\\_official\\_LAC.xls](#) we present extreme and moderate poverty headcount ratios reported by official sources in each country. We identify the source of information, and report the published poverty statistics.

### *Relative lines*

Some countries (*e.g.* those in the European Union) use a relative rather than an absolute measure of poverty. According to this view, since social perceptions of poverty change as the country develops and living standards go up, the poverty line should increase along with economic growth. Probably the most popular relative poverty line is the one set at 50% of the median of the household per capita income distribution. As the economy grows this line increases, and poverty is more likely to increase than with a fixed poverty line.

The second stage in poverty measurement –the aggregation– is relatively simpler. In this paper for each poverty line we compute three of the most used poverty indicators: the headcount ratio, the poverty gap, and the FGT (2).<sup>12</sup> In all cases we compute poverty based on the distribution of individuals, not households.

### **Basic needs (NBI)**

Income poverty measures have two important limitations. First, as *monthly* income is used as the welfare indicator, some people are incorrectly considered as poor if having a temporary negative shock, or due to seasonality reasons. Second, there are convincing

---

<sup>12</sup> See Foster, Greer and Thornbecke (1984) for references.

arguments for considering poverty as a multidimensional issue.<sup>13</sup> Insufficient income is just one of the manifestations of a more complex phenomenon.

These two reasons lead to the search for other variables to measure poverty. Researchers, agencies and NSOs have used different measures of housing, education, health, employment and access to social services to define non-income or *structural* poverty. Given current practices in some countries and the availability of information in all LAC countries, we construct an indicator of poverty according to the following conditions:

- (i) more than 4 persons per room
- (ii) the household lives in “poor” places (e.g. street, shanty towns)
- (iii) the dwelling is made of low-quality materials (see section 7)
- (iv) the dwelling does not have access to water
- (v) the dwelling does not have an hygienic restroom (see section 7)
- (vi) there are children aged 7 to 11 not attending school
- (vii) the household head does not have a primary school degree
- (viii) the household head does not have a high-school degree, and there are more than 4 household members for each income earner.

All persons in a household are considered as poor if they meet at least one of the above conditions. This indicator is similar to the popular NBI (Necesidades Básicas Insatisfechas) indicator. We also combine this approach with that of income poverty by simultaneously using the NBI indicator with the USD 2.5-a-day poverty measure: only if an individual is poor under both criteria, she is considered as “chronically” poor.

We restrict the analysis to urban areas, since arguably the conditions for the NBI indicator should differ between urban and rural areas (*e.g.* access to sanitation).

### Sample variability

Measures of the different dimensions of a distribution are subject to sample variability problems, since they come from surveys, not census data. If our sample size were only two individuals, indicators would surely widely vary over time, even when the population remained completely unchanged, as we randomly selected two different individuals each year. This problem is alleviated with larger samples but it is not completely eliminated. We tackle this point by computing standard errors and confidence intervals for some poverty and inequality indicators. Standard errors are estimated by bootstrapping

---

<sup>13</sup> Bourguignon (2003) discusses the need and the problem of going from income poverty to a multidimensional approach of endowments. Attanasio and Székely (eds.) (2001) show evidence of poverty as lack of certain assets for LAC countries.

techniques, which provide interval estimations and dispersion measures for the distributional indicators in a simple and efficient way.<sup>14</sup> For a given indicator we compute its bootstrapped standard error, the coefficient of variation, and the corresponding confidence interval for a 95% of significance.

### **Latin American means**

The file *poverty\_LA\_CEDLAS.xls* reports the poverty headcount ratio for Latin America and its regions (Mercosur, Andean region, Central America and Mexico) using two alternative poverty lines – 2.5 and 4 USD a day. The estimates are for 1992, 1998, 2003 and 2008. For the countries where no observation is available in some time period, we estimate it from other sources (*e.g.* official reports of poverty figures) or by applying poverty-growth elasticities.<sup>15</sup> Average statistics are shown both weighting and not weighting by country population.

Latin America is divided into three regions:

*Mercosur*: Argentina, Brazil, Chile, Paraguay, Uruguay

*Andean region*: Bolivia, Colombia, Ecuador, Peru, Venezuela

*Central America*: Costa Rica, El Salvador, Honduras, Guatemala, Mexico, Nicaragua, Panama and Dominican Republic.

---

<sup>14</sup> The implementation of the bootstrap method will follow Sosa Escudero and Gasparini (2000). For more theoretical references on the subject see Biewen (2002), Davidson and Duclos (2000) and Mills and Zandvakili (1997).

<sup>15</sup> Details available upon request.

## 4. INEQUALITY

This section contains one file (*Inequality\_LAC.xls*) with several tables:

### *deciles pci*

- Share of deciles and income ratios. Distribution of household per capita income.

### *indices pci*

- Inequality indices. Distribution of household per capita income.

### *intervals pci*

- Standard error, coefficient of variation and 95% confidence interval for the Gini coefficient of the household per capita income distribution. Computed by bootstrap with 100 replications.

### *deciles ei*

- Share of deciles and income ratios. Distribution of household equivalized income.

### *indices ei*

- Inequality indices. Distribution of household equivalized income.

### *intervals ei*

- Standard error, coefficient of variation and 95% confidence interval for the Gini coefficient of the household equivalized income distribution. Computed by bootstrap with 100 replications.

### *deciles lmi*

- Share of deciles and income ratios. Distribution of household equivalized labor monetary income in urban areas.

### *indices lmi*

- Inequality indices. Distribution of household equivalized labor monetary income in urban areas.

### *intervals lmi*

- Standard error, coefficient of variation and 95% confidence interval for the Gini coefficient of the household equivalized labor monetary income distribution in urban areas. Computed by bootstrap with 100 replications.

### *deciles ni*

- Share of deciles and income ratios. Distribution of income/consumption variable used for poverty estimates with national lines.

### *indices ni*

- Inequality indices. Distribution of income/consumption variable used for poverty estimates with national lines.

### *intervals ni*

- Standard error, coefficient of variation and 95% confidence interval for the Gini coefficient of the income/consumption variable used for poverty estimates with national lines. Computed by bootstrap with 100 replications.

### *gini1*

- Gini coefficient. Distribution among individuals of different household income variables
  - per capita income
  - equivalized income (different scales)
  - total household income
  - equivalized income restricted to age groups

### *gini2*

- Gini coefficient. Distribution among individuals of different household income variables
  - Household per capita income. Only urban areas
  - Household per capita income. Only rural areas
  - Household equivalized income. Only urban areas
  - Household equivalized income. Only rural areas
  - Household per capita income. Includes income only from labor sources.
  - Household per capita income. Includes only monetary income
  - Household per capita income. Includes only monetary income from labor sources.
  - Household per capita income. Includes only monetary income from labor sources in urban areas.

### *gini3*

- Gini coefficient. Distribution among individuals of different household income variables. Including and excluding zeros from the computations.

### *polarization*

- Bipolarization indices computed over the distribution of household per capita income, household equivalized income and the income/consumption variable used for poverty estimates with national lines (“official income”). Esteban, Gradin and Ray and Wolfson indices.

## *Methodological issues*

### **Deciles**

All deciles have 10% of the population. Deciles have an equal number of individuals (not households).

### **Income ratios**

10/1: mean income of decil 10/mean income of decil 1

90/10: mean income of percentile 90/mean income of percentile 10

95/80: mean income of percentile 95/mean income of percentile 80

### **Indices**

The tables show statistics on the following inequality indices (see Lambert, 1993 or Cowell, 1995 for references)

- the Gini coefficient
- the Theil index
- the coefficient of variation
- the Atkinson index with parameters 0.5, 1 and 2
- the generalized entropy index with parameters 0 and 2 (The Theil index is the entropy with parameter 1).

### **Equivalized income**

Equivalized household income is obtained by dividing total household income by  $(A + \alpha_1 \cdot K_1 + \alpha_2 \cdot K_2)^\theta$ , where A is the number of adults, K1 the number of children under 5

years old and  $K2$  the number of children between 6 and 14. In the benchmark case we take  $\alpha_1=0.5$ ,  $\alpha_2=0.75$  and  $\theta=0.9$ .

In the sheet *gini1* we compute equivalized income with different scales:

A =  $\theta=0.9$ ,  $\alpha_1=0.5$ ,  $\alpha_2=0.75$  (the benchmark case)

B =  $\theta=0.75$ ,  $\alpha_1=0.5$ ,  $\alpha_2=0.75$

C =  $\theta=0.9$ ,  $\alpha_1=0.3$ ,  $\alpha_2=0.5$

D =  $\theta=0.75$ ,  $\alpha_1=0.3$ ,  $\alpha_2=0.5$

E = Amsterdam scale ( $\theta=1$ , males aged 14 to 17=.98, females older than 14=.9 and children under 14=.52)

## Polarization

We compute for each country/year two bipolarization indices: the Wolfson Index, which cuts the distribution at the median income, and the EGR Index, which finds the optimal income cut-off.<sup>16</sup> The table *polarization* shows the results for these bipolarization measures for the distribution of household per capita income, the distribution of equivalized household income and the distribution of the income/consumption variable used for poverty estimates with national lines.

---

<sup>16</sup> EGR refers to Esteban, Gradín and Ray (1999).

---

## **BOX 1: Grouping the population**

The following sections show statistics on a wide range of variables. These statistics are usually presented by dividing the population in groups. In this box we list the main groups used throughout the tables.

### *Equivalized income quintiles*

All individuals are sorted according to their equivalized household income (*i.e.* total household income adjusted for demographics – see the benchmark case above for details) and divided in 5 groups of equal size (each one with 20% of the population). Quintile 1 includes the poorest 20% of the population. Only individuals with valid equivalized income are included in the tables.

### *Education*

Three groups are formed according to years of formal education. Low=0 to 8 years, Medium=9 to 13 years, and High=more than 13 years.

### *Area*

The population is divided into urban or rural, according to the classification made by the National Statistical Office included in the household survey original dataset.

We also present the information grouped by gender and age. Definitions are straightforward.

---

## 5. DEMOGRAPHICS

This section contains two files: *demographics\_LAC.xls* and *regions&migration\_LAC.xls*

The file *demographics\_LAC.xls* has the following sheets

### *hh size*

- Household size: number of persons in a household, excluding “external” members (renters, domestic servants and their families).
- Tables by equivalized income quintiles and by education of the household head.

### *children*

- Number of children under 12 years per household
- Tables by parental income quintiles and by education of the household head.
- “Parental income” is total income earned by the household head and her/his spouse. In most cases (but not in all) they are the parents of the child.
- Children of “external” members of the household (renters, domestic servants) are excluded.

### *dependency*

- Dependency rate: household size over the number of income earners in the household.
- Tables by equivalized income quintiles and by education of the household head.

### *age structure*

- Proportion of total population in each age bracket: [0,14], [15,24], [25,40], [41,64], and [65+).

### *age*

- Mean age by equivalized income quintile.

### *mating*

- Indicators of assortative mating: linear correlation coefficients between couples in years of education, hourly wages and hours of work.
- Analysis restricted to household heads and their spouses.
- Hourly wages are total labor income from the main activity divided by hours of work in that activity. The correlation coefficient is computed for all couples in which both members have positive wages.
- The two last columns in the table report the correlation coefficient for total hours of work (in all activities). The correlation coefficient in the last column is computed for all couples in which total worked hours are positive.

The file *regions&migrations\_LAC.xls* has the following sheets

### *area*

- Share of population in rural and urban areas. The information comes from household surveys, not Census data. Shares are obtained using population weights (as in all statistics in the database). The definition urban-rural is the one adopted by each National Statistical Office.

### *regions*

- Share of population (covered in the survey) in each region. The information comes from household surveys, not Census data. Shares are obtained using population weights. The regions are those used by each National Statistical Office.

### *migration*

- Share of migrants in the population. In most countries *migrants* are defined as those individuals not born in the city where they are currently living. Share of foreign migrants and new-comers. Characterization of migrant population by income, education and area.

## 6. EDUCATION

This section has four files with information on years of education, literacy, school enrollment and educational mobility.

File 1: *years\_edu\_LAC.xls*

Includes several sheets with information on years of formal education.

### *structure*

- Educational structure: share of adults aged 25-65 in each skill group (defined in terms of years of education).
- Skill groups: Low=0 to 8 years of formal education, Medium=9 to 13 years, and High=more than 13 years.
- Tables for all adults, males, females, working males and working females.

### *years*

- Distribution of years of formal education in the adult population (aged 25 to 65)

### *age\_gender*

- Average years of formal education by age and gender group.

### *income*

- Average years of formal education of adult population (aged 25 to 65) by equivalized income quintile.

### *income\_age*

- Average years of formal education by age group and equivalized income quintile.

### *area*

- Average years of formal education by area (urban-rural). Tables for all population, adults (25-65), male adults and female adults.

### *inequality*

- Gini coefficient for the distribution of years of formal education. Tables by age groups.

File 2: *literacy\_LAC.xls*

It includes three sheets. We consider an individual as illiterate if she reports not having reading or writing skills (see below).

*age\_gender*

- Literacy rates by age and gender.

*income*

- Literacy rates by equivalized income quintiles. Youths (15-24) and adults (25-65).

*area*

- Literacy rates by area (urban-rural) and age group.

File 3: *enrollment\_LAC.xls*

This file contains several sheets with information on gross and net school enrollment rates. Gross rates refer to the share of a given population attending any educational level. Net rates refer to the share of individuals in a certain educational level age who are actually attending that level.

*gender*

- Gross enrollment rates by age group and gender

*income*

- Gross enrollment rates by age group and equivalized income quintile

*area*

- Gross enrollment rates by age group and area (urban-rural)

*primary*

- Net enrollment rate in primary school by gender, equivalized income quintile and area

*secondary*

- Net enrollment rate in secondary school by gender, equivalized income quintile and area

### *tertiary*

- Net enrollment rate in superior education by gender, equivalized income quintile and area

### *primary completion*

- Primary completion rates, *i.e.* share of youths aged 15-24 with a primary school degree. Tables by gender, quintiles and areas.

### *public*

- Share of students in public schools. By educational level. Tables by quintiles and areas.

File 4: *mobility\_LAC.xls*

### *mobility*

- Educational Mobility Index for teenagers (13 to 19) and young adults (20 to 25) (see below for details).

## ***Methodological issues***

### **Years of education**

Most surveys record years of formal education. In those cases where this variable does not exist it is possible to approximate it from the maximum educational level attained by the person and her age. For some countries the distribution is truncated in the upper tail, since surveys do not ask on years of graduate education (e.g. Argentina).

### **Skill groups**

We divide the adult population into three groups according to years of schooling: low education (less than 9 years of education), medium (between 9 and 13 years of education) and high (more than 13). Given that the educational systems differ across countries and over time, we may be including people with different educational levels (*e.g.* secondary incomplete and complete) into the same category. We consider that an acceptable cost to be paid in order to present simple statistics of educational levels in a similar format for all countries/years.

## Gini coefficients

We compute the Gini coefficient on the distribution of years of education as one measure of inequality in education. The use of educational-Ginis has been increasing in the last years.<sup>17</sup> However, it is not obvious that inequality in education should be measured by an index of *relative* rather than *absolute* differences among individuals.<sup>18</sup>

## Literacy

Nearly all LAC surveys report whether the individual has reading and writing skills. We consider the individual as illiterate if she reports not having reading or writing skills (or both). If a specific literacy question is missing in the survey, we consider as illiterate a person who has less than two years of formal education.

## Educational levels

A typical Latin American formal educational system starts at the age of 6. There are around 7 years of primary school and 5 years of high school. The duration of the college education depends on the field and the college type, usually varying between 3 and 6 years. Most countries depart from this benchmark in different directions. In some countries formal education starts at the age of 5. In others primary school is longer and high school shorter. In these cases primary school is usually divided into three sections.

When making decisions on the allocation of individuals to educational levels we prefer to keep within-country comparability at the cost of losing cross-country comparability. If the educational system of a country has varied, we make efforts to use the same educational classification of the latest survey available. The classification process will be documented in a report to be included in the web page.

## Enrollment rates

School enrollment rates can be defined in alternative ways. A simple definition of enrollment implies computing the share of people in an age group who are attending school, regardless of the educational level they are attending. Although imperfect, this indicator of *gross* school enrollment provides useful information on those who are in the formal education system. Statistics for gross enrollment rates for children and youngsters from 3 to 23 years-old are presented in the tables *gender*, *income* and *area* in *enrollment\_LAC.xls*.

---

<sup>17</sup> For instance, Thomas, Wang and Fan (2002) calculate Ginis over the distribution of years of education for 140 countries in the period 1960-2000.

<sup>18</sup> The Gini coefficient, as most of the inequality indices, is scale-invariant (see Lambert, 1993).

Another possibility is to report the share of people in a certain age group who are attending the educational level that corresponds to their age (*net enrollment*). This alternative faces several problems, since there is not a unique mapping from age to educational levels (e.g. in Argentina youngsters aged 18 could be in college or in technical high schools). Second, as it was mentioned above educational systems differ across countries, and even within countries. For instance, a young Argentinean of 14 years old should be in high school if she lives in Buenos Aires city, but in the EGB (a 9-year program that starts at the age of 6) if she lives in the Province of Buenos Aires. Despite these problems we tried to compute measures of net enrollment in primary and secondary school with the information in the household surveys. The sheets *primary*, *secondary*, and *tertiary* show statistics on net enrollment.

### **Educational mobility**

Statistics on educational mobility are computed following the methodology developed in Andersen (2001). The dependent variable is the schooling gap, defined as the difference between (i) years of education that a child would have completed had he entered school at normal age and advanced one grade each year, and (ii) the actual years of education. In other words, the schooling gap measures years of missing education. The Educational Mobility Index (EMI) is defined as 1 minus the proportion of the variance of the school gap that is explained by family background.<sup>19</sup> In an economy with very low mobility, family background would be important and thus the index would be near zero.<sup>20</sup>

---

<sup>19</sup> The regression, which includes several controls, is run for youths who live with their parents (they are the only ones with information of parental income and education).

<sup>20</sup> For technical details see Andersen (2001).

## 7. EMPLOYMENT

This section has three files: *employment\_LAC.xls*, *wages&hours\_LAC.xls* and *labor\_benefits.xls*

The file *employment\_LAC.xls* has the following sheets

### *labor force*

- Labor force participation by gender, age, education, and area.

### *employment*

- Share of adults employed by gender, age, education, and area.

### *unemployment*

- Unemployment rates by gender, age, education, and area.

### *duration*

- Duration of unemployment (in months) by gender, age, education, and area.

### *change*

- Share of workers willing to change employment and/or increase hours of work by gender, age, education, and area.

### *structure*

- Distribution of workers by gender, age, education and area.

### *stru\_region*

- Distribution of workers by region.

### *stru\_type*

- Distribution of workers by labor relationship, type of firm and labor category

### *stru\_sector*

- Distribution of workers by economic sector

### *informal\_1*

- Share of adults in informal jobs. Definition 1: Informal=salaried workers in small firms, non-professional self-employed and zero-income workers.

### *informal\_2*

- Share of salaried workers in informal jobs. Definition 2: Absence of social security rights.

### *child*

- Share of children aged 10-14 who are employed.

### *MDG*

- Share of women in non-agricultural employment.

The file *wages&hours\_LAC.xls* has the following sheets

### *wage\_1*

- Hourly wages in main activity in nominal LCU by gender, age, education, and area.

### *wage\_2*

- Hourly wages in main activity in nominal LCU by type of job.

### *wage\_3*

- Hourly wages in main activity in nominal LCU by informality status.

### *wage\_4*

- Hourly wages in main activity in nominal LCU by sector.

### *hours\_1*

- Weekly hours of work by gender, age, education, and area.

### *hours\_2*

- Weekly hours of work by type of job.

### *hours\_3*

- Weekly hours of work by informality status.

#### *hours\_4*

- Weekly hours of work by sector

#### *li\_1*

- Monthly labor income in LCU by gender, age, education, and area.

#### *li\_2*

- Monthly labor income in LCU by type of job.

#### *li\_3*

- Monthly labor income in LCU by informality status.

#### *li\_4*

- Monthly labor income in LCU by sector

#### *Ginis*

- Gini coefficients for the distribution of wages and earnings

#### *correlation*

- Linear correlation coefficients between hours of work and hourly wages

#### *wagegaps*

- Ratio of hourly wages by educational groups. Prime-age males.

#### *Mincer*

- Coefficients of educational dummies in Mincer equations. Males and females.

#### *Mincer2*

- Dispersion in unobservables and conditional gender wage gaps

The file *labor\_benefits\_LAC.xls* has the following sheets

#### *contract*

- Share of salaried workers with labor contracts by age, gender, education and area

*pensions*

- Share of salaried workers with right to pensions when retired by age, gender, education and area

*health*

- Share of salaried workers with right to health insurance linked to the job by age, gender, education and area

*others*

- Share of salaried workers with access to 13th month and holidays. Share of salaried workers unionized.

## Methodological issues

### Type of firm

Workers are classified into three groups according to whether they work in small firms, large firms or the public sector. The latter includes jobs in stated-owned firms, public schools, hospitals and other services, and public administration. The following table reports the cut-off point in each country/year to define a small private firm.

Country	# employees
Argentina	5
Bolivia	
1993-1997	5
2000-2002	4
Brazil	5
Chile	
1996	4
Rest of years	5
Colombia	n.a
Costa Rica	5
Dominican Rep.	4
Ecuador	5
El Salvador	
1991	n.a
Rest of years	5
Guatemala	5
Haiti	n.a
Honduras	5
Jamaica	4
Mexico	5
Nicaragua	5
Panama	4
Paraguay	5
Peru	5
Suriname	5
Uruguay	4
Venezuela	4

### Informality

There are at least two different concepts that are referred by the term *labor informality*. The “productive” definition pictures informal workers as those in low-productivity, unskilled, marginal jobs, while the “legalistic” or “social protection” definition stresses the lack of labor protection and social security benefits. The productive definition is concerned with the type of job (*e.g.* salaried vs. self-employed, large vs. small firms), while the legalistic definition is concern with the compliance of the labor relationship with some rules, mainly labor protection.

The empirical implementation of the productive notion of informality has been linked to (i) the type of job (salaried, self-employment), (ii) the type of economic unit (small, large, public sector), (iii) and the worker's skills. Following this practice we divide the working population into 7 groups:

1. Entrepreneurs (*patrones*)
2. Salaried workers in large private firms
3. Salaried workers in the public sector
4. Salaried workers in small private firms
5. Skilled self-employed
6. Unskilled self-employed
7. Zero-income workers

To implement this classification we include as *unskilled* all individuals without a tertiary or superior education degree, and we define as *small* all firms with 5 or fewer employees.<sup>21</sup> Given that an individual could have more than one job, we apply the classification only to his/her main occupation. We implement the following definition of labor informality:

**Definition 1 (productive definition):** *An individual is considered an informal worker if (s)he belongs to any of the following categories: (i) unskilled self-employed, (ii) salaried worker in a small private firm, (iii) zero-income worker.*

A second strand of the literature has stressed the “legalistic” or “social protection” notion of informality. Informal firms are those not complying with the norms in terms of labor contracts, labor taxes, and labor regulations, and then their workers have no rights to labor protection or social benefits linked to employment.

The right to receive a pension when retired is the social security benefit most asked in LAC household surveys. However, not all countries have questions on this item, and in those that have, questions are different (see table below). Moreover, in most countries the questions apply only to salaried workers, leaving all the self-employed as missing.

We implement the following legalistic/social-protection definition of informality:

**Definition 2 (legalistic or social protection definition):** *A salaried worker is informal if s(he) does not have the right to a pension linked to employment when retired.*

---

<sup>21</sup> Given differences in surveys, the cut-off point is not 5 employees in all countries. See the table above.

<b>Country</b>	<b>A worker is formal if she ..</b>
<b>Argentina</b>	has the right to a pension when retired
<b>Bolivia (since2000)</b>	is affiliated with a AFP (Administradora de Fondos de Pensiones)
<b>Brazil</b>	contributes to the Social Security system
<b>Chile</b>	is affiliated with any social security system
<b>Colombia (ENH)</b>	has the right to a pension when retired
<b>Ecuador (ECV)</b>	has the right to a pension when retired
<b>El Salvador</b>	is affiliated with any social security system (no information for domestic servants)
<b>Guatemala</b>	contributes to the IGSS (Instituto Guatemalteco de Seguridad Social)
<b>Mexico (since 2000)</b>	has the right to a pension when retired
<b>Nicaragua</b>	contributes to the INSS (Instituto Nicaragüense de Seguridad Social)
<b>Paraguay</b>	is affiliated with any social security system
<b>Peru (since 1999)</b>	is affiliated with any social security system
<b>Uruguay (since 2001)</b>	has the right to a pension when retired
<b>Venezuela</b>	
<b>1995-1998</b>	has the right to social benefits or social insurance IVSS
<b>2000-2003</b>	has the right to social benefits

## Mincer equations

We run regressions of the logarithm of the hourly wage in the main occupation for adults aged 25 to 55. As regressors we include educational dummies, age, age squared, an urban dummy and regional dummies. We estimate the model separately for men and women by Heckman maximum likelihood methods. The selection equation includes in addition school enrollment and the number of children. In the tables we report the marginal “returns” to completing each educational level.

The Mincer equation is also informative on two interesting factors: the role of unobservable variables and the gender wage gap. The error term in the Mincer regression is usually interpreted as capturing the effect on hourly wages of factors that are unobservable in household surveys, like natural ability, contacts and work ethics. An increase in the dispersion of this error term may reflect an increase in the returns to these unobservable factors in terms of hourly wages. We show the standard deviation of the error term of each Mincer equation.

The coefficients in the Mincer regressions are usually different for men and women, indicating that they are paid differently even when having the same observable characteristics (education, age, location). To further investigate this point we simulate the counterfactual wage that men would earn if they were paid like women. The last column in sheet *Mincer2* reports the ratio between the average of this simulated wage and the actual average wage for men. Cases in which this ratio is less than one reflect the fact that women earn less than men even when controlling for observable characteristics.

## 8. HOUSING

The file *housing\_LAC.xls* has the following sheets

### *housing*

National statistics by equivalized income quintiles of:

- Share of housing owners
- Number of rooms in each dwelling
- Number of persons per room in a household
- Share of poor dwellings (i.e. located in poor neighborhoods or inconvenient places -see below)
- Share of dwellings of low-quality materials (see below)

### *housing urban*

- Same statistics than above but only for urban areas.

### *housing age*

- Same statistics by age of the household head.

### *housing area*

- Same statistics by area (rural and urban)

### *housing edu*

- Same statistics by education of the household head

### *Methodological issues*

#### **Housing ownership**

Housing is probably the main asset that most people own. Several household surveys in LAC report whether the house is owned by the family who lives in, and some also report the rental value of the dwelling. Survey questions usually differentiate between the ownership of the dwelling and the ownership of the lot. We construct a variable that takes the value 1 when the household owns both the lot and the dwelling.

### **Number of rooms**

The size of the house is an indicator of wealth and a possible proxy for poverty. Although the size is not reported in surveys, there is usually information on the number of rooms. We construct a variable that records the number of rooms in the dwelling that are not shared with other households. In a report to be included in the web page we will specify the construction of this variable in each country. In the SEDLAC tables we show the number of rooms, and also that variable normalized by household size.

### **Poor housing**

This variable captures whether the dwelling is located in a shanty town or other clearly identifiable poor neighborhood, or the individual/household lives in inconvenient places (e.g. street). Definitions widely vary across countries. We provide details in a methodological report.

### **Quality of the house**

This variable records low quality of the materials used for the walls, roof and floor. Countries significantly differ in the materials used for houses and in the coverage of these questions in the surveys. Materials that are a clear indicator of poverty in one country (or region) may not be related to poverty in other country. Comparisons based on these variables should be made with care and preferably only within countries. These indicators are very country-specific. The methodological report has more information on this issue.

## 9. INFRASTRUCTURE

The file *infrastructure\_LAC.xls* has the following sheets

### *infrastructure*

- *Water*: Share of households with access to a source of water in the house or lot.
- *Hygienic restrooms*: Share of households with a restroom with a toilet connected to a sewerage system or to a septic tank.
- *Sewerage*: share of households with a dwelling connected to a public sewerage system.
- *Electricity*: Share of households with access to electricity in the house from any source.
- *Telephone*: Share of households with a fixed phone or a cellular phone.
- All statistics by equivalized income quintiles. National statistics.

### *infrastructure urban*

- Same statistics than above but only for urban areas.

### *infrastructure areas*

- Same statistics by area (rural and urban)

## *Methodological issues*

### **Water**

The easy access to a safe source of water is one of the fundamental indicators of development. Most LAC surveys do not ask about potable water, but on the location of the water source. We construct a variable that takes the value 1 if the household has access to a source of water (safe water if recorded in the survey) in the house or lot.

### **Hygienic restrooms**

This variable takes the value 1 if the household has a restroom with a toilet connected to a sewerage system or to a septic tank.

### **Sewerage**

This variable is 1 when the house is connected to a public sewerage system.

### **Electricity**

The access to electricity is recorded in some surveys. Our variable electricity is equal to 1 if the household has access to electricity in the house from any source.

### **Telephone**

The access to a telephone is included in a variable that takes the value 1 if the household has a fixed phone or someone in the household has a cellular phone. The second possibility, increasingly important, is in some surveys captured in the consumption questionnaires. This fact introduces a difference between those countries that have budget questions and capture spending in cellular phones, and those with no expenditure questions.

## 10. DURABLE GOODS AND SERVICES

The file *durables\_LAC.xls* has the following sheets

### *home*

- Share of households with a refrigerator, a washing machine, air conditioner, and fixed heating, by area and income quintiles.

### *phone*

- Share of households with a fixed telephone, a cell phone, or any telephone, by area and income quintiles.  
Surveys with only information on fixed phones are marked with \*.

### *TV*

- Share of households with a television set, access to cable or satellite TV, and VCR or DVD, by area and income quintiles.

### *PC-Internet*

- Share of households with a personal computer and access to Internet at home, and households with at least one member who uses Internet, by area and income quintiles.

### *transportation*

- share of households with a car, a motorcycle, and a bicycle. Number of years of the car.

## 11. POVERTY-ALLEVIATION PROGRAMS

The file *pap\_LAC.xls* presents statistics for the poverty-alleviation programs reported in the household surveys. The file has the following sheets:

### *income*

- Share of households beneficiaries of a public poverty-alleviation program by equivalized income quintiles.

### *education*

- Share of households beneficiaries of a public poverty-alleviation program by education of the household head.

### *area*

- Share of households beneficiaries of a public poverty-alleviation program by area.

### *shares*

- Distribution of beneficiaries of a public poverty-alleviation program by income quintiles. Households and individuals.

### *share*

- Public social program transfers as a share of labor income and total income.

### *benefits*

- Monthly transfers in LCU from a poverty-alleviation program by equivalized income quintiles.

### *dist\_income*

- Distribution of transfers from a poverty-alleviation program by equivalized income quintiles.

### *dist\_area*

- Distribution of transfers from a poverty-alleviation program by area.

### *Methodological issues*

Unfortunately, household survey information on the access to the safety net and other poverty alleviation programs (PAPs) is not common in many LAC countries. For those countries with detailed information we present some basic statistics, indicating the specific questions on PAPs in the survey. So far we have identified information for:

- Argentina: Programa Jefes de Hogar
- Chile: Pasis, Asignaciones familiares, Subsidio Único Familiar, Subsidio de Cesantía, and Subsidio de Agua Potable
- Colombia: Comedores escolares
- Guatemala: a wide range of social programs (food, health, transportation)
- Jamaica: food stamps
- Mexico: Procampo, and Progresa
- Nicaragua: Programa de empleo, Programa de donación de alimentos
- Peru: assistance in food, education and health
- Uruguay: asignaciones familiares

Some of these programs are not strictly poverty-alleviation ones. However, we decided to include them at this stage of the project.

## 12. AGGREGATE WELFARE

The table *welfare\_LAC.xls* reports the annual growth rates in four alternative aggregate welfare functions.

### *Methodological issues*

When equity considerations are deemed irrelevant, the performance of the economy is evaluated in terms of the unweighted average of individual income. This implies that one additional peso in the hands of a rich person is equally valuable (from a social point of view) than an additional peso in the hands of a poor. However, most people have equity concerns and aversion to inequality, implying welfare functions that attach more weight to the poor.

The table *welfare\_LAC.xls* reports the annual growth rates in four alternative aggregate welfare functions: the utilitarian (also known as Bentham function), one proposed by the Nobel Laureate Amartya Sen (equal to the mean times 1 minus the Gini coefficient), and two from the family of Atkinson functions.<sup>22</sup> The first one coincides with mean income. The following three introduce equity considerations into the assessment of the growth process. They can be thought as weighted growth rates of the economy, where the weights are larger for the income changes of the poor. The Atkinson function with parameter 2 approximates Rawlsian value judgments, according to which social welfare increases only if the well-being of the poorest group goes up. The Sen and Atkinson (1) functions reflect intermediate value judgments.

In columns (i) to (iv) of the table in *welfare\_LAC.xls* we compute all the distributional weighted-growth rates using only information from household surveys. Instead, in columns (v) to (viii) we take real per capita GDP from national accounts as the average income measure, and combine it with the inequality indices computed from household surveys. Given that most assessments of the performance of an economy are made by looking at per capita GDP, we use this variable and complement it with inequality indices to obtain rough estimates of the value of aggregate welfare according to different value judgments.

---

<sup>22</sup> See Lambert (1993) for technical details.

## 13. PRO-POOR GROWTH

This section has two files with information on pro-poor growth.

The file *pro-poor growth.xls* has three sheets.

### PPG\_1

- Ravallion and Chen's pro-poor growth rate with national and international lines.

### PPG\_2

- Indicator of progressive growth: difference between the growth rate for the poor and the overall growth rate.

### PEGR

- Poverty-equivalent growth rate: difference between PEGR and overall growth rate.

The file *growth\_incidence.xls* displays growth-incidence curves for all LAC countries.

## Methodological issues

### Pro-poor growth

Define  $\alpha$  to be a weighted sum of the individual income growth rates  $g_i$ , i.e.  $\alpha = \sum_i \sigma_i g_i$ , where  $\sigma_i$  are the weights, which are non-increasing in individual income  $x_i$ . In a typical poverty analysis the social weights of the non-poor are zero, i.e.  $\sigma_i = 0$  if  $x_i \geq z$ , where  $z$  is the poverty line. In this context growth is said to be pro-poor if  $\alpha > 0$ . In particular, if  $\sigma_i$  is the same for all the poor people and equal to  $1/NH$ , where  $N$  stands for total population, then  $\alpha$  is just the average of the growth rates of the poor. Ravallion and Chen (2003) argue for the use of this average as a measure of pro-poor growth. They show that this indicator is equal to the change in the Watts poverty index per unit time divided by the headcount index.<sup>23</sup>

---

<sup>23</sup> The Watts index is the population mean of the log of the ratio of the poverty line to censored income, where the latter is actual income for those individuals below the poverty line, and the poverty line for those individuals above it. The Watts index is the only poverty measure complying with the set of axioms for an ideal poverty measure in Zheng (1993).

Define  $\beta$  as a weighted sum of the difference between the individual income growth rate  $g_i$  and the growth rate of the mean  $g_\mu$ , i.e.  $\beta = \sum_i \sigma_i (g_i - g_\mu)$ . Growth is said to be *progressive* if  $\beta > 0$ . In the case where  $\sigma_i = 0$  if  $x_i \geq z$ , and  $\sigma_i = 1/NH$  if  $x_i < z$ , then  $\beta$  is just the difference between the average of the growth rates of the poor and the growth rate of the mean. In sheet [PPG\\_2](#) table [welfare\\_LAC.xls](#) we compute this measure of progressive growth for four alternative poverty lines.

It is argued that the concept of “pro-poor growth” should make reference to situations where growth is associated to a proportionally larger income increase for the poor than for the rest of the population. According to this view growth is “pro-poor” if poverty falls more than it would have if all incomes had grown at the same rate (Baulch and McCulloch (2000), Kakwani and Pernia (2000); Kakwani *et al.* (2003), Son (2004)). Kakwani *et al.* (2003) propose computing the distribution neutral growth rate that achieves the observed change in poverty. Growth is considered pro-poor when this *poverty equivalent growth rate* is higher than the actual global growth rate. In sheet [PEGR](#) we show the poverty equivalent growth rates for all countries in the dataset, and the difference between this rate and the growth rate of the mean.

### Growth incidence curves

These curves introduced by Ravallion and Chen (2003) are simple and illustrative instruments to analyze growth rates along the income distribution. Specifically, they show the proportional income change at each percentile of the income distribution. They are frequently used to study the extent to which different segments of the population participate in the growth process (or suffer from a recession). The interpretation of such a simple instrument, however, should be made with cautious. There are multiple factors that affect income changes, and all are reflected at the same time in a growth-incidence curve. Some of them may have nothing to do with the “growth process”, and some may have complex interactions.<sup>24</sup>

---

<sup>24</sup> Suppose half of the population lives in the countryside and is poor, while the other half lives in the cities and is not poor. Suppose in a given period the government has made investments in schools and infrastructure that allowed productivity to increase 50% in rural areas and 20% in urban areas. However, in the same period the international price of the main crop collapsed, reducing the price received by local farmers by 50%. In that scenario, the growth-incidence curve for the country will show stagnant incomes for the poor, and increasing incomes for the non-poor. Growth in this country is then typically characterized as “not pro-poor” in any of its definitions, despite the fact the increase in productivity driven by the government policies benefited (especially) the poor.

The growth-incidence curves are known to be very volatile at the extremes, especially in the bottom percentiles. For this reason we have computed confidence intervals, and deleted from the figures those points where estimates seem unreliable.<sup>25</sup>

---

<sup>25</sup> Growth-incidence curves with confidence intervals are available upon request.

## References

- Andersen, L. (2001). Social mobility in Latin America: links with adolescent schooling. *IADB Research Network Working Paper #R-433*.
- Attanasio, O. and Székely, M. (eds.) (2001). Portrait of the poor. An assets-based approach. IADB.
- Baulch, B. and McCulloch, N. (2000). Tracking pro-poor growth. ID21 insights 31, Sussex, Institute of Development studies.
- Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics* 108, 317-342.
- Bourguignon, F.(2003). From income to endowments: the difficult task of expanding the income poverty paradigm. Delta WP 2003-03.
- Buhmann, B., Rainwater, G., Schmaus, G. and Smeeding, T. (1988). Equivalence scales, well being, inequality and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study database. *Review of Income and Wealth* 34, 115-142.
- Chen, S. and Ravallion, M (1997). How did the world's poor fare in the 1990s? *The World Bank Economic Review*.
- Chen, S. and Ravallion., M. (2001). How did the world's poorest fare in the 1990s? World Bank working paper.
- Davidson, R. and Duclos, J.Y (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* 68 (6) nov, 1435-1464.
- Deaton, A. (1997). *The analysis of household surveys. Microeconomic analysis for development policy*. Washington D.C.: The World Bank.
- Deaton, A. (2003). How to monitor poverty for the Millennium Development Goals. Research Program in Development Studies, Princeton University.
- Deaton, A. and Zaidi, S. (2002). Guidelines for constructing consumption aggregates for welfare analysis. *LSMS Working Paper* 135.

- Kakwani, N. and Pernia, E. (2000). What is pro-poor growth?. *Asian Development Review* 18, 1-16.
- Kakwani, N., Khandker, S. and Son, H. (2003). Poverty equivalent growth rate: with applications to Korea and Thailand. Economic Commission for Africa, mimeo.
- Mills, J. and Zandvakili, S. (1997). Statistical inference via bootstrapping for measures of inequality. *Journal of Applied Econometrics* 12, 133-150.
- Piketty, T. and Saez, E. (2003). Income inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, CXVII, 1, February, 1-39.
- Ravallion, M. and Chen, S. (1997). What can new survey data tell us about recent changes in distribution and poverty? *The World Bank Economic Review* 11 (2), 357-82.
- Ravallion, M., Datt, G. and van de Walle, D. (1991). Quantifying absolute poverty in the Developing World. *Review of Income and Wealth* 37, 345-361.
- Ravallion, M. and S. Chen (2003). Measuring Pro-Poor Growth, *Economic Letters* 78: pp.93-99.
- Ravallion, M. S. Chen, and P. Sangraula (2008). Dollar a day revisited. World Bank Policy Research Working paper 4620.
- Son, H. (2004). Pro-poor growth: Asian experience. Mimeo.
- Sosa Escudero, W. and Gasparini, L. (2000). A note on the statistical significance of changes in inequality. *Económica* XLVI (1). Enero-Junio.
- Székely, M. and Hilgert, M (1999). What's behind the inequality we measure: an investigation using Latin American data. *IADB Working Paper* # 409.
- Székely, M. Lustig, N., Cumpa, M. and Mejía, J. (2000). Do we know how much poverty there is?, *IADB Working Paper* # WP-437
- Székely, M. (2001). The 1990s in Latin America: another decade of persistent inequality, but with somewhat lower poverty. *IADB Working Paper* # 454.

World Bank (1990). World Bank's World Development Report 1990: Poverty.

World Bank (2003). World Bank's World Development Report 2003: Poverty.

World Bank (2004). *Breaking with history? Inequality in Latin America and the Caribbean.*