DEVELOPING & TESTING POVERTY ASSESSMENT TOOLS

RESULTS FROM ACCURACY TESTS IN PERU

NOVEMBER 2005

This publication was produced for review by the United States Agency for International Development.
It was prepared by the IRIS Center, at the University of Maryland.

DEVELOPING & TESTING POVERTY ASSESSMENT TOOLS

RESULTS FROM ACCURACY TESTS IN PERU

SUBMITTED BY

The IRIS Center, University Research Corporation International

SUBMITTED TO

Scott Kleinberg

CONTRACT NUMBER

GEG-I-02-02-00029-00
Task Order No. 2: Developing Poverty Assessment Tools

AUTHORS

At the time of this writing, Manfred Zeller, Ph.D., was a professor for Socioeconomics of Rural Development at the Georg-August University of Göttingen's Institute of Rural Development in Germany. He is currently Professor for Rural Development Theory and Policy at the University of Hohenheim, Germany.

Gabriela Alcaraz V. and Julia Johannsen are Ph.D. candidates at the Institute of Rural Development, Georg-August University of Göttingen, Germany.

SERIES INFORMATION

This series presents a number of technical reports about USAID's program to develop Poverty Assessment Tools, implemented by the IRIS Center at the University of Maryland. These reports are made available to microenterprise practitioners, donor agency representatives, researchers and various stakeholders interested in the development of accurate and user-friendly tools to assess the poverty level of poor microenterprise clients in developing and transition economies.

CONTACT INFORMATION

IRIS Center
University of Maryland
Department of Economics
2105 Morrill Hall
College Park, MD 20742
USA

E-mail: info@iris.econ.umd.edu
Phone: +1.301.405.3110
Fax: +1.301.405.3020
Web: www.iris.umd.edu

ABSTRACT

This report presents the results of poverty assessment tool accuracy tests conducted by IRIS in Peru in 2004. The report first describes the design of the field research and the computation of the applicable poverty line, followed by an overview of the analytical methods chosen. A number of increasingly complex econometric methods are used to increase the accuracy of the estimation. This analysis results in the identification of groups of indicators that identify the poverty status of the 800 households included in the sample. In addition, the report analyzes the poverty outreach performance of six purposefully selected microfinance institutions (MFIs) on the basis of an additional sample of 1175 households.

For more information on the project, please visit *www.povertytools.org.*

For more information on AMAP and related publications, please visit *www.microLINKS.org*.

For more information about the IRIS Center, visit *www.iris.umd.edu*.

Contents

Chapter one: Introduction

USAID has commissioned the IRIS Center to develop, test and disseminate poverty assessment tools which will meet U.S. Congressional requirements for accuracy and cost of implementation. Accuracy tests of poverty indicators have been implemented by IRIS in Bangladesh, Peru, Uganda, and Kazakhstan. Comprehensive information on the project is available at *www.povertytools.org*, and will not be summarized in this report.

The purpose of this report is to present the results of the accuracy tests in Peru[i]. In the remaining part of Chapter 1, we provide an overview of the design of the field research for the accuracy test, and the computation of the applicable poverty line. Chapter 2 provides an overview of the analysis presented in this report.

In Chapter 3, we present the results on selected poverty indicators from nine regression models. Each of these models can be viewed as a potential, newly designed poverty assessment tool which is calibrated for Peru based on a nationally representative sample. The regression models are run in SAS, using the function MAXR that seeks to maximize the explained variance of the dependent variable (per-capita daily expenditure) by a set of BEST5, BEST10, and BEST15 regressors. Any set of five, ten, or fifteen poverty indicators can be considered a poverty assessment tool for purposes of identifying the poverty status of a household. The first 6 regression models differ with respect to the set of poverty indicators allowed in the model, starting from a model with a full set of potential regressors, and gradually restricting the set of regressors on the basis of practicality in implementation. A seventh model is run as an example of a tool that considers only those poverty indicators that were rated as highly verifiable by Instituto Cuánto, the survey firm in Peru. A subsequent model compiles these indicators with powerful subjective as well as monetary indicators. Finally, the last model makes use of poverty indicators usually available in the World Bank's Living Standards Measurement (LSMS) surveys. Thus, the first eight models can be considered alternative best combinations of poverty indicators which were mainly derived from existing practitioner tools for poverty assessment, while Model 9 is a tool derived from poverty indicators usually available in LSMS surveys.

Chapter 4 presents results from an alternative estimation approach, the so-called "two-step" models. Compared to the models presented in Chapter 3, the performance of models presented in Chapter 4 is overall much better.

Chapter 5 presents results on the poverty outreach of six microfinance institutions in Peru. The six institutions were purposely selected, and about two hundred new clients were randomly selected from each of the six institutions. The purpose of this sample is to investigate the poverty outreach of different types of microfinance institutions. Chapter 6 summarizes the results.

1.1 Field survey for accuracy tests in Peru

The survey firm Instituto Cuánto[ii] in Lima, Peru carried out the survey and completed double entry of data using the ISSA (Integrated System for Survey Analysis) software. Cuánto was founded in 1988 by Dr. Richard Webb, former president of the Central Bank, and Dr. Graciela Fernández B., former head of the National Statistical Systems in Peru. In total, six different geographical routes were identified and covered by three to eight interviewers each to implement a composite questionnaire survey with 800 households, followed two weeks later by a benchmark questionnaire. Training of the interviewers began on June 7, 2004. The survey was carried out from June 21 to August 26, 2004 and double entry of all data was completed by October 6, 2004.

The questionnaires can be downloaded from *www.povertytools.org*. The composite and benchmark questionnaires were adapted to the Peruvian context. In the case of the composite questionnaire, this meant including poverty indicators specific to the different geographical macro-regions, such as the number of lamas owned, or the inclusion of certain inferior foods in Section E (see questions E151 thru 159). Useful sources for the identification of country- and region-specific poverty indicators include the official statistical report by Webb and Fernández (2003), as well as the long-standing experience of the professional interviewers and researchers in Peru who were involved in adapting the questionnaire. Adapting the benchmark questionnaire mainly involved selecting major food items. For this, we used the results of the most-recent National Living Standard Measurement Survey (Encuesta Nacional sobre Medición de Niveles de Vida, or ENNIV, 2000), as well as a report published by the International Food Policy Research Institute (Zeller et al., 2002).

The adaptation and processing of the two questionnaires has benefited greatly from the long-term expertise of the personnel of Instituto Cuánto, particularly Moises Ventocilla A., the director general; Luis Castillo Q., the project director; and Pedro Llontop L., and Mario Reyna Farje E., senior researchers, as well as their supervisors and interviewers; in conducting economic, social, and market studies during the past 17 years in Peru's diverse cultural and socioeconomic settings.

1.2 Sampling Frame

Requirements for sampling. For the general population sample, we decided to select 800 households. The sample had to be nationally representative. Given the regional diversity in terms of agro-climatic, cultural, and socioeconomic conditions resulting from the north-south extension of the Andes, the sampling in Peru had to consider various criteria to be truly representative at the national level. The first criteria are the four "macro-regions" in east-west direction:

Metropolitan Lima, the rest of the coastal region, the Andean highlands, and the lowlands (jungle region).

These macro-regions are the common basis for the government's calculation of poverty lines in Peru and were controlled for in our multi-stage cluster sampling technique used to select a random sample of households. With this technique, we draw successive samples at each administrative level. The highest administrative unit in Peru are called departments, of which there are 24. These departments are further disaggregated into provinces and districts. Each district holds a number of cities and villages.

The first stage of sampling was conducted at the department level, based on centrally available and published population data. In view of logistical and budget constraints, we decided to randomly select 8 of the 24 departments. The probability of selecting a certain department was equal to its share of population in the country. The selected departments were: Arequipa, Cajamarca, Cusco, La Libertad, Lima (twice), Loreto, and Piura.

For each of the departments selected, we allocated a sample size of 100. This is referred to as probability-proportionate-to-size sampling (PPS), and we repeated the procedure at subsequent sampling stages at the district and community level. A further criterion for sampling is a representative distribution of rural and urban areas within each macro-region (except for the Lima Metropolitan region, which is entirely urban). Combining the macro-regional and rural/ urban criteria, Peru is commonly divided into seven "sampling areas." For each of the seven areas, the Government of Peru calculates a specific poverty line in order to account for differences in living costs. Table 1 shows the distribution of the sample of 800 households and the national population, broken down by each of the seven official sampling areas.

Table 1: Distribution of sample and national population, by sampling area

Macro-regions, Sampling area	No. of sample households	Percentage of national population	Percentage of sample population
TOTAL	**800**	**100**	**100**
Coast and Lima	**400**	**51.8**	**50.0**
Lima Metropolitan	200	28.9	25.0
Urban Coast	132	17.8	16.5
Rural Coast	68	5.2	8.5
Highlands	**266**	**35.0**	**33.3**
Urban Highlands	99	12.6	12.4
Rural Highlands	167	22.4	20.9
Lowlands	**134**	**13.1**	**16.8**
Urban Lowlands	66	6.0	8.3
Rural Lowlands	68	7.1	8.5

In the second stage three districts (six in the Lima department) in each of the eight selected departments were randomly chosen proportionate to the size of the districts compared with the total department population. Each of the twenty-four districts was then randomly determined to provide either only rural or only urban sample households (see Annex A-1).

For the next sampling step at district level, Instituto Cuánto had access to detailed urban and rural maps that included a pre-established division into (A) survey segments in urban district maps and (B) rural survey areas (RSA) in rural maps, as defined by the National Statistic Institute. These two types of survey segments constituted the next sampling cluster. We randomly selected one segment or RSA in each district in addition to an appropriate number of urban street blocks suitable for the survey work in the case of an urban district assignment (we excluded extended industrial and natural areas within the urban segment). In rural districts, we chose one community in each district proportionate to size. In districts that do not host large enough communities to randomly select from for the required number of sample households, we selected more than one community proportionate to size compared to the total district population.

Finally, in each of the selected communities, the random walk method (see Henry et al., 2003) was applied to select a sample of 33 or 34 survey households per district. Thus, the total sample of 800 is a self-weighing, nationally representative sample which contains the best possible set of districts taking into account criteria such as macro-regional diversity, urban/ rural domain, costs of transport and survey personnel, as well as timetable of overall survey operations.

1.3 Poverty line

The legal text by U.S. Congress refers to two alternative poverty lines in defining the "very-poor." The term "very-poor" refers to individuals

(A) living in the bottom 50% below the poverty line established by the national government; or
(B) lliving on the equivalent of less than $1/day.

Through the above term "or", the legislation implies that a person could be considered very-poor if he/she was *either* living on less than a dollar a day, *or* was in the bottom half of the distribution of those below the national poverty line. The legislation thus identifies two alternative measures of extreme poverty, relating to two commonly used poverty lines:

National Poverty Line (A): *the bottom 50 percent* of those classified as poor by any national poverty line. In Peru, the national poverty line is expressed in Soles, the local currency. However, due to the geographic diversity, there is no single national poverty line for the whole country. Instead, there are seven regional poverty lines that reflect the consumer basket (based on the regional consumption habits and prices) in each of the seven sampling areas.

International Poverty Line (B): one dollar income per day per capita (equal to $1.08 per day in purchasing power parity (PPP) dollars at 1993 prices).

Because the benchmark questionnaire (see *www.povertytools.org*) enumerates per-capita expenditures in current Soles (the local currency in Peru) as of the survey date, we converted the national and international poverty line into Soles values as of July 2004. The starting point for these calculations with respect to the national poverty line are the income percentiles from 1 to 100 in each of the seven sampling areas, expressed as annual expenditures per person in May 2000, according to Peru's most recent National Living Standard Measurement Survey (ENNIV, 2000).

As illustrated in Table 2, between 44 and 69 percent of all households in 2000 fall below the respective national poverty line in each of the seven sampling areas or regions. The weighted average at the national level results in a total headcount of 54.1 percent poor population in Peru (Webb and Fernández, 2003). According to the U.S. Congressional legislation, half (i.e., the bottom 50 percent) of these population shares below the disaggregated national poverty line can be considered as "very-poor."

Table 2: National poverty line and poverty headcount indices for the year 2000, by region

Expenditures May 2000 Region	**Annual national poverty line** (Soles/ person/ year)*	**Daily national poverty line** (Soles/ pers./day)	**Poverty headcount** (percent)**
Lima Metrop.	2,810.7	7.7	45.2%
Urban Coast	2,335.3	6.4	53.1%
Rural Coast	1,574.8	4.3	64.4%
Urban Highland	2,001.4	5.5	44.3%
Rural Highland	1,315.2	3.6	65.5%
Urban Lowland	1,934.7	5.3	51.5%
Rural Lowland	1,327.6	3.6	69.2%
TOTAL Poor (National aggregate of poverty headcount)			**54.1%**

Source: Calculations based on data from ENNIV (2000); and Webb and Fernández (2003), for the poverty headcount.

* These figures refer to the expenditure cut-off that represents the national poverty line in each of the seven regions, and are extracted from Annex A-2,

** The poverty headcount corresponds to the official figures based on ENNIV data of the year 2000, as published in Webb and Fernández (2003)

For the international poverty line, we converted $1 into Soles using purchasing power parity rates. In March 2003, $1 was equivalent in purchasing power to 1.86 Soles[iii]. To compare the national with the international poverty line in 2000, we adjusted the $1.08 poverty line by the loss in purchasing power (due to national inflation) between May 2000 and March 2003. Similar to the procedure of the regional poverty lines, this required adjusting the 1.86 Soles value by the increase in the national consumer price index (CPI) for the period from May 2000 to March 2003. Using CPI data, we calculated a total inflation of 5.38 percent over the 35-month period.[iv] We therefore adjusted the value of 1.86 Soles by 5.38 percent, a multiplication with the factor 100/(100 + 5.38). The result was 1.77 Soles, equivalent to the purchasing power of $1 for May 2000. Finally, multiplying this value by 1.08 (the international poverty line of $1 is equivalent to $1.08 at 1993 prices) yields 1.91 Soles. This amount is the international poverty line expressed in Soles for May 2000.

Using these calculations for a comparison of the poverty incidence according to the two poverty lines in 2000, in absolute terms, if one would take the bottom 50 percent below the *national* poverty line for defining the very-poor in the Rural Coast region, for example, an absolute 32.2 percent of the population would be counted as very-poor in this region. On the other hand, only 13.7 percent of the population in the Rural Coast region falls below the international poverty line of $1 a day (see Table 3). Hence, the national poverty line (concept A) defines a higher percentage as very-poor when compared to the international poverty line (concept B), not only in the Rural Coast but in all of the seven regions (compare shadowed columns in Table 3). The 'or' definition in the text by the US Congress suggests using the poverty line that yields a higher headcount index of "very-poor." Thus, the applicable poverty line for the accuracy tests in Peru is the respective national poverty line.

Table 3: Comparison between the international poverty line and the poverty headcount according to the US Congress definition of the "very-poor," based on expenditure data of May 2000

Expenditures May 2000 Region	**Annual income equivalent to 50% < national poverty line** (Soles/ person/year)*	**Daily income equivalent to 50% < national poverty line** (Soles/person/day)	**Poverty headcount** (percent)**
Lima Metrop.	1,999.9	5.5	22.6%
Urban Coast	1,564.2	4.3	26.6%
Rural Coast	1,015.6	2.8	32.2%
Urban Highland	1,351.8	3.7	22.2%
Rural Highland	795.0	2.2	32.8%
Urban Lowland	1,281.2	3.5	25.8%
Rural Lowland	871.3	2.4	34.6%
TOTAL Very-poor (Bottom 50 percent of Total Poor)			**27.1%**
Expenditures May 2000 Region	**Annual int. poverty line** (Soles/person/year)	**Daily int. poverty line** (Soles/person/day)	**Poverty headcount** (percent)
Lima Metrop.	695.8	1.91	1.1%
Urban Coast	695.8	1.91	2.5%
Rural Coast	695.8	1.91	13.7%
Urban Highland	695.8	1.91	3.9%
Rural Highland	695.8	1.91	23.0%
Urban Lowland	695.8	1.91	4.1%
Rural Lowland	695.8	1.91	21.4%

Source: Calculations based on data from ENNIV (2000).

* These figures refer to the expenditure cut-off that represents the bottom 50 percent below the national poverty line in each of the seven regions, and are extracted from Annex A-2.

** This poverty headcount is based on the definition of "very-poor" by the US Congress and corresponds to the 50 percent cut-off of the official national poverty headcount, as published in Webb and Fernández (2003) and presented in the last column of Table 2.

The expenditure amount equivalent to the bottom 50 percent share of the population in each region is then adjusted by the inflation rate between May 2000 and July 2004. To do this, we adjusted the respective *annual* Soles value of May 2000 (second column in Table 2 and 3), converted into *daily* expenditures per person (third column in Table 2 and 3) by the increase in the national consumer price index (CPI) during this period. Using published data on CPI in Lima for the period May 2000 to May 2004, and using the average monthly CPI in the two months after May 2004 as an estimate of the CPI change for the period May to July 2004, we calculated a total inflation of 9.3 percent over the 51-month period from May 2000 to July 2004 (see Table 4).

Table 4: Evolution of the Consumer Price Index (CPI) in Lima

Date	May 2000	March 2003	May 2004	July 2004, imputed
CPI Lima (Dic 2001 = 100%)	98.09	103.37	106.57	107.20

Source: Instituto Nacional de Estadística e Informática (2004)

To accommodate the accuracy test survey implemented by IRIS in July 2004, we again adjusted the $1 poverty line by the loss in purchasing power between March 2003 and July 2004. Similar to the procedure described for the period 2000 to 2003, this required adjusting the value of 1.86 Soles equivalent to $1 in March 2003 by the increase in the national CPI for the period from March 2003 to July 2004. Using CPI data for the period March 2003 to May 2004, and the average monthly CPI in the two months after May 2004 as an estimate of the CPI change for the period May to July 2004, we calculated a total inflation of 3.7 percent over the 17 month period. We therefore increased the value of 1.86 Soles by 3.7 percent (i.e., a multiplication with (100 + 3.7)/100). The result is a value of 2.08 Soles. This value is the international poverty line expressed in Soles for the survey month July 2004.

As for the expenditure values in 2000, also in 2004, in every single sampling area, the 50 percent cut-off value below the national poverty line (in Soles) is higher than the corresponding Soles value of the international poverty line (see Table 5). Therefore, we confirm the use of the national poverty line, disaggregated into seven regional lines. We define households having a per-capita daily expenditure level below the resulting daily expenditure value in Soles as of July 2004 in each of the seven sampling areas (see shadowed column in Table 5) as being very-poor.

Table 5 Regional cut-off values for the identification of the "very-poor" in the seven sampling areas, based on the disaggregated national poverty line, as of July 2004

Adjusted expenditures as of July 2004	**Annual income equivalent to 50% < national poverty line** (Soles/person/year)	**Daily income equivalent to 50% < national poverty line** (Soles/person/day)	**Annual int. poverty line (1 US-$ PPP)** (Soles/person/year)	**Daily int. poverty line (1 US-$ PPP)** (Soles/person/day)
Lima Metrop.	2,182.0	5.98	759.2	2.08
Urban Coast	1,706.6	4.68	759.2	2.08
Rural Coast	1,108.0	3.04	759.2	2.08
Urban Highland	1,474.9	4.04	759.2	2.08
Rural Highland	867.4	2.38	759.2	2.08
Urban Lowland	1,397.9	3.83	759.2	2.08
Rural Lowland	950.6	2.60	759.2	2.08

Source: Calculations based on data from ENNIV (2000)

In the sample of the IRIS accuracy test, 26.875 percent of households were found to be very-poor. This headcount index is very close to the bottom 50 percent cut-off of the published headcount index of 54.1 percent (yielding a headcount index of very-poor of 27.05 percent), derived from Peru's 2000 National Living Standard Measurement Survey (Webb and Fernández, 2003).

To stay true to the language of the legislation, throughout this report we will use the term "very-poor" or "VP" for those households with expenditures that fall below the bottom 50 percent cut-off of the respective national poverty line, and the term "not very-poor" or "NVP" for those with expenditures equal to or above the bottom 50 percent cut-off of the respective national poverty line. Readers should bear in mind that such binomial, either/or labels tend to distort the underlying reality, which is continuous — the standard of living of a household just above the line is not that much different than that of a household just below the line. Thus, the term "not very-poor" is simply shorthand for "estimated to have per capita daily consumption expenditures more than the dollar equivalent of the bottom 50 percent cut-off of the respective national poverty line," as shown in Table 5. We wish to note that a considerable share of these so-called "not very-poor" are actually categorized as poor by the national poverty line. Moreover, even among those households above the national poverty line, there exist a considerable share that are so vulnerable to poverty that, for example, a bad harvest, an illness of a family member, or a social obligation may drive them into poverty.

Chapter two: Overview of Regression Analysis

2.1 Introduction

In Chapter 3, we analyze the accuracy of newly designed tools and develop nine regression models for generating tools. These models consider all the poverty indicators that were compiled in the composite questionnaire, based on submissions of practitioner tools to IRIS in late 2003, and reviewed by Zeller (2004). In addition, indicators have been included based on recent poverty assessment studies published in academic literature. Thus, with the exception of Model 9 that uses LSMS type indicators only, the newly designed tools considered in Chapter 3 seek *best combinations* from poverty indicators of existing practitioners' tools.

2.2 Composite Questionnaire

The structure of the composite questionnaire is as follows. The full questionnaire can be downloaded from http://www.povertytools.org.

A. Identification of household (location, client status, etc.)
B. Household roster/demography, including individual and household-level indicators (derived from all practitioner tools)
C. Household expenditures by category (adapted from FINCA and ACCION tools)
D. Housing indicators (generic questions adapted from tools by AIM, ASA, CASHPOR, CIMS-OI, PRIZMA, and TUP), plus poverty indicators concerning minimum wages acceptable to respondents
E. Food consumption/food-security scales (adapted from tools by CGAP, Freedom from Hunger, and the World Food Program Food Security and Hunger Questionnaire)
F. Asset-based indicators (adapted from GRAMEEN Network and most other tools)
G. Social capital, voice, and vulnerability (adapted from recent advancements in social science research)
H. Estimates of objective and subjective poverty (adapted from recent advancements in social science research)
I. Information on client status of individual household members in programs and institutions supporting micro-finance or business development services (including information on size of loans and outstanding debt)
K. Monetary voluntary savings by individual household members (WOCCU)

The pretest of the composite questionnaire revealed that it was not possible to ask the questions in section K without jeopardizing the willingness of the respondents to cooperate for the subsequent benchmark visit. The questions on monetary savings as well as informal debt are

highly sensitive in Peru, possibly caused by the relatively high level of crime. It was therefore decided to ask section K questions at the end of the second interview (the benchmark questionnaire which captured household expenditures), conducted fourteen days after the first interview (the composite questionnaire).

2.3 Selection of indicators

In Chapter 3, we present results from nine models that were run with ordinary least squares (OLS) using SAS software. The models differ by the type of regressors used. While Model 1 includes 250 regressors, the seventh model has the most restrictive list, with 104 potential poverty indicators.

As one can see from the results for Model 1 in Chapter 3, the set of best poverty indicators is dominated by different expenditure and asset categories, apart from household demographic characteristics.[v] In Model 1, there are only a few poverty indicators from other dimensions and sections of the composite questionnaire. In a gradual process starting with Model 2, we reduce the number of regressors so as to allow indicators from other dimensions and sections of the questionnaire to enter among the best set of poverty indicators.

The overriding principle is to narrow down the list of poverty indicators with respect to two criteria:

Difficulty of indicators. Information on some indicators is easy to obtain, while for others it is not. Difficulty can be expressed in terms of time, money, and social costs expended for obtaining information. Social costs are especially important when addressing culturally sensitive questions. The difficulty of an indicator will therefore vary with the socio-economic and cultural context. It will also depend on the skill level and quality of training of interviewers. Furthermore, difficulty is strongly affected by the educational level and intellectual skills of the respondent, and by the interview situation (whether in private at home, or among peers and/or strangers in public — where certain type of questions may incur high social costs for the respondent). For example, the value of total assets is very difficult and tedious to obtain, and therefore relatively unsuitable for an operational poverty assessment tool. Another example is question C2 in the composite questionnaire — the value of food that is home-produced and consumed by the household in an average week — as well as several other expenditure indicators.

Verifiability of indicator. Another useful characteristic of an indicator, with respect to operational use, is ease of verification (in terms of time, monetary, and social costs). Here, we distinguish between objective and subjective indicators. Subjective indicators include any self-assessment (perception, feeling, attitudes) by the respondents (e.g., Section E9 onwards and Section H, regarding perceived adequacy of livelihood); or any assessment done by the interviewer (e.g., rating the poverty status of a household on a scale from 1 to 5, as in Section A). While some subjective indicators are among the more powerful poverty indicators, as will be

shown later, they are hardly verifiable, as the scales used are subjective and not disclosed to others. Objective indicators are characterized by using scales for measurement that can be — at least in principle — verified by consistent standards of measurement metrics. Examples of objective indicators include the age of a person (in years), the size of the rooms (in square meters), or whether the roof is made of natural fibers — these are all directly measurable through conventional and universally comparable scales. Measurability using comparable scales is a prerequisite for direct verifiability. Objective indicators, however, may also vary in their degree of verifiability. An example of an objective but unverifiable indicator is the number of luxury food eaten in the past seven days, or the money received from migrant relatives, or how many days a child was sick in the past 12 months. Common to this group of hardly verifiable objective indicators is the fact that actions or states occurred in the past.

Because the difficulty and verifiability of an indicator cannot be generalized across different socio-economic and cultural contexts, it might appear arbitrary to classify a particular indicator (or a group of indicators) as being more or less difficult to ask, or more or less verifiable. Therefore, we understand that not all readers will agree with our selection of progressively smaller subsets of regressors for defining Models 1 through 6. Our aim is to develop a variety of tools that differ in the dimensions of poverty that are considered. Moreover, this approach should be understood as a first attempt to address the practicality issue by presenting different models with perhaps increasingly simple and verifiable indicators. In Models 7 and 8, we use Instituto Cuánto's (the survey firm) subjective assessment of verifiability as an alternative attempt to address the practicality issue. To get more information on the practicality of poverty indicators, the IRIS Center also had microfinance (MF) and business development services (BDS) organizations conduct practicality tests.

Our sequence of regression models with progressively fewer poverty indicators (from Models 1 to 6) aims to generate different poverty assessment tools that gradually become less accurate but hopefully also more practical (easy to implement), less costly, and less prone to falsification by respondents or survey intermediaries.

For each model presented in Chapter 3, we present a set of BEST 5, BEST 10, and BEST 15 poverty indicators. Each of these three sets can be considered a poverty assessment tool in itself, and we document for each tool its level of Total Accuracy, Poverty Accuracy, and Non-Poverty Accuracy, as well as the degree of Undercoverage and Leakage. From an operational point of view — and everything else being the same— a tool derived only from the five best indicators presents an easier, more practical poverty assessment tool than one that uses the BEST15 (or even more) poverty indicators.[vi] This is quite obvious: fewer questions are necessary to ask and to analyze with a BEST5 tool compared to a BEST15 tool. However, fewer poverty indicators in a tool usually also tends to imply a lower degree of accuracy.

This highlights the important trade-off between accuracy and practicality of a poverty assessment tool. Cutting the right balance here requires us to carefully consider the trade-offs

between accuracy (and residual errors) and practicality, and this will ultimately determine the choice and certification of certain tools.

2.4 Specification of regression models

The following nine models were run as ordinary least squares in SAS. In all regressions, the sample size is 800. The dependent variable is the natural logarithm of per-capita daily expenditures in Soles, the currency in Peru.

Table 2.2.1 Dependent variable *per capita daily expenditures*

Variable	N	Minimum	Maximum	Mean	Standard deviation
Per capita daily expenditures	800	0.56	41.73	7.55	5.89
Ln expenditures per capita (natural logarithm)	800	-0.57	3.73	1.75	0.76

In all regressions, an INCLUDE statement always includes the following nine regressors as control variables:

Table 2.2.2 Description of the nine control variables

Variable	N	Minimum	Maximum	Mean	Standard deviation
Household size	800	1	14	4.68	2.04
Household size squared	800	1	196	26.06	23.99
Age of household head	800	18	94	47.72	16.10
Lima Metropolitan	800	0	1	0.25	0.43
Coast Rural	800	0	1	0.04	0.20
Highland Urban	800	0	1	0.13	0.33
Highland Rural	800	0	1	0.21	0.41
Lowland Urban	800	0	1	0.08	0.28
Lowland Rural	800	0	1	0.08	0.28

The first three control variables take into account the influence of important demographic factors that — in previous research — have been found to be powerful variables in explaining per-capita expenditures at the household level. As pointed out above, a sampling area combines the highest administrative unit of the macro-regions within Peru with the urban/rural domain. The six dummy variables seek to capture regional agro-ecological, cultural, and socioeconomic differences. The inclusion of these six dummy variables ensures that the estimated regression coefficients are controlled for regional differences.

All variables that are defined in monetary values (such as for expenditures and assets) are converted into natural logarithms since the dependent variable is also expressed in natural logarithm.[vii] All ordinal variables (for example type of roof, with lower values indicating inferior materials and higher values indicating superior materials) have been converted into dummy variables that reflect the different subtypes. For example, if the database has three types of roof

(1=natural material, 2=metal, 3=superior, such as tile), then dummy variables for two of the three different types of roof were formulated and tested in the statistical analysis for their potential of being a significant poverty indicator.

The nine different models were run in SAS using the MAXR technique that seeks to obtain a model with a high R-square. The R-square (R^2) is the ratio of the variance in the dependent variable that is explained by the model and its regressors, divided by the overall observed variance of the dependent variable. The coefficient ranges between 0 and 1. R^2 takes on the value of 1 when predicted values for the dependent variable for all households are the same as the observed values. A coefficient of 0.6 for R^2 implies that 60 percent of the observed variance in the dependent variable is explained by the model and its regressors.

High explanatory power of a model is a prerequisite for good predictions of the dependent variable per-capita daily expenditures (and thereby poverty status). The maximum R^2 improvement technique (MAXR) is a subcommand for regressions in SAS. The MAXR technique seeks to maximize explained variance (i.e., R^2), and considers all combinations among pairs of regressors to move from one step to the next. In the first step, the MAXR method begins by finding the one-variable model producing the highest R^2. In the second step, another variable, the one that yields the greatest increase in R^2, is added. Once the two-variable model is obtained, each of the variables in the model is compared to each of the variables not in the model. For each comparison of single pairs of variables, MAXR demonstrates whether removing one variable and replacing it with the other one increases R^2. After comparing all possible switches, MAXR makes the switch that produces the largest increase in R^2. Comparisons then begin again in the third step and so forth, and the process continues until MAXR finds that no switch can increase R^2. This limit may not be reached at 15 variables, but may include many more regressors. Thus, the MAXR technique allows us to identify the best model in each category: with only one, only 5 (termed in this paper the BEST5 model), only 10 (BEST10 model), or only 15 (BEST15) indicators.

2.5 Differences between the models

From the composite questionnaire, we computed 871 poverty indicators and related variables for their computation. Prior to using SAS software with the function MAXR, we dropped all of the original monetary variables in Soles that had been replaced by their natural logarithms, as well as the original ordinal and nominal variables that had been converted into dummies, and all of the variables necessary for computation and comparison that did not serve as direct poverty indicators. The remaining 553 poverty indicators composed the basic regression file used to analyze for each variable its potential as regressor. Similar to the subsequent analysis of the nine models, the SAS MAXR routine (as explained in Chapter 2.4) was used to select the best 250 potential regressors (in addition to the nine control variables) for the regression models 1 to 9. All of the dimensions of poverty (as well as all submitted poverty indicators from practitioner tools) from the total number of composite questionnaire sections were represented not only in the initial 553 indicators but also considered in the final regression file of the best 250

indicators, and hence in the generation of tools. Special care was given to the generation and testing of gender-specific poverty indicators. Annex C separately lists the gender-specific indicators that were selected for the final regression analysis (i.e., subset of 250 regressors).

Next, we describe the differences between the models (see also Figure 2.5.1).

Model 1: This model includes all 259 regressors considered in the regression analysis using SAS software (including nine control variables contained in every model). As will be shown later, this model contains mainly regressors that are derived from indicators on expenditures or value of assets.

Model 2: In this model, we drop all expenditure-related variables, except *monthly household expenditures on utilities, i.e. electricity, phone, water etc.,* (see section C of the composite questionnaire) and *clothing expenditures per capita in past 12 month* (see section B of the same questionnaire). These variables were the best two expenditure categories among 13 tested using SAS MAXR technique.[viii] A reduction of the number of expenditure variables is a first step towards a more operational set of poverty indicators. As pointed out, self-reported expenditures by respondent — irrespective of whether the recall period for expenditures is one week, one month or one year— are impossible to verify directly.

While the variables *clothing expenditure* and *expenditures on utilities* are two of the easier ones to recall among the expenditure group, the remaining questions contained in section C (question C1 to C12) require more intensive interviewer training as they are prone to high measurement error in practice. The interviewer needs to facilitate the interview by asking prompting questions on major elements of the different expenditure categories. For example, a particularly difficult expenditure category is home-produced food — especially for interviewers unfamiliar with traditional (or metric) measures used for crop yields in agriculture and food subsistence production (see question C2). Furthermore, the interviewer needs to provide special assistance to respondents with no or low school education for even simple calculations such as adding up expenses, especially since some elements of a certain expenditure category are recalled by the respondent on a monthly basis, and others are best remembered on a weekly basis (1 bag of potatoes per month, but a basket of rice per week). While these questions did not pose any significant difficulties for the experienced Instituto Cuánto interviewers, they may pose difficulties for less-experienced interviewers.

In total, Model 2 has 241 regressors that were retained from Model 1 (see Annex B).

Model 3: The set of regressors for this model is similar to Model 2. The only difference is the exclusion of the variable *total value of household assets* as a regressor. This variable is the natural logarithm of the total value of all assets possessed by the household. The total asset value is a powerful poverty indicator, and its exclusion allows other variables for single assets (or subgroups) to enter among the best regressors. The variable has been calculated from the value of all assets (from section D and F of the composite questionnaire). This variable is considered a

costly and therefore less practical poverty indicator, since it would require asking many of the questions from section D and F.

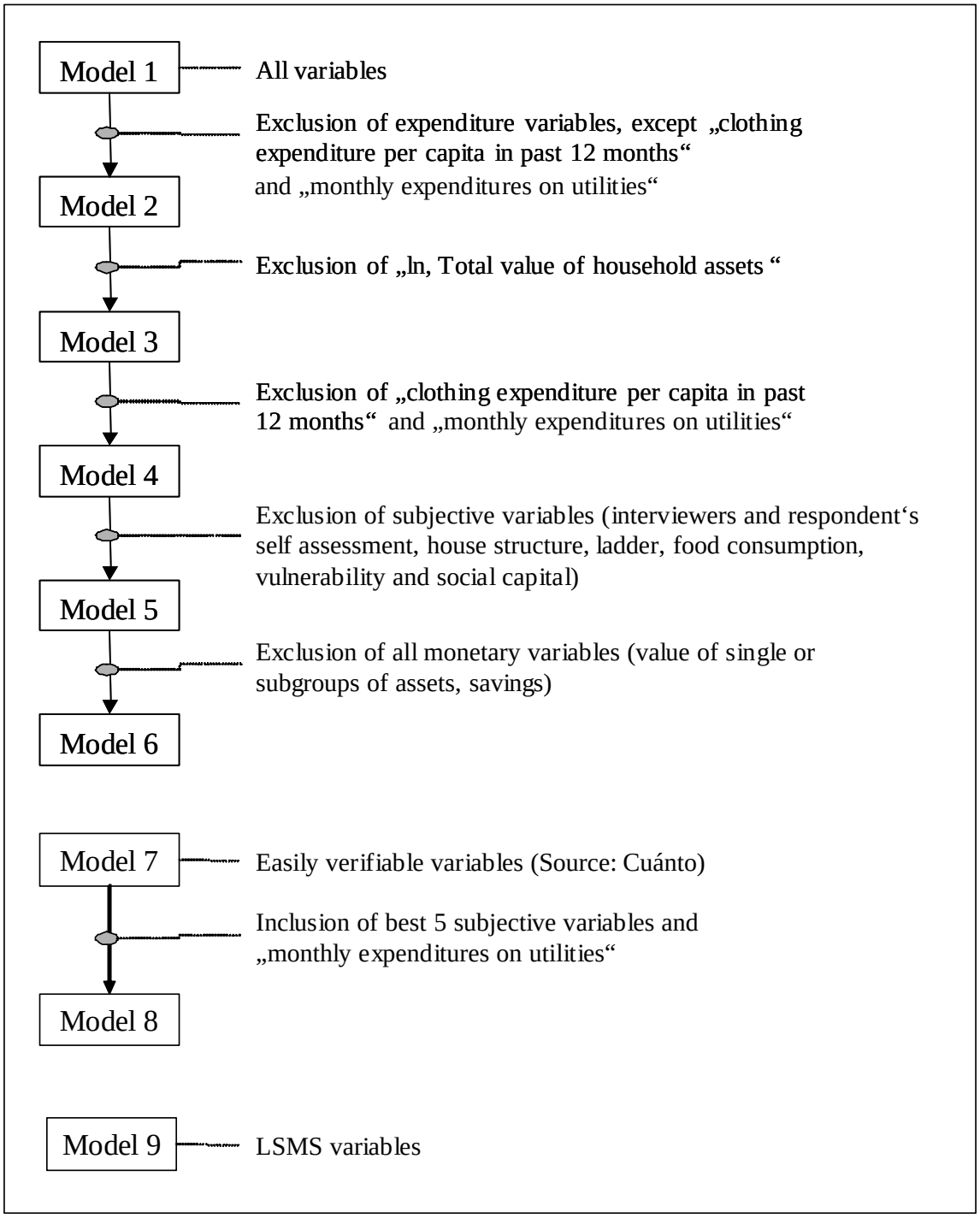
Model 4: The set of regressors for this model is similar to Model 3. The only difference is the exclusion of the variables *monthly household expenditures on utilities* and *clothing expenditures per capita in past 12 months* (both in form of the natural logarithm of the original expenditures in Soles). As they were the most powerful poverty indicators among the expenditure group, their exclusion allows other poverty indicators to enter into the best set of regressors.

Model 5: This is similar to Model 4, but all subjective poverty indicators are excluded. Such indicators include all ordinal rankings either done by the interviewer (such as those at the beginning of the interview in Section A, or the assessment of the structure of the house), and all ordinal rankings concerning feelings or self-assessment of the respondent (for example, the ladder questions in Section H). While these subjective indicators can be powerful poverty indicators, they can hardly be verified, at least not in a direct way. Thus, such indicators allow strategic answers by the respondent depending on his or her expectations for the interview. For example, if the respondent feels that by making herself poorer than she is, he or she would have a higher chance of being accepted by program or to get a loan, he or she may strategically alter his or her response accordingly.[ix] The subjective poverty indicators that were excluded in Model 5 (compared to Model 4) are presented accordingly in Annex B.

Model 6: This model is similar to Model 5, but excludes all monetary variables from the remaining subset of regressors. With this approach, we now solely base the model on demographic characteristics and the number and the type of assets possessed.

Model 7: Compared to Model 6, this model is more restrictive with respect to the criteria verifiability, and incorporates 104 indicators which were rated by Instituto Cuánto (see Annex D) as “easily verifiable” and easily obtained from the respondent.[x] The model contains many poverty indicators that are used in the housing index, as well as variables on asset ownership, and other observable indicators.

Figure 2.5.1 . Schematic representation of the models' construction process.



Model 8: This model is similar to Model 7, but includes *monthly expenditures on utilities* (as best single expenditure indicator, in addition to the *average clothing expenditures* indicator contained in the Model 7 variables) and five powerful subjective variables:[xi]

- Days in past seven days with main meal consisting of plain rice and any vegetables
- Household always ate enough from what they wanted (past 12 months)
- Household feels that their housing expenses are below need
- Household rates itself above national poverty line
- How much does the household need per month to live?

Model 8 is an example of a combination of indicators that are deemed easily verifiable by survey experts in Peru (some of the indicators are directly observable) with powerful subjective and objective indicators that are not directly verifiable. However, this model may allow indirect verifiability of clothing and utility expenditures (as well as subjective indicators) through a comparison with the answers to the readily verifiable indicators.

Model 9. This model incorporates variables that are usually available in LSMS surveys. It includes 127 regressors related to demographic, asset, expenditures, housing, and credit and financial asset information.

Annex B presents a description of the 259 regressors entered into the different models. For each model, the corresponding column (M*) indicates the specific regressors included in the model type. Figure 2.5.1 presents an overview of the nine regression models tested.

In conclusion, the models differ in their sets of poverty indicators being submitted to regression analysis. The result of the regression analysis, i.e. the identified set of best regressors (be it 5, 10, or 15), could be potentially used as a tool in nationally representative surveys in Peru for assessing whether a household is below or above the poverty line. The nine models differ in the number and type of regressors that are considered, and models 1 to 7 represent increasingly simple tools that appear progressively less prone to risks such as strategic answers and verification problems.

Chapter three: Results from Regression Models

In the following, the results are summarized by listing

- The regressors that were among the best five (BEST5), best ten (BEST10), and best fifteen (BEST15)
- The adjusted R^2 achieved (e.g., an R^2 of 0.6 indicates that 60 percent of the observed variance in the dependent variable is explained by the regressors)

We note that the set of BEST regressors is statistically determined by SAS's MAXR technique, which searches for the best model fit. The term BEST regressors should not be misunderstood as a value statement that implies that the set is best in terms of the Total Accuracy of a regression model, or for any of the other measures of performance listed below. Rather, the set of BEST5, BEST10, or BEST15 regressors refers to the best model fit, given the constraints on the set of available regressors and on the maximum number of regressors for inclusion (for example five regressors in a BEST5 model).

In order to assess the prediction power of each regression model (or tool) for poverty assessment, we used the following seven measures of accuracy performance (IRIS, 2005):

- **Total Accuracy —** the percentage of the total sample of 788 households whose poverty status is correctly predicted by the regression model.
- **Poverty Accuracy —** the accuracy among the very-poor, which refers to the households correctly predicted as very-poor, expressed as a percentage of the total very-poor.
- **Non-Poverty Accuracy —** the accuracy among the not very-poor, which refers to the households correctly predicted as not very-poor, expressed as percentage of the total number of not very-poor.
- **Undercoverage — t**he error of predicting very-poor households as being not very-poor, expressed as a percentage of the total number of very-poor.
- **Leakage —** the error of predicting not very-poor households as very-poor, expressed as a percentage of the total number of very-poor.
- **Poverty Incidence Error** (PIE) — the difference between the predicted and the actual (observed) poverty incidence, measured in percentage points.
- **Balanced Poverty Accuracy Criterion** (BPAC) — Poverty Accuracy minus the absolute difference between Undercoverage and Leakage, each expressed as a percentage of the total number of very-poor. When Undercoverage and Leakage are equal, the BPAC is equal to the Poverty Accuracy. BPAC is measured in percentage points. The application of the BPAC

criteria is based on three assumptions about the valuation of errors concerning the predictions of the very-poor and the not very-poor (see IRIS, 2005).

The Leakage and Undercoverage measures are often used in the literature to assess the poverty targeting performance of development and safety net policies, institutions, or projects. The PIE measure indicates the precision of a model (or poverty assessment tool) in correctly predicting the poverty incidence. Ideally, the value of PIE should be zero, implying that the predicted poverty rate equals the observed poverty rate. Positive PIE values indicate an underestimation of the poverty incidence, whereas negative values imply an overestimation of the poverty headcount index. PIE is useful if the objective of the poverty assessment is to measure the poverty outreach of an entire institution that provides microfinance or business development services. Hence, the evaluation question is: What percentage of institution X's clients are very-poor? Note that a satisfactory PIE value may be reached through a combination of a low Poverty Accuracy and a low Non-Poverty Accuracy. This is because errors in predicting the very-poor may cancel out with errors made in predicting the not very-poor, which will result in a satisfactory PIE value. Thus, a good PIE value (close to zero) may be achieved by a model that has a low Poverty Accuracy and a low Non-Poverty Accuracy and a high Leakage and Undercoverage. Hence, selecting a model on the basis of PIE entails the risk of choosing a model with a low Poverty Accuracy and high Undercoverage and Leakage. The balanced poverty assessment criterion BPAC considers these three accuracy measures, and models with a higher positive BPAC value indicate a higher Poverty Accuracy, adjusted by the absolute difference between Leakage and Undercoverage. There may exist trade-offs between PIE and BPAC in the selection of models. A perfect prediction model would have a value of zero for PIE and a value of 100 for BPAC. In such a perfect model, Leakage and Undercoverage would have a value of zero, and Total Accuracy, Poverty Accuracy, and Non-poverty Accuracy a value of 100.

In section 3.1 through 3.9, we present results from models using the OLS regression technique, using the different sets of regressors identified as Model 1 thru 9. In section 3.10, we use three alternative regression techniques. For each of the regression models presented in Chapter 3, we provide the above performance measures for predicting the very-poor and not very-poor.

The above-mentioned measures of model performance are illustrated next, using the results from Model 1.

3.1 Model 1

This model includes all 250 regressors available for the regression analysis, in addition to the nine control variables. Table 3.1.1 presents the number of households classified as very-poor and not very-poor by the international poverty line, as well as the predicted poverty status of the households within both groups.

Table 3.1.1 Observed vs. Predicted poverty status for the BEST 5 regressors set.

Poverty status (as measured by benchmark questionnaire in survey)	**Predicted poverty status**		
	Not very-poor	**Very-poor**	**Total**
Not very-poor	539	46	585
Very-poor	78	137	215
Total	617	183	800

Observed poverty status:

- Percentage of very-poor = (215 / 800) * 100 = 26.875 %
- Percentage of not very-poor = (585 / 800) * 100 = 73.1 %

Predicted poverty status:

- Percentage of predicted very-poor = (183 / 800) * 100 = 22.9 %
- Percentage of predicted not very-poor = (617 / 800) * 100 = 77.1 %

Model performance:

- Total Accuracy = (539 + 137) / 800) * 100 = 84.5 %
- Poverty Accuracy = (137 / 215) * 100 = 63.72 %
- Non-Poverty Accuracy = (539 / 585) * 100 = 92.14 %
- Undercoverage = (78 / 215) * 100 = 36.28 %
- Leakage = (46 / 215) * 100 = 21.40 %
- PIE = (22.9-26.87) = -3.97
- BPAC = 63.72 - abs(36.28-21.40) = 48.84

From Table 3.1.2, it can be observed that the highest performance in terms of Total Accuracy and BPAC of Model 1 is actually achieved in the BEST15 set (87.38 and 59.07 percentage points, respectively). Furthermore, monetary variables (being expenditures or asset values) account for approximately half of the indicators incorporated in each set. This model has a tendency to focus on aspects related to food security, assets, and expenditures.

Table 3.1.2 Summary of accuracy results for Model 1

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.793	
• Days in past seven days with main meal consisting of plain rice and any vegetables • Share of food expenditures from total household expenditures • Annualized total household expenditures • Total value of household assets • Household has electricity	Total Accuracy: 84.50 Poverty Accuracy: 63.72 Non-Poverty Accuracy: 92.14 Undercoverage: 36.28 Leakage: 21.40 PIE: -3.97 BPAC: 48.84
Best 10 indicators: R^2 adjusted = 0.820	
Next best five indicators: • Days in past seven days with main meal consisting of plain rice only • Number of cars owned by the household • Number of steps above step identified as international poverty line • Wood as exterior-walls' material • Distance to department capital • Average daily per-capita clothing expenditures Removed indicators: • Days in past seven days with main meal consisting of plain rice and any vegetables	Total Accuracy: 87.00 Poverty Accuracy: 70.23 Non-Poverty Accuracy: 93.16 Undercoverage: 29.77 Leakage: 18.60 PIE: -3.00 BPAC: 59.06
Best 15 indicators: R^2 adjusted =0.834	
Next best five indicators: • Days in past seven days with main meal consisting of plain rice and any vegetables • Household ate less food from what they wanted for more than 10 days, but less than 30 days, during past 12 months • Value of remittances sent to relatives in last 12 months • Value of motor tillers owned by the household • Community access to subsidized food ("glass of milk — vaso de leche") in past 24 months	Total Accuracy: 87.38 Poverty Accuracy: 70.70 Non-Poverty Accuracy: 93.50 Undercoverage: 29.30 Leakage: 17.67 PIE: -3.13 BPAC: 59.07

The BEST15 set in Model 1 achieves the highest adjusted R^2 value (0.834) and the highest BPAC value (59.07%) among all models tested in section 3.1 through 3.9. Model 1 allows the selection of all possible indicators from the composite questionnaire and, therefore, is expected to present the most powerful set of regressors. Remember, however, that "powerful" in terms of the SAS MAXR routine refers to the highest R^2 — and not necessarily to the best accuracy performance with respect to all seven accuracy measures. For example, Model 3 achieves the

highest Total Accuracy of all nine models with 88.13 percent, as will be shown in section 3.3, but its BPAC value is below that of Model 1.With respect to PIE, we observe that Model 1 tends to underestimate the proportion of very-poor households by three percentage points.

It is worth noting that already in Model 1, the indicator related to the widely spread food aid programs in Peru appears among the BEST15 set of regressors. This result suggests that these programs target poorer communities fairly well. In general, the selected indicators of Model 1 may not be viewed as optimal in terms of practicality, i.e. the difficulty of obtaining information on and verifying the indicators. For example, the indicators *Total value of household assets* and *Share of food expenditures from total household expenditures* would require intensive and detailed questioning about the assets owned by the households (and their valuation) and about their expenditure level in the last 12 months. In addition, this type of information is difficult to verify.

3.2 Model 2

This model excludes all expenditure or expenditure-derived variables (section C of the composite questionnaire), with the exception of *monthly household expenditures on utilities* and *clothing expenditures per capita in the past 12 months.*

The highest accuracy levels, as well as the lowest Undercoverage and Leakage measures, is achieved by the BEST15 regressor set. All figures improve substantially from the BEST5 to the BEST15 option.

In comparison to Model 1, this model shows a generally lower Total Accuracy in the BEST5 and BEST10 sets and a slightly higher Total Accuracy in the BEST15 set. The Poverty Accuracy and the Undercoverage results even improve in the BEST5 set of Model 2, as compared to Model 1. The remaining performance indicators of the BEST5, BEST10, and BEST 15 set decrease in Model 2, except for the Poverty Accuracy (70.70%) and the Undercoverage figures (29.3%), which stay the same as in Model 1.

With respect to PIE, the BEST5 set achieves the best performance (-3.00 percentage points). The highest BPAC and the second best PIE are registered in the BEST15 set (58.61 and -3.25 percentage points respectively).

In terms of poverty dimensions, this model incorporates variables related to the household's housing characteristics and informal savings activities while reducing the number of food security and expenditure indicators, resulting in a more pronounced multidimensional set of indicators than in Model 1.

Table 3.2.1 Summary of the accuracy results Model 2

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.763	
• Median education level of adult household members • Number of steps above step identified as international poverty line • Household monthly expenditure on utilities (electricity, phone, water, etc) • Total value of household assets • Average daily per-capita clothing expenditures	Total Accuracy: 83.75 Poverty Accuracy: 64.19 Non-Poverty Accuracy: 90.94 Undercoverage: 35.81 Leakage: 24.65 PIE: -3.00 BPAC: 53.03
Best 10 indicators: R^2 adjusted = 0.796	
Next best five indicators: • Number of rooms in the dwelling • Days in past seven days with main meal consisting of plain rice only • Number of cars owned by the household • Wood as exterior-walls' material • Sum of distances to department, provincial and district capitals	Total Accuracy: 85.25 Poverty Accuracy: 65.12 Non-Poverty Accuracy: 92.65 Undercoverage: 34.88 Leakage: 20.00 PIE: -4.00 BPAC: 50.24
Best 15 indicators: R^2 adjusted = 0.810	
Next best five indicators: • Household feels that their health care expenses are below need • Value of metal pots owned by the household • Household declares not to be able to save anything • Value of debt owed by other households to the household • Community access to subsidized food ("glass of milk — vaso de leche") in past 24 months	Total Accuracy: 87.50 Poverty Accuracy: 70.70 Non-Poverty Accuracy: 93.68 Undercoverage: 29.30 Leakage: 17.21 PIE: -3.25 BPAC: 58.61

3.3 Model 3

This model is based on Model 2, but excludes the variable for *value of total household assets*. All performance indicators strongly improve from the BEST5 to the BEST15 option.

In terms of accuracy results, Model 3 has a performance similar to or even better than Model 1 and 2. This holds particularly true for the BEST15 set of indicators, whose Total Accuracy of 88.13 percent is the highest one among all tools presented in this report. While the Poverty Accuracy and the related Undercoverage figures stay the same as in Model 1 and 2, the results for Leakage and Non-Poverty Accuracy are better.

As well, the best performance in PIE and BPAC are obtained in the BEST15 set, which registered a considerable improvement from the levels obtained in the BEST5 set. The results show an increase on PIE of 1.5 percentage points and on BPAC of 17.21 percentage points.

When we compare the values for BPAC for the BEST15 set of Model 3 with 56.28 percent with those of Model 1 (59.07%), we observe a relatively small decline of about 3 percentage points. The results from Model 3 imply good news with respect to the practicality concerns of reliable poverty assessments in Peru. The good results of Model 3 show that (in the case of Peru) neither extensive expenditure data nor summarized information on all household assets are necessary to achieve a reasonably accurate prediction.

Table 3.3.1 Summary of the accuracy results Model 3

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.741	
• Median education level of adult household members • Number of cars owned by the household • Number of color TV's owned by the household • Household rates itself above national poverty line • Average daily per-capita clothing expenditures	Total Accuracy: 83.83 Poverty Accuracy: 59.07 Non-Poverty Accuracy: 92.31 Undercoverage: 40.93 Leakage: 20.93 PIE: -5.38 BPAC: 39.07
Best 10 indicators: R^2 adjusted = 0.789	
Next best five indicators: • Number of rooms in the dwelling • Days in past seven days with main meal consisting of plain rice only • Wood as exterior-walls' material • Value of metal pots owned by the household • Sum of distances to department, provincial and district capitals	Total Accuracy: 85.75 Poverty Accuracy: 65.58 Non-Poverty Accuracy: 93.16 Undercoverage: 34.42 Leakage: 18.60 PIE: -4.25 BPAC: 49.76
Best 15 indicators: R^2 adjusted = 0.808	
Next best five indicators: • Number of steps above step identified as international poverty line • Household feels that their health care expenses are below need • Household monthly expenditure on utilities (electricity, phone, water, etc) • Value of tractors owned by the household • Household declares not to be able to save anything • Value of debt owed by other households to the household • Community access to subsidized food ("glass of milk — vaso de leche") in past 24 months Removed indicators: • Number of color TV's owned by the household • Household rates itself above national poverty line	Total Accuracy: 88.13 Poverty Accuracy: 70.70 Non-Poverty Accuracy: 94.53 Undercoverage: 29.30 Leakage: 14.88 PIE: -3.88 BPAC: 56.28

While in the BEST10 option, asset and housing variables together still represent half of the indicator set, their relative importance drops to one third in the BEST15 option, which now includes four indicators related to food security and subjective poverty assessment that appear somewhat more difficult to verify. Hence, the best set of indicators of all tools presented includes not only a great variety of different poverty dimensions but might also imply a trade-off between practicability and verifiability compared to other models.

3.4 Model 4

This model is similar to Model 3, but excludes the variables *monthly household expenditures on utilities* and *clothing expenditures per capita in the past 12 months*. Although, at this step, all of the expenditure variables have been excluded from the set of possible indicators, the adjusted R^2 levels are only slightly lower than in Model 3, and the accuracy performance, in particular in the BEST5 and BEST10 option, have further improved compared to Model 3.

Within Model 4, the Total Accuracy increases from 84.5 percent in the BEST5 set to 87 percent in the BEST10 set. The same applies to the remaining accuracy measures, as well as Undercoverage and Leakage. In contrast, the BEST15 tool does not further improve the performance, except for a slightly higher Non-Poverty Accuracy and a lower Leakage figure when compared to the BEST10 in the same Model 4 and the BEST15 in Model 3.

PIE and BPAC perform at best in the BEST10 set. PIE reaches a value of -4.50 percentage points (lower than in the previous models), while BPAC achieves 50.69 percentage points, 4.17 percentage points higher than the BPAC from the BEST15 set. Compared to Model 1, we observe a loss of about 9 percentage points with respect to the BPAC value. This is due to the more limited set of regressors.

While there is little difference in the selected BEST5 and BEST10 variables sets between Model 3 and Model 4, the combination of indicators in the BEST15 has become less balanced in Model 4. Still, various dimensions, including demographic and housing characteristics, food security, asset ownership, subjective indicators and community variables, are represented. But single asset indicators alone account for one third of the variable set, and also the number of subjective indicators and housing variables have increased, at the expense of informal savings variables that have disappeared completely in Model 4. Following the trend observed from Model 1 up to this point, this model still accounts for a high proportion of subjective and non-verifiable indicators.

Table 3.4.1 Summary of the accuracy results Model 4

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.741	
• Median education level of adult household members • Days in past seven days with main meal consisting of plain rice and any vegetables • Number of cars owned by the household • Number of color TV's owned by the household • Household rates itself above national poverty line	Total Accuracy: 84.50 Poverty Accuracy: 60.93 Non-Poverty Accuracy: 93.16 Undercoverage: 39.07 Leakage: 18.60 PIE: -5.50 BPAC: 40.46
Best 10 indicators: R^2 adjusted = 0.781	
Next best five indicators: • Number of rooms in the dwelling • Days in past seven days with main meal consisting of plain rice only • Household feels that their housing expenses are below need • Wood as exterior-walls' material • Value of metal pots owned by the household • Sum of distances to department, provincial and district capitals Removed indicators: • Days in past seven days with main meal consisting of plain rice and any vegetables	Total Accuracy: 87.00 Poverty Accuracy: 67.44 Non-Poverty Accuracy: 94.19 Undercoverage: 32.56 Leakage: 15.81 PIE: -4.50 BPAC: 50.69
Best 15 indicators: R^2 adjusted = 0.800	
Next best five indicators: • Availability of telephone (fixed land line) in the house • Days in past seven days with main meal consisting of plain rice and any vegetables • Household owns microwave • Number of beds owned by the household • Amount that household needs per month to live • Value of tractors owned by the household • Household has electricity Removed indicators: • Number of rooms in the dwelling • Number of color TV's owned by the household	Total Accuracy: 87.00 Poverty Accuracy: 66.05 Non-Poverty Accuracy: 94.70 Undercoverage: 33.95 Leakage: 14.42 PIE: -5.25 BPAC: 46.52

3.5 Model 5

Model 5 is based on Model 4, but excludes all subjective variables. With this, all variables related to subjective self-assessment of the adequacy of food consumption (i.e. Food Security Scale variables from Freedom from Hunger), vulnerability, the respondents' own poverty assessment, as well as the interviewers' assessment of poverty and the condition of the house were dropped, leaving some important dimensions out of consideration.

This model experienced a further slight decrease in the adjusted R^2. The best performance was achieved by the BEST15 set.

Table 3.5.1 Summary of the accuracy results Model 5

Variables	Model performance (%,%pt)	
Best 5 indicators: R^2 adjusted = 0.723		
• Median education level of adult household members • Number of cars owned by the household • Number of color TV's owned by the household • Value of food processing assets • Household has electricity	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE: BPAC:	82.50 57.21 91.79 42.79 22.33 -5.50 36.75
Best 10 indicators: R^2 adjusted = 0.765		
Next best five indicators: • Number of rooms in the dwelling • Availability of telephone (fixed land line) in the house • Wood as exterior-walls' material • Number of days in past seven days consuming any of six superior food items • Value of metal pots owned by the household • Sum of distances to department, provincial and district capitals Removed indicators: • Number of color TV's owned by the household	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE: BPAC:	84.88 63.26 92.82 36.74 19.53 -4.63 46.05
Best 15 indicators: R^2 adjusted = 0.784		
Next best five indicators: • Ratio male adult household members/ female adult household members • Total household members participating in water/ waste group • Value of tractors owned by the household • Household declares not to be able to save anything	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE:	85.13 64.19 92.82 35.81 19.53 -4.38

• Community access to subsidized food ("glass of milk — vaso de leche") in past 24 months	BPAC: 47.91

The exclusion of subjective variables allowed additional housing, savings, and demographic variables to enter into the best combinations, making this model strongly reliant on asset information (ownership and value) and housing characteristics.

The average Total Accuracy level and the Non-Poverty Accuracy for the three sets of Model 5 decreased by 2 percent compared to Model 4, while the Poverty Accuracy decreased even further. Also, with respect to the Leakage figures, Model 5 shows a clear drop in performance when compared to Model 4.

In terms of PIE and BPAC, the BEST15 set also registered the best performance. PIE increased from -5.5 percentage points in the BEST5 set to -4.38. BPAC increased 11.16 percentage points from the BEST5 to the BEST15 set.

On the other hand, in terms of the difficulty for obtaining and verifying information regarding the poverty indicators, this Model could be considered better than the previous models, due to the exclusion of the subjective variables and to the incorporation of asset, housing, and demographic variables which appear to be more verifiable.

3.6 Model 6

This model excludes all monetary variables, leaving 159 variables in the analysis. The adjusted R^2 ranges from 0.720 to 0.777.

As in the previous model, this model incorporates a high proportion of nearly two thirds of asset and housing-related variables. The main difference compared to Model 5 is the replacement of value-based asset variables by those related to the incidence of ownership of a particular asset (formulated as a dummy variable 0= no and 1=yes) or the number of a specific asset(s) owned.

In all of the five performance indicators, the BEST5 set of Model 6 is inferior to all other tools presented up to this point.

The best overall performance was obtained in the BEST15 set. Total Accuracy reached 85.13 percent and Poverty-Accuracy 64.65 percent. In terms of PIE and BPAC, this set registered -4.13 and 49.30 percentage points, respectively.

Table 3.6.1 Summary of the accuracy results Model 6

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.720	
• Median education level of adult household members • Number of rooms in the dwelling • Number of cars owned by the household • Number of color TV's owned by the household • Household has electricity	Total Accuracy: 81.50 Poverty Accuracy: 55.35 Non-Poverty Accuracy: 91.11 Undercoverage: 44.65 Leakage: 24.19 PIE: -5.50 BPAC: 34.89
Best 10 indicators: R^2 adjusted = 0.760	
Next best five indicators: • Availability of telephone (fixed land line) in the house • Number of metal pots owned by the household • Wood as exterior-walls' material • Rooms per person • Number of days on past seven days consuming any of six superior food items • Sum of distances to department, provincial and district capitals Removed indicators: • Number of rooms in the dwelling	Total Accuracy: 85.13 Poverty Accuracy: 62.79 Non-Poverty Accuracy: 93.33 Undercoverage: 37.21 Leakage: 18.14 PIE: -5.13 BPAC: 43.72
Best 15 indicators: R^2 adjusted = 0.777	
Next best five indicators: • Household received in-kind services from food aid programs in last 3 years • Total household members participating in water/ waste group • Household owns a tractor • Security key lock or metal frame with padlock in main entrance door • Household declares not to be able to save anything	Total Accuracy: 85.13 Poverty Accuracy: 64.65 Non-Poverty Accuracy: 92.65 Undercoverage: 35.35 Leakage: 20.00 PIE: -4.13 BPAC: 49.30

3.7 Model 7

This model incorporates the 104 poverty indicators that have been rated as highly verifiable (score 4 or 5) by the staff and interviewers of Instituto Cuánto, based on their long-term experience in conducting field research and surveys in Peru. In Annex G, we list the ratings of the survey firm for all 259 regressors employed in Model 1. The performance of the model is lower than Model 6 in terms of adjusted R^2.

The best performance is obtained in the BEST 15 set, with a Total Accuracy level of 84.63 percent and a Poverty Accuracy level of 63.72 percent. Considering that the 15 indicators are fairly easy to obtain and to verify as deemed by Cuánto, the Total Accuracy levels (of only less than 1 percent lower than in Model 6) are still high. In terms of PIE and BPAC, this set achieved similar values than those found in Model 6, with a PIE of -4.13 and BPAC of 48.37 percentage points. Model 7, therefore, is able to generate tools that appear to have a high level of practicality.

Table 3.7.1 Summary of the accuracy results Model 7

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.708	
• Number of rooms in the dwelling • Availability of telephone (fixed land line) in the house • Household has electricity • Remittances sent • Average daily per-capita clothing expenditures	Total Accuracy: 82.38 Poverty Accuracy: 56.28 Non-Poverty Accuracy: 91.97 Undercoverage: 43.72 Leakage: 21.86 PIE: -5.88 BPAC: 34.42
Best 10 indicators: R^2 adjusted = 0.755	
Next best five indicators: • Household owns a color TV • Household owns a microwave • Household owns a suit • Wood as exterior-walls' material • Distance to department capital	Total Accuracy: 83.63 Poverty Accuracy: 61.86 Non-Poverty Accuracy: 91.62 Undercoverage: 38.14 Leakage: 22.79 PIE: -4.13 BPAC: 46.51
Best 15 indicators: R^2 adjusted = 0.773	
Next best five indicators: • Percent of adult household members who can read and write • Household owns a Motocab • Household owns a tractor • No lock on entrance door or wood/ metal bar to close from inside • Household head sleeps on something other than bed • Community access to subsidized food ("glass of milk — vaso de leche") in past 24 months • Sum of distances to department, provincial and district capitals Removed indicators: • Wood as exterior-walls' material • Distance to department capital	Total Accuracy: 84.63 Poverty Accuracy: 63.72 Non-Poverty Accuracy: 92.31 Undercoverage: 36.28 Leakage: 20.93 PIE: -4.13 BPAC: 48.37

In reviewing the above poverty indicators selected by the regression model, we observe that many of them are highly verifiable, simply by observation when visiting the client's home. Probably the most difficult questions to ask refer to the value of remittances and clothing expenditures. However, the staff of the survey firm rated these questions as relatively easy to ask (score 4 out of 5), although some readers may take issue with their score of 4 for verifiability.

3.8 Model 8

This model is similar to Model 7, but includes *monthly expenditures on utilities* (as the best single expenditure indicator, in addition to *average clothing expenditures*) as well as five powerful subjective variables:[xii]

- Days in past seven days with main meal consisting of plain rice and any vegetables
- Household always ate enough from what they wanted (past 12 months)
- Household feels that their housing expenses are below need
- Household rates itself above national poverty line
- Amount that household needs per month to live

The inclusion of these variables only slightly increases the model's performance compared to Model 7. We observed that four of these subjective variables were selected already in the best regressor sets of various previous models. This reflects the importance of incorporating subjective variables for poverty assessment in Peru. However, in practice, actual accuracy may be lower with subjective variables as they tend to be more difficult to ask and are particularly difficult to verify (see Cuánto's rating of verifiability and difficulty of asking in Annex D).

The adjusted R^2 values range between 0.725 and 0.792. The best performance is achieved by the BEST15 set with 85.38 percent Total Accuracy, a PIE of -4.13 percentage points, and a BPAC of 49.77 percentage points. Compared to Model 7, the Total Accuracy and BPAC increased only by about 1 percentage point. A greater improvement can be observed in the Poverty Accuracy of Model 8's BEST10 and BEST15 options.

Table 3.8.1 Summary of the accuracy results Model 8

Variables	Model performance (%,%pt)
Best 5 indicators: R^2 adjusted = 0.725	
• Average daily per-capita clothing expenditures • Household monthly expenditure on utilities (electricity, phone, water, etc) • Days in past seven days with main meal consisting of plain rice and any vegetables • Household always ate enough from what they wanted (past 12 months) • Household rates itself above national poverty line	Total Accuracy: 83.38 Poverty Accuracy: 57.67 Non-Poverty Accuracy: 92.82 Undercoverage: 42.33 Leakage: 19.53 PIE: -6.13 BPAC: 34.87
Best 10 indicators: R^2 adjusted = 0.773	
Next best five indicators: • Number of rooms in the dwelling • Household owns a microwave • No lock on entrance door or wood/ metal bar to close from inside • Sum of distances to department, provincial and district capitals • Remittances sent • Household feels that their housing expenses are below need Removed indicators: • Days in past seven days with main meal consisting of plain rice and any vegetables	Total Accuracy: 84.88 Poverty Accuracy: 64.19 Non-Poverty Accuracy: 92.48 Undercoverage: 35.81 Leakage: 20.47 PIE: -4.13 BPAC: 48.85
Best 15 indicators: R^2 adjusted = 0.792	
Next best five indicators: • Availability of telephone (fixed land line) in the house • Household owns a Motocab • Household owns a tractor • Days in past seven days with main meal consisting of plain rice and any vegetables • Amount that household needs per month to live	Total Accuracy: 85.38 Poverty Accuracy: 65.12 Non-Poverty Accuracy: 92.82 Undercoverage: 34.88 Leakage: 19.53 PIE: -4.13 BPAC: 49.77

3.9 Model 9

Model 9 uses a set of 127 regressors that are usually found in World Bank LSMS surveys. Model 9 shows higher adjusted R^2 levels than Model 5, 6, 7, and 8. Compared to Model 7, we observed an increase in the performance indicators, even though the LSMS indicators in Model 9 are similar to the indicators (demography, asset ownership, and housing) that are deemed "easy-to-verify" in Model 7. However, Model 9 also includes a complicated indicator of the overall value of assets. Thus, practitioners might prefer a set of indicators that achieve similar accuracy results, like those in Model 8.

The best performance in terms of maximization of BPAC (55.81 percentage points) is observed in the BEST10 set. Also, this set achieves 84.50 percent Total Accuracy, 66.05 percent Poverty Accuracy, and -4.12 PIE percentage points.

Table 3.9.1 Summary of the accuracy results Model 9

Variables	Model performance (%,%pt)	
Best 5 indicators: R^2 adjusted = 0.747		
• Total value of household assets • Household has electricity (autobattery, own generator included) • Remittances sent/total household expenditures • Average daily per-capita clothing expenditures • Availability of telephone (fixed land line) in the house	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE: BPAC:	82.75 62.33 86.70 37.67 26.51 -3.00 51.16
Best 10 indicators: R^2 adjusted = 0.784		
Next best five indicators: • Education level of household head is high school, university or post graduate • Percentage of adult household members who read and write • Number of cars owned by the household • Value of metal pots owned by the household • Leaves or straw as roof material	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE: BPAC:	84.50 66.05 87.97 33.95 23.72 -2.75 55.81
Best 15 indicators: R^2 adjusted = 0.798		
Next best five indicators: • Maximum education level of all members is incomplete primary • Poultry ownership • Wood as exterior-walls' material • Household head is single • Value of food processing assets	Total Accuracy: Poverty Accuracy: Non-Poverty Accuracy: Undercoverage: Leakage: PIE: BPAC:	86.37 66.98 93.50 33.02 17.67 -4.12 51.63

3.10 Results from other single-step regression techniques: Quantile, Probit, and Linear Probability Model

In contrast to the two-step models presented in the next chapter, which consist of two regression runs, we present single-step regressions in this chapter (i.e., regressions consisting of only one regression run).

The previous nine sections presented single-run models estimated with the Ordinary Least Squares (OLS) regression technique using the continuous dependent variable logarithm of daily per-capita expenditures. Annex E.1 summarizes their results, whereas Annex F.1 shows the BEST15 regressors for each of the nine models.

Alternative single-step regression techniques include Probit, Quantile, and Linear Probability Models (LPMs). The LPM and the Probit models use a dummy variable (coded one if the household is very-poor and zero otherwise) as dependent variable. The LPM model is also estimated with OLS using the SAS package, and the selection of BEST15 regressors is done using the MAXR procedure.

In the Probit and Quantile regressions, it is not feasible to use the MAXR procedure to select the BEST15 set of regressors. In order to test the accuracy performance of the Quantile regression model (which uses the log of daily per-capita expenditures as the dependent variable), the BEST15 regressors set (as it was determined by the OLS-MAXR regression) is used. The Quantile regression models are estimated with STATA package, whereas the Probit model is estimated with SAS. For the Probit model, where the dependent variable is a dummy variable similar to the LPM, we use the BEST15 regressors that were identified in the LPM model with the SAS MAXR procedure. The Probit model (like the LPM model) estimates the probability of a household being below the poverty line.

Annex E.3 presents the accuracy performance for these alternative single-step regression techniques. We restrict the testing of these alternative regression techniques to four sets of regressors, namely Model 1, 4, 7, and 9, and estimate only the models with a set of fifteen regressors.

For Model 1, we show the complete accuracy performance results in the following table.

Table 3.10.1 Summary of the accuracy results from the single-step regression techniques for Model 1

Model 1 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.834	87.38	70.70	29.30	17.67	-3.13	59.07
Quantile regression (estimation point: 45)		87.75	77.67	22.33	23.26	0.25	**76.74**
Linear Probability	0.523	89	68.37	31.63	9.30	-6.01	46.05
Probit		89	74.88	25.12	15.81	-2.5	65.58

For Model 1, the best single-step regression technique in terms of maximizing BPAC is the Quantile regression. Through an iterative procedure involving a series of regressions with the given set of BEST15 regressors, alternative percentile points of estimation for the quantile model are tested in order to maximize BPAC. With an optimal point of estimation set at the 45th percentile, the Quantile regression achieves a PIE of 0.25 percentage points. In other words, this model almost perfectly predicts the observed poverty rate. Moreover, the value for Poverty Accuracy is 77.67 percent, and for BPAC it is 76.74 percentage points. Compared to the other single-step regression techniques, there are considerable gains in the BPAC using the Quantile regression technique.

Similar results concerning the differences in accuracy performance between the four alternative single-step regression techniques are shown next. Tables 3.10.2, 3.10.3, and 3.10.4 present the results for Model 4, 7, 9, respectively.

Table 3.10.2 Summary of the accuracy results from the single-step regression techniques for Model 4

Model 4 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.800	87	66.05	33.95	14.42	-5.25	46.51
Quantile regression (estimation point: 38)		85.25	73.02	26.98	27.91	0.25	**72.09**
Linear Probability	0.489	88.13	68.37	31.63	12.56	-5.13	49.30
Probit		87.88	71.16	28.84	16.28	-3.38	58.61

Model 4 represents the set of regressors that excludes the total value of household assets and all expenditure variables. It includes all subjective poverty indicators and most indicators from the practitioner tools. Table 3.10.2 compares the accuracy performance of four single-step regression techniques for the set of regressors in Model 4. The best regression technique is the single-step Quantile. This technique achieves a value for BPAC of 72.09 and a value of PIE of 0.25 percentage points. Compared to the other three regression techniques, this result constitutes a significant improvement.

Table 3.10.3 Summary of the accuracy results from the single-step regression techniques for Model 7

Model 7 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.773	84.63	63.72	36.28	20.93	-4.13	48.37
Quantile regression (estimation point: 41)		84.88	72.09	27.91	28.37	0.13	**71.63**
Linear Probability		86.25	62.33	37.67	13.49	-6.5	38.15
Probit		85.75	66.98	33.02	20	-3.5	53.95

Model 7 represents the set of regressors that only includes poverty indicators that have been rated as easily verifiable (score 4 or 5) by the survey firm. Annex D provides the ratings for all 250 regressors. Table 3.10.3 compares the accuracy performance of four single-step regression techniques for the set of regressors termed Model 7. The best regression technique in terms of BPAC is again the Quantile model. This technique achieves a value for BPAC of 71.63 percentage points and a value of PIE of 0.13 percentage points. Compared to the other three single-step regression techniques, the Quantile result again constitutes a considerable improvement with respect to BPAC.

Table 3.10.4 Summary of the accuracy results from the single-step regression techniques for Model 9

Model 9 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.798	86.38	66.98	33.02	17.67	-4.13	51.63
Quantile regression (estimation point: 41)		85.13	72.56	27.44	27.91	0.13	**72.09**
Linear Probability	0.417	85.50	62.33	37.67	16.28	-5.75	40.93
Probit		85.63	65.58	34.42	19.07	-4.12	50.23

Model 9 represents the set of regressors usually contained in World Bank LSMS surveys. Table 3.10.4 compares the accuracy performance of four single-step regression techniques for the set of regressors termed Model 9. As before, the best regression technique in terms of BPAC is the Quantile model. This technique achieves a BPAC value of 72.09 percentage points and a PIE value of 0.13 percentage points.

In conclusion, the use of the Quantile regression technique allows us to substantially improve the accuracy performance (especially with respect to BPAC and PIE) compared to the single-step Ordinary Least Squares (OLS) models. This result holds true for all four sets of regressors (i.e. Model 1, 4, 7, and 9) which were tested in this section. The Probit and Linear Probability (LPM) techniques did not perform as well as the Quantile regression technique. In the next chapter, we describe the accuracy performance of the two-step methods.

Chapter Four: Two-step models

4.1 Introduction

The accuracy measures presented in Models 1 through 9 (Chapter 3) refer to the accuracy of the models in predicting the average poverty status for all expenditure percentiles in the full sample. However, they do not take into account the differences in accuracy observed at different levels of expenditure (benchmark indicator "daily expenditures per capita"). The models presented in Chapter 3 exhibit a high Total Accuracy, but a significantly lower Poverty Accuracy. They underestimated the Actual Poverty Incidence by a margin of three to six percentage points. The relatively low performance of the OLS models in Chapter 3 is partly driven by the level of the actual incidence of the very-poor (at 26.87 percent). The more a country's poverty rate deviates from a level of 50 percent, the more the single-step OLS models tend to show a weaker performance.

In order to improve the estimation of poverty status, we employed a two-step approach (see Grootaert et al., 1998) which breaks down the differences in the accuracy measures by percentiles of the benchmark indicator. The first step consists of estimating the original model for the full sample. The predicted daily expenditures per capita are compared against a wide range of benchmark cut-offs at different percentile levels. In step two, the model is estimated using a subsample that only includes those households whose predicted expenditures fall below the different cut offs, in order to identify the best regressor set for that subsample. The estimation in step two is repeated with OLS, using SAS's MAXR routine. Finally, the combined accuracy level of the two models is calculated by considering the predicted status from step one for the households with predicted expenditures above the different cut offs and the predicted status from step two for the subsample of predicted expenditures below the corresponding cut offs. The percentile cut-off that achieves the highest combined BPAC measure is then considered as the best two-step model.

In the remaining part of section 4.1, we present in detail first the results of the two-step OLS approach for the BEST15 regressor set of Model 1. Sections 4.2 to 4.4 present the results for similar two-step OLS approaches but using alternative sets of regressors, as defined by Models 4, 7 and 9. Overall, the two-step OLS models have a much better performance than the single-step OLS models presented in Chapter 3. The models lead to an improvement in Poverty Accuracy and a reduction of the Poverty Incidence Error (PIE). Also, the Balanced Poverty Accuracy Criterion (BPAC) increases noticeably (see Annex E.2). Section 4.5 combines the two-step approach with the three alternative regression techniques, i.e. Quantile, Probit, and Linear Probability Model.

4.1.1 First step: Model 1 — BEST15 set on full sample

We first evaluate the performance of Model 1 with the BEST15 regressors. Table 4.1.1.1 presents the results, which correspond to the results already shown in in Table 3.1.2.

Table 4.1.1.1 Accuracy level for the BEST 15 regressor set.

Measure	Level
Total Accuracy	87.38 %
Poverty Accuracy	70.70 %
Non-Poverty Accuracy	93.5 %
Undercoverage	29.30 %
Leakage	17.67 %
PIE	-3.13 %points
BPAC	59.07 %points

Table 4.1.1.2 presents a comparison of the predicted expenditures and the actual expenditures, both expressed in Soles. It can be observed that while the model tended to overestimate the daily expenditures per capita in the lower extreme of the distribution, it underestimated them in the higher levels of the distribution and around the mean. As a consequence, the predicted poverty incidence was 3.13 percentage points lower than the actual poverty incidence (negative PIE).

Table 4.1.1.2 Comparison between predicted and actual expenditures.

Variable	Minimum	Maximum	Mean	Std. Deviation
Actual daily expenditures per capita, Soles. (benchmark)	0.56	41.73	7.55	5.89
Predicted daily expenditures per capita, Soles.	0.97	38.19	7.24	5.19

4.1.2 Second step and combined accuracy of the two-step model

By testing the set of variables from Model 1 on the different subsamples (e.g. all expenditure percentiles above the headcount rate), the new BEST15 regressor sets were identified. Afterwards, the combined accuracy measures for all subsamples were determined and the optimal subsample was selected. For this, the main evaluation criterion was the maximization of BPAC.

Following this approach, the highest BPAC level was found when using the 81st percentile as the cutoff point for the subsample estimated in the second step. The combined accuracy measures from the two-step model are presented in Table 4.1.2.1.

It can be observed for the subsample that while the adjusted R^2 value was lower than in the BEST15 set from the first step, the model performed better. Total Accuracy and Poverty Accuracy increased from 87.38 to 88.63 percent and from 70.70 to 75.81 percent, respectively. While Undercoverage decreased by 5.11 percentage points, Leakage increased by 0.47 percentage points. The predicted poverty incidence increased, deriving on a PIE level of -1.63 percentage points, 1.5 points higher than in the single-step model.

BPAC increased from 59.07 percent in the single-step model to 69.77 percent, an increase of 10.7 percentage points.

Table 4.1.2.1. Combined accuracy from two-step estimation — Model 1

Measure	**Percentile 81st**
Number of observations in the subsample	661
Adjusted R^2 for the subsample	0.761
Total Accuracy (%)	88.63 %
Poverty Accuracy (%)	75.81 %
Undercoverage (%)	24.19 %
Leakage (%)	18.14 %
PIE (% points)	-1.63 % points
BPAC (% points)	69.77 % points

Table 4.1.2.2 presents the BEST15 regressor set obtained for the subsample at the 81st percentile. The BEST15 set from the first step (corresponding to BEST15 in Table 3.1.2) is presented for comparison. Eight variables (shaded in gray) appear to be important in both steps of the model. These were:

- Household ate less food for less than 30 days but more than 10 days during past 12 months
- In the last seven days, how many days did a main meal consist of plain rice only?
- Number of steps above step identified as int poverty line, if minus below
- How much household sent to relatives in last 12 months
- Value of motor tillers
- Annualized total household expenditures
- Total value of household assets
- In past 24 months, community had access to subsidized food (vaso de leche)

It is important to mention that only the variable *In past 24 months, community had access to subsidized food (vaso de leche)* was considered as easily verifiable by Instituto Cuánto.

Table 4.1.2.2 BEST15 regressor sets derived from the second step

Variables	Full sample	Percentile 81st
Distance to union headquarter?	X	
Household ate less food for less than 30 days but more than 10 days during past 12 months	X	X
Exterior walls: wood	X	
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	X	
In the last seven days, how many days did a main meal consist of plain rice only?	X	X
Household has electricity (autobattery, own generator included)	X	
Food expenditure share, C, in percentage	X	
Number of steps above step identified as int poverty line, if minus below	X	X
Average daily per-capita clothing expenditures	X	
How much household sent to relatives in last 12 months	X	X
Value of motor tillers	X	X
Annualized total household expenditures	X	X
Total value of household assets	X	X
Car number	X	
In past 24 months, community had access to subsidized food (vaso de leche)	X	X
Sum of distances to department, provincial and district capital		X
No lock in main entrance door or wood or metal bar to close from inside		X
Median education of adult household members		X
Household monthly expenditure on utilities (electricity, phone, water, etc)		X
How much are households usual monthly expenditures for transport		X
Rooms per person		X
Household head is single		X
Total number of new regressors	**15**	**7**

The last row in Table 4.1.2.2 shows the number of new regressors that substitute for some of the original regressors used in step 1. The BEST15 set for the subsample rely mostly on demographic and expenditure-related variables.

A practitioner tool based on a two-step model would have to include questions which obtain information on the 15 regressors selected by the BEST15 model of the first step. In addition, the practitioner tool would need to obtain information about the new additional poverty indicators that have been identified among the BEST15 regressors (percentile 81st) of the second step (see Table 4.1.2.2).

In practice, all questions related to the first and second step (15 plus 7 indicators) can be integrated into a single interview with each household. The interviewer could begin with the BEST15 indicators in the first step and then compute an estimated per capita daily expenditure. If the estimated expenditure falls above the cutoff value for the 81st percentile, the household is rated as not very-poor and the interview can be terminated. If, however, the predicted per capita expenditure value falls below this cut off, the interview needs to be continued by asking questions related to the seven additional regressors of the second step. Based on the values obtained for the seven regressors (plus the original regressors from the first step), a second value for predicted per capita daily expenditures is computed. If this second value is below the applicable poverty line, the household is rated as very-poor. In practice, however, it is recommended not to interrupt the interview for the calculation based on the first fifteen indicators, but to continue with the questions for the remaining seven poverty indicators. In this case, the calculations of one (or two) expenditure values are done after the interview.

4.2 Two-step Model 4

As mentioned in section 2.5, Model 4 excluded the variable *Total value of household assets* as well as all expenditure variables. With this, it was possible to create a regressors set containing all subjective poverty indicators and most indicators from the practitioners' tools. Table 4.2.1 presents the performance of the two-step approach for this set of regressors.

Table 4.2.1. Accuracy results for Model 4

Measure	Model 4	Percentile 88th
Number of observations	800	727
Adjusted R^2 for the sample/subsample	0.800	0.759
Total Accuracy (%)	87	87.75
Poverty Accuracy (%)	66.05	71.16
Undercoverage (%)	33.95	28.84
Leakage (%)	14.42	16.74
PIE (% points)	-5.25	-3.25
BPAC (% points)	46.51	59.07

Among the subsamples, the highest combined BPAC was found at the 88th percentile. Total Accuracy increased 0.75 percentage points. Poverty Accuracy registered an increase of 5.11 percentage points (meaning a similar reduction in Undercoverage). Leakage increased by only 2.32 percentage points.

The two-step approach predicted a higher incidence of poverty than that predicted by the single-step model. PIE changed from -5.25 to -3.25 percentage points, reducing the difference between predicted and observed poverty headcount. The gains in BPAC derived from this approach reach 12.56 percentage points.

4.3 Two-step Model 7

As explained in Chapter 2, Model 7 was constructed using the variables which were rated as easily verifiable by Instituto Cuánto. Table 4.3.1 presents the performance of the two-step approach for this set of regressors.

Table 4.3.1. Accuracy results for Model 7

Measure	Model 7	Percentile 54th
Number of observations	800	424
Adjusted R^2 for the sample/subsample	0.773	0.608
Total Accuracy (%)	84.63	85.88
Poverty Accuracy (%)	63.72	68.84
Undercoverage (%)	36.28	31.16
Leakage (%)	20.93	21.4
PIE (% points)	-4.13	-2.63
BPAC (% points)	48.37	59.07

Among the subsamples, the highest combined BPAC was found at the 54th percentile. Total Accuracy decrease slightly, by 1.25 percentage points, but Poverty Accuracy increased by 5.12 percentage points. Leakage increased by 0.47 percentage points.

The two-step approach predicted a higher incidence of poverty than that predicted by the single-step model. PIE improved from -4.13 to -2.63 percentage points, reducing the difference between predicted and observed poverty headcount. The gains in BPAC derived from this approach reaches 10.7 percentage points.

4.4 Two-step Model 9

As presented in the previous chapter, Model 9 incorporated a set of variables which are commonly found in World Bank LSMS datasets.

Table 4.4.1. Accuracy results for Model 9

Measure	Model 9	Percentile 87th
Number of observations	800	713
Adjusted R^2 for the sample/subsample	0.798	0.750
Total Accuracy (%)	86.38	85.88
Poverty Accuracy (%)	66.98	69.77
Undercoverage (%)	33.02	30.23
Leakage (%)	17.67	22.33
PIE (% points)	-4.13	-2.13
BPAC (% points)	51.63	61.86

The highest combined BPAC was found at the 87th percentile. For the second step, the subsample consisted on 713 households and the model yielded an adjusted R^2 of 0.750.

With respect to the accuracy measures, a slight improvement can be observed. While Total Accuracy decreased 0.5 percentage points, Poverty Accuracy increased 2.79 percentage points. As well, Leakage increased 4.66 percentage points. PIE improved reaching -2.13 percentage points and the gain in BPAC was 10.23 percentage points.

4.5 Results from other two-step regression techniques: Quantile, Probit, and Linear Probability Model

The previous three sections have presented models that were estimated with the Ordinary Least Squares (OLS) regression technique using the continuous dependent variable logarithm of daily per-capita expenditures. Annex E.2 summarizes their results, whereas Annex F.2 shows the BEST15 regressors for each of the four sets of regressors (i.e. Model 1, 4, 7, and 9).

Similar to single-step regression techniques, alternative formulations of the two-step approach again consist of using the Probit, Quantile, and Linear Probability Models as alternative regression techniques. For example, in a two-step modeling framework, a two-step Probit model consists of running two Probit regressions. Similar to the above OLS models, the first run includes the full sample, whereas the second subsample includes a subset of poorer households.

As already mentioned in section 3.10, the Linear Probability Model (LPM) and the Probit model have as dependent variable a dummy variable that is coded one if the household is very-poor and zero otherwise. Similar to the OLS regression technique presented in sections 4.1 to 4.4, the Quantile regression model uses the log of daily per-capita expenditures as the dependent variable. Similar to the single-step models, the regressors used in the two-step Quantile regressions are the same as those identified by SAS MAXR for the two-step OLS regressions. In addition, the percentile cutoff point for the second-step subsample in Quantile regressions is the same as the one determined in the two-step OLS model. Moreover, the point of estimation for the first step Quantile regression is similar to the one found optimal for the single-step Quantile model presented in Chapter 3. To identify the optimal second point of estimation for the second-step Quantile regression, we again employ an iterative procedure that runs a series of regressions with the given set of BEST15 regressors (as determined by the second-step OLS regression). Also similar to the single-step models, the regressors used in the two-step Probit regressions are the same as those identified by SAS MAXR for the two-step LPM regressions, and the cutoff point for the subsample in the two-step Probit is the same as in the LPM model.

We restrict the testing of the three alternative two-step regression techniques to four sets of regressors, namely Model 1, 4, 7, and 9. Again, the models are estimated with a set of best fifteen regressors.

For Model 1, the results concerning the accuracy performance of the four two-step regression techniques are shown in Table 4.5.1.

Table 4.5.1 Summary of the accuracy results of two-step regression techniques for Model 1

Model 1 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Two-step methods -MAXR variable selection							
OLS Percentile 81st	0.761 subsample	88.63	75.81	24.19	18.14	-1.63	69.77
Quantile regression (estimation points 45, 39) — 81 perc. Cutoff		89.25	79.53	20.47	19.53	-0.25	78.60
Linear Probability Percentile 66th	0.485 subsample	90.13	79.53	20.47	16.28	-1.13	75.35
Probit Percentile 66th		90.63	81.40	18.60	16.28	-0.63	**79.07**

Table 4.5.1 shows the accuracy performance of the four alternative two-step regression techniques. The OLS model is similar to the one presented in section 4.1. For the set of regressors as identified by Model 1, the above table shows that the best two-step regression technique in terms of maximizing BPAC is the Probit model.

This model has its cut off at the 66th percentile and yields a PIE of -0.63 percentage points and a BPAC of 79.07 percentage points. In spite of not having the best PIE (in two-step Quantile), this approach achieved a value close to zero and therefore the predicted poverty rate is close to the observed poverty rate. Poverty Accuracy was 81.40 percent. Compared with the two-step OLS regression technique, the gains in BPAC from using the Probit technique are considerable. Nevertheless, in comparison with the two-step Quantile and Linear Probability techniques, this technique shows only minor improvements.

For Model 4, the results concerning the accuracy performance of the four two-step regression techniques are shown in Table 4.5.2.

Table 4.5.2 Summary of the accuracy results of two-step regression techniques for Model 4

Model 4 Poverty rate: 26,88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Two-step methods -MAXR variable selection							
OLS Percentile 88th	0.759 subsample	87.75	71.16	28.84	16.74	-3.25	59.07
Quantile regression (estimation points 38, 35) — 88 perc.Cutoff		86.88	75.35	24.65	24.19	-0.13	**74.88**
Linear Probability Percentile 63rd	0.470 subsample	88.63	73.95	26.05	16.28	-2.63	64.19
Probit Percentile 63rd		88.5	73.02	26.98	15.81	-3	61.86

For the set of regressors in Model 4, Table 4.5.2 shows that the best two-step regression technique in terms of maximizing BPAC is the two-step Quantile. With points of estimation set at the 38th percentile for the first step and at the 35th percentile for the second step, this model achieves a PIE of -0.13 percentage points. In other words, it almost perfectly predicts the observed poverty rate. The value for Poverty Accuracy is 75.35 percent, and for the BPAC it is 74.88 percentage points. Compared with the two-step OLS regression technique, the two-step Quantile technique provides considerable gains in Poverty Accuracy and BPAC.

For Model 7, the results concerning the accuracy performance of the four two-step regression techniques are shown in Table 4.5.3.

Table 4.5.3 Summary of the accuracy results of two-step regression techniques for Model 7

Model 7 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Two-step methods -MAXR variable selection							
OLS Percentile 54th	0.608 subsample	85.88	68.84	31.16	21.4	-2.63	59.07
Quantile regression (estimation points 45, 25)							

— 54 perc. Cutoff		85.13	72.56	27.44	27.91	0.13	**72.09**
Linear Probability Percentile 64th	0.415 subsample	87.75	72.56	27.44	18.14	-2.5	63.26
Probit Percentile 64th		87	71.63	28.37	20	-2.25	63.25

For the set of regressors in Model 7, Table 4.4.3 shows that the best two-step regression technique in terms of maximizing BPAC is again Quantile regression. With points of estimation set at the 45th percentile for the first step and at the 25th percentile for the second step, and using the poorest 54th percent (similar to the two-step OLS) as the subsample for the second step, the Quantile regression achieves a PIE of 0.13 percentage points. In other words, this model almost perfectly predicts the observed poverty rate. Moreover, the value for Poverty Accuracy is 72.56 percent and the BPAC value is 72.09 percentage points. In comparison with the two-step OLS regression technique, using the two-step Quantile regression results in considerable BPAC gains.

For Model 9, the results concerning the accuracy performance of the four two-step regression techniques are shown in Table 4.5.4.

Table 4.5.4 Summary of the accuracy results of two-step regression techniques for Model 9

Model 9 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Two-step methods -MAXR variable selection							
OLS Percentile 87th	0.750 subsample	85.88	69.77	30.23	22.33	-2.13	61.86
Quantile regression (estimation points 41,37) — 87 perc.Cutoff		85.25	73.49	26.51	28.37	0.50	**71.63**
Linear Probability Percentile 59th	0.405 subsample	86.87	69.30	30.69	18.13	-3.37	56.74
Probit Percentile 59th		86.50	68.84	31.16	19.07	-3.25	56.74

For the set of regressors in Model 9 (i.e. the regressors usually contained in LSMS data sets), Table 4.5.4 shows that Quantile regression is the best two-step regression technique in terms of maximizing BPAC. With points of estimation set at the 41st percentile for the first step and at the37th percentile for the second step, and using the poorest 87 percent (similar to the two-

step OLS) as the subsample for the second step, the Quantile regression achieves a PIE of 0.5 percentage points. In other words, this model almost perfectly predicts the observed poverty rate. Moreover, the value for Poverty Accuracy is 73.49 percent, and for BPAC it is 71.63 percentage points. In comparison with the two-step OLS regression technique, using the two-step Quantile regression results in considerable gains in Poverty Accuracy and BPAC.

Annex E.3 presents the accuracy performance for these alternative two-step regression techniques and shows the results for the four single-step regression techniques that were already presented in section 3.10. The four summary tables in Annex E.3 show that the two-step Quantile regression technique achieves the highest BPAC for the two sets of regressors, namely Model 4 and Model 7. In Model 1, the two-step Probit achieved the highest BPAC. In Model 9, the single-step Quantile technique achieves the highest BPAC. Annex F.3. shows the set of regressors that were used by the regression technique (out of a total of eight techniques) that yielded the highest BPAC value.

Chapter five: Poverty outreach of institutions providing financial and business development services

5.1 Introduction

In this chapter, we present results on the poverty outreach of financial institutions in Peru. Section I1 of the composite questionnaire of the IRIS accuracy tests asked each adult household member (i.e. 18 years or older) about his or her client relationship with banks and microfinance institutions. Current clients were asked whether they received business development services from their provider of financial services. The questions in section I2 asked each adult household member separately whether they received business development services during the past five years.

In section 5.2, we analyze the poverty outreach performance of institutions providing financial or business development services on the basis of a nationally representative sample of 800 households. In Peru, an additional sample of 1175 households—supported by a grant from the Consultative Group to Assist the Poorest (CGA)—was randomly selected among clients from six purposefully selected microfinance institutions (MFIs). In Chapter 5.3, we report on the results of the depth of outreach of these six MFIs.

5.2 Outreach in the nationally representative sample

In the sample of 800 households, there were 2312 household members who were adults (18 years or older) and there were 142 households (160 adults) who were current clients of financial institutions. Of these client households, 93 had had loan transactions in the past with their financial institution(s) and they provided data on their most recent loans with these financial institutions. Furthermore, we asked if any previous loans from these institutions were still partly or fully to be repaid. In total, 112 loans were reported by the 93 borrowing households. The 112 loans were borrowed by 100 adults — i.e., 100 adults out of 160 clients were borrowers. Some households had more than one of their adult members borrowing from a financial institution, and some respondents had more than one loan to be repaid at the time of the survey.

We first show the poverty status of all clients of financial institutions, and then the poverty status of those clients who had borrowed in the past at least once. We term this latter group “borrowers” and in the tables below we disaggregate them by type of financial institution.

Table 5.2.1 Poverty status of non-clients and clients, by type of financial institution

Type of institution	Mean per capita daily expenditures (in Soles)	Very-poor (VP): Among bottom 50 percent poorest below national poverty line (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
Public bank (N=45)	10.44	22.22	28.89	4.44	24.44
Private bank (N=70)	12.28	7.14	18.57	0	2.86
Municipal savings and loan bank (N=37)	9.58	2.70	27.03	0	2.70
Rural savings and loan bank (N=5)	10.96	20.00	20.00	0	20.00
Cooperatives (N=5)	15.03	0	0	0	20.00
Micro-and small-enterprise development entity (Microbank) (N=4)	8.36	0	50.00	0	0
NGOs (N=9)	6.49	22.22	55.56	22.22	44.44
Other government entity providing financial services (N=4)	12.37	0	0	0	0
Not a client (N=2152)	7.15	29.37	53.90	9.67	33.69
Total (N=2331)	7.441	27.93	51.65	9.10	31.96

Note: This table includes multiple client relationships (i.e. a person has a client relationship with more than one financial institution).

The results show that 20 to 22 percent of clients of publicly owned banks, rural savings and loan banks, as well as non-government organizations belong to very-poor households (see third column in Table 5.2.1). Poverty outreach performance is lower for other types of financial institutions. We should point out, however, that disaggregation by type of financial institutions leads to too few cases (for example, the N for cooperatives is only 5). This means that we cannot make statistically valid generalizations about poverty outreach performance by type of institution.

However, when we use Peru's national poverty line to define poverty status, the poverty outreach of financial institutions noticeably improves (see fourth column in Table 5.2.1). Here, the headcount indices range from 0 (other government owned programs) to 55.6 percent (NGO credit institutions).

The table above shows that non-clients are on average poorer than clients. This is even more evident from the following table, which compares the mean of per-capita expenditures and of the three alternative headcount indices by client status.

Table 5.2.2 Poverty status of clients (n=160) compared to non-clients (n=2152)

Person is a client of a financial institution	Average per capita daily expenditures (in Soles)	Very-poor (VP): bottom 50 percent poorest (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
No (N=2152)	7.15	29.37	53.90	9.67	33.69
Yes (N=160)	10.74	11.25	25.63	2.50	12.50
All adults in sample of 800 households (N=2312)	7.40	28.11	51.95	9.17	32.22

Note: A t-test rejects the Null-Hypothesis of equal means in the two groups for all five variables at a probability of error of one percent.

Table 5.2.3 below compares poverty levels of borrowing clients versus non-borrowing clients.

Table 5.2.3 Poverty status of clients, by borrower status

Person has borrowed in the past	Average per-capita daily expenditures, in Soles	Very-poor (VP): bottom 50 percent poorest (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
No (N=60)	11.67	15.00	30.00	1.67	13.33
Yes (N=100)	10.18	9.00	23.00	3.00	12.00
Total number of clients (N=160)	10.74	11.25	25.63	2.50	12.50

Note: Multiple borrower or client relationships are excluded. Of the 160 clients, 100 persons have reported to have borrowed at least once. The differences between borrowers and non-borrowers are not statistically significant at a probability of error of ten percent.

Table 5.2.3 suggests that the headcount index for the three poverty measures is somewhat lower for non-borrowers than for borrowers. However, the observed differences are not statistically significant.

Table 5.2.4 considers the poverty status of clients at the household level. We distinguish three, non-mutally exclusive groups:

- households which at the time of survey had borrowed at least once from a formal institutions since becoming a client (i.e., current and previous borrowers combined)
- households which had received a business development service during the past five years
- households which have a savings (passbook) or a fixed term deposit account (i.e., that save with a formal institution)

Table 5.2.4 Poverty status of households, by type of client relationship

Household received / has	Per-capita daily expenditures, in Soles	Very-poor (VP): bottom 50 percent poorest (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
.. loans (N=93)	10.0	9.7	24.7	3.2	12.9
.. BDS services in past 5 years (N=54)	6.5	35.2	53.7	27.8	50.0
.. a savings or a fixed term deposit account (N=68)	12.5	5.9	22.1	2.9	11.8

While borrowers and savers appear similar in their poverty status, past and present clients of BDS services are much poorer compared to those receiving financial services, and also compared to the average household in Peru. About 35 percent of them are very-poor, and half of them earn less than two dollars a day per capita. As the three categories above are not mutually exclusive, we further differentiate these client households into seven mutally exclusive groups. Table 5.2.5 shows the poverty status of these nine groups, compared to non-clients.

Table 5.2.5 Poverty status of households, by type of client relationship

Household received …	Mean per-capita daily expenditures, in Soles	Very-poor (VP): bottom 50 percent poorest (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
a loan , BDS service and a savings service (N=1)	7.48	0	0	0	0
a loan and BDS	8.56	13.3	20.0	13.3	13.3

service (N=15)					
a loan and a savings service (N=15)	13.11	6.7	20.0	0	13.3
BDS service and a savings service (N=6)	5.97	16.7	50.0	16.7	50.0
only a loan (N=62)	9.68	9.7	27.4	1.6	12.9
only a BDS service (N=32)	5.64	50.0	71.9	37.5	68.8
only a savings service (N=46)	13.24	4.3	19.6	2.2	6.5
any financial or BDS services (N=623)	6.87	30.0	55.4	10.1	35.3
Total (N=800)	7.55	26.9	50.4	10.0	32.5

Some of the groups in the table have too few observations to draw any conclusions. Yet, one can observe a glaring difference in poverty status between those households that only received a BDS service compared to those that receive only a savings or only a loan service. Half of those receiving only BDS services belong to the very-poor whereas the respective percentages for borrowers and savers are 9.7 and 4.3 percent.

5.3 Outreach of six selected micro-finance institutions

Six micro-finance institutions were purposely selected to encompass a range of different types of MFIs (coops, microbanks, rural savings banks, NGOs) across urban and rural locations. Within the MFIs, only new clients within a confined geographical area were sampled, considering reasonable costs and other survey logistics. Sampling criteria included:

- MFIs should represent different institutional types (savings and credit cooperatives, NGOs, micro-banks, etc.)
- Some MFIs should have significant rural outreach, and should aim to target the poorer segments of the population
- The size of the MFI should be large enough to allow for a sample size of 200 new incoming clients
- The 200 new clients should be sampled from a complete list of new clients provided by the MFI for a smaller geographical area of Peru (i.e., one or few districts) in order to reduce logistical costs of the survey

A number of MFIs were approached by the survey firm to participate in the survey and some refused. Eventually, six MFIs volunteered to cooperate and provided a list of new clients along with information on how to find them. These six institutions are:

- EDYFICAR, a registered NGO (EDPYME)
- CRAC Cruz de Chalpon (a rural savings bank)
- CMAC Chinca (a municipal savings bank)
- Coop San Isidro Huaral (a cooperative)
- Coop San Pedro Andahuaylas (a cooperative)
- CARITAS (an NGO)

For a number of MFIs, it was difficult to obtain complete and correct new client lists, and, in some cases (in particular CMAC Chinca, Coop San Pedro, and CARITAS), 50 percent or more of clients had already been with the institution for more than one year. In one of the MFIs, only 175 instead of 200 clients could be surveyed for a variety of reasons beyond the survey firm's control. Moreover, Mibanco, the largest pro-poor financial service provider in Peru, was unwilling to participate.

In the sample of 1175 households, there were 3530 adult members. Among these, 1515 were current clients of financial institutions. The majority of the households (n=1047) had loan transactions in the past with their financial institution(s) and provided data on their most recent loans. In total, 1253 adults had borrowed in the past. We observed, once more, that in some households more than one adult borrowed from a financial institution, some respondents had more than one outstanding loan, and some respondents were clients of more than one institution.

The following table shows the poverty status of all clients, differentiated by type of financial institution. The first six institutions in Table 5.3.1 are those which were purposefully sampled. As in the national sample, several members of a household may be clients, and may work with more than one financial institution. The remaining three institutions listed (Mibanco, Banco del Travajo, and Banco de la Nation) include those complementary institutions with the largest market share which also showed up in the sample.

Table 5.3.1 Poverty status of clients, by financial institution

Financial institution	Mean per capita daily expenditures (in Soles)	Very-poor (VP): bottom 50 percent poorest (%)	Below national poverty line (%)	Below 1 US-$ PPP (%)	Below 2 US-$ PPP (%)
Edyficar (N=200)	10.7	16.5	41.0	0	2.5
CRAC Cruz de Chalpon (N=200)	11.5	12.6	23.4	1.1	9.7
CMAC Chinca (N=199)	10.3	8.0	38.1	0	6.0
Coop San Isidro Huaral (N=199)	12.2	4.0	15.6	0	1.5

Coop San Pedro Andahuaylas (N=200)	6.4	16.0	43.5	13.5	44.5
Caritas (N=198)	10.3	5.6	22.2	0.5	6.0
Mibanco (N=67)	11.8	13.4	31.3	0	3.0
Banco del Trabajo (N=45)	11.3	2.2	17.8	0	2.2
Banco de la Nacion (N=63)	12.0	3.1	9.5	0	4.8
Client of other financial institution (N=205)	11.9	2.9	15.6	0	3.4
Total (N=1551)	10.6	9.0	27.6	1.9	9.7

Note: This table includes multiple client relationships (i.e. the household has a client relationship with more than one financial institution).

The results show that Edyficar (a micobank) has the highest share of very-poor clients, followed by San Pedro (a savings and credit cooperative), and then Mibanco. Coop San Isidro Huaral had the lowest outreach to the very-poor among the purposefully selected institutions, with only 4 percent of very-poor clients. As expected, the large banks Banco del Trabajo and Banco de la Nation do not reach many very-poor. When one uses the national poverty line of Peru to define poverty status, the poverty outreach of financial institutions noticeably improves. In this case, the headcount indices range from 9.5 percent (in the case of Banco de la Nacion) to 43.5 percent (in the case of Cooperative San Pedro). When focusing on those falling below the international poverty line of one dollar per day, only the cooperative San Pedro achieved a noticeable outreach among the very-poor. The same holds true for the two-dollar poverty line.

The strong differences between the San Pedro cooperative and Edyficar, with respect to poverty outreach when using the international poverty lines of one and two dollars, occur because San Pedro works in rural areas outside Lima. In these areas, national poverty lines are low compared to international poverty lines, which do not account for differences in costs of living within a country. Edyficar, on the other hand, works in above-average areas with higher living costs, which therefore exhibit higher regional poverty lines. (Remember that the national poverty line is actually disaggregated in seven regional lines corresponding to the seven macro-regions.) In these areas, it reaches a considerable share of the very-poor — but most of these still earn incomes above the international poverty lines.

Chapter six: Summary

This report first presented nine single-step regression models, each with a set of best 5, best 10, and best 15 regressors. These models were run with the Ordinary Least Squares (OLS) regression technique. Because of the relatively low poverty headcount index (26.8%), the OLS models did not accurately predict the percentage of very-poor households, in spite of Total Accuracy exceeding 80 percent. Whereas tests of the single-step OLS regression technique are performed for nine different sets of regressors (Model 1 through Model 9), the alternative single-step and two-step regression techniques are limited to four sets of regressors — Model 1, 4, 7, and 9. As with the other three test countries (Bangladesh, Kazakhstan, and Uganda), Model 1 includes all poverty indicators enumerated in the field countries. Compared to Model 1, Model 4 excludes all expenditure categories and the total value of household assets but still contains most poverty indicators from practitioners' tools as well as subjective poverty indicators. Model 7 includes only indicators that are deemed by experienced survey firm staff as being highly verifiable and easy to ask — it is thus the most practical model. Model 9 uses indicators similar to those found in World Bank Living Standards Measurement Survey (LSMS) data sets.

The report also contains tests of three alternative single-step regression techniques — Probit, Quantile, and Linear Probability Model (LPM). Among those, the single-step Quantile regression technique yielded the highest value for the Balanced Poverty Accuracy Criterion (BPAC) for all four sets of regressors. In addition, two-step models were tested using the four different regression techniques.

The nine single-step OLS models show satisfactory levels of Total Accuracy. However, all nine models show lower Poverty Accuracy levels, and all consistently underestimate the poverty headcount, yielding negative Poverty Incidence Error (PIE) measures. Annex E.1 provides a summary of accuracy results for all nine single-step OLS models. Annex F.1 summarizes the variables used as best 15 regressors in the different models.

Considerable improvements were achieved by using Quantile regression techniques in a single-step framework. For Model 1, 4, 7, and 9, positive BPAC values of over 70 percentage points were achieved, while the value of PIE was close to zero — indicating an almost perfect prediction of observed poverty rate. These results constitute large improvements compared to the results obtained with single-step OLS, Probit, or LPM techniques.

Annex E.1 shows that all models estimated with a single-step using OLS were less accurate for the very-poor than for the not very-poor. This implies that the inaccuracies in prediction are not equally distributed over all expenditure percentiles. Rather, they are systematically higher for the very-poor than for the not very-poor. This problem of unbalanced accuracies can be potentially reduced by the use of two-step models, following a method

pioneered by Grootaert et al. (1998). The computational costs of these models are higher than single-step models.

The results of the two-step OLS models presented in Chapter 4 compare favorably with the single-step OLS models presented in Chapter 3. While the Total Accuracy of the two-step models is only marginally higher than for the single-step OLS models (and in certain cases it is even slightly lower), the two-step models have a clear advantage in estimating the proportion of the population that is very-poor.

The three regression techniques were explored in a two-step regression framework for the four different sets of regressors. Again, these are:

- Model 1 — full set of regressors
- Model 4 — all regressors except total value of assets and expenditure categories
- Model 7 — the model designed to be most practical
- Model 9 — with a set of regressors usually contained in LSMS data sets

When considering all eight different regression techniques tested (i.e. the four single-step and the four two-step techniques), the two-step Quantile regression model achieved the highest BPAC for two sets of regressors (Model 4 and Model 7), while two-step Probit achieved the highest BPAC in Model 1, and the single-step Quantile in Model 9. Annex F.3 lists the set of best regressors that were used in the model achieving the highest BPAC value.

In conjunction with tests in Bangladesh, Kazakhstan, and Uganda, the accuracy tests in Peru show that the choice of a suitable regression technique is an empirical issue. The choice is influenced by the level of the poverty rate. In countries with a low poverty rate, two-step techniques and/or Quantile regression appear to yield better results in terms of the BPAC.

Bibliography

ENNIV (2000). Encuesta Nacional sobre Medición de Niveles de Vida. Processed expenditure data of the National Living Standard Measurement Survey, provided by Instituto Cuánto.

Instituto Nacional de Estadística e Informática (INEI), available [online]: www.inei.gob.pe (accessed in October 2004)

Grootaert, Christiaan, and Jeanine Braithwaite. 1998. "Poverty correlates and indicator-based targeting in eastern Europe and the Former Soviet Union." Poverty Reduction and Economic Management Network, Environmentally and Socially Sustainable Development Network. Washington D.C.: The World Bank.

Henry, Carla, Manohar Sharma, Cécile Lapenu, and Manfred Zeller. 2003. Microfinance poverty assessment tool. Technical Tools Series No. 5, September 2003. Published by the Consultative Group to Assist the Poor (CGAP) and The World Bank, Washington, D.C. ISBN-0-8213-5674-7 (PDF-File at http://www.cgap.org/).

IRIS. 2005. Note on Assessment and Improvement of Tool Accuracy. Mimeograph, Revised version from June 2, 2005. IRIS center, University of Maryland.

Webb, Richard, Graciela Fernández B.. 2003. Anuario Estadístico — Perú en Números 2003. Lima, Peru: Instituto Cuánto.

Zeller, Manfred, Manohar Sharma, Akhter Ahmed, and Shahidur Rashid. 2002. Group-based financial institutions for the rural poor in Bangladesh: An institutional-and household-level analysis. Research report No. 120. Washington, D.C.: International Food Policy Research Institute.

Zeller, Manfred. 2004. Review of Poverty Assessment Tools. Report submitted to IRIS and USAID as part of the Developing Poverty Assessment Tools Project. January 2004. (Available [online] at http://www.povertytools.org).

Zeller, Manfred, Gabriela V. Alcaraz, and Julia Johannsen. 2005. Developing and testing poverty assessment tools: Results from accuracy tests in Bangladesh. IRIS Center, University of Maryland, College Park. First report in November 2004. Final report in August 2005 (Available [online] at http://www.povertytools.org).

Zeller, Manfred, and Gabriela V. Alcaraz. 2005. Developing and testing poverty assessment tools: Results from accuracy tests in Uganda. IRIS Center, University of Maryland, College Park. First report in May 2005. Final report in August 2005 (Available [online] at http://www.povertytools.org).

Zeller, Manfred, and Gabriela V. Alcaraz. 2005. Developing and testing poverty assessment tools: Results from accuracy tests in Kazakhstan. IRIS Center, University of Maryland, College Park. First report in June 2005, and final report in August 2005. (Available [online] at http://www.povertytools.org).

Zeller, Manfred, and Julia Johannsen. 2005. Is there a difference in poverty outreach by type of microfinance institution: The case of Peru. Report submitted to the Consultative Group to Assist the Poorest. IRIS Center, University of Maryland, College Park. April 2005. (Available [online] at http://www.cgap.org).

Annexes

Annex A-1: Distribution and size of the sample

	Total	Coast/ Lima	Highland	Lowland
TOTAL	**800**	**400**	**266**	**134**
Urban	**497**	**332**	**99**	**66**
Rural	**303**	**68**	**167**	**68**
Arequipa	**100**	**-**	**100**	**-**
Urban	**66**	**-**	**66**	**-**
Mariano Melgar	33	-	33	-
Tiabaya	33	-	33	-
Rural	**34**	**-**	**34**	**-**
Cerro Colorado	34	-	34	-
Cajamarca	**100**	**-**	**100**	**-**
Rural	**100**	**-**	**100**	**-**
Cajamarca	34	-	34	-
Encañada	33	-	33	-
Querocoto	33	-	33	-
Cusco	**100**	**-**	**66**	**34**
Urban	**33**	**-**	**33**	-
Wanchaq	33	-	33	-
Rural	**67**	**-**	**33**	**34**
Echarate	34	-	-	34
Quiquijana	33	-	33	-
La Libertad	**100**	**100**	**-**	**-**
Urban	**66**	**66**	**-**	**-**
La Esperanza	33	33	-	-
Trujillo	33	33	-	-
Rural	**34**	**34**	**-**	**-**
Chao	34	34	-	-

Lima	**200**	**200**	**-**	**-**
Urban	**200**	**200**	**-**	**-**
Ate	33	33	-	-
El Agustino	33	33	-	-
Lima	33	33	-	-
Rímac	34	34	-	-
San Juan de Miraflores	34	34	-	-
Santiago de Surco	33	33	-	-
Loreto	**100**	**-**	**-**	**100**
Urban	**66**	**-**	**-**	**66**
Iquitos	33	-	-	33
Punchana	33	-	-	33
Rural	**34**	**-**	**-**	**34**
Yurimaguas	34	-	-	34
Piura	**100**	**100**	**-**	**-**
Urban	**66**	**66**	**-**	**-**
Pariñas	33	33	-	-
Sullana	33	33	-	-
Rural	**34**	**34**	**-**	**-**
Chulucanas	34	34	-	-

Annex A-2: Mean annual expenditures per person and region, by expenditure percentile

National poverty line (in Soles), by region
Bottom 50 percent cut-off value below the national poverty line in the region

ENNIV 2000: Mean annual expenditures per person and region, by expenditure percentile

Region Per-centile	Lima Metrop.	Urban Coast	Rural Coast	Urban Highland	Rural Highland	Urban Lowland	Rural Lowland
1	684.2	455.5	368.0	443.4	266.0	377.5	261.7
2	915.1	608.5	428.9	556.4	308.6	567.0	389.0
3	1.028.4	778.9	443.8	582.8	346.6	635.5	422.7
4	1.094.4	877.5	481.8	709.1	372.5	692.3	442.9
5	1.203.3	946.5	576.3	798.8	408.2	758.7	467.7
6	1.294.2	986.1	600.2	855.3	435.1	790.1	493.6
7	1.359.1	1.045.0	612.8	909.5	458.0	839.7	506.9
8	1.425.1	1.074.4	625.4	929.1	486.5	878.7	510.1
9	1.452.7	1.085.9	651.8	958.0	497.4	905.2	517.8
10	1.495.2	1.113.5	665.1	967.9	506.7	921.7	530.5
11	1.543.5	1.156.0	669.8	994.6	516.7	930.1	542.3
12	1.579.7	1.174.9	678.8	1.026.6	528.5	950.2	560.9
13	1.638.5	1.192.4	690.0	1.047.9	565.0	1.007.3	566.4
14	1.688.3	1.256.2	698.5	1.073.0	579.6	1.021.7	577.2
15	1.725.1	1.278.2	716.4	1.117.4	592.4	1.048.4	606.3
16	1.773.3	1.298.8	743.0	1.164.6	599.3	1.079.7	621.2
17	1.796.8	1.315.1	766.6	1.191.3	610.4	1.100.9	638.1
18	1.810.2	1.336.2	780.7	1.243.4	626.5	1.116.0	654.1
19	1.846.4	1.356.4	799.2	1.275.6	642.8	1.131.2	665.0
20	1.874.8	1.408.3	816.2	1.303.9	659.4	1.156.4	671.0
21	1.933.7	1.449.4	845.4	1.328.4	678.9	1.176.8	692.9
22	1.989.0	1.476.2	860.8	1.351.8	686.3	1.201.5	699.9
23	2.010.8	1.486.0	890.0	1.395.1	695.5	1.219.2	707.4
24	2.059.7	1.511.9	904.2	1.427.0	704.3	1.245.3	728.6
25	2.093.1	1.526.1	913.8	1.457.7	716.1	1.264.3	758.3
26	2.117.5	1.553.0	929.5	1.487.2	722.4	1.298.2	784.6
27	2.145.1	1.575.4	936.9	1.503.4	737.9	1.315.5	795.6
28	2.166.5	1.593.2	962.9	1.532.9	753.7	1.335.1	808.1
29	2.206.6	1.611.9	969.5	1.562.0	764.0	1.350.3	819.6
30	2.256.9	1.624.6	977.4	1.606.8	775.9	1.370.3	829.4
31	2.285.6	1.650.9	1.007.5	1.628.1	784.2	1.391.2	838.1
32	2.307.5	1.680.1	1.023.6	1.657.3	791.0	1.410.9	850.0
33	2.361.9	1.711.9	1.035.1	1.693.1	799.1	1.433.0	866.2
34	2.419.9	1.755.7	1.049.3	1.730.4	811.4	1.446.3	871.3
35	2.452.2	1.785.8	1.070.3	1.748.4	826.5	1.474.5	878.2

National poverty line (in Soles), by region
Bottom 50 percent cut-off value below the national poverty line in the region

ENNIV 2000: Mean annual expenditures per person and region, by expenditure percentile

Region Per-centile	Lima Metrop.	Urban Coast	Rural Coast	Urban Highland	Rural Highland	Urban Lowland	Rural Lowland
36	2.473.0	1.798.3	1.087.0	1.762.2	856.6	1.505.3	899.0
37	2.511.3	1.812.7	1.106.8	1.777.9	875.5	1.536.9	919.4
38	2.552.0	1.829.8	1.123.6	1.798.5	900.4	1.550.8	942.2
39	2.596.1	1.847.4	1.132.1	1.820.2	917.7	1.564.4	953.9
40	2.631.1	1.865.3	1.134.8	1.861.4	925.7	1.593.0	960.9
41	2.661.6	1.892.2	1.179.7	1.900.0	938.4	1.643.1	973.8
42	2.699.0	1.913.5	1.200.0	1.929.3	952.4	1.663.1	986.2
43	2.733.6	1.951.6	1.210.6	1.965.0	973.7	1.674.3	993.2
44	2.768.9	1.987.5	1.229.9	2.001.4	986.3	1.693.2	1.009.8
45	2.810.7	1.999.4	1.253.2	2.043.1	998.6	1.709.7	1.023.2
46	2.861.5	2.033.0	1.279.2	2.082.5	1.003.5	1.745.1	1.030.1
47	2.922.9	2.079.8	1.291.2	2.133.4	1.013.7	1.782.9	1.047.8
48	2.960.7	2.126.6	1.311.1	2.157.3	1.031.0	1.817.2	1.058.9
49	2.991.3	2.169.4	1.319.9	2.188.8	1.045.0	1.866.2	1.065.3
50	3.043.6	2.233.1	1.325.7	2.227.8	1.055.1	1.911.7	1.071.0
51	3.110.2	2.268.4	1.342.4	2.278.1	1.074.2	1.934.7	1.078.6
52	3.152.0	2.322.5	1.358.9	2.311.0	1.084.8	1.947.7	1.085.8
53	3.215.9	2.335.3	1.368.8	2.344.6	1.095.3	1.982.2	1.100.9
54	3.261.3	2.353.2	1.382.8	2.385.6	1.100.7	2.005.2	1.114.2
55	3.296.8	2.391.4	1.400.9	2.435.2	1.113.7	2.037.0	1.122.2
56	3.356.4	2.442.2	1.410.8	2.473.9	1.141.0	2.080.6	1.144.4
57	3.423.2	2.459.2	1.435.5	2.502.4	1.157.0	2.102.5	1.160.1
58	3.465.9	2.501.1	1.463.4	2.540.1	1.182.0	2.132.7	1.169.8
59	3.498.6	2.552.3	1.472.1	2.613.0	1.189.9	2.185.9	1.179.2
60	3.563.7	2.599.6	1.490.2	2.656.9	1.212.7	2.232.3	1.189.9
61	3.617.5	2.648.1	1.515.7	2.681.1	1.255.3	2.281.2	1.210.6
62	3.675.2	2.687.2	1.550.5	2.739.4	1.275.8	2.350.7	1.230.5
63	3.745.2	2.719.4	1.574.8	2.802.5	1.289.0	2.416.6	1.246.0
64	3.796.7	2.752.6	1.591.5	2.839.4	1.303.9	2.472.7	1.265.1
65	3.842.0	2.777.4	1.634.1	2.872.8	1.315.2	2.527.4	1.278.8
66	3.898.9	2.812.4	1.662.5	2.971.7	1.337.2	2.577.5	1.290.1
67	4.039.0	2.857.8	1.677.1	3.020.7	1.381.7	2.609.7	1.301.0
68	4.147.4	2.906.5	1.690.9	3.093.9	1.409.6	2.643.8	1.327.6
69	4.242.3	2.933.5	1.707.9	3.171.6	1.429.5	2.712.7	1.353.4
70	4.308.3	2.960.5	1.739.5	3.220.0	1.443.4	2.746.8	1.378.6
71	4.397.4	3.007.6	1.771.7	3.264.8	1.456.5	2.814.3	1.406.7

National poverty line (in Soles), by region
Bottom 50 percent cut-off value below the national poverty line in the region

ENNIV 2000: Mean annual expenditures per person and region, by expenditure percentile

Region Per-centile	Lima Metrop.	Urban Coast	Rural Coast	Urban Highland	Rural Highland	Urban Lowland	Rural Lowland
72	4.473.4	3.068.6	1.800.9	3.285.5	1.487.0	2.868.1	1.445.5
73	4.584.4	3.154.8	1.818.7	3.405.3	1.517.6	2.965.5	1.471.7
74	4.713.5	3.209.3	1.846.5	3.501.8	1.552.2	3.012.5	1.514.6
75	4.873.7	3.256.3	1.865.7	3.624.0	1.590.2	3.048.9	1.557.9
76	5.017.3	3.358.5	1.919.6	3.667.8	1.627.4	3.091.9	1.597.1
77	5.145.0	3.448.6	1.982.6	3.718.9	1.652.1	3.149.7	1.627.0
78	5.262.0	3.542.0	2.016.6	3.801.3	1.685.7	3.208.5	1.663.7
79	5.489.1	3.612.0	2.084.4	3.864.5	1.719.6	3.293.3	1.696.1
80	5.721.9	3.702.5	2.147.5	4.032.8	1.764.7	3.398.4	1.735.2
81	5.920.8	3.767.9	2.243.7	4.147.4	1.795.0	3.481.4	1.780.3
82	6.107.4	3.887.8	2.301.8	4.307.8	1.813.3	3.547.8	1.807.7
83	6.349.7	3.983.2	2.339.7	4.437.2	1.854.9	3.652.0	1.826.0
84	6.507.7	4.071.7	2.373.9	4.525.4	1.880.5	3.753.6	1.872.0
85	6.753.0	4.158.0	2.434.7	4.610.3	1.916.0	3.864.0	1.926.0
86	7.097.6	4.278.1	2.487.0	4.659.6	1.943.5	3.985.1	1.980.5
87	7.346.3	4.457.3	2.538.8	4.724.4	2.032.4	4.151.2	2.007.5
88	7.695.0	4.644.3	2.610.1	4.867.6	2.116.7	4.240.6	2.042.1
89	8.152.0	4.856.9	2.703.4	5.049.6	2.169.1	4.400.2	2.085.3
90	8.693.4	5.118.0	2.796.4	5.304.3	2.266.5	4.557.2	2.144.4
91	9.158.8	5.455.9	2.952.9	5.655.5	2.397.2	4.754.2	2.198.6
92	9.732.7	5.928.8	3.060.0	5.863.6	2.529.7	4.956.1	2.282.6
93	10.301.8	6.334.4	3.347.9	6.377.9	2.644.5	5.230.2	2.347.1
94	10.910.5	6.864.3	3.494.3	6.738.9	2.751.1	5.670.9	2.432.5
95	11.676.3	7.202.5	3.785.9	7.313.0	2.971.0	6.081.4	2.618.3
96	13.230.5	7.955.7	4.091.4	8.093.9	3.314.8	6.774.1	2.776.6
97	14.599.6	9.935.2	4.917.4	9.179.7	3.774.7	7.500.2	3.070.3
98	19.539.1	12.014.2	6.455.4	10.946.9	4.551.6	8.550.2	3.431.9
99	32.372.6	18.442.4	8.909.6	14.848.9	6.143.3	12.843.9	5.513.9

Annex B: Descriptives of all regressors (n= 259), by type of model

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household size	1	14	4.68	2.04	X	X	X	X	X	X	X	X	X
Household size squared	1	196	26.06	23.99	X	X	X	X	X	X	X	X	X
Age of household head	18	94	47.72	16.11	X	X	X	X	X	X	X	X	X
Lima Metropolitan	0	1	0.25	0.43	X	X	X	X	X	X	X	X	X
Coast Rural	0	1	0.04	0.20	X	X	X	X	X	X	X	X	X
Highland Urban	0	1	0.13	0.33	X	X	X	X	X	X	X	X	X
Highland Rural	0	1	0.21	0.41	X	X	X	X	X	X	X	X	X
Lowland Urban	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Lowland Rural	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Age of youngest household member	0	90	12.37	15.56	X	X	X	X	X	X	X	X	
Age of oldest household member	18	94	50.28	17.09	X	X	X	X	X	X			
Median education of all household members	0	6	2.83	1.49	X	X	X	X	X	X			X
Minimum education level of any household member	0	6	1.21	1.41	X	X	X	X	X	X			X
Maximum education level of any household member	0	7	3.95	1.53	X	X	X	X	X	X			X
Median education of adult household members	0	7	3.24	1.57	X	X	X	X	X	X			X
Maximum education level of any adult household member	0	7	3.87	1.63	X	X	X	X	X	X			X
Spouse can read only	0	1	0.01	0.11	X	X	X	X	X	X	X	X	X
Number of household members who can read only	0	3	0.06	0.27	X	X	X	X	X	X	X	X	X
Number of adult household members who can read only	0	50	0.65	4.60	X	X	X	X	X	X	X	X	X
Household head can read and write	0	1	0.92	0.27	X	X	X	X	X	X	X	X	X
Number of adult household members who can read and write	0	9	2.65	1.49	X	X	X	X	X	X	X	X	X
Percentage of adult household members who read and write	0	100	59.23	25.90	X	X	X	X	X	X	X	X	X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household head is male	0	1	0.82	0.39	X	X	X	X	X	X	X	X	
Number of female adult household members	0	6	1.50	0.85	X	X	X	X	X	X	X	X	
Percentage of household members being chronically ill	0	100	9.72	19.35	X	X	X	X	X	X			X
Percentage of chronically ill adults (in relation to household size)	0	100	8.69	18.68	X	X	X	X	X	X			X
Household head is chronically ill	0	1	0.15	0.36	X	X	X	X	X	X			X
Percentage of household members with any disability (in relation to household size)	0	50	1.09	5.37	X	X	X	X	X	X	X	X	X
Percentage of adults with any disability (in relation to household size)	0	50	0.87	4.93	X	X	X	X	X	X	X	X	X
Percentage of dependents younger than 15 and older than 64 years (in relation to household size)	0	100	38.53	24.20	X	X	X	X	X	X	X	X	
Percentage of dependents younger than 14 and older than 60 years (in relation to household size)	0	100	38.59	24.85	X	X	X	X	X	X	X	X	
Head of household has non-agricultural self-employment	0	1	0.31	0.46	X	X	X	X	X	X			X
Head of household is non-agricultural daily worker	0	1	0.10	0.30	X	X	X	X	X	X			X
Head of household is retired	0	1	0.08	0.27	X	X	X	X	X	X			X
Head of household is occupied in housework	0	1	0.07	0.25	X	X	X	X	X	X			X
Head of household chooses leisure	0	1	0.00	0.05	X	X	X	X	X	X			X
Median education level of females	0	6	2.65	1.57	X	X	X	X	X	X			X
Maximal education level males	0	7	3.67	1.56	X	X	X	X	X	X			X
Number of male adults in household	0	6	1.39	0.89	X	X	X	X	X	X	X	X	X
Number of literate female adults in household	0	6	1.32	0.94	X	X	X	X	X	X	X	X	X
Number of females with some disability	0	1	0.02	0.13	X	X	X	X	X	X	X	X	X
Number of males with some disability	0	1	0.03	0.17	X	X	X	X	X	X	X	X	X
Number of females with some chronic illness	0	3	0.20	0.46	X	X	X	X	X	X			X
Number of males with some chronic illness	0	4	0.18	0.43	X	X	X	X	X	X			X
Average number of days sick by females	0	365	3.79	18.38	X	X	X	X	X	X			X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Total number of days sick by females	0	365	7.02	28.93	X	X	X	X	X	X			X
Ratio male adults/female adults	0	5	1.12	0.89	X	X	X	X	X	X	X	X	
How many rooms does the dwelling have?	1	12	3.17	1.90	X	X	X	X	X	X	X	X	X
Do you have Telephone (fixed land line) in the house?	0	1	0.28	0.45	X	X	X	X	X	X	X	X	X
Do you have Mobile (cell phone) in the house?	0	1	0.15	0.35	X	X	X	X	X	X	X	X	X
How many meals were served to the household members during the last 2 days?	2	8	5.70	0.77	X	X	X	X					
In the last seven days, how many days other red meat served by the household in a main meal eaten	0	7	0.80	1.32	X	X	X	X					
In the last seven days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	0	4	0.09	0.38	X	X	X	X					
In the last seven days, how many days Butter (urban) or margarina (rural) served by the household in a main meal eaten	0	7	1.46	2.37	X	X	X	X					
In the last seven days, how many days chicken eggs served by the household in a main meal eaten	0	7	3.00	2.33	X	X	X	X					
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	0	7	0.39	1.02	X	X	X	X				X	
In the last seven days, how many days did a main meal consist of plain rice only?	0	7	0.24	0.71	X	X	X	X					
In the last 30 days, for how many days did your household not have enough to eat everyday?	0	30	2.86	5.20	X	X	X	X					
Did you or any other adult household member lose weight in last 12 months because you did not have enough money to buy food?	0	1	0.21	0.41	X	X	X	X					
In the last seven days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	0.03	1000	1.28	35.35	X	X	X	X	X	X			

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
In last 3 years, how many marriages of a first degree relative to household head or spouse?	0	7	0.14	0.52	X	X	X	X	X	X			
Total number of children adopted, last 3 years	0	1	0.00	0.06	X	X	X	X	X	X	X	X	
Total number of months of serious (not chronic) illness of working adult member, last 3 years	0	36	0.42	2.40	X	X	X	X	X	X			
Occurrence of a serious chronic illness or major disability of any household member In last 3 years	0	1	0.02	0.15	X	X	X	X	X	X	X	X	
Has it occurred that a major working, income-earning adult member left the household for ever in last 3 years	0	1	0.03	0.16	X	X	X	X	X	X			
Total number of deaths of dependent household members, last 3 years	0	4	0.22	0.60	X	X	X	X	X	X	X	X	
Did your household have a very serious problem or failure in your own animal production in last 3 years?	0	1	0.18	0.38	X	X	X	X	X	X	X	X	
Did your household have a very serious problem or failure in your own micro-enterprise in last 3 years?	0	1	0.02	0.15	X	X	X	X	X	X	X	X	
During last 3 years, have you or any of your household members received in-kind services from food aid programs	0	1	0.34	0.47	X	X	X	X	X	X	X	X	
For how many months have you participated in school feeding during last three years?	0	36	3.28	9.44	X	X	X	X	X	X			
For how many months have you received subsidized food during last three years?	0	36	6.95	12.97	X	X	X	X	X	X			
For how many months have you participated in social kitchens during last three years?	0	36	0.45	3.54	X	X	X	X	X	X			
For how many months have you participated in other food aid programs during last three years?	0	36	0.08	1.48	X	X	X	X	X	X			

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Have you or members of household are denied service or only limited opportunity to job training/ employment	0	1	0.15	0.36	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to transportation	0	1	0.11	0.31	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to water distribution	0	1	0.09	0.28	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to sanitation services	0	1	0.06	0.23	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to agricultural extension	0	1	0.05	0.22	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to justice/conflict resolution	0	1	0.16	0.36	X	X	X	X					
Have you or members of household are denied service or only limited opportunity to security/police services	0	1	0.18	0.38	X	X	X	X					
Does the household own the house?	0	1	0.70	0.46	X	X	X	X	X	X	X	X	X
Total agricultural area (irrigated or not), square meters	5	800000	6939.60	33090.88	X	X	X	X	X	X			X
Household in Traders association	0	1	0.01	0.07	X	X	X	X	X	X	X	X	
Total household members in professional association	0	2	0.01	0.13	X	X	X	X	X	X	X	X	
Total household members in trade union	0	2	0.02	0.20	X	X	X	X	X	X	X	X	
Household in water/waste group	0	1	0.09	0.29	X	X	X	X	X	X	X	X	
Total household members in Water/waste group	0	2	0.10	0.32	X	X	X	X	X	X	X	X	

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Total of household members in NGO for BD services	0	1	0.01	0.08	X	X	X	X	X	X	X	X	
Household in other NGO	0	1	0.02	0.13	X	X	X	X	X	X	X	X	
Household in religious group	0	1	0.18	0.38	X	X	X	X	X	X	X	X	
Total of household members in religious group	0	8	0.35	0.91	X	X	X	X	X	X	X	X	
Household in youth group	0	1	0.01	0.08	X	X	X	X	X	X	X	X	
Total household members in youth group	0	2	0.01	0.12	X	X	X	X	X	X	X	X	
Total household members in women's group	0	2	0.06	0.25	X	X	X	X	X	X	X	X	
Household in parents group	0	1	0.14	0.35	X	X	X	X	X	X	X	X	
Total household members in parents group	0	2	0.16	0.40	X	X	X	X	X	X	X	X	
Household in sports group	0	1	0.03	0.16	X	X	X	X	X	X	X	X	
Total household members in sports group	0	2	0.03	0.17	X	X	X	X	X	X	X	X	
Total household members in other groups	0	2	0.04	0.21	X	X	X	X	X	X	X	X	
Cattle ownership	0	1	0.06	0.23	X	X	X	X	X	X	X	X	X
Milkcows ownership	0	1	0.08	0.27	X	X	X	X	X	X	X	X	X
Lamas ownership	0	1	0.01	0.09	X	X	X	X	X	X	X	X	X
Sheep/goats ownership	0	1	0.11	0.31	X	X	X	X	X	X	X	X	X
Pigs ownership	0	1	0.12	0.33	X	X	X	X	X	X	X	X	X
Poultry ownership	0	1	0.37	0.48	X	X	X	X	X	X	X	X	X
Motorcycle ownership	0	1	0.02	0.12	X	X	X	X	X	X	X	X	X
Motocab ownership	0	1	0.01	0.12	X	X	X	X	X	X	X	X	X
Tractor ownership	0	1	0.01	0.08	X	X	X	X	X	X	X	X	X
Other vehicle ownership	0	1	0.02	0.15	X	X	X	X	X	X	X	X	X
Color TV ownership	0	1	0.52	0.50	X	X	X	X	X	X	X	X	X
VCR ownership	0	1	0.12	0.32	X	X	X	X	X	X	X	X	X
Electric or gas cooker ownership	0	1	0.55	0.50	X	X	X	X	X	X	X	X	X
Microwave ownership	0	1	0.06	0.23	X	X	X	X	X	X	X	X	X
Fan ownership	0	1	0.14	0.34	X	X	X	X	X	X	X	X	X
Bed ownership	0	1	0.98	0.14	X	X	X	X	X	X	X	X	X
Suit/ Jacket ownership	0	1	0.34	0.47	X	X	X	X	X	X	X	X	X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Leather shoes ownership	0	1	0.53	0.50	X	X	X	X	X	X	X	X	X
Milkcow number	0	40	0.46	2.96	X	X	X	X	X	X			X
Pigs number	0	10	0.23	0.81	X	X	X	X	X	X			X
Horses number	0	8	0.22	0.75	X	X	X	X	X	X			X
Poultry number	0	120	4.49	9.43	X	X	X	X	X	X			X
Car number	0	3	0.09	0.31	X	X	X	X	X	X			X
Motorcycles number	0	2	0.02	0.14	X	X	X	X	X	X			X
Radios number	0	8	0.89	0.68	X	X	X	X	X	X			X
Colour TVs number	0	5	0.65	0.77	X	X	X	X	X	X			X
Video recorders number	0	3	0.13	0.36	X	X	X	X	X	X			X
Refrigerators number	0	2	0.38	0.50	X	X	X	X	X	X			X
Electric/ gas kitchens number	0	2	0.56	0.52	X	X	X	X	X	X			X
Beds number	0	10	3.21	1.75	X	X	X	X	X	X			X
Shoes number	0	30	1.85	3.04	X	X	X	X	X	X			X
Skirts number	0	20	1.15	2.39	X	X	X	X	X	X			X
Metal pots number	0	43	4.66	3.15	X	X	X	X	X	X			X
Wooden plows number	0	6	0.14	0.53	X	X	X	X	X	X			X
Food expenditure share in Percentage	0.11	1.93	1.21	0.30	X								
Average age of household members, except head	0.5	78	18.44	10.57	X	X	X	X	X	X			
Number of steps above step identified as int poverty line, if minus below	-9	6	1.24	1.87	X	X	X	X					
Household rates itself below the step reflecting the int. poverty line	0	1	0.14	0.35	X	X	X	X					
Household rates itself below the step reflecting the respective national poverty line	0	1	0.37	0.48	X	X	X	X					
Inferior food consumed at least sometimes, urb/rural calculation	0	1	0.17	0.37	X	X	X	X					
Inferior food consumed at least sometimes, macroregional calc	0	1	0.28	0.45	X	X	X	X					

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household participated at least in one food aid program during last 3 months	0	1	0.31	0.46	X	X	X	X	X	X			
Household participated in school feeding during last 3 months	0	1	0.14	0.34	X	X	X	X	X	X			
Household received subsidized food during last 3 months	0	1	0.25	0.43	X	X	X	X	X	X			
Household participated in social kitchen during last 3 months	0	1	0.02	0.13	X	X	X	X	X	X			
Household participated in any other food aid program during last 3 months	0	1	0.01	0.09	X	X	X	X	X	X			
Agree that you feel accepted as a member of this village/neighborhood?	0	1	0.94	0.24	X	X	X	X					
Agree that if you loose your goat (rural) or purse (urban) someone will give it back to you?	0	1	0.34	0.47	X	X	X	X					
Household feels that clothing expenses are below need	0	1	0.46	0.50	X	X	X	X					
Household feels that clothing expenses are above need	0	1	0.01	0.08	X	X	X	X					
Household feels that health care expenses are below need	0	1	0.37	0.48	X	X	X	X					
Household feels that health care expenses are above need	0	1	0.01	0.09	X	X	X	X					
Household feels that housing expenses are below need	0	1	0.38	0.48	X	X	X	X				X	
Household rates itself above intl. poverty line on ladder	0	1	0.71	0.45	X	X	X	X				X	
Household rates itself above national poverty line on ladder	0	1	0.41	0.49	X	X	X	X					
House size: small	0	1	0.23	0.42	X	X	X	X	X	X	X	X	
Quality of walls: poor	0	1	0.14	0.34	X	X	X	X	X	X	X	X	

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
No lock in main entrance door or wood or metal bar to close from inside	0	1	0.18	0.39	X	X	X	X	X	X	X	X	
Security key lock/metal frame with padlock in main entrance door	0	1	0.28	0.45	X	X	X	X	X	X	X	X	
Roof with leaves, straw of bamboo/wood	0	1	0.10	0.30	X	X	X	X	X	X	X	X	X
Roof with Cl sheet	0	1	0.37	0.48	X	X	X	X	X	X	X	X	X
Exterior walls: wood	0	1	0.08	0.27	X	X	X	X	X	X	X	X	X
Floor is wood or brick/stone	0	1	0.05	0.22	X	X	X	X	X	X	X	X	X
Cooking fuel is bamboo/wood/sawdust collected	0	1	0.23	0.42	X	X	X	X	X	X	X	X	X
Cooking fuel is bamboo/wood/sawdust purchased	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Toilet: pit toilet	0	1	0.15	0.35	X	X	X	X	X	X	X	X	X
Rooms per person	0.08	8	0.81	0.69	X	X	X	X	X	X	X	X	X
Public borehole/spring or public well	0	1	0.09	0.28	X	X	X	X	X	X	X	X	X
Untreated piped/river water	0	1	0.19	0.39	X	X	X	X	X	X	X	X	X
Head of household sleeps on something else than bed (e.g. floor, mat, mattress, hammock)	0	1	0.06	0.24	X	X	X	X	X	X	X	X	
Household cooks in one of the rooms in the house	0	1	0.13	0.33	X	X	X	X	X	X	X	X	
Number of days in past seven days any of six superior food eaten (max. 42)	0	25	8.81	5.26	X	X	X	X	X	X			
Household always ate enough from what they wanted (12 months)	0	1	0.10	0.30	X	X	X	X				X	
Household often did not have enough food (12 months)	0	1	0.05	0.23	X	X	X	X					
Household borrows from corner shop rarely	0	1	0.19	0.39	X	X	X	X					
Household borrows from corner shop often	0	1	0.15	0.35	X	X	X	X					
Household borrows from corner shop mostly	0	1	0.07	0.25	X	X	X	X					
Household borrows from neighbors/relatives rarely	0	1	0.11	0.31	X	X	X	X					
Household borrows from neighbors/relatives sometimes, often or mostly	0	1	0.10	0.30	X	X	X	X					

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household ate less food for less than 30 days but more than 10 days during past 12 months	0	1	0.12	0.33	**X**	**X**	**X**	**X**					
Household ate less food for less than 10 days during past 12 months	0	1	0.16	0.37	**X**	**X**	**X**	**X**					
Household had to skip meals less than 30 days but more than 10 days during past 12 months	0	1	0.06	0.23	**X**	**X**	**X**	**X**					
annualized total household expenditures	6.12	10.5	8.75	0.66	**X**								
Sum of household clothing expenditures in past 12 months	1.61	9.21	6.05	1.09	**X**						**X**	**X**	**X**
Annualized food expenditures recall average week	1.65	9.81	8.24	0.68	**X**								**X**
Annualized non-food expenditures (services, transport)	0.91	9.6	7.27	0.94	**X**								
Minimum wage male worker would accept during low income season if offered 8 hours work	0.69	4.61	2.81	0.89	**X**	**X**	**X**	**X**					
Minimum wage male worker would accept for next working day if offered 8 hours work	1.1	5.3	3.12	1.04	**X**	**X**	**X**	**X**					
Value food produced by household in farm or garden, or gathers and consumes, per week	-5.16	5.01	-2.88	3.58	**X**								
Household monthly expenditure on utilities (electricity, phone, water, etc)	-2.49	6.4	2.71	2.70	**X**	**X**	**X**					**X**	
How much household usual monthly expenditures for transport	-2.46	6.59	2.83	2.51	**X**								
How much household usual monthly expenditures for fuel	-3.38	5.86	2.59	2.12	**X**								
How much household usual monthly expenditures for other goods	-5.46	5.01	-2.98	3.84	**X**								
How much household spent last 12 months on school/ education	-0.37	9.48	4.19	2.91	**X**								
How much household spent last 12 months on health	-1.05	8.7	4.07	2.31	**X**								
Expenditures on furniture, last 12 months	-1.64	8.22	-0.64	2.59	**X**								

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
How much household sent to relatives in last 12 months	-1.55	8.52	-0.80	2.24	X								
Household expenditure on other expenditures in last 12 mo (social events, gifts, taxes)	-1.97	8.01	1.18	3.33	X								
Value of agricultural area, irrigated	1.1	11.29	1.90	2.40	X	X	X	X	X				
Total value of household assets	3	12.21	6.83	1.40	X	X							X
How much second person did send you from somewhere else, past 12 months	-3.66	5.7	-3.63	0.55	X	X	X	X	X				
How much does your household need per month to live	5.01	8.7	6.77	0.65	X	X	X	X				X	
Religion of household head is other than catholic	0	1	0.18	0.38	X	X	X	X	X	X			
Household usually purchases rice twice a week	0	1	0.05	0.21	X	X	X	X					
Household usually purchases rice weekly	0	1	0.27	0.44	X	X	X	X					
Household usually purchases rice fortnightly	0	1	0.08	0.27	X	X	X	X					
Household usually purchases rice monthly or less frequent than that	0	1	0.13	0.33	X	X	X	X					
Household owns any of motor tiller, wooden plow, tube irrigation or husking machine	0	1	0.10	0.30	X	X	X	X	X	X	X	X	X
Number of memberships out of 22 institutions	0	7	0.77	0.98	X	X	X	X	X	X	X	X	
Value of lamas	-8.05	4.87	-7.96	1.10	X	X	X	X	X				X
Value of milkcows	-2.9	8.52	-2.17	2.52	X	X	X	X	X				X
Value of sheep and goats	-4.57	7.4	-3.59	2.84	X	X	X	X	X				X
Value of pigs	-4.55	6.4	-3.48	2.90	X	X	X	X	X				X
Value of horses	-3.88	7.65	-2.80	3.02	X	X	X	X	X				X
Value of motorcabs	-1.53	9.62	-1.39	1.16	X	X	X	X	X				X
Value of tractors	-1.82	10.46	-1.75	0.93	X	X	X	X	X				X
Value of other vehicles	-4.65	8.16	-4.42	1.50	X	X	X	X	X				X
Value of radios	-2.12	8.01	2.60	2.80	X	X	X	X	X				X
Value of electric/ gas cooking	-1.86	7.31	1.71	3.38	X	X	X	X	X				X
Value of food processing assets	-3.12	6.21	-0.13	3.51	X	X	X	X	X				X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Value of fans	-4.54	5.48	-3.41	2.87	X	X	X	X	X				X
Value of bed/hammocks	-1.86	8.16	4.23	1.67	X	X	X	X	X				X
Value of sunday jackets	-2.93	7.94	-0.49	3.53	X	X	X	X	X				X
Value of leather shoes	-3.03	7.47	0.59	3.58	X	X	X	X	X				X
Value of metal pots	-3.2	7.35	2.18	2.51	X	X	X	X	X				X
Value of motor tillers	-6.91	4.38	-6.88	0.55	X	X	X	X	X				X
Value of tubes for irrigation	-2.64	11.72	-2.61	0.61	X	X	X	X	X				X
Value of radio, TV, VCR and cdplayer	-0.61	8.56	4.68	2.13	X	X	X	X	X				X
Total value of agricultural assets (motortiller, plow, irrigation, husking machine)	-2.56	11.72	-1.95	1.90	X	X	X	X	X				X
Total value of all animals	-1.84	9.81	0.87	3.58	X	X	X	X	X				X
Any household member has a passbook savings account	0	1	0.08	0.28	X	X	X	X	X	X			X
Any household member has a life insurance	0	1	0.04	0.20	X	X	X	X	X	X			X
Spouse has any account	0	1	0.05	0.21	X	X	X	X	X	X			X
Household has borrowed for food and emergencies from informal sector in past 3 years	0	1	0.23	0.42	X	X	X	X					
Household has lent money to others in past 3 years	0	1	0.15	0.36	X	X	X	X					
Household has borrowed from informal sector in past 3 years	0	1	0.31	0.46	X	X	X	X					
Household declares to not be able to save anything	0	1	0.83	0.38	X	X	X	X	X	X			
Value of jewelry	-1.81	11.85	-1.32	1.93	X	X	X	X	X				X
Value of largest loan for food/emergency in past 3 years	-1.33	10.04	0.19	2.96	X	X	X	X	X				
Value of debt owed by other households to household	-1.04	9.71	-0.04	2.40	X	X	X	X	X				X
Value of formal savings of spouse	-2.08	10.46	-1.98	0.76	X	X	X	X	X				X
Value of informal savings deposited at home or somewhere else	-3.76	6.4	-3.75	0.36	X	X	X	X	X				
Do you have secondary school ?	0	1	0.59	0.49	X	X	X	X	X	X	X	X	

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Do you have market/ bazaar?	0	1	0.42	0.49	X	X	X	X	X	X	X	X	
How far away the community center (km)?	0	4	0.22	0.70	X	X	X	X	X	X	X	X	
How far away the access to mainline phone (km)?	0	28	2.73	6.46	X	X	X	X	X	X	X	X	
Distance to department capital?	0	33	2.07	6.72	X	X	X	X	X	X	X	X	
In past 24 months, community had access to subsidized food (vaso de leche)	0	1	0.88	0.33	X	X	X	X	X	X	X	X	
Percentage of households that had access to children immunization programs in past 24 months	5	95	39.58	28.71	X	X	X	X	X	X	X	X	
Sum of distances to department, provincial and district capital	0	48	6.44	13.27	X	X	X	X	X	X	X	X	
Education level of other household members, excluding head	0	6	2.45	1.57	X	X	X	X	X	X			X
Squared age of household head	324	8836	2536.23	1705.18	X	X	X	X	X	X			X
Household has electricity (autobattery, own generator included)	0	1	0.76	0.43	X	X	X	X	X	X	X	X	X
Household has piped water (treated or untreated)	0	1	0.83	0.37	X	X	X	X	X	X	X	X	X
Remittances sent	-0.25	0.89	-0.10	0.25	X						X	X	X
Household head is single	0	1	0.07	0.25	X	X	X	X	X	X	X	X	X
Subjective ranking scale 1 to 5 compared to community	0	1	0.01	0.07	X	X	X	X					
Subjective ranking scale 1 to 5 compared to community	0	1	0.30	0.46	X	X	X	X					
Share of daily clothing expenditures in total daily expenditures	0	0.67	0.07	0.08	X								
Average daily per-capita clothing expenditures, Soles	-4.98	1.7	-1.29	1.07	X	X	X				X	X	X
Maximum education level of all household members is none or kindergarden	0	1.00	0.01	0.12									X
Maximum education level of all household members is incomplete primary	0	1.00	0.08	0.28									X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Maximum education level of all household members is complete primary	0	1.00	0.07	0.260									X
Maximum education level of all household members is incomplete secondary	0	1.00	0.11	0.320									X
Maximum education level of all household members is high school/university/post-graduate	0	1.00	0.34	0.470									X
Education level of household head is none or kindergarden	0	1.00	0.05	0.220									X
Education level of household head is incomplete primary	0	1.00	0.19	0.390									X
Education level of household head is complete primary	0	1.00	0.12	0.320									X
Education level of household head is incomplete secondary	0	1.00	0.13	0.33									X
Education level of household head is high school/univ/post-grad	0	1.00	0.19	0.39									X
Number of household members (w/o household head) with no education level or kindergarden	0	4.00	0.44	0.68									X
Number of household members (w/o household head) with incomplete primary	0	6.00	0.89	1.06									X
Number of household members (w/o household head) with complete primary	0	4.00	0.26	0.57									X
Number of household members (w/o household head) with incomplete secondary	0	4.00	0.59	0.87									X
Number of household members (w/o household head) with complete secondary	0	6.00	0.76	1.09									X
Number of household members (w/o household head) with high school level (superior)	0	5.00	0.24	0.58									X
Number of household members (w/o household head) with university level	0	4.00	0.19	0.56									X

Variable label	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Number of household members (w/o household head) with post-graduate level	0	1.00	0.00	0.06									**X**

Annex C: Gender-specific variables used in regression analysis

Variable label	N	Min.	Max.	Mean	Std. Dev.
Spouse can read only	800	0	1	0.01	0.11
Head of household is male	800	0	1	0.82	0.39
Number of female adult household members	800	0	6	1.50	0.85
Median education level of females	800	0	6	2.65	1.57
Maximal education level males	800	0	7	3.67	1.56
Number of male adults in household	800	0	6	1.39	0.89
Number of literate female adults in household	800	0	6	1.32	0.94
Number of females with some disability	800	0	1	0.02	0.13
Number of males with some disability	800	0	1	0.03	0.17
Number of females with some chronic illness	800	0	3	0.20	0.46
Number of males with some chronic illness	800	0	4	0.18	0.43
Average number of days sick by females	800	0	365	3.79	18.38
Total number of days sick by females	800	0	365	7.02	28.93
Ratio male adults/female adults	800	0	5	1.12	0.89
Total household members in women's group	800	0	2	0.06	0.25
Minimum wage main male income earner would accept during low income season if offered 8 hours work	800	0.69	4.61	2.81	0.89
Minimum wage main male income earner would accept for next working day if offered 8 hours work	800	1.1	5.3	3.12	1.04
Spouse has any account	800	0	1	0.05	0.21
Value of formal savings of spouse	800	-2.08	10.46	-1.98	0.76

Note: This list does not include gender-specific poverty indicators among the first set of 553 regressors that were submitted to the first MAXR analysis but to the set of the best 250 indicators that came out of that regression.

Annex D: Verifiability scores provided by Instituto Cuánto

Variable assessment scale: 1 very hard — 5 easily verifiable
Note: The shadowed indicators (with verifiability scores of 4, 5) have been included in Model 7 and Model 8

Table D.1 Verifiability score of the variables

Source: Communication via email in January 2005 with Instituto Cuánto, Peru (based on Cuántos own assessment and the results of a group discussion with the interviewers after their survey field work in August 2004)

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Household size	whole section B	4	4
Household size squared	whole section B	4	
Age of household head	B4	4	4
Lima Metropolitan	cover sheet	5	5
Coast Rural	cover sheet	5	
Highland Urban	cover sheet	5	
Highland Rural	cover sheet	5	
Lowland Urban	cover sheet	5	
Lowland Rural	cover sheet	5	
Age of youngest household member	B4	4	
Age of oldest household member	B4		
Median education of all household members	B8	3	2
Minimum education level of any household member	B8		
Maximum education level of any household member	B8		
Median education of adult household members	B8		
Maximum education level of any adult household member	B8		
Spouse can read only	B7	4	5
Number of household members who can read only	B7	4	
Number of adult household members who can read only	B7	4	
Household head can read and write	B7	4	
Number of adult household members who can read and write	B7	4	
Percentage of adult household members who read and write	B7	4	
Household head is male	B3	5	5
Number of female adult household members	whole section B	4	

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Percentage of household members being chronically ill	B13	3	3
Percentage of chronically ill adults (in relation to household size)	B13	3	
Household head is chronically ill	B13	3	
Percentage of household members with any disability (in relation to household size)	B14	5	5
Percentage of adults with any disability (in relation to household size)	B14	5	5
Percentage of dependents younger than 15 and older than 64 years (in relation to household size)	whole section B	4	
Percentage of dependents younger than 14 and older than 60 years (in relation to household size)	whole section B	4	
Head of household has non-agricultural self-employment	B11	3	5
Head of household is non-agricultural daily worker	B11	3	
Head of household is retired	B11	3	
Head of household is occupied in housework	B11	3	
Head of household chooses leisure	B11	3	
Median education level of females	B8	3	
Maximal education level males	B8	3	
Number of male adults in household	whole section B	4	
Number of literate female adults in household	B7	4	
Number of females with some disability	B14	5	
Number of males with some disability	B14	5	
Number of females with some chronic illness	B13	3	
Number of males with some chronic illness	B13	3	
Average number of days sick by females	B12	2	3
Total number of days sick by females	B12	2	
Ratio male adults/female adults	whole section B	4	
How many rooms does the dwelling have?	D4	4	5
Do you have Telephone (fixed land line) in the house?	D17C	5	5
Do you have Mobile (cell phone) in the house?	D17D	4	5
How many meals were served to the household members during the last 2 days?	E01A	2	3
In the last seven days, how many days Other red meat served by the household in a main meal eaten	E03B	2	4

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
In the last seven days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	E03C	2	4
In the last seven days, how many days Butter (urban) or margarina (rural) served by the household in a main meal eaten	E03E	2	4
In the last seven days, how many days Chicken eggs served by the household in a main meal eaten	E03F	2	4
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	E04	2	4
In the last seven days, how many days did a main meal consist of plain rice only?	E05	2	4
In the last 30 days, for how many days did your household not have enough to eat everyday?	E07	2	1
Did you or any other adult household member lose weight in last 12 months because you did not have enough money to buy food?	E14	2	2
In the last seven days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	E3C	2	3
In last 3 years, how many marriages of a first degree relative to household head or spouse?	G011	3	5
Total number of children adopted, last 3 years	G013	4	5
Total number of months of serious (not chronic) illness of working adult member, last 3 years	G018	3	5
Occurrence of a serious chronic illness or major disability of any household member In last 3 years	G019	4	4
has it occurred that a major working, income-earning adult member left the household for ever In last 3 years	G110		
Total number of deaths of dependent household members , last 3 years	G111	4	5
Did your household have a very serious problem or failure in your own animal production In last 3 years?	G115	4	4
Did your household have a very serious problem or failure in your own micro-enterprise In last 3 years?	G116	4	4
During last 3 years, have you or any of your household members received in-kind services from food aid programs	G119	4	4
For how many months have you participated in school feeding during last three years?	G120_1	2	3

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
For how many months have you received subsidized food during last three years?	G120_2	2	3
For how many months have you participated in social kitchens during last three years?	G120_3	2	3
For how many months have you participated in other food aid programs during last three years?	G120_5	2	3
Have you or members of household are denied service or only limited opportunity to job training/ employment	G3D	3	3
Have you or members of household are denied service or only limited opportunity to transportation	G3F	3	3
Have you or members of household are denied service or only limited opportunity to water distribution	G3G	3	3
Have you or members of household are denied service or only limited opportunity to sanitation services	G3H	3	3
Have you or members of household are denied service or only limited opportunity to agricultural extension	G3I	3	3
Have you or members of household are denied service or only limited opportunity to justice/conflict resolution	G3J	3	3
Have you or members of household are denied service or only limited opportunity to security/police services	G3K	3	3
Does the household own the house?	D1	4	4
Total agricultural area (irrigated or not), square meters	F1A	1	3
Household in Traders association	G2A2	4	4
Total household members in professional assoc.	G2A2	4	
Total household members in trade union	G2A2	4	
Household in water/waste group	G2A2	4	
Total household members in Water/waste group	G2A2	4	
Total of Household members in NGO for BD services	G2A2	4	
Household in other NGO	G2A2	4	
Household in religious group	G2A2	4	
Total of household members in religious group	G2A2	4	
Household in youth group	G2A2	4	
Total household members in youth group	G2A2	4	
Total household members in women's group	G2A2	4	
Household in parents group	G2A2	4	
Total household members in parents group	G2A2	4	

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Household in sports group	G2A2	4	
total household members in sports group	G2A2	4	
Total household members in other groups	G2A2	4	
Cattle ownership	section F2, quantity	4	4
Milkcows ownership	section F2, quantity	4	4
Lamas ownership	section F2, quantity	4	4
Sheep/goats ownership	section F2, quantity	4	4
Pigs ownership	section F2, quantity	4	4
Poultry ownership	section F2, quantity	4	4
Motorcycle ownership	section F2, quantity	4	4
Motocab ownership	section F2, quantity	4	4
Tractor ownership	section F2, quantity	4	4
Other vehicle ownership	section F2, quantity	4	4
Color TV ownership	section F2, quantity	4	4
VCR ownership	section F2, quantity	4	4
Electric or gas cooker ownership	section F2, quantity	4	4
Microwave ownership	section F2, quantity	4	4
Fan ownership	section F2, quantity	4	4
Bed ownership	section F2, quantity	4	4
Suit/ Jacket ownership	section F2, quantity	4	4
Leather shoes ownership	section F2, quantity	4	4
Milkcow number	section F2, quantity	2	
Pigs number	section F2, quantity	2	
Horses number	section F2, quantity	2	4
Poultry number	section F2, quantity	2	
Car number	section F2, quantity	2	4
Motorcycles number	section F2, quantity	2	
Radios number	section F2, quantity	2	4
Colour TVs number	section F2, quantity	2	
Video recorders number	section F2, quantity	2	
Refridgerators number	section F2, quantity	2	4
Electric/ gas kitchens number	section F2, quantity	2	
Beds number	section F2, quantity	2	
Shoes number	section F2, quantity	2	
Skirts number	section F2, quantity	2	4

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Metal pots number	section F2, quantity	2	4
Wooden plows number	section F2, quantity	2	4
Food expenditure share in percentage	C1, C2	2	4
Average of household members, except head	B4	3	3
Number of steps above step identified as int poverty line, if minus below	H7	1	
Household rates itself below the step reflecting the int. poverty line	H7	1	
Household rates itself below the step reflecting the respective national poverty line	H7	1	
Inferior food consumed at least sometimes, urb/rural calculation	E15	2	1
Inferior food consumed at least sometimes, macroregional calc	E15	2	
Household participated at least in one food aid program during last 3 months	E17	3	3
Household participated in school feeding during last 3 months	E17	3	
Household received subsidized food during last 3 months	E17	3	
Household participated in social kitchen during last 3 months	E17	3	
Household participated in any other food aid program during last 3 months	E17	3	
Agree that you feel accepted as a member of this village/neighborhood?	DG2B4	3	4
Agree that if you loose your goat (rural) or purse (urban) someone will give it back to you?	DG2B5/6	3	5
Household feels that clothing expenses are below need	H2	3	5
Household feels that clothing expenses are above need	H2	3	
Household feels that health care expenses are below need	H3	3	5
Household feels that health care expenses are above need	H3	3	
Household feels that housing expenses are below need	H5	3	5
Household rates itself above subjective intl. poverty line on ladder	H8b	1	
Household rates itself above subjective natl. poverty line on ladder,	H8a	1	
House size: small	D2.2	4	4
quality of walls: poor	D2.3	5	5

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
No lock in main entrance door or wood or metal bar to close from inside	D3	5	5
Security key lock/metal frame with padlock in main entrance door	D3	5	
Roof with leaves, straw of bamboo/wood	D6	5	
Roof with Cl sheet	D6	5	
Exterior walls: wood	D7	5	5
Floor is wood or brick/stone	D8	5	5
Cooking fuel is bamboo/wood/sawdust collected	D9	5	5
Cooking fuel is bamboo/wood/sawdust purchased	D9	5	
Toilet: pit toilet	D12	5	5
Rooms per person	D4	4	
Public borehole/spring or public well	D11	5	
Untreated piped/river water	D11	5	
Head of household sleeps on something else than bed (e.g. floor, mat, mattress, hammock)	D15	4	5
Household cooks in one of the rooms in the house	D16	5	5
Number of days in past seven days any of six superior food eaten (max. 42)	E3	3	3
Household always ate enough from what they wanted (12 months)	E9	3	4
Household often did not have enough food (12 months)	E9	3	
Household borrows from corner shop rarely	E12a	3	3
Household borrows from corner shop often	E12a	3	
Household borrows from corner shop mostly	E12a	3	
Household borrows from neighbors/relatives rarely	E12b	3	3
Household borrows from neighbors/relatives sometimes, often or mostly	E12b	3	
Household ate less food for less than 30 days but more than 10 days during past 12 months	E13a	2	2
Household ate less food for less than 10 days during past 12 months	E13a	2	
Household had to skip meals less than 30 days but more than 10 days during past 12 months	E13b	2	2
Annualized total household expenditures	whole section C		
Sum of household clothing expenditures in past 12 months	B15	4	4

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Annualized food expenditures recall average week	C1, C2	1	3
Annualized nonfood expenditures (services, transport)	C3, C4	3	
Minimum wage male worker would accept during low income season for 8 hours working	D18b	1	1
min wage male working memmber would accept for next working day for 8 hours working	D18a	1	1
Value of food produced by household in farm or garden, or gathers and consumes, per week	C2	1	2
Household monthly expenditure on utilities (electricity, phone, water, etc)	C3	3	4
How much household usual monthly expenditures for transport	C4	3	4
How much household usual monthly expenditures for fuel	C5	3	4
How much household usual monthly expenditures for other goods	C6	3	3
How much household spent last 12 months on school/ education	C7	3	4
How much household spent last 12 months on health	C8	3	4
expenditures on furniture, last 12 months	C10	3	4
How much household sent to relatives in last 12 months	C11	3	4
Household expenditure on other expenditures in last 12 mo (social events, gifts, taxes)	C12	3	4
Value of agricultural area, irrigated	F1A	1	
Total value of household assets	Section F2, value	2	
How much second person did send you from somewhere else, past 12 months	B20	3	4
How much does your household need per month to live	H6	1	
Religion of household head is other than catholic		3	5
Household usually purchases rice twice a week	E6	3	5
Household usually purchases rice weekly	E6	3	
Household usually purchases rice fortnightly	E6	3	
Household usually purchases rice monthly or less frequent than that	E6	3	
Household owns any of motor tiller, wooden plow, tube for irrigation or husking machine	F2	4	
Number of memberships out of 22 institutions	G2A2	4	
Value of lamas	Section F2, value	2	3

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Value of milkcows	Section F2, value	2	3
Value of sheep and goats	Section F2, value	2	3
Value of pigs	Section F2, value	2	3
Value of horses	Section F2, value	2	3
Value of motorcabs	Section F2, value	2	3
Value of tractors	Section F2, value	2	3
Value of other vehicles	Section F2, value	2	3
Value of radios	Section F2, value	2	3
Value of electric/ gas cooking	Section F2, value	2	3
Value of food processing assets	Section F2, value	2	3
Value of fans	Section F2, value	2	3
Value of bed/hammocks	Section F2, value	2	3
Value of sunday jackets	Section F2, value	2	3
Value of leather shoes	Section F2, value	2	3
Value of metal pots	Section F2, value	2	3
Value of motor tillers	Section F2, value	2	3
Value of tubes for irrigation	Section F2, value	2	3
Value of radio, TV, VCR and cdplayer	Section F2, value	2	3
Total value of agricultural assets (motortiller, plow, irrigation, huskmach)	Section F2, value	2	3
Total value of all animals	Section F2, value	1	1
Any household member has a passbook savings account	102.3	1	1
Any household member has a life insurance	102.3	1	
Spouse has any account	102.3	1	
Household has borrowed for food and emergencies from informal sector in past 3 years	G118	3	4
Household has lent money to others in past 3 years	101.5	3	4
Household has borrowed from informal sector in past 3 years	K14, G118		
Household declares to not be able to save anything	102.2	1	2
Value of jewelry	101.2	1	1
Value of largest loan for food/emergency in past 3 years	G118	1	
Value of debt owed by other households to household	101.5	3	4
Value of formal savings of spouse	102.3	1	
Value of informal savings deposited at home or somewhere else	101.1	1	1

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Do you have secondary school ?	Community: B9	4	5
Do you have market/ bazaar?	Community: B13	4	5
How far away the community center (km)?	Community: B15	5	5
How far away the access to mainline phone (km)?	Community: B22	5	5
Distance to union headquarter?	Community: B25	5	5
In past 24 months, community had access to subsidized food (vaso de leche)	Community: D2	5	5
Percentage of households that had access to children immunization programs in past 24 months	Community: D5	2	5
Sum of distances to department, provincial and district capital	Community: B23-25	5	
Education level of other household members, excluding head	B8	3	
Squared age of household head	B4	4	
household has electricity (autobattery, own generator included)	D10	5	5
Household has piped water (treated or untreated)	D11	5	
Remittances sent	C11	4	5
Household head is single	B5	4	5
Subjective ranking scale 1 to 5 compared to community	A11a	3	4
subjective ranking scale 1 to 5 compared to community	A11a	3	
Share of daily clothing expenditures in total daily expenditures	B15	3	
Average daily per-capita clothing expenditures, Soles	B15	4	

Annex E: Accuracy performance of regression models

Annex E.1: Single-step OLS models with per-capita daily expenditures as continuous dependent variable (selection of regressors by MAXR)

Model	Description	Type	Adj. R^2	Total Accuracy (%)	Poverty Accuracy (%)	Under-coverage (%)	Leakage (%)	PIE (% points)	BPAC (% points)
1	All 260 regressors *(Ref. Table 3.1.2)*	B-5	0.793	84.5	63.72	36.28	21.4	-4.00	48.84
		B-10	0.82	87	70.23	29.77	18.6	-3.00	59.06
		B-15	0.834	87.38	70.7	29.3	17.67	-3.13	59.07
2	Exclusion of expenditure *monthly household expenditures on utilities, i.e. electricity, phone, water etc.* and *clothing expenditures per capita in past 12 month (Ref. Table 3.2.1)*	B-5	0.763	83.75	64.19	35.81	24.65	-3.00	53.03
		B-10	0.796	85.25	65.12	34.88	20	-4.00	50.24
		B-15	0.81	87.5	70.7	29.3	17.21	-3.25	58.61
3	Exclusion of *total value of household assets (Ref. Table 3.3.1)*	B-5	0.741	83.83	59.07	40.93	20.93	-5.38	39.07
		B-10	0.789	85.75	65.58	34.42	18.6	-4.25	49.76
		B-15	0.808	88.13	70.7	29.3	14.88	-3.88	56.28
4	Exclusion of *monthly household expenditures on utilities, i.e. electricity, phone, water etc.* and *clothing expenditures per capita in past 12 month (Ref. Table 3.4.1)*	B-5	0.741	84.5	60.93	39.07	18.6	-5.50	40.46
		B-10	0.781	87	67.44	32.56	15.81	-4.50	50.69
		B-15	0.8	87	66.05	33.95	14.42	-5.25	46.52

Model	Description	Type	Adj. R^2	Total Accuracy (%)	Poverty Accuracy (%)	Under-coverage (%)	Leakage (%)	PIE (% points)	BPAC (% points)
5	Exclusion subjective variables *(Ref. Table 3.5.1)*	B-5	0.723	82.5	57.21	42.79	22.33	-5.50	36.75
		B-10	0.765	84.88	63.26	36.74	19.53	-4.63	46.05
		B-15	0.784	85.13	64.19	35.81	19.53	-4.38	47.91
6	Exclusion monetary variables *(Ref. Table 3.6.1)*	B-5	0.72	81.5	55.35	44.65	24.19	-5.50	34.89
		B-10	0.76	85.13	62.79	37.21	18.14	-5.13	43.72
		B-15	0.777	85.13	64.65	35.35	20	-4.13	49.30
7	Easily verifiable variables (Cuánto) *(Ref. Table 3.7.1)*	B-5	0.708	82.38	56.28	43.72	21.86	-5.88	34.42
		B-10	0.755	83.63	61.86	38.14	22.79	-4.13	46.51
		B-15	0.773	84.63	63.72	36.28	20.93	-4.13	48.37
8	Model 7 plus strong subjective and expenditure regressors *(Ref. Table 3.8.1)*	B-5	0.725	83.38	57.67	42.33	19.53	-6.13	34.87
		B-10	0.773	84.88	64.19	35.81	20.47	-4.13	48.85
		B-15	0.792	85.38	65.12	34.88	19.53	-4.13	49.77
9	LSMS-type regressors (Ref. Table 3.9.1)	B-5	0.747	82.75	62.33	37.67	26.51	-3.00	51.16
		B-10	0.784	84.50	66.05	33.95	23.72	-2.75	55.81
		B-15	0.798	86.37	66.98	33.02	17.67	-4.12	51.63

Annex E.2: Two-step models with a continuous dependent variable (OLS estimation) for Models 1, 4, 7 and 9

OLS 2-Step Poverty rate: 31.41%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Model 1 Percentile 81th	0.761 subsample	88.63	75.81	24.19	18.14	-1.63	69.77
Model 4 Percentile 88th	0.759	87.75	71.16	28.84	16.74	-3.25	59.07
Model 7 Percentile 54th	0.608 subsample	85.88	68.84	31.16	21.4	-2.63	59.07
Model 9 Percentile 87th	0.750 subsample	85.88	69.77	30.23	22.33	-2.13	61.86

Annex E.3: Summary results for all single and two-step regressions (Models 1, 4, 7 and 9).

Model 1 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.834	87.38	70.70	29.30	17.67	-3.13	59.07
Quantile regression (estimation point: 45)		87.75	77.67	22.33	23.26	**0.25**	76.74
Linear Probability	0.523	89	68.37	31.63	9.30	-6.01	46.05
Probit		89	74.88	25.12	15.81	-2.5	65.58
Two-step methods -MAXR variable selection							
OLS Percentile 81th	0.761 subsample	88.63	75.81	24.19	18.14	-1.63	69.77
Quantile regression (estimation points 45, 39) — 81 perc. cutoff		89.25	79.53	20.47	19.53	**-0.25**	*78.60*
Linear Probability Percentile 66th	0.485 subsample	90.13	79.53	20.47	16.28	-1.13	75.35
Probit Percentile 66th		90.63	81.40	18.60	16.28	-0.63	**79.07**

Model 4 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.800	87	66.05	33.95	14.42	-5.25	46.51
Quantile regression (estimation point: 38)		85.25	73.02	26.98	27.91	*0.25*	*72.09*
Linear Probability	0.489	88.13	68.37	31.63	12.56	-5.13	49.30
Probit		87.88	71.16	28.84	16.28	-3.38	58.61
Two-step methods -MAXR variable selection							
OLS Percentile 88th	0.759	87.75	71.16	28.84	16.74	-3.25	59.07
Quantile regression (estimation points 38, 35) — 88 perc. cutoff		86.88	75.35	24.65	24.19	**-0.13**	**74.88**
Linear Probability Percentile 63rd	0.470	88.63	73.95	26.05	16.28	-2.63	64.19
Probit Percentile 63rd		88.5	73.02	26.98	15.81	-3	61.86

Model 7 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.773	84.63	63.72	36.28	20.93	-4.13	48.37
Quantile regression (estimation point: 41)		84.88	72.09	27.91	28.37	**0.13**	*71.63*
Linear Probability		86.25	62.33	37.67	13.49	-6.5	38.15
Probit		85.75	66.98	33.02	20	-3.5	53.95
Two-step methods -MAXR variable selection							
OLS Percentile 54th	0.608 subsample	85.88	68.84	31.16	21.4	-2.63	59.07
Quantile regression (estimation points 45, 25) — 54 perc. cutoff		85.13	72.56	27.44	27.91	**0.13**	**72.09**
Linear Probability Percentile 64th	0.415 subsample	87.75	72.56	27.44	18.14	-2.5	63.26
Probit Percentile 64th		87	71.63	28.37	20	-2.25	63.25

Model 9 Poverty rate: 26.88%	**Adj. R^2**	**Total Accuracy (%)**	**Poverty Accuracy (%)**	**Under-coverage (%)**	**Leakage (%)**	**PIE (% point)**	**BPAC (% point)**
Single-step methods -MAXR variable selection							
OLS	0.798	86.38	66.98	33.02	17.67	-4.13	51.63
Quantile regression (estimation point: 41)		85.13	72.56	27.44	27.91	**0.13**	**72.09**
Linear Probability	0.417	85.50	62.33	37.67	16.28	-5.75	40.93
Probit		85.63	65.58	34.42	19.07	-4.12	50.23
Two-step methods -MAXR variable selection							
OLS Percentile 87th	0.750 subsample	85.88	69.77	30.23	22.33	-2.13	61.86
Quantile regression (estimation points 41, 37) - 87 perc. cutoff		85.25	73.49	26.51	28.37	*0.50*	*71.63*
Linear Probability Percentile 59th	0.405 subsample	86.87	69.30	30.69	18.13	-3.37	56.74
Probit Percentile 59th		86.50	68.84	31.16	19.07	-3.25	56.74

Annex F: Variables included in the BEST15 models

Annex F.1: Variables included in the single-step OLS models (BEST 15 sets)

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Distance to union headquarter?	X								
Household ate less food for less than 30 days but more than 10 days during past 12 months	X								
Exterior walls: wood	X	X	X	X	X	X			X
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	X			X				X	
In the last seven days, how many days did a main meal consist of plain rice only?	X	X	X	X					
Household has electricity (autobattery, own generator included)	X			X	X	X	X		X
Food expenditure share, C, in percentage	X								
Number of steps above step identified as int poverty line, if minus below	X	X	X						
Average daily per-capita clothing expenditures	X	X	X				X	X	X
Hw much did household sent to relatives in last 12 months	X								
Value of motor tillers	X								
Annualized total household expenditures	X								
Total value of household assets	X	X							X
Car numbers	X	X	X	X	X	X			X
In past 24 months, community had access to subsidized food (vaso de leche)	X	X	X		X		X		
How many rooms does the dwelling have?		X	X		X		X	X	
Household feels that health care expenses are below need		X	X						
Sum of distances to department, provincial and district capital		X	X	X	X	X	X	X	
Median education of adult household members		X	X	X	X	X			

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household monthly expenditure on utilities (electricity, phone, water, etc)		X	X					X	
Value of debt owed by other households to household		X	X						
Value of metal pots		X	X	X	X				X
Household declares to not be able to save anything		X	X		X	X			
Value of tractors			X	X	X				
Do you have a telephone (fixed land line) in the house?				X	X	X	X	X	X
Microwave ownership				X			X	X	
Bed numbers				X					
Household feels that housing expenses are below need				X				X	
Household rates itself above subjective national poverty line on ladder				X				X	
How much does your household need per month to live				X				X	
Ratio male adults to female adults					X				
Total household members in Water/waste group					X	X			
Number of days in past seven days any of six superior food eaten (max. 42)					X	X			
Value of food processing assets					X				X
During last 3 years, have you or any of your household members received in-kind services from food aid programs						X			
Tractor ownership						X	X	X	
Colour TVs number						X			
Metal pots number						X			
Security key lock/metal frame with padlock in main entrance door						X			
Rooms per person						X			
Percentage of adult household members who read and write							X		X
Motocab ownership							X	X	
Color TV ownership							X		
Suit/ Jacket ownership							X		
No lock in main entrance door or wood or metal bar to close from inside							X	X	

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Head of household sleeps on something else than bed (e.g. floor, mat, mattress, hammock)							X		
Remittances sent/ total household expenditures							X	X	X
Household always ate enough from what they wanted (12month)								X	
Maximum education level of all members is incomplete primary									X
Education level of household head is high school/university/post-graduate									X
Poultry ownership									X
Roof with leaves, straw of bamboo or wood									X
Household head is single									X

Annex F.2: Variables included in the two-step OLS regressions (Models 1, 4, 7 and 9)

Variables Two-step OLS	Model 1		Model 4		Model 7		Model 9	
	1st step	2nd step	1st step	2nd step	1st step	2nd step	1st step	2nd step
Distance to union headquarter?	X							
Household ate less food for less than 30 days but more than 10 days during past 12 months	X	X						
Exterior walls: wood	X		X	X			X	
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	X		X	X				
In the last seven days, how many days did a main meal consist of plain rice only?	X	X	X	X				
Household has electricity (autobattery, own generator included)	X		X	X	X	X	X	X
Food expenditure share, C, in percentage	X							
Number of steps above step identified as int poverty line, if minus below	X	X						
Average daily per-capita clothing expenditures	X				X		X	X
How much household sent to relatives in last 12 months	X	X						
Value of motor tillers	X	X						
Annualized total household expenditures	X	X						
Total value of household assets	X	X					X	X
Car number	X		X	X			X	X
In past 24 months, community had access to subsidized food (vaso de leche)	X	X		X	X			
Sum of distances to department, provincial and district capital		X	X	X	X	X		
No lock in main entrance door or wood or metal bar to close from inside		X			X	X		
Median education of adult household members		X	X					
Household monthly expenditure on utilities (electricity, phone, water, etc)		X						
How much are households usual monthly expenditures for transport		X						
Rooms per person		X		X		X		X
Household head is single		X					X	X
Do you have Telephone (fixed land line) in the house?			X	X	X	X	X	X

Variables Two-step OLS	Model 1		Model 4		Model 7		Model 9	
	1st step	2nd step	1st step	2nd step	1st step	2nd step	1st step	2nd step
Microwave ownership			X	X	X			
Beds number			X					
Household feels that housing expenses are below need			X	X				
Household rates itself above subjective national poverty line on ladder,			X	X				
How much does your household need per month to live			X					
Value of tractors			X					
Value of metal pots			X	X			X	X
Maximum education level of any household member				X				
Colour TVs number				X				
Percentage of adult household members who read and write					X	X	X	X
How many rooms does the dwelling have?					X			
Motocab ownership					X			
Tractor ownership					X			
Color TV ownership					X			
Suit/ Jacket ownership					X			
Head of household sleeps on something else than bed (e.g. floor, mat, mattress)					X	X		
Remittances sent/ total household expenditures					X	X	X	X
Age of youngest household member						X		
Total number of children adopted in last 3 years						X		
Does the household own the house?						X		
Total household members in trade union						X		
Electric or gas cooker ownership						X		
Leather shoes ownership						X		
Sum of household clothing expenditures in past 12 months						X		
Maximum education level of all members is incomplete primary							X	
Education level of household head is high school/university/post-graduate							X	X
Poultry ownership							X	
Roof with leaves, straw of bamboo/wood							X	X
Value of food processing assets							X	X
Fan ownership								X
Cooking fuel is collected bamboo/wood/sawdust								X

Annex F.3: Poverty indicators used in the best model (in terms of maximization of BPAC)

Variable	Model 1	Model 4	Model 7	Model 9
	2-step Probit	2-step Quan.	2-step Quan.	1-step Quan.
Household expenditures[1]				
Per-capita daily average clothing expenditures	X		X	X
Annualized total household expenditures	X			
Monthly expenditures on utilities	X			
Monthly expenditures on transport	X			
Share of food expenditures from total household expenditures	X			
Sum of household clothing expenditures in past 12 months			X	
Education				
Household Head				
Education level of household head is high school/university/post-graduate				X
Household Members				
Median education of adult household members	X	X		
Maximum education level of any household member		X		
Proportion of adult household members who read and write			X	X
Maximum education level of all members is incomplete primary				X
Housing Characteristics				
Exterior walls material: wood	X	X		X
Rooms per person	X	X	X	
Household has electricity (auto battery, own generator included)	X	X	X	X
No lock in main entrance door or wood or metal bar to close from inside	X		X	
Do you have Telephone (fixed land line) in the house?		X	X	X
Number of rooms in the dwelling			X	
Roof material is leaves, jute stick or straw				X

Variable	Model 1	Model 4	Model 7	Model 9
Assets[1]				
Consumer durables				
Number of cars owned by the household	X	X		X
Value of motor tillers	X			
Total value of household assets	X			X
Microwave ownership		X	X	
Number of beds owned by the household		X		
Value of metal pots		X		X
Number of color TV's owned by the household		X		
Leather shoes ownership			X	
Motocab ownership			X	
Color TV ownership			X	
Suit / Jacket ownership			X	
Household head sleeps on something else than bed (e.g. floor, mat, mattress, hammock)			X	
Electric or gas cooker ownership			X	
Value of food processing assets				X
Agriculture				
Value of motor tillers	X			
Value of tractors		X		
Tractor ownership			X	
Poultry ownership				X
Fiancial				
Remittances sent to relatives in past 12 months	X			
Remittances sent/total household expenditures in past 12 months			X	X
Other				
Household head is single	X			X
Age of youngest household member			X	
Total household members belonging to trade union			X	
Total number of children adopted, last 3 years			X	
Does the household own the house?			X	
Subjective variables				
In the last seven days, how many days did a main meal consist of plain rice and any vegetables only?	X	X		
In the last seven days, how many days did a main meal consist of plain rice only?	X	X		
Household ate less food for less than 30 days but more than 10 days during past 12 months	X			
Number of steps above step identified as poverty line	X			
Amount that household needs per month to live		X		

[1] For analysis, all monetary variables were used in terms.

Variable	Model 1	Model 4	Model 7	Model 9
Household feels that housing expenses are below need		X		
Household rates itself above subjective national poverty line (on ladder)		X		
Community				
In the past 24 months, the community had access to subsidized food (vaso de leche)	X	X	X	
Distance to union headquarter?	X			
Sum of distances to department, provincial and district capital	X	X	X	

Endnotes

[i] This report consists of original work and data analysis. Citations of entire paragraphs or tables in published material by other authors is only permitted after prior consent with the authors and the IRIS Center. The cleaning and processing of data, as well as the entire analysis presented in this report, was carried out at the Institute of Rural Development, Georg-August-University of Göttingen, Germany. The costs of the survey of the additional 1175 microfinance clients and related analysis were financed by a grant from the Consultative Group to Assist the Poorest (CGAP), whose assistance is gratefully acknowledged. We gratefully acknowledge the valuable comments and support given by the IRIS project members Thierry van Bastelaer, Tresja Denysenko, Kate Druschel, and Anthony Leegwater; by Advisory Panel members Lauren Hendricks (CARE), Jonathan Murdoch (Princeton University), and Laura Foose (SEEP, PAWG), and by Stefan Schwarze and Norbert Binternagel of the Institute of Rural Development at the University of Göttingen. The input by the SEEP Network and its Poverty Assessment Working Group (PAWG), the Advisory Panel for the Developing Poverty Assessment Tools project, and USAID is gratefully acknowledged. In particular, Christian Grootaert provided valuable comments and advice during all phases of the field research and data analysis, especially also with regard to the choice of regression technique, in particular the alternative estimation method presented in Chapter 4. We gratefully acknowledge also the excellent cooperation with the Instituto Cuánto in Lima, Peru. All remaining errors are ours.

[ii] Contact Information: Instituto Cuánto, Baltazar La Torre 1115 - San Isidro. Lima/ Peru, Phone: (+51) (1) - 264-3505/ 264-1695/ 264-1699, E-mail: gproyectos@cuanto.org, Cuánto on Web: http://www.cuanto.org

[iii] Purchasing power parity exchange rates between US-Dollar and other currencies are available www.worldbank.org/povmonitor/ppp1993.htm.

[iv] For monthly inflation rates, we use those published by the Peru Bureau of Statistics. They are based on Lima only because due to the lack of data, there are no national CPI data available in Peru.

[v] The best sets of poverty indicators identified on each of the nine models refer to the combination of 5, 10 or 15 indicators selected by the SAS-MAXR procedure.

[vi] The terms regressor and poverty indicator are interchangeably used in this document. Literally speaking, they refer to a certain type of variable used in the regression. The regressors can be derived from one or many questions from the composite questionnaire. For example, some regressors or poverty indicators are directly computed from the variable obtained in the survey, such as the age of the household head. Other regressors require computation (using info from one or several questions) as they are not directly asked but are derived from the responses to the questions asked. An example is the size of the household (which is calculated from the information given in section B of the questionnaire).

[vii] For the case of zeroes as original monetary values, these were replaced by the value of one pro mille of the mean in order to be able to compute the natural logarithm.

[viii] Using the MAXR function of SAS, we selected in a prior model the best two regressors among 13 expenditure categories (referring to questions C1 to C12 as well as clothing expenditures of section B of the composite questionnaire). The inclusion of only the best two of the expenditure categories was done so as to avoid dominance of expenditure variables in subsequent models.

[ix] It is therefore important to consider the framework of incentives for when, where, and by whom a poverty assessment is carried out (incentives for the respondent as well as the interviewer). The following quote taken from an email by Jan Maes (Trickle Up Program) highlights some of the issues involved here: "One way of preventing clients from exaggerating their poverty or otherwise responding in a way they think 'would help their case,' is to conduct the poverty assessment survey after loan approval rather than to use it as part of the approval process. In other words, this implies that the USAID certified tools will be ex post poverty assessment tools rather than ex ante poverty targeting tools"… "If you use the assessment as part of the loan application or selection process, you will have to interview all potential clients, including of course those who 'fail the poverty test'. On the downside, since you only get your poverty results after clients have already entered the program, you might learn when it is already too late that you are not reaching the poorest."

[x] The project directors of Instituto Cuánto were asked to rate the verifiability of each of the indicators on a scale from 1 to 5 where 1 is very difficult or impossible to verify, and 5 stands for easy verifiability. In Annex D, we list the rating given by the survey firm Cuánto in Peru. In addition, Cuánto rated the corresponding questions contained in the questionnaire according to their difficulty to ask. In model 7, we include only the regressors that have been rated as easily verifiable (i.e. a score of 4 or 5), and easy to ask (i.e. a score of 4 or 5, where a rating is available).

[xi] These variables were identified by the SAS-MAXR procedure as the strongest variables among all subjective variables which were excluded in Model 5.

[xii] These variables were identified by the SAS-MAXR procedure as the strongest variables among all subjective variables which were excluded in Model 5.