microREPORT

Developing and testing poverty assessment tools: Results from accuracy tests in Peru

Manfred Zeller
Julia Johannsen
Gabriela Alcaraz V.

Final report

March, 2005

Accelerated Microenterprise Advancement Project

Accelerated Microenterprise Advancement Project (AMAP) is a 4-year contracting facility that USAID/Washington and Missions can use to acquire technical services to design, implement, or evaluate microenterprise development, which is an important tool for economic growth and poverty alleviation.

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

Contract Number: GEG-I-02-02-00029-00
Task Order No. 2: Developing Poverty Assessment Tools
Contractor: **The IRIS Center at the University of Maryland**

Manfred Zeller, Ph.D., is full professor for Socioeconomics of Rural Development at the Georg-August-University of Göttingen, Germany. http://www.gwdg.de/~uare

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

The IRIS Center is a research and technical assistance provider housed at the University of Maryland, College Park.

NOTE: Additional funding for this report provided by the Consultative Group to Assist the Poor (CGAP).

Contents

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[1]. 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 best 5, best 10, and best 15 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 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 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 5 summarizes the results.

1.1 Field survey for accuracy tests in Peru

The survey firm Instituto Cuánto[2] 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 the composite questionnaire survey with 800 households, followed two weeks later by the 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 at www.povertytools.org. The composite as well as the benchmark questionnaire required adaptation to the country-specific context. In the case of the composite questionnaire, this entailed the inclusion of 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 professional interviewers and researchers in Peru that were intensively involved in the questionnaire adaptation. The adaptation of the benchmark questionnaire mainly involved the selection of major food items. For this, we referred to results from the most recent National Living Standard Measurement Survey (Encuesta Nacional sobre Medición de Niveles de Vida (ENNIV)) from the year 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, its director general, Moises Ventocilla A., the project director Luis Castillo Q. and the senior researchers Pedro Llontop L., and Mario Reyna Farje E., as well as their supervisors and interviewers, in conducting economic, social and market studies during the past 17 years in the diverse cultural and socioeconomic settings in Peru.

1.2 Sampling Frame

Requirements for sampling. For the general population sample, it was determined to choose a sample size of 800 households. The sample was required 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).

They are the common basis for the official calculation of poverty lines in Peru and were controlled for in our multi-stage cluster sampling approach used in drawing up a random sample of households. This sampling procedure allows us to draw successive samples at lower administrative units. The highest administrative unit in Peru are so-called departments. The twenty-four departments are further disaggregated into provinces and districts. Each district holds a number of cities and villages.

The first stage of sampling was at the department level based on centrally available and published population data. In view of logistical and budget constraints, it was determined to randomly select eight departments out of the total of 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 8 chosen departments, a sample size of 100 was allocated. This so-called probability-proportionate-to-size sampling (PPS) was repeated 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 Lima Metropolitan 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, differentiated by the seven official “sampling areas” (see Table 1).

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

Macro-regions, Sampling area	No. of sample households	% of national population	% 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, out of the total of districts in each of the eight selected departments, three districts (and six in the Lima department) were randomly chosen proportionate to size of the districts compared to total population size in the department. 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 constitute 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 case of an urban assignment of the district (excluding extended industrial and natural areas within the urban segment). In contrast, in rural districts, we chose one community in each district proportionate to size. An exception are districts that do not host large enough communities to randomly select from for the required number of sample households. In these cases, more than one community was selected proportionate to size compared to the total population size in the district.

Finally, in each of the randomly selected communities, the random walk method (see Henry et al. 2003) was applied to select a random sample of 33 or 34 survey households per district. Thus, the total sample size is 800, and the sample is a self-weighing, nationally representative sample which sought to come up with the best possible set of districts considering

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) living 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, the national poverty line is disaggregated into 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, it is necessary to convert 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 in annual expenditures per person in May 2000, according to Peru's most recent National Living Standard Measurement Survey (ENNIV) from the year 2000.

As illustrated in Table 2 below, 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 nat. poverty line** (Soles/ pers./ year)*	**Daily nat. poverty line** (Soles/ pers./ day)	**Poverty headcount** (percent)**
Lima Metrop.	2810.7	7.7	45.2%
UrbanCoast	2335.3	6.4	53.1%
Rural Coast	1574.8	4.3	64.4%
Urban Highland	2001.4	5.5	44.3%
Rural Highland	1315.2	3.6	65.5%
Urban Lowland	1934.7	5.3	51.5%
Rural Lowland	1327.6	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, it is necessary to convert 1 US-dollar into Soles using purchasing power parity rates. In March 2003, 1 US dollar was equivalent in purchasing power to 1.86 Soles[3]. To compare the national with the international poverty line in 2000, it is necessary to adjust the 1.08-dollar 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 requires 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 for the period May 2000 to March 2003, we calculate a total inflation of 5.38 percent over the 35 month period.[4] We therefore adjust the value of 1.86 Soles by 5.38 percent, a multiplication with the factor 100/(100 + 5.38). This amounts to 1.77 Soles, equivalent to the purchasing power of 1 US-dollar for May 2000. Finally, multiplying this value by the factor 1.08 (international poverty line of 1 US-dollar is equivalent to 1.08 US-dollar at 1993 prices) yields an amount of 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 dollar 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 area 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

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% < nat. pov. line (Soles/ pers./ year)*	Daily income equivalent To 50% < nat. pov. line (Soles/ pers./ day)	Poverty headcount (percent)**
Lima Metrop.	1999.9	5.5	22.6%
Urban Coast	1564.2	4.3	26.6%
Rural Coast	1015.6	2.8	32.2%
Urban Highland	1351.8	3.7	22.2%
Rural Highland	795.0	2.2	32.8%
Urban Lowland	1281.2	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/ pers./ year)	Daily int. poverty line (Soles/ pers./ 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. This requires adjusting 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 2 months after

May 2004 as an estimate of the CPI change for the period May to July 2004, we calculate a total inflation of 9.3 percent over the total 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, it was necessary again to adjust the 1 dollar 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 requires adjusting the value of 1.86 Soles equivalent to 1 US-dollar in March 2003 by the increase in the national consumer price index (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 2 months after May 2004 as an estimate of the CPI change for the period May to July 2004, we calculate a total inflation of 3.7 percent over the 17 month period. We therefore increase the value of 1.86 Soles by 3.7% (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% < nat. pov. line** (Soles/ pers./ year)	**daily income equivalent to 50% < nat. pov. line** (Soles/ pers./ day)	**Annual int. poverty line (1 US-$ PPP)** (Soles/ pers./ year)	**Daily int. poverty line (1 US-$ PPP)** (Soles/ pers./ day)
Lima Metrop.	2182.0	5.98	759.2	2.08
Urban Coast	1706.6	4.68	759.2	2.08
Rural Coast	1108.0	3.04	759.2	2.08
Urban Highland	1474.9	4.04	759.2	2.08
Rural Highland	867.4	2.38	759.2	2.08
Urban Lowland	1397.9	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 are 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 (i.e. yielding a headcount index of very poor of 27.05 percent) that is derived from the most recent National Living Standard Measurement Survey of Peru in the year 2000 (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 having an expenditure falling below the bottom 50 percent cut-off of the respective national poverty line, and the term "not very poor" or "NVP" for those having an expenditure equal or above the bottom 50 percent cut-off of the respective national poverty line. Readers should bear in mind that ANY 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 being poor by the national poverty line, and that even among those above the national poverty line there exist a considerable share of households that are vulnerable to poverty such that, for example, a bad harvest, an illness of a family member, or a social obligation may drive them into poverty.

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 that are 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 as well as household-level indicators (derived from all practitioner tools)
C. Household expenditures by category (adapted from FINCA and ACCION tool)
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 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 on 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 also caused by the relatively high level of crime. It was therefore decided to ask section K 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 using the software SAS. The models differ by the type of regressors used. While Model 1 includes 250 regressors, the seventh model has the most restrictive list of 104 potential poverty indicators.

As one can see from the results for Model 1 in Chapter 3, the set of best[5] poverty indicators is dominated by different expenditure and asset categories, apart from household demographic characteristics. 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 is 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, and several other expenditure indicators.

Verifiability of indicator. Another useful characteristic of an indicator for its operational use is its 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 of natural fibers; these are 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 hardly verifiable indicator is the number of luxury food eaten in the past 7 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.

Having recognized in the above that the difficulty and verifiability of an indicator cannot be generalized across different socio-economic and cultural contexts, we acknowledge that it might appear rather 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 our selection of progressively smaller subsets of regressors for defining Model 1 thru Model 6 might not be agreeable to all readers. This approach mainly aims 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 Model 7 and 8, we use the subjective assessment of verifiability of the survey firm Instituto Cuánto as an alternative attempt to address the practicality issue. To get more information on the practicality of poverty indicators, the IRIS project includes practicality tests carried out by microfinance (MF) and business development services (BDS) organizations.

Our sequence of regression models with progressively fewer poverty indicators (from Model 1 to Model 6) aims to generate different poverty assessment tools that gradually become less accurate but hopefully also more practicable, 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 overall accuracy, accuracy among the very poor and the not very poor, 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 best 15 (or even more) poverty indicators[6]. 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 the tool usually also tend 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 poverty assessment tools.

2.4 Specification of regression models

The following nine model types 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 national 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 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[7] since the dependent variable is also expressed in natural logarithm. 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 do 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[8]. 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 relatively 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 interviewers of Instituto Cuánto (the survey firm in Peru), 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.

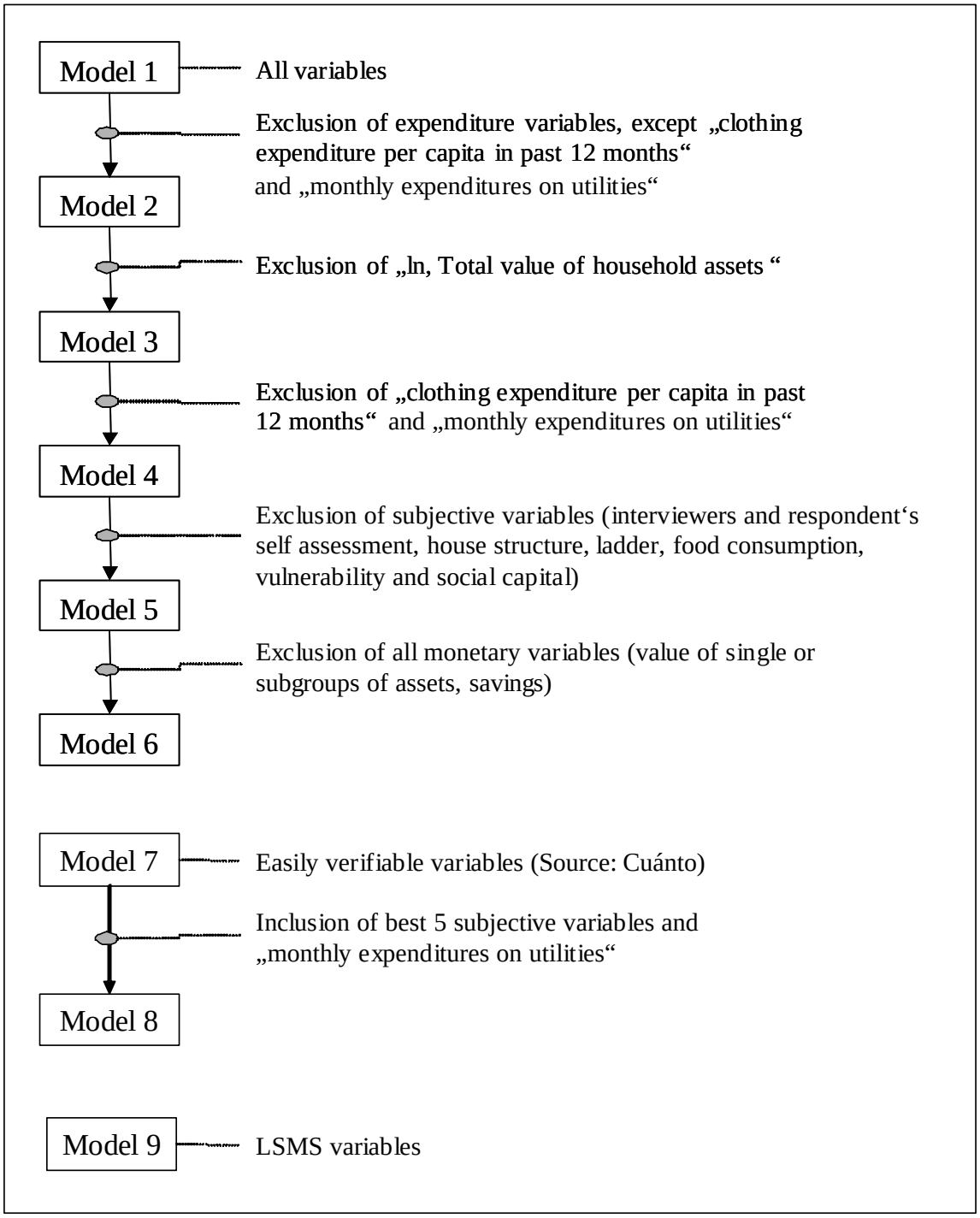
Model 4: The set of regressors for this model is similar to Model 3. The only difference is the exclusion of the variables *monthly household expenditures on utilities* and *clothing expenditures per capita in past 12 months* (both in form of the natural logarithm of the original expenditures in Soles). 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[9]. The subjective poverty indicators that were excluded in Model 5 (compared to Model 4) are presented accordingly in the 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[10]. 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 the monthly expenditures on utilities (as best single expenditure indicator, in addition to average clothing expenditures contained in the Model 7 variables), plus five powerful subjective variables[11]:

- Days in past 7 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 or poverty assessment tool may allow indirect verifiability of the clothing and utility expenditures and the subjective indicators through comparing them 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.

Results from Regression Models

In the following, the results are summarized by listing

- the regressors that were among the best5, best10, and best15 models
- the adjusted R-square achieved (e.g., an R-square of 0.6 indicates that 60 percent of the observed variance in the dependent variable is explained by the regressors).

For purposes of assessing the prediction power of a regression model (or tool) for poverty assessment, we also present the following five measures of performance for each model:

- the **overall accuracy** (Accuracy). This is the percentage of the total sample of 799 households that are correctly predicted in their poverty status by the regression model
- the **accuracy among the very poor** (Acc. among VP), which refers to the households correctly predicted as very poor, expressed as percentage of the total very poor
- the **accuracy among the not very poor** (Acc. among NVP), which refers to the households correctly predicted as not very poor, expressed as percentage of the total not very poor
- the **undercoverage** (Undercoverage). This measure represents the error of predicting very poor households as being not very poor, expressed as percentage of the total very poor
- the **leakage** (Leakage), which reflects the error of predicting not very poor households as very poor, expressed as percentage of the total very poor.

The latter two measures, leakage and undercoverage, are often used in the literature on assessing the poverty targeting performance of development and safety net policies, institutions or projects.

We note that the set of BEST regressors is statistically determined by the MAXR technique of SAS which searches for the best model fit. The term BEST regressors should not be misunderstood as a value statement that implies as being best for the overall accuracy of a regression model or for any of the other four measures of performance listed above. The set of BEST 5, BEST 10, or BEST 15 regressors simply 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 BEST 5 model).

The above mentioned measures of model performance are exemplified with the results of Model 1 which are presented next.

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 = 26.9 %
- 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:

- Accuracy = (539 + 137) / 800) * 100 = 84.5 %
- Accuracy among the very poor = (137 / 215) * 100 = 63.72 %
- Accuracy among the not very poor = (539 / 585) * 100 = 92.14 %
- Undercoverage = (78 / 215) * 100 = 36.28 %
- Leakage = (46 / 215) * 100 = 21.40 %

From Table 3.1.2, it can be observed that the highest performance in terms of accuracy of Model 1 is actually achieved in the BEST15 set (87.38%). 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 (%)
Best 5 indicators: R^2 adjusted = 0.793	
• Days in past 7 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	Accuracy: 84.50 Acc. among VP: 63.72 Acc. among NVP: 92.14 Undercoverage: 36.28 Leakage: 21.40
Best 10 indicators: R^2 adjusted = 0.820	
Next best five indicators: • Days in past 7 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 7 days with main meal consisting of plain rice and any vegetables	Accuracy: 87.00 Acc. among VP: 70.23 Acc. among NVP: 93.16 Undercoverage: 29.77 Leakage: 18.60
Best 15 indicators: R^2 adjusted =0.834	
Next best five indicators: • Days in past 7 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	Accuracy: 87.38 Acc. among VP: 70.70 Acc. among NVP: 93.50 Undercoverage: 29.30 Leakage: 17.67

Interestingly though, Model 1 does not show the best accuracy results of all the tools presented in this report. While the BEST15 set in Model 1 achieves the highest adjusted R^2 value among all tools (0.834), Model 3 achieves a higher overall accuracy (88.13%) and Model 4 a higher accuracy among the not very poor (94.70%) and the lowest leakage figure (14.42%). With respect to the highest accuracy among the very poor (70.70%) and the lowest undercoverage (29.30%), Models 2 and 3 achieve the same results as Model 1.

This result is surprising only at first sight, as 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. If we remember, however, that powerful in terms of the SAS MAXR routine used refers to the highest R^2 and not necessarily to the highest accuracy with respect to the expenditure cut-off representing the bottom 50 percent below the national poverty line, the results are consistent. Model 1 has the highest R^2 of all models.

It is worth to note that already in model 1, the first 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 overall accuracy in the BEST5 and BEST10 sets and a slightly higher overall accuracy in the BEST15 set. The accuracy among the very poor 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 accuracy among the very poor (70.70%) and the undercoverage figure (29.3%) that stay the same as in model 1.

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 (%)
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	Accuracy: 83.75 Acc. among VP: 64.19 Acc. among NVP: 90.94 Undercoverage: 35.81 Leakage: 24.65
Best 10 indicators: R^2 adjusted = 0.796	
Next best five indicators: • Number of rooms in the dwelling • Days in past 7 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	Accuracy: 85.25 Acc. among VP: 65.12 Acc. among NVP: 92.65 Undercoverage: 34.88 Leakage: 20.00
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	Accuracy: 87.50 Acc. among VP: 70.70 Acc. among NVP: 93.68 Undercoverage: 29.30 Leakage: 17.21

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 overall accuracy of 88.13% is the highest one among all tools presented in this report. While the accuracy among the very poor and the related undercoverage figures stay the same as in model 1 and 2, the results for leakage and accuracy among the not very poor are better.

The results from Model 3 (as compared to the inferior results from the more complex models 1 and 2) 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 the most accurate prediction of the poverty status (very poor or not) of households.

Table 3.3.1 Summary of the accuracy results Model 3

Variables	Model performance (%)
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	Accuracy: 83.83 Acc. among VP: 59.07 Acc. among NVP: 92.31 Undercoverage: 40.93 Leakage: 20.93
Best 10 indicators: R^2 adjusted = 0.789	
Next best five indicators: • Number of rooms in the dwelling • Days in past 7 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	Accuracy: 85.75 Acc. among VP: 65.58 Acc. among NVP: 93.16 Undercoverage: 34.42 Leakage: 18.60
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	Accuracy: 88.13 Acc. among VP: 70.70 Acc. among NVP: 94.53 Undercoverage: 29.30 Leakage: 14.88

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 that 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 overall accuracy significantly increases from 84.5% in the BEST5 set to 87% 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 accuracy among the not very poor and a lower leakage figure when compared to both, BEST10 in the same model 4 and BEST15 in model 3.

No other model in this report presents better accuracy results for the BEST5 and BEST10 tool, which again confirms that reliable poverty assessment in Peru does not have to include expenditure data.

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 (%)
Best 5 indicators: R^2 adjusted = 0.741	
• Median education level of adult household members • Days in past 7 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	Accuracy: 84.50 Acc. among VP: 60.93 Acc. among NVP: 93.16 Undercoverage: 39.07 Leakage: 18.60
Best 10 indicators: R^2 adjusted = 0.781	
Next best five indicators: • Number of rooms in the dwelling • Days in past 7 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 7 days with main meal consisting of plain rice and any vegetables	Accuracy: 87.00 Acc. among VP: 67.44 Acc. among NVP: 94.19 Undercoverage: 32.56 Leakage: 15.81
Best 15 indicators: R^2 adjusted = 0.800	
Next best five indicators: • Availability of telephone (fixed land line) in the house • Days in past 7 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	Accuracy: 87.00 Acc. among VP: 66.05 Acc. among NVP: 94.70 Undercoverage: 33.95 Leakage: 14.42

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 and a clear drop in accuracy levels. The best performance was achieved by the BEST15 set.

Table 3.5.1 Summary of the accuracy results Model 5

Variables	Model performance (%)
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	Accuracy: 82.50 Acc. among VP: 57.21 Acc. among NVP: 91.79 Undercoverage: 42.79 Leakage: 22.33
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	Accuracy: 84.88 Acc. among VP: 63.26 Acc. among NVP: 92.82 Undercoverage: 36.74 Leakage: 19.53
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 • Community access to subsidized food ("glass of milk – vaso de leche") in past 24 months	Accuracy: 85.13 Acc. among VP: 64.19 Acc. among NVP: 92.82 Undercoverage: 35.81 Leakage: 19.53

The exclusion of subjective variables caused 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 overall accuracy level and the accuracy among the not very poor for the three sets of model 5 decreased by 2% compared to model 4, while the accuracy among the very poor decreased even further. Also with respect to the leakage figures, model 5 shows a clear drop in performance when compared to model 1 to 4.

On the other hand, in terms of the difficulty for obtaining information and the verifiability of the 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, i.e., lower in all indicators sets than all of the previous models.

As in the previous model, this model incorporates a high proportion of nearly two thirds of asset and housing-related variables. The main difference to model 5 constitutes 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.

Table 3.6.1 Summary of the accuracy results Model 6

Variables	Model performance (%)
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	Accuracy: 81.50 Acc. among VP: 55.35 Acc. among NVP: 91.11 Undercoverage: 44.65 Leakage: 24.19
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	Accuracy: 85.13 Acc. among VP: 62.79 Acc. among NVP: 93.33 Undercoverage: 37.21 Leakage: 18.14
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	Accuracy: 85.13 Acc. among VP: 64.65 Acc. among NVP: 92.65 Undercoverage: 35.35 Leakage: 20.00

3.7 Model 7

This model incorporates those 104 poverty indicators that have been rated as highly verifiable (score 4 or 5) by the staff and interviewers of 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. 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, having an overall accuracy level of 84.63% and an accuracy level among the very poor of 63.72%. Considering that the 15 indicators are fairly easy to obtain and to verify as deemed by Cuánto, the overall accuracy levels (of only less than 1% lower than in Model 6) are still high. 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 (%)
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	Accuracy: 82.38 Acc. among VP: 56.28 Acc. among NVP: 91.97 Undercoverage: 43.72 Leakage: 21.86
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	Accuracy: 83.63 Acc. among VP: 61.86 Acc. among NVP: 91.62 Undercoverage: 38.14 Leakage: 22.79
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	Accuracy: 84.63 Acc. among VP: 63.72 Acc. among NVP: 92.31 Undercoverage: 36.28 Leakage: 20.93

By inspection of the above poverty indicators selected by the regression model, we observe that many of them are highly verifiable, simply by observation when visiting the home of the client household. Probably the most difficult questions to ask refer to the value of remittances sent as well as the clothing expenditures. However, the staff of the survey firm rated these questions as relatively easy to ask (score 4 out of 5) although their score of 4 for verifiability of these two variables may be doubted by some readers.

3.8 Model 8

This model is similar to Model 7, but includes the monthly expenditures on utilities (as best single expenditure indicator, in addition to average clothing expenditures contained in the Model 7 variables), plus five powerful subjective variables[12]:

- Days in past 7 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?

The incorporation of these variables only slightly increases the model's performance to a level that is somewhat higher than Model 5. It can be 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 if one wishes to maximize accuracy performance. However, in practice, actual accuracy may be lower with subjective variables as they are in tendency more difficult to ask and especially difficult to verify (see Cuánto's rating of verifiability and difficulty of asking) in Annex D.

The adjusted R-squared values range between 0.725 and 0.792. The best performance is achieved by the BEST15 set (85.38% overall accuracy). In comparison with Model 7, the overall accuracy increased only to about 1%. A greater improvement can be observed in the accuracy among the very poor in the Best 10 and Best 15 option of Model 8.

Table 3.8.1 Summary of the accuracy results Model 8

Variables	Model performance (%)
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 7 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	Accuracy: 83.38 Acc. among VP: 57.67 Acc. among NVP: 92.82 Undercoverage: 42.33 Leakage: 19.53
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 7 days with main meal consisting of plain rice and any vegetables	Accuracy: 84.88 Acc. among VP: 64.19 Acc. among NVP: 92.48 Undercoverage: 35.81 Leakage: 20.47
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 7 days with main meal consisting of plain rice and any vegetables • Amount that household needs per month to live	Accuracy: 85.38 Acc. among VP: 65.12 Acc. among NVP: 92.82 Undercoverage: 34.88 Leakage: 19.53

3.9 Model 9

Model 9 uses a set of 127 regressors which are usually found in LSMS surveys of the World Bank. Model 9 shows higher adjusted R^2 levels than model 5, 6, 7 and 8. We observe further an increase in the performance indicators compared to model 7, altbeit the LSMS indicators in model 9 are more or less similar to the indicators (demography, asset ownership and housing) that are deemed easy to verify in model 7. However, model 9 also contains the complicated indicator of the overall value of assets. Thus, a set of indicators like the one

contained in model 8 that achieves similar accuracy results than the LSMS model might be preferred by practitioners.

The best performance is observed in the BEST 15 set, with 85.50% overall accuracy and 66.98% accuracy among the very poor.

Table 3.9.1 Summary of the accuracy results Model 9

Variables	Model performance (%)
Best 5 indicators: R^2 adjusted = 0.758	
• Median education level of adult household members • Total value of household assets • Household has electricity • Remittances sent • Average daily per-capita clothing expenditures	Accuracy: 83.38 Acc. among VP: 61.40 Acc. among NVP: 91.45 Undercoverage: 38.60 Leakage: 23.26
Best 10 indicators: R^2 adjusted = 0.786	
Next best five indicators: • Availability of telephone (fixed land line) in the house • Number of cars owned by the household • Leaves or straw as roof material • Wood as exterior-walls' material • Value of metal pots owned by the household	Accuracy: 85.13 Acc. among VP: 64.19 Acc. among NVP: 92.82 Undercoverage: 35.81 Leakage: 19.53
Best 15 indicators: R^2 adjusted = 0.799	
Next best five indicators: • Number of rooms in the dwelling • Household owns a microwave • Collected wood/ sawdust/ bamboo as cooking fuel • Household head is single • Household owns sheep/ goats • Number of horses owned by the household Removed indicators: • Leaves or straw as roof material	Accuracy: 85.50 Acc. among VP: 66.98 Acc. among NVP: 92.31 Undercoverage: 33.02 Leakage: 20.93

Two step models

4.1 Introduction

The accuracy measures presented through Models 1 to 9 (Chapter 3) refer to the average accuracy of the models to predict the poverty status using 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”). In order to improve the estimation of poverty status of the models, a two-step approach (see Grootaert et al. (1998)) was performed by considering the differences in the accuracy measures by deciles of the benchmark indicator. The original model is evaluated by the level of accuracy obtained in the different deciles (first step) for the full sample. Afterwards, the model is estimated with a subsample, now only including the deciles with low accuracy levels in order to identify the best regressor set for that subsample. The estimation in the second step is again performed with OLS, using the MAXR routine of SAS. Finally, the combined accuracy level of the two models is calculated by considering the predicted status from the first step for the high-accuracy deciles and the predicted status from the second step for the selected deciles included in the subsample for the second step.

In the following, we present results of the two step approach for the best 15 regressor set of Model 1. Similar two-step models could potentially be estimated for the sets of regressors of Model 1 thru 9.

4.1.1 First step: Model 1 - Best 15 set on full sample

Next, we evaluate the performance of model 1 with the best 15 regressors (see Table 3.1.2) with respect to the level of accuracy achieved for each of the ten deciles of the observed per-capita daily expenditures. Table 4.1.1.1 presents the results obtained. The row “Total average” present the accuracy results already shown in Table 3.1.2 for the best 15 regressors set.

From the table, it can be observed that Model 1 with the best 15 regressors achieved the highest accuracy levels in decile 8, 9 and 10 by correctly predicting the poverty status of all NVP households. As well, it is observed that accuracy drops 2.5% from decile 8th to decile 7th and it decreased further by 10% to decile 6.

In comparison, the accuracy level achieved in the first deciles is 15% to 30% lower than the accuracy level observed in the higher deciles, a situation which justifies the use of subsamples composed by the lower deciles in order to find a best 15 regressor set that best predict the poverty status of the households located in them.

Table 4.1.1.1 Accuracy level by deciles of daily expenditures per capita observed for the best 15 regressor set.

Decile of benchmark "daily expenditures per capita"	Accuracy (%)	Accuracy among VP (%)	Accuracy among NVP (%)	Under-coverage (%)	Leakage (%)
1	85.00	85.00	-	15.00	17.67
2	77.50	76.92	78.57	23.08	17.67
3	78.75	72.50	85.00	27.50	17.67
4	70.00	20.88	89.09	72.00	17.67
5	77.50	43.75	85.94	56.25	17.67
6	87.50	50.00	88.46	50.00	17.67
7	97.50	-	97.50	-	-
8	100.00	-	100.00	-	-
9	100.00	-	100.00	-	-
10	100.00	-	100.00	-	-
Total average	**87.38**	**70.70**	**93.50**	**29.30**	**17.67**
N	800	215	585	215	215
Percent of population being very poor				26.87	
Percent of population being predicted as very poor				23.75	

Four different versions (subsamples) were created for the second step. These were:

A. Model 1 tested on deciles 1 to 7
B. Model 1 tested on deciles 1 to 6
C. Model 1 tested on deciles 1 to 5
D. Model 1 tested on deciles 1 to 4

4.1.2 Second step: Model 1 - Best 15 set on subsamples

After testing the set of variables corresponding to Model 1 on the different subsamples, the four new best 15 regressor sets were identified and were evaluated in terms of the five accuracy measures. Table 4.1.2.1 presents a summary of the results obtained for each subsample.

Table 4.1.2.1 Summary of accuracy results for the Best 15 regressor set on each subsample based on Model 1.

Measure	Version			
	A Deciles 1-7	**B Deciles 1- 6**	**C Deciles 1-5**	**D Deciles 1-4**
Number of observations	560	480	400	320
R^2 adjusted	0.741	0.720	0.679	0.655
Accuracy (%)	84.10	81.45	83.50	82.81
Accuracy among VP (%)	77.67	79.53	85.91	86.29
Accuracy among NVP (%)	88.11	83.01	80.74	77.23
Undercoverage (%)	22.32	20.46	14.08	13.70
Leakage (%)	19.06	20.93	16.90	14.21
Population being VP (%)	38.39	44.79	53.25	61.56
Population predicted as VP (%)	37.14	45.00	54.75	61.87

From Table 4.1.2.1 it can be observed for the four subsamples, that while the adjusted R2 and the overall accuracy were lower than in the best 15 set from the first step (0.834 and 87.38% respectively), the accuracy among the very poor increased from 70.70% (full sample) in the first step to a maximum of 86.29 % in version (subsample) D. On the contrary, the accuracy among the non very poor decreased to less than 90%.

An interesting result is the percentage of population being predicted as very poor when compared to the percentage of population being very poor. The best 15 set resulting from the subsample A (deciles 1 to 7) underestimated the actual proportion of households being very poor on those deciles, while the best 15 sets from the subsamples B, C and D overestimated the proportion of very poor households. At the decile level, version C provided the most accurate prediction of the household's poverty status (Accuracy among VP), presenting the highest improvement in decile 4, from 20.88 % in the first step to 64% in the combined accuracy measurement. Detailed combined accuracy measures at the decile level are shown in Annex G.

Table 4.1.2.2 presents the Best 15 regressor set obtained for each of the subsamples. The best 15 set from the first step (corresponding to Best 15 in Table 3.1.2) is presented for comparison. Only three variables (shaded in gray) appear to be important on all versions, these were:

- Days in past 7 days with main meal consisting on plain rice and any vegetables
- Annualized total household expenditures
- Value of motor tillers owned by the household

The variables "Days in past 7 days with main meal consisting on plain rice only" and "Household has electricity" appeared to be important in the first step as well as in the subsamples A, B and C during the second step, but were replaced by other variables in subsample D.

The variables "Household's expenditures on transport, per month" and "Sum of distances to department, provincial and district capital" were selected within the best 15 set in the four

subsamples, but were not included within the best 15 set in the first step.

Table 4.1.2.2 Best 15 regressor sets derived from the second step

Variables	First step	Second step			
	Model 1	A Deciles 1-7	B Deciles 1- 6	C Deciles 1-5	D Deciles 1-4
Days in past 7 days with main meal consisting on plain rice and any vegetables	X	X	X	X	X
Days in past 7 days with main meal consisting on plain rice only	X	X	X	X	
Numbers of cars owned by the household	X				
Share of food expenditures from total household expenditures	X				
Number of steps above step identified as international poverty line	X			X	X
Wood as exterior-walls' material	X	X			
Household ate less food from what they wanted for more than 10 days, but less than 30 days, during past 12 months	X				
Annualized total household expenditures	X	X	X	X	X
Value of remittances sent to relatives in last 12 months	X		X	X	
Total value of household assets	X				
Value of motor tillers owned by the household	X	X	X	X	X
Distance to department capital	X				
Community access to subsidized food ("glass of milk-vaso de leche") in past 24 months	X				
Household has electricity	X	X	X	X	
Average daily per-capita clothing expenditures	X				
Total number of children adopted in last 3 years		X			
Motocab ownership		X			
Number of pigs owned by the household		X	X		
Number of beds owned by the household		X			
Household feels that housing expenditures are below need		X	X	X	
Household rates itself above national poverty line on ladder		X	X		
Number of rooms per person		X			
Household's expenditures on transport, per month		X	X	X	X

Variables	First step	Second step			
	Model 1	A Deciles 1-7	B Deciles 1- 6	C Deciles 1-5	D Deciles 1-4
Sum of distances to department, provincial and district capital		X	X	X	X
Fan ownership			X	X	X
Number of refrigerators owned by the household			X		
No lock in main entrance door or wood/metal bar to close door			X	X	X
Education level of household members, excluding household head			X		
Percentage of dependents younger than 15 and older than 64 years old, in relation to household size				X	
Number of household members belonging to women's group				X	X
Number of days in past 7 days with any of six superior food eaten				X	
Age of youngest household member					X
Median education of household members					X
Number of days in past 7 days with other red meat served in main meal					X
Cattle ownership					X
Household expenditures on health in past 12 months					X
Religion of household head is other than catholic					X
Number of additional regressors in second step		9	9	8	11

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. In model version D, 11 new regressors are being selected, while in model version C only 9 regressors are listed.

4.1.3 Combined accuracy of the two step models

The combined accuracy level from the two step model is presented in Table 4.1.3.1. In general, the two step models yielded a higher overall accuracy (above 88.62 %) than the original Model 1 (first step – 87.38%). The highest accuracy level was achieved by version C (90.25 %), for which the first step provided the predicted poverty status for deciles 6 to 10 and the second step for deciles 1 to 5. As well, version C registered the highest accuracy among the very poor (85.58 %) bring this level clearly higher than the level achieved by the original Model 1, Best 15 (70.70 %).

The accuracy among the not very poor decreased by about 1.5 % in comparison with the first step. For all versions, undercoverage decreased by more than 7%. The lowest level of 14.41% is achieved by version C. As expected, leakage increased in all model versions up to a maximum of 22.32 in version D (i.e. this is 4.65% higher than in the first step).

Table 4.1.3.1. Combined accuracy (N = 800) Model 1, Best 15 set, by model version

Measure	A Deciles 1-7	B Deciles 1- 6	C Deciles 1-5	D Deciles 1-4
Accuracy (%)	88.87	88.62	90.25	89.37
Accuracy among VP (%)	77.67	79.53	85.58	82.79
Accuracy among NVP (%)	92.99	91.96	91.96	91.79
Undercoverage (%)	22.32	20.46	14.41	17.20
Leakage (%)	19.06	21.86	21.86	22.32
Population predicted as VP (%)	26.00	27.25	28.87	28.25

Finally, the population predicted as very poor increased as a result of the two step approach. The proportions of VP obtained on all versions (27.5% in average) were closer to the actual proportion of population being VP (26.87 %) than the initial proportion of VP as predicted in the first step of model 1 (23.75 %).

These results suggest that while the original Best 15 Model 1 (first step) tends to underestimate the proportion of very poor households, the two step approach tends to slightly overestimate the very poor. This overestimation of the headcount index of the very poor is explained by the increase in leakage, i.e. the increase in misclassifying the not very poor as being very poor.

A practitioner tool based on a two-step model would have to include questions to 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 best 15 regressors of the second step. For example, in model version C, there were nine additional regressors (see Table 4.1.2.2).

The use of a two-step model as a practitioner tool is further exemplified, using model version C as example. In practice, all questions related to the first and second step (15 plus 9 indicators) can be integrated into one single interview with the client. The interviewer would begin with the best 15 indicators of the first step, and then compute an estimated per-capita daily expenditure. If the predicted per-capita daily expenditure falls above the cut-off value for the 6^{th} 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 the 6^{th} percentile, the interview needs to be continued by asking questions related to the nine additional regressors of the second step. Based on the values obtained for the 9 regressors (plus the remaining six 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 client is rated as being very poor, otherwise not. In practice, however, it is recommended to not interrupt the interview for the calculation based on the first fifteen indicators, but to continue with the questions for the remaining nine poverty

indicators. In this case, the calculations of one (or two) expenditure values are done after the interview.

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. In section I1 of the composite questionnaire of the IRIS accuracy tests, a question was asked for each adult household member (i.e. 18 years or older) about his or her client relationship with banks and microfinance institutions. For any current client, we also ask whether he or she received business development services from their provider of financial services. In section I2, we ask for each adult household member separately whether they received business development services during the past 5 years.

In section 5.2, we analyze the poverty outreach performance of institutions providing financial or business development services on the basis of the nationally representative sample of 800 households. In Peru, an additional sample of 1175 households was randomly selected among clients from six purposefully selected microfinance institutions (MFIs). In chapter 5.3, we report results on the depth of outreach of these six MFIs.

5.2 Outreach in the nationally representative sample

In the sample of 800 households, there are 2312 adult members 18 years or older. Of the total sample of households, in 142 households are 160 adults who are current clients of financial institutions. Of these client households, 93 households have had loan transactions in the past with their financial institution(s), and 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 these 93 borrowing households. The 112 loans were borrowed by 100 adults, i.e. 100 adults out of 160 clients were borrowers. Some households have more than one of their adult members borrowing from a financial institution, and some persons 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 those of those clients who had borrowed in the past at least once. We term this latter group ‘borrowers’ and the tables are differentiated 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 20-22% of clients of publicly owned banks, rural savings and loan banks, as well as non-government organizations belong to very poor households. For the other types of financial institutions, the poverty outreach performance is lower. One needs to point out, however, that the disaggregation by type of financial institutions leads to too few cases (e.g., for cooperatives, n=5), so that statistically valid generalizations about the poverty outreach performance by type of institution cannot be drawn. When one considers the national poverty line of Peru for defining the poverty status, the poverty outreach of financial institutions noticeably improves. Here, the headcount indices range from 0 (other government owned programs) to 55.6 percent (NGO credit institutions).

The above table clearly 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.

The above table 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 5 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 2 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 service (N=15)	8.56	13.3	20.0	13.3	13.3
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 that are shown in the table have too few observations for drawing 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. Criteria applied in this sampling 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, Cuánto, to participate in the survey and some did refuse to participate in the study. Eventually, six volunteered to cooperate by providing a list of new clients as well as information on how to find those clients. These sixe 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)

It proved in a number of MFIs difficult to obtain complete and correct lists of new clients, 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. Moreover, MIBANCO as one as the largest pro-poor financial service provider was not willing to cooperate. In one of the MFIs, only 175 instead of 200 clients could be surveyed for a variety of reasons beyond the control of Cuanto. Moreover, Mibanco, the largest pro-poor financial service provider in Peru, was unwilling to participate in the survey.

In the sample of 1175 households, there are 3530 adult members being 18 years or older. Of these adults, 1515 adults are current clients of financial institutions. The majority of the households (n=1047) have had loan transactions in the past with their financial institution(s), and provided

data on their most recent loans with these financial institutions. In total, 1253 adults had borrowed in the past. We again observe that in some households more than one adult borrows from a financial institution, some persons had more than one outstanding loan, and some 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 the microbank Edyficar has the highest share of very poor clients, followed by the savings and credit cooperative San Pedro and then Mibanco. The lowest outreach to the very poor among the purposefully selected institutions is achieved by Coop San Isidro Huaral, where 4.0% of clients were very poor. As expected, the large banks Banco del Trabajo and Banco de la Nation do not reach many very poor. When one considers the national poverty line of Peru for defining the poverty status, the poverty outreach of financial institutions noticeably improves.

Here, the headcount indices range from 9.5 % (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 1 US-Dollar, only the cooperative San Pedro achieves a noticeable outreach among the the very poor. The same holds true for the two-dollar poverty line.

The stark differences between the cooperative San Pedro cooperative and Edyficar with respect to poverty outreach when using the international poverty lines of one and two dollars occur because the Cooperative San Pedro works outside Lima in rural areas where the national poverty lines are low in comparison with the 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 national poverty lines. In these areas, it reaches a considerable share of the very poor but most of these still earn incomes above the international poverty lines.

Summary

In this report, we presented 9 regression models, each with a set of best 5, best 10, and best 15 regressors. Thus, in total, there are 27 potential, newly developed tools that could be used for poverty assessment in Peru. We also presented a two-step model as an alternative estimation approach, and estimated four versions using this approach. Other tools can be generated from the data set by varying the choice of subsets of regressors, e.g. by using additional information on the difficulty or verifiability of indicators.

6.1 Synthesis of results

The nine models show satisfactory levels of overall accuracy, i.e. the percentage of household correctly predicted as being very poor or not very poor. In the case of tools using only 5 poverty indicators (i.e., the BEST 5 models presented in chapter 3), levels in overall accuracy range from 81.50 percent for Model 6 (which has a restrictive set of regressors excluding subjective indicators and all monetary values for assets and expenditures) to 84.50 percent in Model 1 (which includes the full set of 259 regressors).

When one increases the number of poverty indicators from 5 to 10, and further to 15, the accuracy, undercoverage and leakage improve in most cases. Maximum overall accuracy is reached in the BEST 15 set from Model 3, with more than 88 percent. In other words, only about 12 percent of households are wrongly predicted in their poverty status by this tool. All of the first four models achieve overall accuracies of at least 87 percent. Near-maximum accuracy and lowest undercoverage and leakage levels can thus be reached with a set of 10 to 15 poverty indicators (holding other factors equal)[13].

Surprisingly, the performance does not decrease when excluding expenditure and summary asset values from the regressor set (as it happens in Model 1 to Model 4). Direct expenditure information is not necessary to correctly predict the poverty status in Peru. This, of course, is good news for the development of practitioners' tools, as the omission of expenditure indicators means greater practicality.

Model 7 includes only indicators that are deemed by experienced survey interviewers in Peru as highly verifiable and easy to ask. It can be therefore considered as the model with the highest practicability. Model 7 (with 15 poverty indicators) still has an overall accuracy of 84.63 percent and an accuracy among the very poor of 63.72 percent. When we compare these results with Model 3, which achieves the highest accuracy, one loses 3.5 percentage points (88.13 % – 84.63 %) in absolute accuracy and 6.98 percentage points (70.70 % - 63.72 %) in accuracy among the not very poor. This comparison highlights the existing trade-offs between practicality and accuracy. However, these differences in accuracy between model 7 and model 3 will be less in practice (i.e if used by practitioners) as the measurement error for the more complex variables

used in model 3 will be higher than the one for the simpler variables in model 7.

Annex E provides a summary of all accuracy results for each model, while Annex F summarizes the variables in all nine models. From the first model onwards there were various indicators among the best regressors that can potentially identify the very poor, and distinguish them from the not very poor (such as inferior housing material, access to safety net programs and food security indicators). Yet, all models tend to include more indicators that identify the wealthy from the very poor (for example, value of assets, or ownership of a TV or car). A 'being-wealthy' indicator is not the same as a 'being very poor' indicator. If the former dominate the model, as it tends to occur in all of the models, the accuracy of correctly predicting the very poor tends to be much lower than the accuracy of correctly predicting the not very poor. This observation can easily be understood if we remember that the national sample (representing the whole population of Peru) contains 73 percent of 'not very poor' people that dominate the selection of indicators in a one-step regression approach.

Hence, in all nine models we notice that the accuracy among the very poor is lower than the accuracy among the not very poor. To reduce this imbalance between the two types of accuracy, we present two-step regression models following a method pioneered by Grootaert et al. (1998). The computational costs of these models are higher than simple (i.e. one-step) ordinary least squares models.

6.2 The issue of unbalanced accuracies revisited

We note that all models estimated by Ordinary Least Squares (OLS) in Chapter 3 had lower accuracies among the very poor compared to the accuracies among the not very poor. This implies that the inaccuracies in prediction are not equally distributed over all expenditure percentiles but are systematically higher among the very poor compared to the not very poor. This problem of unbalanced accuracies can be significantly reduced by the use of two-step models.

The overall accuracy of the best two-step model (Version C in Chapter 4) is 90.25%, with an accuracy among the very poor of 85.58 % and an accuracy among the not very poor of 91.96 %. This model slightly overestimates the in-sample headcount index by an absolute 1.8 percent, i.e. 28.8 percent instead of 27 percent.

These results compare very favorably with the (one-step) OLS models presented in Chapter 3. Here, the best model (in terms of overall accuracy) is Model 3-BEST 15. This model yields an overall accuracy of 88.13 %, the accuracy among the very poor is 70.70 % while the accuracy among the not very poor is 94.53 %. Compared to this model, the two-step estimation approach achieves improvement in two of the three performance criteria (overall accuracy increases by about 2 percentage points and accuracy among the very poor improves by about 15 percentage points). The trade-off, a loss of only an absolute 2.5 percent in accuracy among the not very poor, appears negligible in comparison. These results clearly support the further use of two-step models

as overall accuracy and accuracy among the very poor improves compared to the one-step OLS method[14].

While the alternative estimation approach introduced in Chapter 4 shows promise in improving the accuracy among the very poor, it is also more time-consuming to perform. Compared to the OLS models in Chapter 3, the method leads to a practitioner tool that will require more questions on additional indicators selected during the second step. For the example of the best model (version C) identified in Chapter 4, instead of 15 indicators, the tool would consist of 9 additional indicators identified in the second step. However, questions on these nine additional indicators can be performed in the same interview. Thus, the gain in accuracy achieved by two-step models comes also at some additional cost for the creation of practical tools.

Apart from the two-step model, there exist a few other alternative estimation methods for addressing the issue of unbalanced accuracies (or low accuracy among the very poor) that will be tested.

In general, for the purpose of developing practitioner tools to assess the outreach of micro-enterprise programs to the very poor, it appears sensible to judge a model A to be superior compared to a model B if A achieves a higher overall accuracy and a higher accuracy among the very poor than model B (even if B has a higher accuracy among the not very poor). Therefore, the two-step model (Version C, Chapter 4) is rated as a better tool than the simple OLS model 3-BEST 15 from Chapter 3. Our analysis with different methods (simple OLS model compared to two-step OLS models) explicitly showed a trade-off between accuracy among the very poor compared to the accuracy among the not very poor. In future analysis, we may also experience trade-offs (among different models) between accuracy among the very poor and overall accuracy. In such cases, the identification of the best model in terms of accuracy might be less obvious. The stakeholders of this project will then need to make a considered judgement of how to value the inherent trade-offs between overall accuracy, accuracy among the not very poor, and accuracy among the very poor.

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/).

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. 2004. Developing and testing poverty assessment tools: Results from accuracy tests in Bangladesh. IRIS Center, University of Maryland, College Park. November 2004. (Available [online] at http://www.povertytools.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	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household size	800	1	14	4.68	2.04	X	X	X	X	X	X	X	X	X
Household size squared	800	1	196	26.06	23.99	X	X	X	X	X	X	X	X	X
Age of household head	800	18	94	47.72	16.11	X	X	X	X	X	X	X	X	X
Dummy if Lima Metropolitan	800	0	1	0.25	0.43	X	X	X	X	X	X	X	X	X
Dummy if Coast Rural	800	0	1	0.04	0.20	X	X	X	X	X	X	X	X	X
Dummy if Highland Urban	800	0	1	0.13	0.33	X	X	X	X	X	X	X	X	X
Dummy if Highland Rural	800	0	1	0.21	0.41	X	X	X	X	X	X	X	X	X
Dummy if Lowland Urban	800	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Dummy if Lowland Rural	800	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Age of youngest household member	800	0	90	12.37	15.56	X	X	X	X	X	X	X	X	
Age of oldest household member	800	18	94	50.28	17.09	X	X	X	X	X	X			
Median education of all household members	800	0	6	2.83	1.49	X	X	X	X	X	X			X
Minimum education level of any household member	800	0	6	1.21	1.41	X	X	X	X	X	X			X
Maximum education level of any household member	800	0	7	3.95	1.53	X	X	X	X	X	X			X
Median education of adult household members	800	0	7	3.24	1.57	X	X	X	X	X	X			X
Maximum education level of any adult household member	800	0	7	3.87	1.63	X	X	X	X	X	X			X
Spouse can read only	800	0	1	0.01	0.11	X	X	X	X	X	X	X	X	X
Number of hh members who can read only	800	0	3	0.06	0.27	X	X	X	X	X	X	X	X	X
Number of adult hh members who can read only	800	0	50	0.65	4.60	X	X	X	X	X	X	X	X	X
Household head can read and write	800	0	1	0.92	0.27	X	X	X	X	X	X	X	X	X
Number of adult hh members who can read and write	800	0	9	2.65	1.49	X	X	X	X	X	X	X	X	X
% of adult hh members who read and write	800	0	100	59.23	25.90	X	X	X	X	X	X	X	X	X
Dummy =1 if male head of hh	800	0	1	0.82	0.39	X	X	X	X	X	X	X	X	
Number of female adult hh members	800	0	6	1.50	0.85	X	X	X	X	X	X	X	X	
% of hh members being chronically ill	800	0	100	9.72	19.35	X	X	X	X	X	X			X
% of chronically ill adults (in relation to hh size)	800	0	100	8.69	18.68	X	X	X	X	X	X			X
Household head is chronically ill	800	0	1	0.15	0.36	X	X	X	X	X	X			X
% of hh members with any disability (in relation to hh size)	800	0	50	1.09	5.37	X	X	X	X	X	X	X	X	X
% of adults with any disability (in relation to hh size)	800	0	50	0.87	4.93	X	X	X	X	X	X	X	X	X
% of dependents lt 15 and gt 64 years (in relation to hh size)	800	0	100	38.53	24.20	X	X	X	X	X	X	X	X	
% of dependents lt 14 and gt 60 years (in relation to hh size)	800	0	100	38.59	24.85	X	X	X	X	X	X	X	X	

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Head of hh has nonagricultural self-employment (1=Yes, 0=no)	800	0	1	0.31	0.46	X	X	X	X	X	X			X
Head of hh is nonagricultural daily worker (1=Yes, 0=no)	800	0	1	0.10	0.30	X	X	X	X	X	X			X
Head of hh is retired (1=Yes, 0=no)	800	0	1	0.08	0.27	X	X	X	X	X	X			X
Head of hh is occupied in housework (1=Yes, 0=no)	800	0	1	0.07	0.25	X	X	X	X	X	X			X
Head of hh chooses leisure (1=Yes, 0=no)	800	0	1	0.00	0.05	X	X	X	X	X	X			X
Median education level of females	800	0	6	2.65	1.57	X	X	X	X	X	X			X
Maximal education level males	800	0	7	3.67	1.56	X	X	X	X	X	X			X
Number of male adults in hh	800	0	6	1.39	0.89	X	X	X	X	X	X	X	X	X
Number of literate female adults in hh	800	0	6	1.32	0.94	X	X	X	X	X	X	X	X	X
Number of females with some disability	800	0	1	0.02	0.13	X	X	X	X	X	X	X	X	X
Number of males with some disability	800	0	1	0.03	0.17	X	X	X	X	X	X	X	X	X
Number of females with some chronic illness	800	0	3	0.20	0.46	X	X	X	X	X	X			X
Number of males with some chronic illness	800	0	4	0.18	0.43	X	X	X	X	X	X			X
Average number of days sick by females	800	0	365	3.79	18.38	X	X	X	X	X	X			X
Total number of days sick by females	800	0	365	7.02	28.93	X	X	X	X	X	X			X
Ratio male adults/female adults	800	0	5	1.12	0.89	X	X	X	X	X	X	X	X	
How many rooms does the dwelling have?	800	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?	800	0	1	0.28	0.45	X	X	X	X	X	X	X	X	X
Do you have Mobile (cell phone) in the house?	800	0	1	0.15	0.35	X	X	X	X	X	X	X	X	X
How many meals were served to the hh members during the last 2 days?	800	2	8	5.70	0.77	X	X	X	X					
In the last 7 days, how many days Other red meat served by the hh in a main meal eaten	800	0	7	0.80	1.32	X	X	X	X					
In the last 7 days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	800	0	4	0.09	0.38	X	X	X	X					
In the last 7 days, how many days Butter (urban) or margarina (rural) served by the hh in a main meal eaten	800	0	7	1.46	2.37	X	X	X	X					
In the last 7 days, how many days Chicken eggs served by the hh in a main meal eaten	800	0	7	3.00	2.33	X	X	X	X					
In the last 7 days, how many days did a main meal consist of plain rice and any vegetables only?	800	0	7	0.39	1.02	X	X	X	X				X	
In the last 7 days, how many days did a main meal consist of plain rice only?	800	0	7	0.24	0.71	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
In the last 30 days, for how many days did your hh not have enough to eat everyday?	800	0	30	2.86	5.20	X	X	X	X					
Did you or any other adult hh member lose weight in last 12 months because you did not have enough money to buy food?	800	0	1	0.21	0.41	X	X	X	X					
In the last 7 days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	800	0.03	1000	1.28	35.35	X	X	X	X	X	X			
In last 3 years, how many marriages of a first degree relative to hh head or spouse?	800	0	7	0.14	0.52	X	X	X	X	X	X			
Total number of children adopted, last 3 years	800	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	800	0	36	0.42	2.40	X	X	X	X	X	X			
Occurrence of a serious chronic illness or major disability of any hh member In last 3 years	800	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 hh for ever In last 3 years	800	0	1	0.03	0.16	X	X	X	X	X	X			
Total number of deaths of dependent household members, last 3 years	800	0	4	0.22	0.60	X	X	X	X	X	X	X	X	
Did your hh have a very serious problem or failure in your own animal production In last 3 years?	800	0	1	0.18	0.38	X	X	X	X	X	X	X	X	
Did your hh have a very serious problem or failure in your own micro-enterprise In last 3 years?	800	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	800	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?	800	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?	800	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?	800	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?	800	0	36	0.08	1.48	X	X	X	X	X	X			
Have you or members of hh are denied service or only limited opportunity to job training/ employment	800	0	1	0.15	0.36	X	X	X	X					
Have you or members of hh are denied service or only limited opportunity to transportation	800	0	1	0.11	0.31	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Have you or members of hh are denied service or only limited opportunity to water distribution	800	0	1	0.09	0.28	X	X	X	X					
Have you or members of hh are denied service or only limited opportunity to sanitation services	800	0	1	0.06	0.23	X	X	X	X					
Have you or members of hh are denied service or only limited opportunity to agricultural extension	800	0	1	0.05	0.22	X	X	X	X					
Have you or members of hh are denied service or only limited opportunity to justice/conflict resolution	800	0	1	0.16	0.36	X	X	X	X					
Have you or members of hh are denied service or only limited opportunity to security/police services	800	0	1	0.18	0.38	X	X	X	X					
does the household own the house?	800	0	1	0.70	0.46	X	X	X	X	X	X	X	X	X
Total agricultural area (irrigated or not), square meters	800	5	800000	6939.60	33090.88	X	X	X	X	X	X			X
HH in Traders association	800	0	1	0.01	0.07	X	X	X	X	X	X	X	X	
Total hh members in professional assoc.	800	0	2	0.01	0.13	X	X	X	X	X	X	X	X	
Total hh members in trade union	800	0	2	0.02	0.20	X	X	X	X	X	X	X	X	
HH in water/waste group	800	0	1	0.09	0.29	X	X	X	X	X	X	X	X	
Total hh members in Water/waste group	800	0	2	0.10	0.32	X	X	X	X	X	X	X	X	
total of HH members in NGO for BD services	800	0	1	0.01	0.08	X	X	X	X	X	X	X	X	
HH in other NGO	800	0	1	0.02	0.13	X	X	X	X	X	X	X	X	
HH in religious group	800	0	1	0.18	0.38	X	X	X	X	X	X	X	X	
Total of hh members in religious group	800	0	8	0.35	0.91	X	X	X	X	X	X	X	X	
HH in youth group	800	0	1	0.01	0.08	X	X	X	X	X	X	X	X	
Total hh members in youth group	800	0	2	0.01	0.12	X	X	X	X	X	X	X	X	
Total hh members in women's group	800	0	2	0.06	0.25	X	X	X	X	X	X	X	X	
HH in parents group	800	0	1	0.14	0.35	X	X	X	X	X	X	X	X	
Total hh members in parents group	800	0	2	0.16	0.40	X	X	X	X	X	X	X	X	
HH in sports group	800	0	1	0.03	0.16	X	X	X	X	X	X	X	X	
total hh members in sports group	800	0	2	0.03	0.17	X	X	X	X	X	X	X	X	
Total hh members in other groups	800	0	2	0.04	0.21	X	X	X	X	X	X	X	X	
Cattle ownership	800	0	1	0.06	0.23	X	X	X	X	X	X	X	X	X
Milkcows ownership	800	0	1	0.08	0.27	X	X	X	X	X	X	X	X	X
Lamas ownership	800	0	1	0.01	0.09	X	X	X	X	X	X	X	X	X
Sheep/goats ownership	800	0	1	0.11	0.31	X	X	X	X	X	X	X	X	X
Pigs ownership	800	0	1	0.12	0.33	X	X	X	X	X	X	X	X	X

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Poultry ownership	800	0	1	0.37	0.48	X	X	X	X	X	X	X	X	X
Motorcycle ownership	800	0	1	0.02	0.12	X	X	X	X	X	X	X	X	X
Motocab ownership	800	0	1	0.01	0.12	X	X	X	X	X	X	X	X	X
Tractor ownership	800	0	1	0.01	0.08	X	X	X	X	X	X	X	X	X
Other vehicle ownership	800	0	1	0.02	0.15	X	X	X	X	X	X	X	X	X
Color TV ownership	800	0	1	0.52	0.50	X	X	X	X	X	X	X	X	X
VCR ownership	800	0	1	0.12	0.32	X	X	X	X	X	X	X	X	X
Electric or gas cooker ownership	800	0	1	0.55	0.50	X	X	X	X	X	X	X	X	X
Microwave ownership	800	0	1	0.06	0.23	X	X	X	X	X	X	X	X	X
Fan ownership	800	0	1	0.14	0.34	X	X	X	X	X	X	X	X	X
Bed ownership	800	0	1	0.98	0.14	X	X	X	X	X	X	X	X	X
Suit/ Jacket ownership	800	0	1	0.34	0.47	X	X	X	X	X	X	X	X	X
Leather shoes ownership	800	0	1	0.53	0.50	X	X	X	X	X	X	X	X	X
Milkcow number	800	0	40	0.46	2.96	X	X	X	X	X	X			X
Pigs number	800	0	10	0.23	0.81	X	X	X	X	X	X			X
Horses number	800	0	8	0.22	0.75	X	X	X	X	X	X			X
Poultry number	800	0	120	4.49	9.43	X	X	X	X	X	X			X
Car number	800	0	3	0.09	0.31	X	X	X	X	X	X			X
Motorcycles number	800	0	2	0.02	0.14	X	X	X	X	X	X			X
Radios number	800	0	8	0.89	0.68	X	X	X	X	X	X			X
Colour TVs number	800	0	5	0.65	0.77	X	X	X	X	X	X			X
Video recorders number	800	0	3	0.13	0.36	X	X	X	X	X	X			X
Refrigerators number	800	0	2	0.38	0.50	X	X	X	X	X	X			X
Electric/ gas kitchens number	800	0	2	0.56	0.52	X	X	X	X	X	X			X
Beds number	800	0	10	3.21	1.75	X	X	X	X	X	X			X
Shoes number	800	0	30	1.85	3.04	X	X	X	X	X	X			X
Skirts number	800	0	20	1.15	2.39	X	X	X	X	X	X			X
Metal pots number	800	0	43	4.66	3.15	X	X	X	X	X	X			X
Wooden plows number	800	0	6	0.14	0.53	X	X	X	X	X	X			X
Food exp share, C, in %	800	0.11	1.93	1.21	0.30	X								
Average of hh members, except head	800	0.5	78	18.44	10.57	X	X	X	X	X	X			
Number of steps above step identified as int pov line, if minus below	800	-9	6	1.24	1.87	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Dummy, if hh rates itself below the step reflecting the int. pov line	800	0	1	0.14	0.35	X	X	X	X					
Dummy, if hh rates itself below the step reflecting the respective nat. pov line	800	0	1	0.37	0.48	X	X	X	X					
Inferior food consumed at least sometimes, urb/rural calculation, 1=yes	800	0	1	0.17	0.37	X	X	X	X					
Inferior food consumed at least sometimes, macroregional calc, 1=yes	800	0	1	0.28	0.45	X	X	X	X					
Household participated at least in one food aid program during last 3 months (e16)	800	0	1	0.31	0.46	X	X	X	X	X	X			
Household participated in school feeding during last 3 months	800	0	1	0.14	0.34	X	X	X	X	X	X			
Household received subsidized food during last 3 months	800	0	1	0.25	0.43	X	X	X	X	X	X			
Household participated in social kitchen during last 3 months	800	0	1	0.02	0.13	X	X	X	X	X	X			
Household participated in any other food aid program during last 3 months	800	0	1	0.01	0.09	X	X	X	X	X	X			
Agree that you feel accepted as a member of this village/neighborhood?	800	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?	800	0	1	0.34	0.47	X	X	X	X					
Dummy: HH feels that clothing expenses are below need	800	0	1	0.46	0.50	X	X	X	X					
Dummy: HH feels that clothing expenses are above need	800	0	1	0.01	0.08	X	X	X	X					
Dummy: HH feels that health care expenses are below need	800	0	1	0.37	0.48	X	X	X	X					
Dummy: HH feels that health care expenses are above need	800	0	1	0.01	0.09	X	X	X	X					
Dummy: HH feels that housing expenses are below need	800	0	1	0.38	0.48	X	X	X	X				X	
Dummy: Household rates itself above intl. poverty line on ladder, source H08b (Note: The international poverty line per capita per day is 2.08 Soles. For the questionnaire, the cut-off value was approximated with 300 Soles on a monthly basis for a household with 2 adults and 3 dependents).	800	0	1	0.71	0.45	X	X	X	X				X	
Dummy: Household rates itself above national poverty line on ladder, source H08a (Note: The national poverty line was adjusted by region, and calculated per month for a household with 2 adults and 3 dependents).	800	0	1	0.41	0.49	X	X	X	X					
House size: small	800	0	1	0.23	0.42	X	X	X	X	X	X	X	X	

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
quality of walls: poor	800	0	1	0.14	0.34	X	X	X	X	X	X	X	X	
No lock in main entrance door or wood or metal bar to close from inside	800	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	800	0	1	0.28	0.45	X	X	X	X	X	X	X	X	
Dummy, roof with leaves, straw of bamboo/wood	800	0	1	0.10	0.30	X	X	X	X	X	X	X	X	X
Dummy, roof with Cl sheet	800	0	1	0.37	0.48	X	X	X	X	X	X	X	X	X
Exterior walls: wood	800	0	1	0.08	0.27	X	X	X	X	X	X	X	X	X
Floor is wood or brick/stone	800	0	1	0.05	0.22	X	X	X	X	X	X	X	X	X
Cooking fuel is bamboo/wood/sawdust collected	800	0	1	0.23	0.42	X	X	X	X	X	X	X	X	X
Cooking fuel is bamboo/wood/sawdust purchased	800	0	1	0.08	0.28	X	X	X	X	X	X	X	X	X
Toilet: pit toilet	800	0	1	0.15	0.35	X	X	X	X	X	X	X	X	X
Rooms per person	800	0.08	8	0.81	0.69	X	X	X	X	X	X	X	X	X
Dummy: Public borehole/spring or public well	800	0	1	0.09	0.28	X	X	X	X	X	X	X	X	X
Dummy: Untreated piped/river water	800	0	1	0.19	0.39	X	X	X	X	X	X	X	X	X
Dummy: Head of hh sleeps on something else than bed (e.g. floor, mat, mattress, hammock)	800	0	1	0.06	0.24	X	X	X	X	X	X	X	X	
Dummy: hh cooks in one of the rooms in the house	800	0	1	0.13	0.33	X	X	X	X	X	X	X	X	
Number of days in past 7 days any of six superior food eaten (max. 42)	800	0	25	8.81	5.26	X	X	X	X	X	X			
Household always ate enough from what they wanted (12mo)	800	0	1	0.10	0.30	X	X	X	X				X	
Household often did not have enough food (12mo)	800	0	1	0.05	0.23	X	X	X	X					
Dummy: hh borrows from corner shop rarely	800	0	1	0.19	0.39	X	X	X	X					
Dummy: hh borrows from corner shop often	800	0	1	0.15	0.35	X	X	X	X					
Dummy: hh borrows from corner shop mostly	800	0	1	0.07	0.25	X	X	X	X					
Dummy: hh borrows from neighbors/relatives rarely	800	0	1	0.11	0.31	X	X	X	X					
Dummy: hh borrows from neighbors/relatives sometimes, often or mostly	800	0	1	0.10	0.30	X	X	X	X					
Household ate less food for less than 30 days but more than 10 days during past 12 months	800	0	1	0.12	0.33	X	X	X	X					
Household ate less food for less than 10 days during past 12 months	800	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	800	0	1	0.06	0.23	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
LOG of annualized tot. summary of expenditures section C	800	6.12	10.5	8.75	0.66	X								
LOG of sum of household clothing expenditures in past 12 months	800	1.61	9.21	6.05	1.09	X						X	X	X
LOG of annualized food expenditures recall average week	800	1.65	9.81	8.24	0.68	X								X
LOG of annualized nonfood expenditures (services, transport)	800	0.91	9.6	7.27	0.94	X								
LOG of min wage he would accept during low income season	800	0.69	4.61	2.81	0.89	X	X	X	X					
LOG of min wage he would accept for next working day	800	1.1	5.3	3.12	1.04	X	X	X	X					
LOG of value food produced by hh in farm or garden, or gathers and consumes, per week	800	-5.16	5.01	-2.88	3.58	X								
LOG of hh monthly expenditure on utilities (electricity, phone, water, etc)	800	-2.49	6.4	2.71	2.70	X	X	X					X	
LOG of how much hh usual monthly expenditures for transport	800	-2.46	6.59	2.83	2.51	X								
LOG of how much hh usual monthly expenditures for fuel	800	-3.38	5.86	2.59	2.12	X								
LOG of how much hh usual monthly expenditures for other goods (c06)	800	-5.46	5.01	-2.98	3.84	X								
LOG of how much hh spent last 12 months on school/ education	800	-0.37	9.48	4.19	2.91	X								
LOG of how much hh spent last 12 months on health	800	-1.05	8.7	4.07	2.31	X								
LOG of expenditures on furniture, last 12 months (c10)	800	-1.64	8.22	-0.64	2.59	X								
LOG of how much hh sent to relatives in last 12 months	800	-1.55	8.52	-0.80	2.24	X								
LOG of hh expenditure on other expenditures in last 12 mo (soc evs, gifts, taxes)	800	-1.97	8.01	1.18	3.33	X								
LOG of value of agricultural area, irrigated	800	1.1	11.29	1.90	2.40	X	X	X	X	X				
LOG of total resale value of assets animals and other assets (trsvalue), in Soles	800	3	12.21	6.83	1.40	X	X							X
LOG of how much second person did send you from somewhere else, past 12 months	800	-3.66	5.7	-3.63	0.55	X	X	X	X	X				
LOG of how much does your household need per month to live	800	5.01	8.7	6.77	0.65	X	X	X	X				X	
Religion of hh head is other than catholic	800	0	1	0.18	0.38	X	X	X	X	X	X			
Household usually purchases rice twice a week	800	0	1	0.05	0.21	X	X	X	X					
Household usually purchases rice weekly	800	0	1	0.27	0.44	X	X	X	X					
Household usually purchases rice fortnightly	800	0	1	0.08	0.27	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household usually purchases rice monthly or less frequent than that	800	0	1	0.13	0.33	X	X	X	X					
Household owns any of motor tiller, wooden plow, tubeirri or husking machine	800	0	1	0.10	0.30	X	X	X	X	X	X	X	X	X
Number of memberships out of 22 institutions	800	0	7	0.77	0.98	X	X	X	X	X	X	X	X	
LOG value of lamas	800	-8.05	4.87	-7.96	1.10	X	X	X	X	X				X
LOG value of milkcows	800	-2.9	8.52	-2.17	2.52	X	X	X	X	X				X
LOG value of sheep and goats	800	-4.57	7.4	-3.59	2.84	X	X	X	X	X				X
LOG value of pigs	800	-4.55	6.4	-3.48	2.90	X	X	X	X	X				X
LOG value of horses	800	-3.88	7.65	-2.80	3.02	X	X	X	X	X				X
LOG value of motorcabs	800	-1.53	9.62	-1.39	1.16	X	X	X	X	X				X
LOG value of tractors	800	-1.82	10.46	-1.75	0.93	X	X	X	X	X				X
LOG value of other vehicles	800	-4.65	8.16	-4.42	1.50	X	X	X	X	X				X
LOG value of radios	800	-2.12	8.01	2.60	2.80	X	X	X	X	X				X
LOG value of electric/ gas cooking	800	-1.86	7.31	1.71	3.38	X	X	X	X	X				X
LOG value of food processing assets	800	-3.12	6.21	-0.13	3.51	X	X	X	X	X				X
LOG value of fans	800	-4.54	5.48	-3.41	2.87	X	X	X	X	X				X
LOG value of bed/hammocks	800	-1.86	8.16	4.23	1.67	X	X	X	X	X				X
LOG value of sunday jackets	800	-2.93	7.94	-0.49	3.53	X	X	X	X	X				X
LOG value of leather shoes	800	-3.03	7.47	0.59	3.58	X	X	X	X	X				X
LOG value of metal pots	800	-3.2	7.35	2.18	2.51	X	X	X	X	X				X
LOG value of motor tillers	800	-6.91	4.38	-6.88	0.55	X	X	X	X	X				X
LOG value of tubes for irrigation	800	-2.64	11.72	-2.61	0.61	X	X	X	X	X				X
LOG of value of radio, TV, VCR and cdplayer	800	-0.61	8.56	4.68	2.13	X	X	X	X	X				X
LOG of total value of agricultural assets (motortiller, plow, irrigation, huskmach)	800	-2.56	11.72	-1.95	1.90	X	X	X	X	X				X
LOG of total value of all animals	800	-1.84	9.81	0.87	3.58	X	X	X	X	X				X
Dummy, if any hh member has a passbook savings account	800	0	1	0.08	0.28	X	X	X	X	X	X			X
Dummy, if any hh member has a life insurance	800	0	1	0.04	0.20	X	X	X	X	X	X			X
Dummy, if spouse has any account	800	0	1	0.05	0.21	X	X	X	X	X	X			X
Dummy, hh has borrowed for food and emergencies from informal sector in past 3 years	800	0	1	0.23	0.42	X	X	X	X					
Dummy, hh has lent money to others in past 3 years	800	0	1	0.15	0.36	X	X	X	X					
Dummy, hh has borrowed from informal sector in past 3 years	800	0	1	0.31	0.46	X	X	X	X					

Variable label	N	Min.	Max.	Mean	St.Dev.	M1	M2	M3	M4	M5	M6	M7	M8	M9
Dummy =1 if hh declares to not be able to save anything	800	0	1	0.83	0.38	X	X	X	X	X	X			
ln, value of jewelry	800	-1.81	11.85	-1.32	1.93	X	X	X	X	X				X
ln, value of largest loan for food/emergency in past 3 years	800	-1.33	10.04	0.19	2.96	X	X	X	X	X				
ln, value of debt owed by other households to hh	800	-1.04	9.71	-0.04	2.40	X	X	X	X	X				X
ln, value of formal savings of spouse	800	-2.08	10.46	-1.98	0.76	X	X	X	X	X				X
ln, value of informal savings deposited at home or somewhere else	800	-3.76	6.4	-3.75	0.36	X	X	X	X	X				
Do you have secondary school ?	800	0	1	0.59	0.49	X	X	X	X	X	X	X	X	
Do you have market/ bazaar?	800	0	1	0.42	0.49	X	X	X	X	X	X	X	X	
How far away the community center (km)?	800	0	4	0.22	0.70	X	X	X	X	X	X	X	X	
How far away the access to mainline phone (km)?	800	0	28	2.73	6.46	X	X	X	X	X	X	X	X	
Distance to department capital?	800	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)	800	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	800	5	95	39.58	28.71	X	X	X	X	X	X	X	X	
Sum of distances to department, provincial and district capital	800	0	48	6.44	13.27	X	X	X	X	X	X	X	X	
Education level of other hh members, excluding head	800	0	6	2.45	1.57	X	X	X	X	X	X			X
Squared age of hh head	800	324	8836	2536.23	1705.18	X	X	X	X	X	X			X
hh has electricity (autobattery, own generator included)	800	0	1	0.76	0.43	X	X	X	X	X	X	X	X	X
HH has piped water (treated or untreated)	800	0	1	0.83	0.37	X	X	X	X	X	X	X	X	X
LOG Remittances sent	800	-0.25	0.89	-0.10	0.25	X						X	X	X
Household head is single	800	0	1	0.07	0.25	X	X	X	X	X	X	X	X	X
Part A, subj ranking scale 1 to 5 compared to community	800	0	1	0.01	0.07	X	X	X	X					
Part A, subj ranking scale 1 to 5 compared to community	800	0	1	0.30	0.46	X	X	X	X					
Share of daily clothing exp in total daily expenditures	800	0	0.67	0.07	0.08	X								
LOG, average daily per-capita clothing expenditures, Soles	800	-4.98	1.7	-1.29	1.07	X	X	X				X	X	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
Dummy =1 if male head of hh	800	0	1	0.82	0.39
Number of female adult hh 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 hh	800	0	6	1.39	0.89
Number of literate female adults in hh	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 hh members in women's group	800	0	2	0.06	0.25
LOG of min wage main male income earner would accept during low income season	800	0.69	4.61	2.81	0.89
LOG of min wage main male income earner would accept for next working day	800	1.1	5.3	3.12	1.04
Dummy, if spouse has any account	800	0	1	0.05	0.21
ln, 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
Dummy if Lima Metropolitan	cover sheet	5	5
Dummy if Coast Rural	cover sheet	5	
Dummy if Highland Urban	cover sheet	5	
Dummy if Highland Rural	cover sheet	5	
Dummy if Lowland Urban	cover sheet	5	
Dummy if 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 hh members who can read only	B7	4	
Number of adult hh members who can read only	B7	4	
Household head can read and write	B7	4	
Number of adult hh members who can read and write	B7	4	
% of adult hh members who read and write	B7	4	
Dummy =1 if male head of hh	B3	5	5
Number of female adult hh members	whole section B	4	

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
% of hh members being chronically ill	B13	3	3
% of chronically ill adults (in relation to hh size)	B13	3	
Household head is chronically ill	B13	3	
% of hh members with any disability (in relation to hh size)	B14	5	5
% of adults with any disability (in relation to hh size)	B14	5	5
% of dependents lt 15 and gt 64 years (in relation to hh size)	whole section B	4	
% of dependents lt 14 and gt 60 years (in relation to hh size)	whole section B	4	
Head of hh has nonagricultural self-employment (1=Yes, 0=no)	B11	3	5
Head of hh is nonagricultural daily worker (1=Yes, 0=no)	B11	3	
Head of hh is retired (1=Yes, 0=no)	B11	3	
Head of hh is occupied in housework (1=Yes, 0=no)	B11	3	
Head of hh chooses leisure (1=Yes, 0=no)	B11	3	
Median education level of females	B8	3	
Maximal education level males	B8	3	
Number of male adults in hh	whole section B	4	
Number of literate female adults in hh	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 hh members during the last 2 days?	E01A	2	3
In the last 7 days, how many days Other red meat served by the hh in a main meal eaten	E03B	2	4
In the last 7 days, how many days Fish, type Cojinova (coast) or Paiche (jungle) or Trucha (Andes)	E03C	2	4
In the last 7 days, how many days Butter (urban) or margarina (rural) served by the hh in a main meal eaten	E03E	2	4

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
In the last 7 days, how many days Chicken eggs served by the hh in a main meal eaten	E03F	2	4
In the last 7 days, how many days did a main meal consist of plain rice and any vegetables only?	E04	2	4
In the last 7 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 hh not have enough to eat everyday?	E07	2	1
Did you or any other adult hh member lose weight in last 12 months because you did not have enough money to buy food?	E14	2	2
In the last 7 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 hh 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 hh member In last 3 years	G019	4	4
has it occurred that a major working, income-earning adult member left the hh for ever In last 3 years	G110		
Total number of deaths of dependent household members , last 3 years	G111	4	5
Did your hh have a very serious problem or failure in your own animal production In last 3 years?	G115	4	4
Did your hh 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
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

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
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 hh are denied service or only limited opportunity to job training/ employment	G3D	3	3
Have you or members of hh are denied service or only limited opportunity to transportation	G3F	3	3
Have you or members of hh are denied service or only limited opportunity to water distribution	G3G	3	3
Have you or members of hh are denied service or only limited opportunity to sanitation services	G3H	3	3
Have you or members of hh are denied service or only limited opportunity to agricultural extension	G3I	3	3
Have you or members of hh are denied service or only limited opportunity to justice/conflict resolution	G3J	3	3
Have you or members of hh 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
HH in Traders association	G2A2	4	4
Total hh members in professional assoc.	G2A2	4	
Total hh members in trade union	G2A2	4	
HH in water/waste group	G2A2	4	
Total hh members in Water/waste group	G2A2	4	
total of HH members in NGO for BD services	G2A2	4	
HH in other NGO	G2A2	4	
HH in religious group	G2A2	4	
Total of hh members in religious group	G2A2	4	
HH in youth group	G2A2	4	
Total hh members in youth group	G2A2	4	
Total hh members in women's group	G2A2	4	
HH in parents group	G2A2	4	
Total hh members in parents group	G2A2	4	
HH in sports group	G2A2	4	
total hh members in sports group	G2A2	4	
Total hh members in other groups	G2A2	4	

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
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
Metal pots number	section F2, quantity	2	4
Wooden plows number	section F2, quantity	2	4
Food exp share, C, in %	C1, C2	2	4

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Average of hh members, except head	B4	3	3
Number of steps above step identified as int pov line, if minus below	H7	1	
Dummy, if hh rates itself below the step reflecting the int. pov line	H7	1	
Dummy, if hh rates itself below the step reflecting the respective nat. pov line	H7	1	
Inferior food consumed at least sometimes, urb/rural calculation, 1=yes	E15	2	1
Inferior food consumed at least sometimes, macroregional calc, 1=yes	E15	2	
Household participated at least in one food aid program during last 3 months (e16)	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
Dummy: HH feels that clothing expenses are below need	H2	3	5
Dummy: HH feels that clothing expenses are above need	H2	3	
Dummy: HH feels that health care expenses are below need	H3	3	5
Dummy: HH feels that health care expenses are above need	H3	3	
Dummy: HH feels that housing expenses are below need	H5	3	5
Household rates itself above subjective intl. poverty line on ladder, source H08b	H8b	1	
Household rates itself above subjective natl. poverty line on ladder, source H08a	H8a	1	
House size: small	D2.2	4	4
quality of walls: poor	D2.3	5	5
No lock in main entrance door or wood or metal bar to close from inside	D3	5	5

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
Security key lock/metal frame with padlock in main entrance door	D3	5	
Dummy, roof with leaves, straw of bamboo/wood	D6	5	
Dummy, 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	
Dummy: Public borehole/spring or public well	D11	5	
Dummy: Untreated piped/river water	D11	5	
Dummy: Head of hh sleeps on something else than bed (e.g. floor, mat, mattress, hammock)	D15	4	5
Dummy: hh cooks in one of the rooms in the house	D16	5	5
Number of days in past 7 days any of six superior food eaten (max. 42)	E3	3	3
Household always ate enough from what they wanted (12mo)	E9	3	4
Household often did not have enough food (12mo)	E9	3	
Dummy: hh borrows from corner shop rarely	E12a	3	3
Dummy: hh borrows from corner shop often	E12a	3	
Dummy: hh borrows from corner shop mostly	E12a	3	
Dummy: hh borrows from neighbors/relatives rarely	E12b	3	3
Dummy: hh 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
LOG of annualized tot. summary of expenditures section C	whole section C		
LOG of sum of household clothing expenditures in past 12 months	B15	4	4
LOG of annualized food expenditures recall average week	C1, C2	1	3

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
LOG of annualized nonfood expenditures (services, transport)	C3, C4	3	
LOG of min wage he would accept during low income season	D18b	1	1
LOG of min wage he would accept for next working day	D18a	1	1
LOG of value food produced by hh in farm or garden, or gathers and consumes, per week	C2	1	2
LOG of hh monthly expenditure on utilities (electricity, phone, water, etc)	C3	3	4
LOG of how much hh usual monthly expenditures for transport	C4	3	4
LOG of how much hh usual monthly expenditures for fuel	C5	3	4
LOG of how much hh usual monthly expenditures for other goods (c06)	C6	3	3
LOG of how much hh spent last 12 months on school/ education	C7	3	4
LOG of how much hh spent last 12 months on health	C8	3	4
LOG of expenditures on furniture, last 12 months (c10)	C10	3	4
LOG of how much hh sent to relatives in last 12 months	C11	3	4
LOG of hh expenditure on other expenditures in last 12 mo (soc evs, gifts, taxes)	C12	3	4
LOG of value of agricultural area, irrigated	F1A	1	
LOG of total resale value of assets animals and other assets (trsvalue), in Soles	Section F2, value	2	
LOG of how much second person did send you from somewhere else, past 12 months	B20	3	4
LOG of how much does your household need per month to live	H6	1	
Religion of hh 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, tubeirri or husking machine	F2	4	
Number of memberships out of 22 institutions	G2A2	4	
LOG value of lamas	Section F2, value	2	3
LOG value of milkcows	Section F2, value	2	3
LOG value of sheep and goats	Section F2, value	2	3

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 5)
LOG value of pigs	Section F2, value	2	3
LOG value of horses	Section F2, value	2	3
LOG value of motorcabs	Section F2, value	2	3
LOG value of tractors	Section F2, value	2	3
LOG value of other vehicles	Section F2, value	2	3
LOG value of radios	Section F2, value	2	3
LOG value of electric/ gas cooking	Section F2, value	2	3
LOG value of food processing assets	Section F2, value	2	3
LOG value of fans	Section F2, value	2	3
LOG value of bed/hammocks	Section F2, value	2	3
LOG value of sunday jackets	Section F2, value	2	3
LOG value of leather shoes	Section F2, value	2	3
LOG value of metal pots	Section F2, value	2	3
LOG value of motor tillers	Section F2, value	2	3
LOG value of tubes for irrigation	Section F2, value	2	3
LOG of value of radio, TV, VCR and cdplayer	Section F2, value	2	3
LOG of total value of agricultural assets (motortiller, plow, irrigation, huskmach)	Section F2, value	2	3
LOG of total value of all animals	Section F2, value	1	1
Dummy, if any hh member has a passbook savings account	102.3	1	1
Dummy, if any hh member has a life insurance	102.3	1	
Dummy, if spouse has any account	102.3	1	
Dummy, hh has borrowed for food and emergencies from informal sector in past 3 years	G118	3	4
Dummy, hh has lent money to others in past 3 years	101.5	3	4
Dummy, hh has borrowed from informal sector in past 3 years	K14, G118		
dummy =1 if hh declares to not be able to save anything	102.2	1	2
ln, value of jewelry	101.2	1	1
ln, value of largest loan for food/emergency in past 3 years	G118	1	
ln, value of debt owed by other households to hh	101.5	3	4
ln, value of formal savings of spouse	102.3	1	
ln, value of informal savings deposited at home or somewhere else	101.1	1	1
Do you have secondary school ?	Community: B9	4	5
Do you have market/ bazaar?	Community: B13	4	5

Variable label	Refers to question	Verifiability (1 - 5)	Difficulty to ask the corresponding question (1 - 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 hh members, excluding head	B8	3	
Squared age of hh head	B4	4	
hh has electricity (autobattery, own generator included)	D10	5	5
HH has piped water (treated or untreated)	D11	5	
LOG Remittances sent	C11	4	5
Household head is single	B5	4	5
Part A, subj ranking scale 1 to 5 compared to community	A11a	3	4
Part A, subj ranking scale 1 to 5 compared to community	A11a	3	
Share of daily clothing exp in total daily expenditures	B15	3	
LOG, average daily per-capita clothing expenditures, Soles	B15	4	

Annex E: Summary table: Results on accuracy for all models

Model	Description	Type*	Adj. R^2	Accuracy (%)	Accuracy among VP (%)	Accuracy among NVP (%)	Under-coverage (%)	Leakage (%)	Predicted headcount (%)	PAC* (% point)
1	All 259 regressors *(See Table 3.1.2)*	B-5	0.793	84.50	63.72	92.14	36.28	21.40	22.88	4.00
		B-10	0.820	87.00	70.23	93.16	29.77	18.60	23.88	3.00
		B-15	0.834	87.38	70.70	93.50	29.30	17.67	23.75	3.13
2	Exclusion of expenditure variables except for *utilities* and *clothing (Table 3.2.1)*	B-5	0.763	83.75	64.19	90.94	35.81	24.65	23.88	3.00
		B-10	0.796	85.25	65.12	92.65	34.88	20.00	22.88	4.00
		B-15	0.810	87.50	70.70	93.68	29.30	17.21	23.63	3.25
3	Exclusion of *total value of household assets (See Table 3.3.1)*	B-5	0.741	83.83	59.07	92.31	40.93	20.93	21.50	5.38
		B-10	0.789	85.75	65.58	93.16	34.42	18.60	22.63	4.25
		B-15	0.808	88.13	70.70	94.53	29.30	14.88	23.00	3.88
4	Exclusion of *expenditures on utilities* and *clothing (See Table 3.4.1)*	B-5	0.741	84.50	60.93	93.16	39.07	18.60	21.38	5.50
		B-10	0.781	87.00	67.44	94.19	32.56	15.81	22.38	4.50
		B-15	0.800	87.00	66.05	94.70	33.95	14.42	21.63	5.25
5	Exclusion of subjective variables *(See Table 3.5.1)*	B-5	0.723	82.50	57.21	91.79	42.79	22.33	21.38	5.50
		B-10	0.765	84.88	63.26	92.82	36.74	19.53	22.25	4.63
		B-15	0.784	85.13	64.19	92.82	35.81	19.53	22.50	4.38
6	Exclusion of monetary variables *(See Table 3.6.1)*	B-5	0.720	81.50	55.35	91.11	44.65	24.19	21.38	5.50
		B-10	0.760	85.13	62.79	93.33	37.21	18.14	21.75	5.13
		B-15	0.777	85.13	64.65	92.65	35.35	20.00	22.75	4.13
7	Inclusion of easily verifiable variables only *(See Table 3.7.1)*	B-5	0.708	82.38	56.28	91.97	43.72	21.86	21.00	5.88
		B-10	0.755	83.63	61.86	91.62	38.14	22.79	22.75	4.13
		B-15	0.773	84.63	63.72	92.31	36.28	20.93	22.75	4.13
8	Like Model 7 plus strong subjective and expenditure regressors *(See Table 3.8.1)*	B-5	0.725	83.38	57.67	92.82	42.33	19.53	20.75	6.13
		B-10	0.773	84.88	64.19	92.48	35.81	20.47	22.75	4.13
		B-15	0.792	85.38	65.12	92.82	34.88	19.53	22.75	4.13
9	Inclusion of LSMS-type regressors only *(See Table 3.9.1)*	B-5	0.758	83.38	61.40	91.45	38.60	23.26	22.75	4.13
		B-10	0.786	85.13	64.19	92.82	35.81	19.53	22.50	4.38
		B-15	0.799	85.50	66.98	92.31	33.02	20.93	23.63	3.25
10	Two-step model based on Model 1 (BEST 15) *(See Table 4.1.3.1)*	Vers. C	0.834 and 0.679	90.25	85.58	91.96	14.41	21.86	28.87	-2.00

Note: B-5, B-10 and B-15 refer to the model with the set of best 5, 10 or 15 regressors, respectively. The predicted headcount index is computed on the basis of the model's predicted expenditures. If predicted expenditures are above the poverty line, the household is rated as not very poor (NVP) and otherwise as very poor (VP). The poverty assessment criteria PAC is defined as the difference between observed and the predicted poverty headcount index (the latter two being measured in percent). A perfect model for assessing the poverty outreach of a program would have a difference of zero percentage points, negative values of PAC indicate an overestimation of the headcount index whereas positive values indicate an underestimation.

Annex F: Variables included in the BEST 15 models

Table F. Best 15 regressors identified by MAXR for each of the nine models

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Share of food expenditures from total household expenditures	X								
Annualized total household expenditures	X								
Total value of household assets	X	X							X
Household has electricity	X			X	X	X	X		X
Days in past 7 days with main meal consisting of plain rice only	X	X	X	X					
Number of cars owned by the household	X	X	X	X	X	X			X
Number of steps above step identified as international poverty line	X	X	X						
Wood as exterior-walls' material	X	X	X	X	X	X			X
Distance to department capital	X								
Average daily per-capita clothing expenditures	X	X	X				X	X	X
Days in past 7 days with main meal consisting of plain rice and any vegetables	X			X				X	
Household ate less food from what they wanted for more than 10 days, but less than 30 days, during past 12 months	X								
Value of remittances sent to relatives in last 12 months	X								
Value of motor tillers owned by the household	X								
Community access to subsidized food ("glass of milk – vaso de leche") in past 24 months	X	X	X		X		X		
Median education level of adult household members		X	X	X	X	X			X
Household monthly expenditure on utilities (electricity, phone, water, etc)		X	X					X	
Number of rooms in the dwelling		X	X		X		X	X	X
Sum of distances to department, provincial and district capitals		X	X	X	X	X	X	X	

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Household feels that their health care expenses are below need		X	X						
Value of metal pots owned by the household		X	X	X	X				X
Household declares not to be able to save anything		X	X		X	X			
Value of debt owed by other households to the household		X	X						
Value of tractors owned by the household			X	X	X				
Household rates itself above national poverty line				X				X	
Household feels that their housing expenses are below need				X				X	
Availability of telephone (fixed land line) in the house				X	X	X	X	X	X
Household owns microwave				X			X	X	X
Number of beds owned by the household				X					
Amount that household needs per month to live				X				X	
Value of food processing assets					X				
Number of days on past seven days consuming any of six superior food items					X	X			
Ratio male adult household members/ female adult household members					X				
Total household members participating in water/ waste group					X	X			
Number of color TV's owned by the household						X			
Rooms per person						X			
Household received in-kind services from food aid programs in last 3 years						X			
Household owns a tractor						X	X	X	
Security key lock or metal frame with padlock in main entrance door						X			
Number of metal pots owned by the household						X			

Variable label	M1	M2	M3	M4	M5	M6	M7	M8	M9
Remittances sent							X	X	X
Household owns a color TV							X		
Household owns a suit							X		
Percent of adult household members who can read and write							X		
Household owns a Motocab							X	X	
No lock on entrance door or wood/ metal bar to close from inside							X	X	
Household head sleeps on something else than bed							X		
Household always ate enough from what they wanted (past 12 months)								X	
Collected wood/ sawdust/ bamboo as cooking fuel									X
Household head is single									X
Household owns sheep/ goats									X
Number of horses owned by the household									X

Annex G : Accuracy results of two-step models, by expenditure decile

Note: Decile of lpbench indicates the deciles for the observed per-capita daily expenditures

Combined accuracy , model version A

Decile of lpbench	Accuracy	Accuracy among VP	Accuracy among NVP	Undercoverage
1	85.00	85.00		15.00
2	82.50	82.69	82.14	17.31
3	80.00	75.00	85.00	25.00
4	76.25	52.00	87.27	48.00
5	87.50	68.75	92.19	31.25
6	85.00	100.00	84.62	
7	92.50		92.50	
8	100.00		100.00	
9	100.00		100.00	
10	100.00		100.00	
Rounded Average	88.88	77.67	92.99	22.33

Combined accuracy , model version B

Decile	Accuracy	Accuracy among VP	Accuracy among NVP	Undercoverage
1	86.25	86.25		13.75
2	82.50	82.69	82.14	17.31
3	83.75	85.00	82.50	15.00
4	72.50	48.00	83.64	52.00
5	83.75	68.75	87.50	31.25
6	80.00	100.00	79.49	0
7	97.50		97.50	
8	100.00		100.00	
9	100.00		100.00	
10	100.00		100.00	
Rounded Average	88.63	79.53	91.97	20.47

Combined accuracy , model version C

Decile	Accuracy	Accuracy among VP	Accuracy among NVP	Undercoverage
1	87.50	87.50		12.50
2	87.50	88.46	85.71	11.54
3	87.50	92.50	82.50	7.50
4	76.25	64.00	81.82	36.00
5	78.75	87.50	76.56	12.50
6	87.50	50.00	88.46	50.00
7	97.50		97.50	
8	100.00		100.00	
9	100.00		100.00	
10	100.00		100.00	
Rounded Average	90.25	85.58	91.97	14.42

Combined accuracy , model version D

Decile	Accuracy	Accuracy among VP	Accuracy among NVP	Undercoverage
1	90.00	90.00		10.00
2	78.75	82.69	71.43	17.31
3	85.00	90.00	80.00	10.00
4	77.50	76.00	78.18	24.00
5	77.50	43.75	85.94	56.25
6	87.50	50.00	88.46	50.00
7	97.50		97.50	
8	100.00		100.00	
9	100.00		100.00	
10	100.00		100.00	
Rounded Average	89.38	82.79	91.79	17.21

Endnotes

[1] 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. 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.

[2] 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

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

[4] 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.

[5] 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.

[6] 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).

[7] 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.

[8] 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.

[9] 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.”

[10] 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).

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

[12] These variables were identified by the SAS-MAXR procedure as the strongest variables among all subjective variables which were excluded in Model 5.
[13] We will further investigate this result in the accuracy tests for Kazakhstan and Uganda.
[14] We tested Model 1-BEST5 of the Bangladesh report, and obtained similarly promising results with the two-step approach.