Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for Colombia
Submitted: June 22, 2007
Revised: June 27, 2007
Revised: September 5, 2012 (formatting only)

1. Overall approach to tool development

The approach used to develop the poverty assessment tool for Colombia is built on the lessons learned and methods refined during the original USAID/IRIS project, Developing Poverty Assessment Tools (September 2003 to October 2006). In the initial phase of the project, the IRIS Center analyzed data from existing national household surveys and from surveys it conducted itself in countries where survey data were not available. The aim was to identify household indicators most closely associated with a household being "very poor" in terms of per-capita expenditures or income. IRIS used statistical methods to identify the 15 indicators that most closely track the per-capita expenditures or income of each household, as revealed by the household survey data. In addition, IRIS compared the performance of 8 different statistical methods in quantifying the statistical links between these 15 indicators and household expenditures/income; the accuracy of each method was assessed using criteria developed especially for this project. In this manner, IRIS identified the best-performing set of indicators (with associated weights) and statistical method for identifying the poverty status of households in each country. Statistical testing for accuracy was carried out for twelve countries in total.

In addition, the indicators that appeared among the "best 15" in at least one of the twelve countries were included in the next part of the project: testing for practicality. USAID selected seventeen microenterprise organizations to conduct field tests of practicality. Each question was rated as to whether the respondent found it to be sensitive, difficult, or that it was perceived that she falsified her answer. The lessons learned from the practicality testing were used to remove impractical indicators from consideration for the final poverty assessment tools.

The end result of this development process was a country-specific poverty assessment tool that estimates—rather than directly measures—household per capita consumption expenditure or income based on a short set of practical indicators. Each country tool is incorporated into a data entry template that allows a microenterprise practitioner to easily enter and store the responses of its sampled clients to indicator questions and will also estimate the percentage of that practitioner's client households who are very poor.

In October 2006, USAID contracted the IRIS Center to build on the statistical methods and practicality information generated during the original project to develop poverty assessment tools for use by microenterprise practitioners in additional countries such as Colombia. As part of this new phase, IRIS will explore the use of existing household survey data sets beyond specifically the LSMS to develop new poverty assessment tools, refine the tool development and testing methodology, and strive to make the tools even simpler and easier to implement. IRIS will also provide a Help Desk to assist practitioners with the implementation of approved USAID poverty assessment tools.

2. Data source

For Colombia, existing data from the 2003 Encuesta de Calidad de Vida (ECV) household survey conducted by the national statistical office, Departamento Administrativo Nacional de Estadística (DANE), was used to construct the poverty assessment tool. The full sample of 22949 households is nationally representative, although residents of the city of Neiva were not sampled. Construction of the tool was done on half of the full sample (or 11475 households) to allow for potential testing of out-of-sample accuracy results on the remaining 11474 households. The benchmark determination of a household's poverty status is based on household per capita income, as suitable data on household consumption expenditures is not available.

3. Process used to select included indicators

Suitable household surveys, such as the LSMS, typically include variables related to education, housing characteristics, consumer durables, agricultural assets, illness and disability, and employment. For Colombia, more than 150 indicators from all categories, save for agricultural assets, were considered. This pool of indicators was screened to remove indicators that might be sensitive, difficult, or susceptible to falsification, based on information collected during the practicality testing from the original project.

The MAXR procedure in SAS was used to select the best poverty indicators (for variables found to be practical) from the pool of potential indicators in an automated manner. MAXR is commonly used to narrow a large pool of possible indicators into a more limited, yet statistically powerful set of indicators. The MAXR technique seeks to maximize explained variance (i.e., R^2) by adding one variable at a time (per step) to the regression model, and then considering all combinations among pairs of regressors to move from one step to the next. Thus, the MAXR technique allows us to identify the best model containing 15 variables (not including control variables for household size, age of the household head, and location).

The MAXR procedure yielded the best 15 variables for the OLS model (also used for the Quantile model) and another set of best 15 variables for the Linear Probability model (also used for the Probit model). The final set of indicators and their weights, therefore depended on selecting one of these four statistical models—OLS, Quantile, Linear Probability, or Probit—as the best model.[1] This selection of the best model was based on the Balanced Poverty Accuracy Criterion (BPAC) and Poverty Incidence Error (PIE) accuracy criteria, along with practicality considerations.[2]

[1] The set of indicators and their weights also depended on the selection of a 1-step or 2-step statistical model.
[2] For a detailed discussion of these accuracy criteria, see "Note on Assessment and Improvement of Tool Accuracy" at www.povertytools.org

4. Estimation methods used to identify final indicators and their weights/coefficients.

The results obtained with each of the eight estimation methods are summarized in Table 1. For Colombia, the most accurate method was 2-stage quantile regression, on the basis of both PIE and BPAC. However, the 1-step Quantile is only slightly less accurate (1 percentage point lower BPAC) and requires 15 rather than 30 indicators. After a consultation with USAID, the 1-step Quantile was selected as the best model, taking into consideration both accuracy and practicality.

Table 1: Accuracy Results Obtained with Different Estimation Methods

Colombia (median) **Poverty Line:** varies by location **Poverty Rate:** 25.7%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	79.82	52.05	47.95	30.67	-4.44	34.76
Quantile regression (estimation point: 39)	**78.87**	**59.41**	**40.59**	**41.63**	**0.27**	**58.37**
Linear Probability	82.26	49.49	50.51	18.59	-8.19	17.57
Probit	82.47	52.59	47.41	20.86	-6.82	26.05
Two-step methods						
OLS – 48 percentile cutoff	80.23	53.75	46.25	30.77	-3.97	38.26
Quantile (estimation points: 39, 20) 48 percentile cutoff	79.60	60.05	39.95	39.43	-0.13	59.54
LP – 37 percentile cutoff	83.58	61.27	38.73	25.24	-3.46	47.78
Probit – 37 percentile cutoff	83.34	61.09	38.91	25.96	-3.33	48.14

5. How coefficients and weights are used to estimate poverty status or household expenditures

For the 1-stage quantile regression method, the estimated regression coefficients directly indicate the weight placed on each of the included indicators in estimating the income of each household in the sample. These estimated coefficients are shown in Table 4. In constructing the Poverty Assessment Tool for each country, these weights are inserted into the "backend" analysis program of the EPI template used to calculate the incidence of extreme poverty among each implementing organization's clients. In constructing the tool for Colombia, IRIS used Bogota as a proxy for the city of Neiva, which was not sampled in the ECV 2003.[3] While a skilled EPI user would be able to locate the model's weights, they would not be seen by the client or the interviewer during the normal course of interviewing, entering the data, or calculating the extreme poverty rate.

6. Decision rule used to classify households as very poor or not very-poor

[3] Bogota was deemed the closest match to Neiva, because of its geographical proximity and similar official poverty line.

The extreme poverty line in the project was the higher of the two potential poverty lines specified in the legislation: $1.08 a day (in PPP terms) OR the bottom (poorest) half of households living below the national poverty line. Two of statistical models (OLS and Quantile) used by the IRIS team predict the per capita consumption expenditures or per capita income for each household, which is then compared to the binding poverty line to decide whether the household is very poor.[4] The other two statistical models (Linear Probability and Probit) predict the probability that a household is very poor (according to the binding, absolute poverty line). If this probability exceeds 0.5, the household is predicted to be very poor.[5]

Colombia actually specifies fifteen poverty lines: one for each of thirteen major cities, one for other urban areas, and a line for rural areas. Examination of the ECV 2003 data reveals the income level of the households living below these poverty lines in each area, along with the income level that divides this group into poorer and less-poor halves – again on an area-specific basis. Among the households included in the ECV 2003 survey, 25.7% fall into the poorest half of those living under these area-specific poverty lines. This is far higher than the 2.7% of households living below the $1.08 a day line. As a result, **the binding poverty line for Colombia is the set of 14 area-specific median poverty lines**, shown in Table 2. The median poverty line for one major city, Neiva, could not be determined, because it was not sampled in the ECV 2003 survey. In order to permit the use of the Poverty Assessment Tool across Colombia, Neiva was matched to Bogota and assigned its extreme poverty line.

Table 2: Monthly Income of the "Very Poor" in Colombia, by Area

Area	Monthly income equivalent to the "median poverty line," below which the poorest 50% of those living under the poverty line live (pesos/ person/ month)	
	As of July 2003	*As of April 2007*
Medellin	148,583	181,562
Barranquilla	126,667	154,781
Bogota & Neiva	134,722	164,624
Cartagena	95,555	116,764
Manizales	149,122	182,220
Monteria	112,682	137,692
Villavicencio	159,889	195,377
Pasto	93,889	114,728
Cucuta	122,292	149,435
Pereira	152,658	186,541
Bucaramanga	165,417	202,131
Cali	137,167	167,611
Other Urban	109,000	133,193
Rural	75,533	92,298

[4] For a 2-step OLS or Quantile model, the decision rule in the 1st-step compares the expenditures predicted for each household to a certain expenditure cutoff.

[5] For a 2-step Linear Probability or Probit model, the decision rule in the 1st-step compares the predicted probability that the households' expenditures exceed a certain cutoff to the 0.5 value.

Because the selected tool is based on a Quantile model, each household whose estimated per capita income according to the tool falls below the median poverty line in its respective city or area is identified as "very poor," and those whose estimated per capita income exceeds the median poverty line is identified as "not very-poor."

Table 3 below compares the poverty status of the sample households as identified by the selected model, versus their true poverty status as revealed by the data from the ECV 2003 household survey. The upper-left and lower-right cells show the number of households correctly identified as "very poor" or "not very-poor," respectively. Meanwhile, the upper-right and lower-left cells show the number of households whose poverty status is incorrectly identified by the model.

Table 3: Poverty Status of Sample Households, as Estimated by the Model and as Revealed by the Benchmark Survey

	Number of households identified as very poor by the tool	**Number of households identified as not very-poor by the tool**
Number of "true" very poor households (as determined by benchmark survey)	1,752 (15.3%)	1,197 (10.4%)
Number of "true" not very-poor households (as determined by benchmark survey)	1,228 (10.7%)	7,299 (63.6%)

Table 4: Regression Output from 1-Stage Quantile Method

COLOMBIA 1-STEP QUANTILE: variables from MAXR/OLS 100 percentile model
Regression results, estimation point of 39 percentile

.39 Quantile regression — Number of obs = 11475
Min sum of deviations 5788.945 — Pseudo R2 = 0.3702

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	12.6697	0.0574	220.6	0.000	12.5571	12.7823
Household lives in Medellin	-0.1724	0.0266	-6.49	0.000	-0.2244	-0.1203
Household lives in Barranquilla	-0.0626	0.0395	-1.59	0.113	-0.1401	0.0148
Household lives in Cartagena	-0.2939	0.0523	-5.61	0.000	-0.3965	-0.1913
Household lives in Manizales	-0.2926	0.0536	-5.46	0.000	-0.3977	-0.1875
Household lives in Monteria	-0.1456	0.1172	-1.24	0.214	-0.3753	0.0841
Household lives in Villavicencio	0.0920	0.0781	1.18	0.239	-0.0611	0.2452
Household lives in Pasto	-0.0823	0.0408	-2.02	0.044	-0.1623	-0.0023
Household lives in Cucuta	-0.0813	0.0501	-1.62	0.105	-0.1794	0.0169
Household lives in Pereira	-0.2198	0.0470	-4.68	0.000	-0.3119	-0.1278
Household lives in Bucaramanga	0.0307	0.0429	0.72	0.474	-0.0533	0.1147
Household lives in Cali	-0.0340	0.0246	-1.38	0.167	-0.0821	0.0142
Household lives in other urban area	-0.2623	0.0144	-18.17	0.000	-0.2906	-0.2340
Household lives in rural area	-0.4351	0.0164	-26.48	0.000	-0.4673	-0.4029
Household size	-0.3723	0.0113	-32.9	0.000	-0.3944	-0.3501
Household size squared	0.0200	0.0011	17.57	0.000	0.0178	0.0222
Household head age	0.0041	0.0022	1.92	0.055	-0.0001	0.0084
Household head age squared	-0.0000	0.0000	-0.54	0.587	-0.0001	0.0000
Share of household members 6-16 who are currently attending school	-0.3927	0.0311	-12.62	0.000	-0.4537	-0.3318
Household head has secondary school education	0.2362	0.0146	16.2	0.000	0.2076	0.2648
Household head has technical education	0.2849	0.0299	9.53	0.000	0.2263	0.3435
Household head has university education	0.6418	0.0230	27.89	0.000	0.5967	0.6869
Share of household members (excluding household head) who have a university education	0.6893	0.0496	13.9	0.000	0.5921	0.7865
Number of rooms in dwelling	0.0977	0.0048	20.47	0.000	0.0883	0.1070
Dwelling is rented	-0.1200	0.0139	-8.62	0.000	-0.1473	-0.0928
Household head's time was mostly spent on actively looking for a job	-0.3886	0.0302	-12.88	0.000	-0.4477	-0.3294
Household head's time was mostly spent on household chores	-0.1661	0.0164	-10.11	0.000	-0.1983	-0.1339
Household owns one or more cellular phones	0.2090	0.0174	12.01	0.000	0.1749	0.2431
Household owns one or more VCR	0.2126	0.0181	11.72	0.000	0.1771	0.2482
Household owns one or more motorcycles	0.2417	0.0218	11.08	0.000	0.1990	0.2845
Household owns one or more cars	0.2818	0.0208	13.58	0.000	0.2411	0.3224
Household owns one or more refrigerators	0.2683	0.0148	18.11	0.000	0.2393	0.2974
Household owns one or more washing machines	0.2161	0.0171	12.66	0.000	0.1827	0.2496