Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for Bosnia and Herzegovina
Submitted: June 29, 2008

The following report is divided into six sections. Section 1 provides a general overview of the tool development process. Section 2 describes the data set used to create Poverty Assessment Tool for Bosnia and Herzegovina. Section 3 details the set of statistical procedures used for selecting indicators and for estimating household expenditure or, for some models, the probability that a household is very poor. Section 4 reports on the in-sample accuracy of each prediction model considered. Sections 5 and 6 explain how regression coefficients are used in poverty prediction and how these predictions are used to classify households into the "very poor" and "not very poor" categories.

Annex 1 to this report provides accuracy results for an additional poverty line beyond that required by the Congressional legislation. Annex 2 supplies a careful consideration of out-of-sample accuracy for Serbia Poverty Assessment Tool.

1. Overall approach to the tool development

The approach used to develop the poverty assessment tool for Bosnia and Herzegovina 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) or the probability that a household is very poor 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 Bosnia and Herzegovina. 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 Bosnia and Herzegovina, existing data from the 2004 LSMS were used to construct the poverty assessment tool. The full sample of 2,836 households is nationally representative. The sample used for tool construction is comprised of a randomly selected 1,418 households (50 percent of the full sample). The remainder, another randomly selected 1,418 households, is reserved for out-of-sample accuracy testing, which investigates the robustness of in-sample poverty estimation.

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 Bosnia and Herzegovina, more than 85 indicators from all categories were considered.

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 Balance Poverty Accuracy Criterion (BPAC) and the Poverty Incidence Error (PIE), along with practicality considerations.[2]

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

As explained more fully in Section 6, the line used to construct the poverty tool for Bosnia and Herzegovina is the "median poverty line" – the level of monthly expenditure that divides the poorest half of those living below the national poverty line from the less-poor half of the officially poor. Table 1 summarizes the accuracy results achieved by each of the eight estimation methods in predicting household poverty relative to this poverty line. For Bosnia and Herzegovina, the most accurate method, on the basis of BPAC, is the 2-step Quantile regression. However, the 1-step Quantile regression is slightly less accurate (roughly five percentage points lower in BPAC) and requires only 15 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: In-sample Accuracy Results for Prediction at the Legislative Poverty Line

BOSNIA AND HERZEGOVINA Median line* Share of "very poor": 7.0%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	92.08	16.54	83.46	21.59	-4.66	-45.32
Quantile regression (estimation point: 34)	**90.91**	**38.76**	**61.24**	**60.89**	**-0.03**	**38.42**
Linear Probability	92.55	1.11	98.89	0	-7.45	-97.78
Probit	93.19	23.17	76.83	13.52	-4.77	-40.15
Two-step methods						
OLS – 20 percentile cutoff	93.58	32.22	67.78	17.39	-3.79	-18.16
Quantile (estimation points: 34, 6) 20 percentile cutoff	91.51	46.45	53.55	60.55	0.52	39.45
LP – 24 percentile cutoff	94.57	40.03	59.97	12.09	-3.61	-7.85
Probit – 24 percentile cutoff	94.30	39.21	60.79	14.82	-3.46	-6.75
* Median poverty line is 1830 marka per capita per year in January 2005 prices. This poverty line is based on the official national poverty line of 2223 marka.[3]						

[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

[3] The poverty line reported above is expressed in the prevailing prices at the time of data collection in December 2004. The poverty line and other monetary variables used in the accuracy results, however,

For Bosnia and Herzegovina, the functionality of predicting the poverty rate at another poverty line—in this case, the national poverty line—has been added. When running the analysis routine with the Epi Info template, the user is presented the option to predict the extreme poverty rate (using the median line), the poverty rate (national line), or both. The methodology and the accuracy results for this prediction are discussed in Annex 1.

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

For the Quantile regression method, the estimated regression coefficients indicate the weight placed on each of the included indicators in estimating the household expenditures of each household in the sample. These estimated coefficients are shown in Table 3. In constructing the Poverty Assessment Tool for each country, these weights are inserted into the "back-end" analysis program of the EPI template used to calculate the incidence of extreme poverty among each implementing organization's clients. While a skilled EPI user would be able to locate the model's weights in the backend, 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 for classifying households as very poor and not very-poor

The legislation governing the development of USAID tools defines the "very poor" as either the bottom (poorest) 50 percent of those living below the poverty line established by the national government or those living on the local equivalent of less than $1/day (in Purchasing Power Parity, or PPP, terms). The applicable poverty line for USAID tool development is the one that yields the higher household poverty rate for a given country.

In Bosnia and Herzegovina the applicable threshold is the median poverty line, the household per capita expenditure value of the 50th percentile below the national poverty line of 2,223 markas per capita per month, at the level of prices prevailing in December 2004 when the household survey data were collected. At 1,830 markas per month, the median poverty line identifies 7.0% of households as "very poor". Alternatively, the $1/day poverty line of 37.43 marka yields a household poverty rate of 0.0%. Hence the decision rule for Bosnia's USAID poverty assessment tool in classifying the "very poor" (and the "not very-poor") is whether that predicted per capita monthly expenditures of a household fall below (or above) the median poverty line.

Because the selected tool is based on a Quantile model, each household whose estimated per capita consumption expenditures according to the tool fall below the median poverty line is identified as "very poor," and each household whose estimated per capita consumption expenditures exceeds the median poverty line is identified as "not very-poor."

---

are inflation-adjusted to the most current time period possible (April 2008), yielding an extreme poverty line of 2,102 marka per capita per month.

An additional requirement for using the median poverty line is that the national poverty line on which it depends is actively used by the local government. This is indeed the case in Bosnia and Herzegovina, where the government uses the national poverty line for poverty monitoring. [can we footnote this?]

Table 2 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 benchmark household survey (in-sample test). 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 indicate the twin errors possible in poverty assessment: misclassifying very poor households as not very-poor; and the opposite, misclassifying not very-poor households as very poor.

Table 2: Poverty Status of Sample Households, as Estimated by Model and 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)	37 (2.7%)	67 (4.8%)
Number of "true" not very-poor households (as determined by benchmark survey)	56 (4.1%)	1,222 (88.4%)

Table 3: Regression Estimates using 1-step Quantile Method for Prediction at the Median Poverty Line

BOSNIA 1 STEP MAXR/QUANT: variables from MAXR/OLS 100 percentile model
Regression results, estimation point of 34 percentile
(sum of wgt is 1.4180e+03)

.34 Quantile regression — Number of obs = 1382
Raw sum of deviations 500.8597 (about 8.1226292)
Min sum of deviations 344.739 — Pseudo R2 = 0.3117

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	7.578	0.182	41.590	0.000	7.220	7.935
Household head age	0.026	0.006	4.060	0.000	0.013	0.038
Household head age squared	0.000	0.000	-3.140	0.002	0.000	0.000
Household size	-0.446	0.047	-9.510	0.000	-0.538	-0.354
Household size squared	0.039	0.007	5.400	0.000	0.025	0.053
Household lives in Republika Srpska	0.013	0.026	0.500	0.619	-0.039	0.065
Household's dwelling has balcony or terrace	0.100	0.031	3.250	0.001	0.040	0.160
Household's dwelling has an attic	0.056	0.027	2.080	0.038	0.003	0.109
Main source of heating is a utility/boiler house	0.205	0.040	5.160	0.000	0.127	0.284
Main source of heating is central heating	0.243	0.058	4.200	0.000	0.130	0.357
Household does not have access to a telephone	-0.126	0.035	-3.620	0.000	-0.194	-0.058
Household head is female	0.120	0.035	3.480	0.001	0.052	0.188
Household head is single	0.110	0.068	1.620	0.105	-0.023	0.242
Highest diploma obtained by household head is university degree	0.149	0.058	2.570	0.010	0.035	0.262
Household owns one or more washing machines	0.081	0.036	2.220	0.026	0.010	0.152
Household owns one or more microwaves	0.118	0.053	2.250	0.025	0.015	0.222
Household owns one or more vacuums	0.155	0.037	4.190	0.000	0.083	0.228
Household owns one or more televisions	0.199	0.059	3.370	0.001	0.083	0.315
Household owns one or more stereos	0.096	0.031	3.110	0.002	0.035	0.157

Household owns one or more automobiles	0.212	0.031	6.870	0.000	0.152	0.273
Household owns one or more calves	0.091	0.043	2.140	0.033	0.008	0.174

Annex 1: Poverty Prediction at the National Poverty Line

Strictly construed, the legislation behind the USAID poverty assessment tools concerns "very poor" and "not very-poor" beneficiaries. Nevertheless, the intended outcome of the legislation is to provide USAID and its implementing partners with poverty measurement tools that they will find useful.

After discussions among USAID, IRIS, and other members of the microenterprise community, a consensus emerged that the tools would benefit from predictive capacity beyond legislatively-defined extreme poverty. To that end, on agreement with USAID, IRIS has used the best indicators and regression type for predicting the "very poor" to also identify the "poor." For $1/day PPP models, this will be the $2/day PPP; for median poverty models, the "poor" threshold will be the national poverty line. Following this logic, then, the "poor" ("not poor") in Bosnia and Herzegovina are defined as those whose predicted incomes fall below (above) the national poverty line.

Table 4 summarizes the predictive accuracy results for the national poverty line using the Quantile model specification from the median poverty line. The indicators are the same as those in the model for the median line, but the percentile of estimation and the coefficients of the model were allowed to change (compare Tables 3 and 6). This methodology allows the content and length of the questionnaire to remain the same, but permits greater accuracy in predicting at the national poverty line.

Table 4: Accuracy Results Obtained for Prediction at the National Poverty Line

Bosnia and Herzegovina National Line Share of Poor: 13.9%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
Quantile regression (estimation point:)	85.43	51.28	48.72	46.31	-0.37	48.87

Table 5 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 benchmark household survey (in-sample test). The upper-left and lower-right cells show the number of households correctly identified as "poor" or "not poor," respectively. Meanwhile, the upper-right and lower-left cells indicate the twin errors possible in poverty assessment: misclassifying poor households as not poor; and the opposite, misclassifying not poor households as poor.

Table 5: Poverty Status of Sample Households, as Estimated by Model and Revealed by the Benchmark Survey, at National Poverty Line

	Number of households identified as poor by the tool	Number of households identified as not poor by the tool
Number of "true" poor households (as determined by benchmark survey)	94 (6.8%)	99 (7.2%)
Number of "true" not poor households (as determined by benchmark survey)	100 (7.2%)	1,089 (78.8%)

Table 6: Regression Estimates using 1-step Quantile Method for Prediction at the National Poverty Line

BOSNIA 1 STEP MAXR/QUANT: variables from MAXR/OLS 100 percentile model
Regression results, estimation point of 38 percentile
(sum of wgt is 1.4180e+03)

.38 Quantile regression — Number of obs = 1382
Raw sum of deviations 522.3676 (about 8.1712456)
Min sum of deviations 359.2463 — Pseudo R2 = 0.3123

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	7.615	0.160	47.700	0.000	7.302	7.928
Household head age	0.027	0.006	4.740	0.000	0.016	0.038
Household head age squared	-0.450	0.046	-9.890	0.000	-0.540	-0.361
Household size	0.000	0.000	-3.860	0.000	0.000	0.000
Household size squared	0.038	0.007	5.450	0.000	0.025	0.052
Household lives in Republika Srpska	0.029	0.025	1.180	0.236	-0.019	0.078
Household's dwelling has balcony or terrace	0.086	0.028	3.060	0.002	0.031	0.141
Household's dwelling has an attic	0.070	0.025	2.790	0.005	0.021	0.119
Main source of heating is a utility/boiler house	0.202	0.037	5.550	0.000	0.131	0.274
Main source of heating is separate central heating	0.250	0.052	4.780	0.000	0.148	0.353
Household does not have access to a telephone	-0.127	0.032	-4.040	0.000	-0.189	-0.066
Household head is female	0.095	0.032	2.980	0.003	0.032	0.157
Household head is single	0.075	0.067	1.120	0.265	-0.057	0.206
Highest diploma obtained by household head is university degree	0.119	0.053	2.240	0.025	0.015	0.223
Household owns one or more washing machines	0.072	0.033	2.200	0.028	0.008	0.136
Household owns one or more microwaves	0.111	0.048	2.310	0.021	0.017	0.205
Household owns one or more vacuums	0.160	0.034	4.730	0.000	0.093	0.226
Household owns one or more televisions	0.197	0.053	3.740	0.000	0.094	0.300
Household owns one or more stereos	0.106	0.029	3.690	0.000	0.049	0.162

Household owns one or more automobiles	0.208	0.028	7.330	0.000	0.152	0.263
Household owns one or more calves	0.069	0.041	1.710	0.087	-0.010	0.149

Annex 2: Out-of-Sample Accuracy Tests

In statistics, prediction accuracy can be measured in two fundamental ways: with in-sample methods and with out-of-sample methods. In the in-sample method, a single data set is used. This single data set supplies the basis for both model calibration and for the measurement of model accuracy. In the out-of-sample method, at least two data sets are utilized. The first data set is used to calibrate the predictive model. The second data set tests the accuracy of these calibrations in predicting values for previously unobserved cases.

The previous sections of this report provide accuracy results of the first type only. The following section presents accuracy findings of the second type, as both a supplement to certification requirements and as an exploration of the robustness of the best model outside of the 'laboratory' setting.

As noted in section 2, the data set used to construct the Bosnia and Herzegovina tool was divided randomly into two data sets of equal size (1,418 and 1,418 households). A naïve method for testing out-of-sample accuracy—or for overfitting—is to simply apply the model calibrated on the first data set to the observations contained in the holdout data set. These results are show in Table 7. The best model (1-step Quantile) does not perform well in terms of BPAC, losing nearly 30 percentage points, but does perform well in predicting the overall extreme poverty rate, with PIE diminishing by roughly a percentage point.

Table 7: Comparison of In-Sample and Out-of-Sample Accuracy Results

	Total Accuracy	Poverty Accuracy	Under-coverage	Leakage	PIE	BPAC
In-Sample Prediction						
	90.91	38.76	61.24	60.89	-0.03	38.42
Out-of-Sample Prediction						
	91.72	26.99	73.00	54.58	-1.19	8.57

Another, more rigorous method for testing the out-of-sample accuracy performance of the tool is to provide confidence intervals for the accuracy measures, derived from 1,000 bootstrapped samples from the holdout sample.[4] Each bootstrapped sample is constructed by drawing observations, with replacement, from the holdout sample. The calibrated model is then applied to each sample to yield poverty predictions; across 1,000 samples, this method provides the sampling distributions for the model's accuracy measures.

[4] This method of out-of-sample testing is used by Mark Schreiner for the PPI scorecards as detailed on www.microfinance.com

Table 8 presents the out-of-sample, bootstrapped confidence intervals for the 1-step Quantile model. The lower bound is defined by the 2.5th percentile of the sample distribution for each accuracy measure; the upper bound is defined by the 97.5th percentile (note that this method does not assume normality).

As with the naïve result in Table 7, the performance of this model is mixed when judged using an out-of-sample approach. The 95% confidence interval is wide for BPAC. On the other hand, the confidence interval is rather narrow for PIE.

Table 8: Bootstrapped Confidence Intervals Computed Empirically from Sampling Distribution Without Normality Assumption

Accuracy Measure	95% Confidence Interval	
	LB	UB
Total Accuracy	89.31	92.83
Poverty Accuracy	24.31	46.19
Undercoverage	53.81	75.69
Leakage	50.49	102.07
PIE	-1.39	2.25
BPAC	-2.49	37.85

One potential explanation for the out-of-sample BPAC performance for Bosnia is the low rate of extreme poverty in that country; only 7.0% of the households can be classified as very poor. Predicting any phenomenon that occurs so infrequently in a population is difficult; errors of misclassification—undercoverage and leakage—can be quite pronounced. It is perhaps also worth noting that the other three regression models (OLS, Linear Probability, and Probit) perform very poorly, even in-sample, in Bosnia.

However, the primary purpose of the PAT is to assess the overall extreme poverty rate across a group of households. The out-of-sample result for PIE in Table 8 indicates that the extreme poverty rate estimated by the Bosnia PAT will fall within 2.25 percentage points of the true value in the population (with 95-percent confidence). By this measure, the predictive model is fairly accurate.