Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for Liberia
Submitted: July 20, 2010

The following report is divided into five sections. Section 1 describes the data used to create the Poverty Assessment Tool for Liberia. Section 2 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 3 reports on the in-sample accuracy of each prediction model considered. Sections 4 and 5 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.

1. Data source

For Liberia, USAID funded, and IRIS implemented, an original data collection exercise in October through December, 2008. This was undertaken after a close examination of the sole available household integrated living standards survey (a CWIQ survey) and the determination that it had significant shortcomings as a data source. After intensive training and a round of pilot-testing, a local survey firm, Subah-Belleh Associates collected expenditure and other living standards information from a nationally-representative sample of 784 households. Given this sample size, the data set was not divided into calibration and holdout samples.

2. 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, and employment. For Liberia, nearly 160 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 the 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]

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

As explained more fully in Section 5, the line used to construct the poverty tool for Liberia is the $1.25/day line. Table 1 summarizes the accuracy results achieved by each of the eight estimation methods in predicting household poverty relative to this poverty line. For Liberia, the most accurate method, on the basis of BPAC, is the 1-step Quantile regression. Following precedent from previous decisions made in 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

LIBERIA $1.25/day line* Share of "very poor": 52.2%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	81.89	88.94	11.06	23.83	6.63	76.17
Quantile regression (estimation point: 56 percentile)	**84.82**	**85.75**	**14.25**	**14.99**	**0.38**	**85.01**
Linear Probability	82.78	86.73	13.27	19.90	3.44	80.10
Probit	83.16	85.75	14.25	18.18	2.04	81.82
Two-step methods						
OLS – 56 percentile cutoff	82.27	85.50	14.50	19.66	2.68	80.34
Quantile (estimation points: 56, 32) 56 percentile cutoff	83.55	84.28	15.72	15.97	0.13	84.03
LP – 56 percentile cutoff	84.06	85.01	14.99	15.73	0.38	84.27
Probit – 56 percentile cutoff	84.69	83.05	16.96	12.54	-2.30	78.62
* $1.25/day poverty line is 1,752 Liberian dollars in 2008 prices.						

For Liberia, the functionality of predicting the poverty rate at other poverty lines—in this case, the $0.75/day, $1.00/day, $2.00/day, and $2.50/day —has been added. This functionality is based on statistical models for prediction at the $1.25/day and $2.50/day lines. The methodology and the accuracy results for this prediction are discussed in Annex 1.

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

[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.

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 CSPro template used to calculate the incidence of extreme poverty among each implementing organization's clients.

5. 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 the international poverty line ($1.25/day in 2005 PPP terms)[3]. The applicable poverty line for USAID tool development is the one that yields the higher household poverty rate for a given country.

In Liberia the applicable threshold is the international poverty line of $1.25/day, at the level of prices prevailing when the household survey data were collected. The value of the line in those prices is 1752.386 Liberian dollars per capita per month.[4] At these values, the $1.25/day poverty line identifies 52.2% of households as "very poor."

Alternatively, the national poverty line of 2,090 Liberian dollars per adult equivalent per month in rural areas and 2,949 in urban areas identifies 57.7% of households as poor and therefore 28.7% as "very poor."

External poverty rate estimates for Liberia are based on the only other recent household living standards survey in Liberia, a CWIQ survey conducted in 2007. Using this data set, various government and multi-lateral institutions report a (population-weighted) national poverty rate of 63.8%. Using population weights to more closely match the methodology of these studies, we find a slightly higher rate of 65.8% for the national poverty. For the PPP line, however, Povcalnet reports an extreme poverty rate of 83.7%

[3] The congressional legislation specifies the international poverty line as the "equivalent of $1 per day (as calculated using the purchasing power parity (PPP) exchange rate method)." USAID and IRIS interpret this to mean the international poverty line used by the World Bank to track global progress toward the Millennium Development Goal of cutting the prevalence of extreme poverty in half by 2015. This poverty line has recently been recalculated by the Bank to accompany new, improved estimates of PPP. The applicable 2005 PPP rate for Liberia is 0.5111, which is expressed in $US dollars (local currency) to the international dollar. The Liberian economy is to a certain extent dollarized (especially for large purchases and in urban areas) but both the CWIQ and our survey collected information in Liberian dollars.

[4] The calculation for the $1.25/day poverty line is (1.25)*(365/12)*(0.511149561746522)*64*(140.89/100) where 64 is the exchange rate between the Liberian and US dollars at the time of the survey. The final term is the CPI adjustment from average 2005 prices to prices at the time of the survey, obtained from the Central Bank of Liberia. We cannot find an online source that cites the PPP line used in the cited figure of more than 80 percent extreme poverty.

for the $1.25/day line, a rate which is substantially higher than for our data (56.1% with population weights).

There are a number of reasons that may explain the difference in observed poverty rates using the $1.25/day line. As stated earlier, there are significant differences in the methodology of the IRIS survey that would result in higher, but more accurate values for household expenditures. These include:

- **Survey administration**. The USAID/IRIS survey was done in two sessions whereas the lengthy CWIQ was done in one single session. Lengthy interviews could lead to under-reporting of expenditures. Also, recall questions are difficult without fixed references in time. The IRIS creates those references in time by two visits. This approach helps recall and, relative to "cold recall", will increase reporting (and be closer to the truth).
- **Questionnaire design**. The USAID/IRIS expenditure questions were drafted such than expenditures were put in the context of the outlay in order to facilitate recall.
    - Health and expenditure questions are asked in three contexts:
        - About particular people (i.e. asked at the individual level)
        - With reference to a particular event (e.g. consultation fees from a doctor's visit)
        - Within the context of other questions on the same topic
        - In contrast, the CWIQ asks expenditure questions as lump sums that do not reference particular people, events, or contexts. Plus, these questions come at the end of a long survey.
    - Food expenditures, the single most important component of consumption, is not handled well in the CWIQ.
        - The CWIQ divides expenditure and auto-consumption into two separate rounds of questions.
        - The USAID/IRIS survey combines them and gets information about how each food item is consumed and how much
        - In practice, the CWIQ approach will lead to under-reporting. First, people may get confused when asked about rice again after talking about many other food items in the interim. Respondents may fail to report auto-consumption, for example, because they believe the interviewer has made a mistake in the survey or has already gotten the information he wants. Second, people have the opportunity to say "no" to something they know from experience will give rise to many more questions. They may under-report in order to end a long interview. The IRIS approach avoids this by asking about "consumption" of all food items first and then seeking detailed information for each of them.

Another likely factor is that the CWIQ survey took place during a relative lean period (the rainy season) in the seasonal calendar. In comparison, USAID/IRIS Center survey took place during the traditional harvest season. In addition, economic growth in Liberia was high between the CWIQ survey and the USAID/IRIS survey.

Hence the decision rule for Liberia's USAID poverty assessment tool in classifying the "very poor" (and the "not very-poor") is whether that predicted per capita daily expenditures of a household fall below (or above) the $1.25/day poverty line.

Because the selected tool is based on a Quantile model, each household whose estimated per capita consumption expenditures according to the tool is less than or equal to the $1.25/day poverty line is identified as "very poor," and each household whose estimated per capita consumption expenditures exceeds the $1.25/day poverty line is identified as "not very-poor."

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)	349 (44.5%)	61 (7.8%)
Number of "true" not very-poor households (as determined by benchmark survey)	58 (7.4%)	316 (40.3%)

Table 3: Regression Estimates using 1-step Quantile Method for Prediction at the $1.25/day Poverty Line

.56 Quantile regression — Number of obs = 784
Min sum of deviations 207.7121 — Pseudo R2 = 0.5180

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	7.8034	0.1013	77.0600	0.0000	7.6046	8.0022
HH lives in Bomi, Gbarpolu, or Grand Cape Mount	0.4180	0.0420	9.9400	0.0000	0.3354	0.5005
HH lives in rural Montserrado, Grand Bassa, or Margibi	0.3178	0.0350	9.0700	0.0000	0.2491	0.3866
HH lives in Grand Gedeh, River Cess, or Sinoe	-0.0093	0.0431	-0.2200	0.8290	-0.0938	0.0752
HH lives in Grand Kru, Maryland, or River Gee	0.1248	0.0461	2.7100	0.0070	0.0343	0.2152
HH lives in urban Montserrado	0.4240	0.0419	10.1300	0.0000	0.3418	0.5062
HH lives in rural area	-0.0736	0.0350	-2.1100	0.0360	-0.1423	-0.0050
Household head age	0.0169	0.0039	4.3000	0.0000	0.0092	0.0246
Household head age squared	-0.0002	0.0000	-4.2500	0.0000	-0.0002	-0.0001
Household size	-0.2325	0.0160	-14.4900	0.0000	-0.2640	-0.2010
Household size squared	0.0083	0.0011	7.7800	0.0000	0.0062	0.0104
HH head is divorced, separated, or widowed	-0.1299	0.0369	-3.5200	0.0000	-0.2024	-0.0575
Share of HH members that can write a one-page letter in English	0.1296	0.0666	1.9500	0.0520	-0.0010	0.2603
Share of HH members with high school certificate as highest diploma completed	0.3777	0.1831	2.0600	0.0390	0.0183	0.7370
Dwelling external walls made of sandcrete or cement blocks	0.1686	0.0346	4.8800	0.0000	0.1008	0.2365
HH toilet facility is flush toilet	0.2439	0.0408	5.9800	0.0000	0.1638	0.3239
HH lighting fuel is palm oil or firewood	-0.0972	0.0288	-3.3800	0.0010	-0.1537	-0.0407
HH cooking fuel is kerosene, oil, animal waste, or other	0.2383	0.0796	2.9900	0.0030	0.0820	0.3946
Number of local hoes HH owns	-0.0386	0.0096	-4.0300	0.0000	-0.0573	-0.0198
Number of local cutlasses HH owns	0.0507	0.0132	3.8500	0.0000	0.0248	0.0766
HH owns one or more coal or charcoal irons	0.0941	0.0323	2.9200	0.0040	0.0308	0.1575
HH owns one or more gas, electric, or kerosene stoves	0.2465	0.1035	2.3800	0.0180	0.0433	0.4497
HH owns one or more generators	0.1460	0.0440	3.3200	0.0010	0.0597	0.2323
HH owns one or more radios	0.0804	0.0246	3.2700	0.0010	0.0321	0.1287
HH owns one or more cassette or CD player	0.1946	0.0471	4.1400	0.0000	0.1022	0.2870
Number of coal pots HH owns	0.0797	0.0166	4.8000	0.0000	0.0471	0.1124

Annex 1: Poverty Prediction at the $2.50/day Poverty Line and Discussion of Additional Poverty Lines

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 $1.25/day PPP models, this will be the $2.50/day PPP; for median poverty models, the "poor" threshold will be the national poverty line. Following this logic, then, the "poor" ("not poor") in Liberia are defined as those whose predicted expenditures fall below (above) the $1.25/day poverty line.

Table 4 summarizes the predictive accuracy results for the $2.50/day poverty line using the Quantile model specification from the $1.25/day poverty line. The indicators are the same as those in the model for the $1.25/day 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 $2.50/day poverty line.

Based on the statistical models underlying prediction at these two lines, IRIS has also introduced the functionality of prediction at five lines to increase the usefulness of the tool to partner organizations. For Liberia, these five lines are the $0.75/day line, $1.00/day line, $1.25/day line, $2.00/day line, and the $2.50/day line. Poverty rates at the first three lines are predicted using the best model for the $1.25/day line, while poverty rates at the last two lines are predicted using the best model for the $2.50/day line. As discussed in this document, accuracy has been tested at the $1.25 and $2.50 lines. Given this, the predictions made at the other lines are intended for indicative use by implementing partners.

The tabulation of poverty prevalence has also been expanded to provide a fuller summary of the incidence of poverty among the implementing organization's clients. Poverty status at the five poverty lines is cross tabulated with regional location, household head's characteristics, household size, and housing conditions. Again, the additional information provided is for indicative purposes rather than statistical inference.

Table 4: Accuracy Results Obtained for Prediction at the $2.50/day Poverty Line

Liberia $2.50/day Line Share of Poor: 84.5%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
Quantile regression (estimation point: 56)	92.22	95.17	4.83	4.37	-0.38	94.72

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 $2.50/day 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)	631 (80.5%)	29 (3.7%)
Number of "true" not poor households (as determined by benchmark survey)	32 (4.1%)	92 (11.7%)

Table 6: Regression Estimates using 1-step Quantile Method for Prediction at $2.50/day Poverty Line

.56 Quantile regression — Number of obs = 784
Min sum of deviations 207.7121 — Pseudo R2 = 0.5180

Variable	Coefficient	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	7.8034	0.1013	77.06	0	7.6050	8.0022
HH lives in Bomi, Gbarpolu, or Grand Cape Mount	0.4180	0.0420	9.9400	0.0000	0.3354	0.5005
HH lives in rural Montserrado, Grand Bassa, or Margibi	0.3178	0.0350	9.0700	0.0000	0.2491	0.3866
HH lives in Grand Gedeh, River Cess, or Sinoe	-0.0093	0.0431	-0.2200	0.8290	-0.0938	0.0752
HH lives in Grand Kru, Maryland, or River Gee	0.1248	0.0461	2.7100	0.0070	0.0343	0.2152
HH lives in urban Montserrado	0.4240	0.0419	10.1300	0.0000	0.3418	0.5062
HH lives in rural area	-0.0736	0.0350	-2.1100	0.0360	-0.1423	-0.0050
Household head age	0.0169	0.0039	4.3000	0.0000	0.0092	0.0246
Household head age squared	-0.0002	0.0000	-4.2500	0.0000	-0.0002	-0.0001
Household size	-0.2325	0.0160	-14.4900	0.0000	-0.2640	-0.2010
Household size squared	0.0083	0.0011	7.7800	0.0000	0.0062	0.0104
HH head is divorced, separated, or widowed	-0.1299	0.0369	-3.5200	0.0000	-0.2024	-0.0575
Share of HH members that can write a one-page letter in English	0.1296	0.0666	1.9500	0.0520	-0.0010	0.2603
Share of HH members with high school certificate as highest diploma completed	0.3777	0.1831	2.0600	0.0390	0.0183	0.7370
Dwelling external walls made of sandcrete or cement blocks	0.1686	0.0346	4.8800	0.0000	0.1008	0.2365
HH toilet facility is flush toilet	0.2439	0.0408	5.9800	0.0000	0.1638	0.3239
HH lighting fuel is palm oil or firewood	-0.0972	0.0288	-3.3800	0.0010	-0.1537	-0.0407
HH cooking fuel is kerosene, oil, animal waste, or other	0.2383	0.0796	2.9900	0.0030	0.0820	0.3946
Number of local hoes HH owns	-0.0386	0.0096	-4.0300	0.0000	-0.0573	-0.0198
Number of local cutlasses HH owns	0.0507	0.0132	3.8500	0.0000	0.0248	0.0766
HH owns one or more coal or charcoal irons	0.0941	0.0323	2.9200	0.0040	0.0308	0.1575
HH owns one or more gas, electric, or kerosene stoves	0.2465	0.1035	2.3800	0.0180	0.0433	0.4497
HH owns one or more generators	0.1460	0.0440	3.3200	0.0010	0.0597	0.2323
HH owns one or more radios	0.0804	0.0246	3.2700	0.0010	0.0321	0.1287
HH owns one or more cassette or CD player	0.1946	0.0471	4.1400	0.0000	0.1022	0.2870
Number of coal pots HH owns	0.0797	0.0166	4.8000	0.0000	0.0471	0.1124