**Poverty Assessment Tool Accuracy Submission**
**USAID/IRIS Tool for Ethiopia**
**Submitted: September 24, 2008**
**Revised (correction to 2005 PPP): December 17, 2009**

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 the Poverty Assessment Tool for Ethiopia. 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.

The report also provides two annexes. Annex 1 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 the Ethiopia Poverty Assessment Tool.

## 1. Overall approach to the tool development

The approach used to develop the Poverty Assessment Tool for Ethiopia 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 or difficult, or whether 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 Ethiopia. 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 also provides a Help Desk to assist practitioners with the implementation of approved USAID Poverty Assessment Tools.

## 2. Data source

For Ethiopia, existing data from the 2004/2005 Welfare Monitoring Survey (WMS) and Household Income and Consumption Expenditure Survey (HICES) were used to construct the Poverty Assessment Tool. The full sample of 21,297 households is nationally representative. The sample used for tool construction is comprised of a randomly selected 10,648 households. The remainder, another randomly selected 10,649 households, is used for out-of-sample accuracy testing, which investigates the robustness of in-sample poverty estimation.[1]

## 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 Ethiopia, more than 120 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).

---

[1] Annex 2 contains a discussion of the methodology employed and results obtained for the out-of-sample testing.

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.[2] 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.[3]

## 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 Ethiopia is the international poverty line developed by the World Bank to track global poverty. The World Bank recently recalculated this line using updated and improved purchasing power parity (PPP) exchange rates and arrived at the value of $1.25/day in 2005 prices. Table 1 summarizes the accuracy results achieved by each of the eight estimation methods in predicting household poverty relative to prediction around this legislative poverty line. For Ethiopia, the most accurate method, on the basis of BPAC, is the 1-step Quantile.

**Table 1: In-sample Accuracy Results for Prediction at the Legislative Poverty Line**

| **Ethiopia**<br>$1.25/day line*****<br>Share of "very poor"**:** 31.5% | **Total Accuracy** | **Poverty Accuracy** | **Under-coverage** | **Leakage** | **PIE** | **BPAC** |
|---|---|---|---|---|---|---|
| **Single-step methods** | | | | | | |
| OLS | 75.20 | 49.01 | 50.99 | 28.57 | -6.99 | 26.58 |
| **Quantile regression (estimation point: 45 percentile)** | **75.04** | **60.38** | **39.62** | **40.47** | **0.26** | **59.53** |
| Linear Probability | 74.68 | 43.04 | 56.96 | 24.29 | -10.18 | 10.36 |
| Probit | 75.23 | 47.81 | 52.19 | 27.29 | -7.76 | 22.91 |
| | | | | | | |
| **Two-step methods** | | | | | | |
| OLS – 48 percentile cutoff | 75.14 | 49.75 | 50.25 | 29.53 | -6.46 | 29.03 |
| Quantile (estimation points: 45, 21.12) 48 | 75.20 | 61.48 | 38.52 | 41.04 | 0.79 | 58.96 |
| LP – 42 percentile cutoff | 75.71 | 53.73 | 46.27 | 31.67 | -4.55 | 39.14 |
| Probit – 42 percentile cutoff | 75.88 | 53.47 | 46.53 | 30.86 | -4.89 | 37.79 |
| * $1.25/day poverty line is 1256.12 Birr per capita per year in 2005 prices. The international poverty line is based on World Bank's calculations and the recent 2005 PPP exchange rates. | | | | | | |

[2] The set of indicators and their weights also depended on the selection of a 1-step or 2-step statistical model.

[3] For a detailed discussion of these accuracy criteria, see "Note on Assessment and Improvement of Tool Accuracy" at www.povertytools.org

For Ethiopia, the functionality of predicting the poverty rate at another poverty line—in this case, the $2.50/day poverty line—has been added. When running the analysis routine with the Epi Info template, the user is presented the options to predict the extreme poverty rate (using the $1.25/day line), the poverty rate ($2.50/day 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 the international poverty line ($1.25/day in 2005 terms)[4]. The applicable poverty line for USAID tool development is the one that yields the higher household poverty rate for a given country.

In Ethiopia the applicable threshold is the $1.25/day poverty line, or 1256.12 Birr per year per person, at the level of prices prevailing in 2005.[5] At 1256.12 Birr per person per year, the international poverty line identifies 31.5% of households as "very poor".[6] Alternatively, the median poverty line of 1197.22 Birr per adult equivalent per year

[4] 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 "$1/day" line established by the World Bank. This poverty line has recently been recalculated by the Bank to accompany new, improved estimates of PPP.
[5] The appropriate 2005 PPP value for Ethiopia is 2.7512605, correcting an error in the previous accuracy submission (that used a different PPP value than the World Bank uses in its global poverty calculations). The international poverty line in per capita per year in average 2005 prices is therefore equal to 2.7512605*1.25*365.25.
[6] PovcalNet lists a *population* poverty headcount of 39.0% using the $1.25/day PPP line. We calculate a population headcount of 39.8% when performing similar calculations using the data.

yields a household poverty rate of 15.0%.[7] Hence the decision rule for Ethiopia's USAID Poverty Assessment Tool in classifying the "very poor" (and the "not very-poor") is whether that predicted per capita annual 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 fall below 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)** | 2,004 (18.8%) | 1,315 (12.3%) |
| **Number of "true" not very-poor households (as determined by benchmark survey)** | 1,343 (12.6%) | 5,986 (56.3%) |

[7] For details on the national poverty line, see http://www.economics.ox.ac.uk/members/stefan.dercon/IFAD1.pdf and http://www.mofaed.org/POVERTY_PROFILE_OF_ETHIOPIA2002.pdf.

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

(sum of wgt is 10,648)

.45 Quantile regression Number of obs = 10,648
Raw sum of deviations 4263.227 (about 7.3157935)
Min sum of deviations 3174.303 Pseudo R2 = 0.2554

| Variables | Coef. | Std. Err. | T | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Household head age | 0.0028 | 0.0018 | 1.5200 | 0.1300 | -0.0008 | 0.0064 |
| Household head age squared | 0.0000 | 0.0000 | -1.7200 | 0.0850 | -0.0001 | 0.0000 |
| Household size | -0.3104 | 0.0080 | -38.7000 | 0.0000 | -0.3261 | -0.2946 |
| Household size squared | 0.0160 | 0.0007 | 23.4800 | 0.0000 | 0.0147 | 0.0174 |
| Household lives in Tigray | -0.1883 | 0.0172 | -10.9400 | 0.0000 | -0.2220 | -0.1546 |
| Household lives in Afar | -0.0728 | 0.0212 | -3.4400 | 0.0010 | -0.1142 | -0.0313 |
| Household lives in Amhara | -0.1528 | 0.0134 | -11.4000 | 0.0000 | -0.1790 | -0.1265 |
| Household lives in Somali | -0.0387 | 0.0195 | -1.9800 | 0.0480 | -0.0770 | -0.0004 |
| Household lives in Beningshangul-Gumuz | 0.0615 | 0.0215 | 2.8700 | 0.0040 | 0.0195 | 0.1035 |
| Household lives in Gambela or Southern Nations | -0.0757 | 0.0135 | -5.6200 | 0.0000 | -0.1022 | -0.0493 |
| Household lives in Harari | 0.0085 | 0.0228 | 0.3700 | 0.7100 | -0.0362 | 0.0531 |
| Household lives in Adis Ababa | -0.2124 | 0.0159 | -13.3700 | 0.0000 | -0.2435 | -0.1813 |
| Household lives in Dire Dawa | -0.1130 | 0.0224 | -5.0500 | 0.0000 | -0.1568 | -0.0692 |
| Household lives in urban area | 0.0612 | 0.0164 | 3.7300 | 0.0000 | 0.0290 | 0.0933 |
| Household head is female | -0.1038 | 0.0121 | -8.6100 | 0.0000 | -0.1275 | -0.0802 |
| Household head never married | 0.1221 | 0.0252 | 4.8500 | 0.0000 | 0.0727 | 0.1714 |
| Share of household members excluding household head who can read and write | 0.2248 | 0.0233 | 9.6600 | 0.0000 | 0.1791 | 0.2704 |
| Number of cattle household owns | 0.0149 | 0.0018 | 8.0800 | 0.0000 | 0.0113 | 0.0186 |
| Number of axes household owns | 0.0351 | 0.0053 | 6.5600 | 0.0000 | 0.0246 | 0.0456 |
| Number of blankets household owns | 0.0137 | 0.0026 | 5.2600 | 0.0000 | 0.0086 | 0.0188 |
| Household owns one or more radios | 0.0810 | 0.0133 | 6.1100 | 0.0000 | 0.0550 | 0.1070 |
| Household owns one or more televisions | 0.1996 | 0.0189 | 10.5900 | 0.0000 | 0.1627 | 0.2366 |
| Household owns one or more videos | 0.1588 | 0.0254 | 6.2400 | 0.0000 | 0.1090 | 0.2087 |
| Number of rooms in dwelling | 0.0438 | 0.0055 | 7.9900 | 0.0000 | 0.0331 | 0.0545 |
| Roof is made of thatch and grass | -0.0826 | 0.0129 | -6.4300 | 0.0000 | -0.1078 | -0.0575 |
| Light is made of firewood, candle, or other | -0.1034 | 0.0142 | -7.2900 | 0.0000 | -0.1312 | -0.0756 |
| Cooking fuel is kerosene, butane gas, electricity, crop residue, don’t use cooking fuel, or other | 0.0669 | 0.0152 | 4.4000 | 0.0000 | 0.0371 | 0.0967 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cooking fuel is mainly purchased firewood | 0.0453 | 0.0166 | 2.7400 | 0.0060 | 0.0129 | 0.0778 |
| Drinking water is tap in compound (shared) | 0.1513 | 0.0168 | 9.0300 | 0.0000 | 0.1185 | 0.1842 |
| Intercept | 8.1782 | 0.0471 | 173.5500 | 0.0000 | 8.0858 | 8.2705 |

## Annex 1: Poverty Prediction at the $2.50/day 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 from 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 Ethiopia are defined as those whose predicted incomes fall below (above) the $2.50/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.

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

| **Ethiopia**<br>$2.50/day Line<br>Share of Poor: 82.3% | **Total Accuracy** | **Poverty Accuracy** | **Under-coverage** | **Leakage** | **PIE** | **BPAC** |
|---|---|---|---|---|---|---|
| **Single-step methods** | | | | | | |
| **Quantile regression (estimation point: 59)** | 85.55 | 91.38 | 8.62 | 8.94 | 0.26 | 91.06 |

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 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)** | 8,009<br>(75.2%) | 755<br>(7.1%) |
| **Number of “true” not poor households (as determined by benchmark survey)** | 783<br>(7.4%) | 1,101<br>(10.3%) |

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

.63 Quantile regression Number of obs = 10648
Raw sum of deviations 4215.43 (about 7.5185719)
Min sum of deviations 3114.927 Pseudo R2 = 0.2611

| Variables | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Household head age | 0.0054 | 0.0031 | 1.7500 | 0.0800 | -0.0007 | 0.0115 |
| Household head age squared | -0.0001 | 0.0000 | -1.8100 | 0.0710 | -0.0001 | 0.0000 |
| Household size | -0.3138 | 0.0133 | -23.5600 | 0.0000 | -0.3400 | -0.2877 |
| Household size squared | 0.0162 | 0.0011 | 14.4900 | 0.0000 | 0.0140 | 0.0184 |
| Household lives in Tigray | -0.1641 | 0.0287 | -5.7300 | 0.0000 | -0.2202 | -0.1079 |
| Household lives in Afar | -0.0918 | 0.0356 | -2.5800 | 0.0100 | -0.1616 | -0.0220 |
| Household lives in Amhara | -0.1307 | 0.0223 | -5.8600 | 0.0000 | -0.1744 | -0.0870 |
| Household lives in Somali | -0.0380 | 0.0320 | -1.1900 | 0.2360 | -0.1007 | 0.0248 |
| Household lives in Beningshangul-Gumuz | 0.0604 | 0.0353 | 1.7100 | 0.0880 | -0.0089 | 0.1296 |
| Household lives in Gambela or Southern Nations | -0.0655 | 0.0224 | -2.9300 | 0.0030 | -0.1094 | -0.0217 |
| Household lives in Harari | 0.0764 | 0.0375 | 2.0400 | 0.0410 | 0.0030 | 0.1498 |
| Household lives in Adis Ababa | -0.1652 | 0.0265 | -6.2400 | 0.0000 | -0.2171 | -0.1133 |
| Household lives in Dire Dawa | -0.1255 | 0.0372 | -3.3800 | 0.0010 | -0.1983 | -0.0526 |
| Household lives in urban area | 0.0083 | 0.0272 | 0.3000 | 0.7600 | -0.0451 | 0.0617 |
| Household head is female | -0.1090 | 0.0201 | -5.4100 | 0.0000 | -0.1484 | -0.0695 |
| Household head never married | 0.1396 | 0.0408 | 3.4200 | 0.0010 | 0.0596 | 0.2196 |
| Share of household members excluding household head who can read and write | 0.2775 | 0.0389 | 7.1300 | 0.0000 | 0.2012 | 0.3538 |
| Number of cattle household owns | 0.0167 | 0.0029 | 5.8600 | 0.0000 | 0.0111 | 0.0223 |
| Number of axes household owns | 0.0376 | 0.0087 | 4.3100 | 0.0000 | 0.0205 | 0.0547 |
| Number of blankets household owns | 0.0164 | 0.0043 | 3.8500 | 0.0000 | 0.0080 | 0.0247 |
| Household owns one or more radios | 0.0635 | 0.0219 | 2.9000 | 0.0040 | 0.0205 | 0.1064 |
| Household owns one or more televisions | 0.2138 | 0.0301 | 7.0900 | 0.0000 | 0.1547 | 0.2729 |
| Household owns one or more videos | 0.1720 | 0.0414 | 4.1600 | 0.0000 | 0.0909 | 0.2531 |
| Number of rooms in dwelling | 0.0388 | 0.0091 | 4.2700 | 0.0000 | 0.0210 | 0.0565 |
| Roof is made of thatch and grass | -0.1017 | 0.0215 | -4.7300 | 0.0000 | -0.1438 | -0.0595 |
| Light is made of firewood, candle, or other | -0.0913 | 0.0236 | -3.8600 | 0.0000 | -0.1377 | -0.0450 |
| Cooking fuel is kerosene, butane gas, electricity, crop residue, don't use cooking fuel, or other | 0.0621 | 0.0251 | 2.4700 | 0.0140 | 0.0128 | 0.1114 |
| Cooking fuel is mainly purchased firewood | 0.0488 | 0.0277 | 1.7600 | 0.0780 | -0.0055 | 0.1031 |
| Drinking water is tap in compound (shared) | 0.0846 | 0.0270 | 3.1400 | 0.0020 | 0.0317 | 0.1375 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Intercept | 8.3310 | 0.0783 | 106.3700 | 0.0000 | 8.1775 | 8.4845 |

## Annex 2: Out-of-Sample Accuracy Tests

**Note: The out-of-sample accuracy tests were not re-run using the corrected, higher $1.25/day poverty line. However, we would expect the tool to exhibit even greater accuracy when predicting from a population with a higher rate of extreme poverty.**

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 Ethiopia tool was divided randomly into two data sets of equal size (10,648 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) performs very well, losing less than two percentage points in BPAC and a negligible amount in PIE.

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

| | Total Accuracy | Poverty Accuracy | Under-coverage | Leakage | PIE | BPAC |
|---|---|---|---|---|---|---|
| **In-Sample Prediction** | | | | | | |
| | 79.13 | 48.28 | 51.72 | 52.43 | 0.14 | 47.57 |
| **Out-of-Sample Prediction** | | | | | | |
| | 78.66 | 47.40 | 52.59 | 54.29 | 0.34 | 45.70 |

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[8]. 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.

[8] 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.  As with the naïve result in Table 7, the performance of this model is quite good.  The confidence interval around the sample mean of BPAC is +/- 4.8 percentage points. For PIE, which measures the difference between the actual poverty rate in the sample and the predicted poverty rate, the confidence interval is especially tight at +/- 1.3 percentage points.

**Table 8: Bootstrapped Confidence Intervals Computed on Assumption of Normality**

| Accuracy Measure | Mean | Std. Err. | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | LB | UB |
| Total Accuracy | 78.66 | 0.59 | 77.49 | 79.83 |
| Poverty Accuracy | 47.40 | 1.58 | 44.31 | 50.49 |
| Undercoverage | 52.59 | 1.58 | 49.50 | 55.69 |
| Leakage | 54.20 | 2.99 | 48.35 | 60.05 |
| PIE | 0.31 | 0.67 | -1.00 | 1.63 |
| BPAC | 44.43 | 2.43 | 39.66 | 49.19 |

The results presented in Table 8 assume a normal distribution for the accuracy measures from the bootstrapped samples.  This ignores the possibility that these estimates may have a skewed distribution.  Table 9 presents alternative 95% confidence intervals. The lower bound is defined by the 2.5$^{th}$ percentile of the sample distribution for each measure; the upper bound is defined by the 97.5$^{th}$ percentile.  On the whole, the results are quite similar between Tables 8 and 9, which adds confidence to the conclusion that the 1-step Quantile model on which the Ethiopia PAT is based performs well out-of-sample.

**Table 9: Bootstrapped Confidence Intervals Computed Empirically from Sampling Distribution Without Normality Assumption**

| Accuracy Measure | 95% Confidence Interval | |
|---|---|---|
| | LB | UB |
| Total Accuracy | 77.53 | 79.86 |
| Poverty Accuracy | 44.51 | 50.49 |
| Undercoverage | 49.50 | 55.49 |
| Leakage | 48.30 | 60.10 |
| PIE | -1.08 | 1.63 |
| BPAC | 38.78 | 48.54 |