Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for Madagascar
Submitted: September 15, 2011

In order to improve the functionality of the existing PAT for Madagascar, the IRIS Center has updated the tool with the following features:

- Re-ran the models at the $1.25/day line, using the new purchasing power parity (PPP) rates lines released by the World Bank
- Calibrated the model to also allow predictions at the $2.50/line
- Incorporated the prediction models into an Epi Info data entry template. This template closely resembles the paper questionnaire and allows the entry, storage, and retrieval of household demographics. The output of the data entry permits poverty prediction at two poverty lines, $1.25/day and $2.50/day.
- Revised the paper questionnaire to reflect best practice in survey design

The data source used for the PAT in Madagascar remains the same as when the tool was originally submitted for certification.

1. 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 Madagascar, more than 90 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]

[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.

2. 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 Madagascar 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 Madagascar, on the basis of BPAC, the 1-step and 2-step Quantile regression models are almost equal in terms of accuracy. However, the 1-step Quantile regression requires only 15 indicators. 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

Madagascar $1.25/day line* Share of "very poor": 74.3%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	84.08	92.23	7.76	16.09	5.55	83.90
Quantile regression (estimation point: 60 percentile)	**84.08**	**88.08**	**11.91**	**11.95**	**.02**	**88.04**
Linear Probability	85.57	93.04	6.96	14.68	5.15	85.32
Probit	85.98	92.42	7.58	13.44	3.91	86.55
Two-step methods						
OLS –97 percentile cutoff	84.82	92.30	7.70	15.06	4.91	84.94
Quantile (estimation points: 60, 58) 97 percentile cutoff	84.35	88.39	11.61	11.85	0.16	88.15
LP – 92 percentile cutoff	85.82	93.26	6.74	14.52	5.19	85.48
Probit –92 percentile cutoff	85.88	92.47	7.53	13.68	4.10	86.32
* The $1.25/day is 3060.695 Malagasy Franc per capita per day in 2001 prices.						

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

4. 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 Madagascar, 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 3060.695 Malagasy Francs per day per capita.[4] At these values, the $1.25/day poverty line identifies 74.4% of households as "very poor." This compares with an estimate from PovcalNet of 76.3%.

Alternatively, the national poverty line of 989563.81 Malagasy Francs per year identifies 62.0% of households as "very poor," implying a poverty rate at the median line of 31.0%/

Hence the decision rule for Madagascar'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.

---

[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 Madagascar is 756.38074.

[4] The calculation for the $1.25/day poverty line is 1.25*(756.38074*5)*(64.744/100) where the final term is the CPI adjustment from average 2005 prices to average 2006 prices. The Malagasy Franc is changed into the Malagasy Ariary in 2005. One Malagasy Ariary is equal to 5 Malagasy Francs.

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)	2,174 (58.8%)	294 (7.9%)
Number of "true" not very-poor households (as determined by benchmark survey)	295 (7.9%)	938 (25.4%)

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

.60 Quantile regression — Number of obs = 3,701
Min sum of deviations 1404.358 — Pseudo R2 = 0.4311

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Household size	-0.3306	0.0174	-19.0000	0.0000	-0.3647	-0.2965
Household head age	0.0131	0.0048	2.7500	0.0060	0.0038	0.0225
Household head age squared	-0.0001	0.0001	-2.0000	0.0450	-0.0002	0.0000
Household size squared	0.0164	0.0014	11.5300	0.0000	0.0136	0.0192
Household lives in a rural area	-0.0092	0.0279	-0.3300	0.7410	-0.0639	0.0455
Household lives in Fianarantsoa	-0.1401	0.0399	-3.5100	0.0000	-0.2184	-0.0618
Household lives in Toamasina	-0.1127	0.0403	-2.7900	0.0050	-0.1918	-0.0336
Household lives in Mahajanga	0.0476	0.0420	1.1300	0.2570	-0.0348	0.1299
Household lives in Toliara	0.0020	0.0411	0.0500	0.9610	-0.0785	0.0825
Household lives in Antsiranana	0.1283	0.0445	2.8800	0.0040	0.0410	0.2157
Household head has no education	-0.2625	0.0403	-6.5200	0.0000	-0.3414	-0.1836
Household head is female	-0.1168	0.0309	-3.7800	0.0000	-0.1773	-0.0562
Number of rooms in dwelling	0.0592	0.0099	5.9600	0.0000	0.0397	0.0787
Roof is made of wood (boards, plywood, hardboard)	0.1548	0.0382	4.0500	0.0000	0.0799	0.2296
Primary source of drinking water is interior plumbing, indoor tap/spigot, or private outside tap/spigot	0.1693	0.0452	3.7400	0.0000	0.0806	0.2579
Primary source of drinking water is river, lake, spring, pond	-0.1221	0.0297	-4.1100	0.0000	-0.1802	-0.0639
Main source of cooking fuel is wood picked	-0.3547	0.0402	-8.8200	0.0000	-0.4336	-0.2759
Main source of cooking fuel is purchased wood	-0.1370	0.0480	-2.8600	0.0040	-0.2311	-0.0430
Household owns one or more tables	0.2278	0.0293	7.7600	0.0000	0.1702	0.2853
Household owns one or more stoves	0.5300	0.0800	6.6300	0.0000	0.3732	0.6868
Household owns one or more stereos	0.1300	0.0288	4.5200	0.0000	0.0735	0.1864
Household owns one or more televisions	0.3643	0.0424	8.5800	0.0000	0.2811	0.4475
Household owns one or more cars	0.4952	0.0963	5.1400	0.0000	0.3064	0.6839

Household owns one or more bicycles	0.2000	0.0406	4.9200	0.0000	0.1204	0.2797
Last level of schooling completed by household head is any level between preschool or CPI and T5 or CM2	-0.1890	0.0402	-4.7100	0.0000	-0.2678	-0.1103
Intercept	8.6952	0.1141	76.2100	0.0000	8.4715	8.9188

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 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 Madagascar are defined as those whose predicted expenditures 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

Madagascar $2.50/day line* Share of "very poor": 91.0%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step method						
Quantile regression (estimation point: 64)	92.58	95.91	4.09	4.33	0.21	95.67
*The $2.50/day line is 6121.40 Malagasy Francs per capita per day in 2001 prices.						

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)	3,127 (84.5%)	133 (3.6%)
Number of "true" not poor households (as determined by benchmark survey)	141 (3.8%)	300 (8.1%)

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

.64 Quantile regression — Number of obs = 3,701
Min sum of deviations 1362.388 — Pseudo R2 = 0.4330

Variable	Coefficient	Standard Error	t	P>|t|	[95% Confidence Interval]	
Household size	-0.3452	0.0138	-25.0100	0.0000	-0.3723	-0.3181
Household head age	0.0166	0.0038	4.4000	0.0000	0.0092	0.0241
Household head age squared	-0.0001	0.0000	-3.4600	0.0010	-0.0002	-0.0001
Household size squared	0.0178	0.0011	15.9700	0.0000	0.0156	0.0200
Household lives in a rural area	-0.0033	0.0215	-0.1500	0.8780	-0.0454	0.0388
Household lives in Fianarantsoa	-0.1327	0.0314	-4.2200	0.0000	-0.1943	-0.0710
Household lives in Toamasina	-0.1242	0.0318	-3.9000	0.0000	-0.1866	-0.0617
Household lives in Mahajanga	0.0474	0.0334	1.4200	0.1550	-0.0180	0.1129
Household lives in Toliara	0.0423	0.0327	1.2900	0.1960	-0.0218	0.1063
Household lives in Antsiranana	0.1062	0.0351	3.0300	0.0020	0.0375	0.1750
Household head has no education	-0.2407	0.0321	-7.5000	0.0000	-0.3037	-0.1778
Household head is female	-0.1244	0.0245	-5.0800	0.0000	-0.1723	-0.0764
Number of rooms in dwelling	0.0585	0.0078	7.5100	0.0000	0.0432	0.0738
Roof is made of wood (boards, plywood, hardboard)	0.1501	0.0298	5.0400	0.0000	0.0917	0.2085
Primary source of drinking water is interior plumbing, indoor tap/spigot, or private outside tap/spigot	0.1944	0.0362	5.3700	0.0000	0.1234	0.2653
Primary source of drinking water is river, lake, spring, pond	-0.1087	0.0233	-4.6500	0.0000	-0.1544	-0.0629
Main source of cooking fuel is wood picked	-0.3436	0.0316	-10.8900	0.0000	-0.4054	-0.2817
Main source of cooking fuel is purchased wood	-0.1496	0.0374	-4.0000	0.0000	-0.2230	-0.0762
Household owns one or more tables	0.2442	0.0233	10.4700	0.0000	0.1985	0.2899
Household owns one or more stoves	0.5564	0.0651	8.5400	0.0000	0.4287	0.6841
Household owns one or more stereos	0.1202	0.0227	5.2900	0.0000	0.0757	0.1647
Household owns one or more televisions	0.3566	0.0336	10.6100	0.0000	0.2907	0.4225
Household owns one or more cars	0.4898	0.0759	6.4500	0.0000	0.3409	0.6386
Household owns one or more bicycles	0.1872	0.0325	5.7600	0.0000	0.1235	0.2510

Last level of schooling completed by household head is any level between preschool or CPI and T5 or CM2	-0.1629	0.0322	-5.0500	0.0000	-0.2261	-0.0996
Intercept	8.6592	0.0903	95.8300	0.0000	8.4821	8.8364

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 1, the data set used to construct the Madagascar tool was divided randomly into two data sets 3,701 households (75 percent of the sample) and 1,237 households (25 percent sample). 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 well in terms of BPAC with a 2.69 difference.

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

	Total Accuracy	Poverty Accuracy	Under-coverage	Leakage	PIE	BPAC
In-Sample Prediction						
	84.05	88.07	11.93	12.03	0.07	87.97
Out-of-Sample Prediction						
	86.69	88.66	11.33	7.95	-2.34	85.28

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.[5] 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.

Table 8 presents the out-of-sample, bootstrapped confidence intervals for the 1-step Quantile model. The performance of this model is very good. The confidence interval around the sample mean BPAC is relatively narrow at +/- 6.04 percentage points. For

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

PIE, which measures the difference between the predicted poverty rate and the actual poverty rate, the confidence interval is +/- 2.63 percentage points.

Table 8: Bootstrapped Confidence Intervals on Assumption of Normality

Variable	Mean	Std. Dev.	Confidence interval	
			LB	UB
Total Accuracy	86.67	1.23	84.26	89.08
Poverty Accuracy	88.62	1.46	85.76	91.48
Undercoverage	11.38	1.46	8.52	14.24
Leakage	7.93	1.24	5.50	10.36
PIE	-2.40	1.34	-5.03	0.23
BPAC	85.12	3.08	79.08	91.16

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 2.5th percentile of the sample distribution for each measure; the upper bound is defined by the 97.5th percentile. On the whole, the results are quite similar between Tables 8 and 9.

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

Accuracy Measure	95% Confidence Interval	
	LB	UB
Total Accuracy	84.17	89.03
Poverty Accuracy	85.64	91.35
Undercoverage	8.65	14.36
Leakage	5.62	10.53
PIE	-5.17	0.21
BPAC	78.74	90.39

The primary purpose of the PAT is to assess the overall extreme poverty rate across a group of households. The out-of-sample results for PIE in Table 8 and Table 9 indicate that the extreme poverty rate estimate produced by the Madagascar PAT appears to be somewhat biased toward underestimating the actual extreme poverty rate, but with a moderately narrow confidence interval (on PIE) of -5.17 to 0.21. By this measure, the predictive model behind the Madagascar PAT is accurate.