Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for Senegal
Submitted: June 20, 2011

The following report on the development of a Poverty Assessment Tool (PAT) for Senegal is divided into five sections. Section 1 describes the data used to create the PAT for Senegal. 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 Senegal, USAID funded, and IRIS implemented, an original data collection exercise in July through September 2009. This was undertaken after a close examination of existing data from the 2001 Enquête Sénégalaise Auprès des Ménages (ESAM-2) and 2005/2006 L'Enquête de Suivi de la Pauvreté au Sénégal (ESPS) and the determination that the two data sets had significant shortcomings as a data source for constructing the poverty assessment tool. A local survey firm, Centre de Recherche pour le Développement Humain, was selected competitively. After intensive training and a round of pilot-testing, the firm collected expenditure and other living standards information from a nationally-representative self-weighted sample of 842 households. Training, pilot-testing, and fieldwork were done with close supervision by IRIS staff and included the use of ultra-mobile personal computers. The full sample of 842 households comprises of 449 urban households and 393 rural households.[1]

2. Process used to select included indicators

Suitable household surveys, such as the LSMS type survey, typically include variables related to education, housing characteristics, consumer durables, agricultural assets, and employment. The data collected in the household roster was used to construct gender indicators such as *household head is female*, *share of female household members of age 6-15*, and *share of female household members of age 16-25*. For Senegal, more than 126 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

[1] Given this sample size, the data set was not divided into calibration and holdout samples.

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.[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]

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 Senegal 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 Senegal, the 1-step Quantile regression is the most accurate method. It has the highest BPAC (closest to 100) and the lowest PIE (closest to 0 in absolute value).

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

SENEGAL ($1.25/day line***)** Share of "Very Poor": 11.5%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	89.43	44.44	55.56	34.34	-2.49	23.23
Quantile regression (estimation point: 39 percentile)	**89.19**	**54.55**	**45.45**	**46.46**	**0.12**	**53.54**
Linear Probability	90.62	30.30	69.70	10.10	-7.01	-29.29
Probit	91.45	42.42	57.58	15.15	-4.99	0
Two-step methods						
OLS –34 percentile cutoff	90.50	53.54	46.46	34.34	-1.43	41.41
Quantile (estimation points: 39, 13) 34 percentile cutoff	89.90	62.63	37.37	48.48	1.31	51.52
LP – 34 percentile cutoff	93.23	58.59	41.41	16.16	-2.97	33.33
Probit –34 percentile cutoff	90.97	42.42	57.58	19.19	-4.51	4.04
*The $1.25/day per capita international poverty line in 2005 Purchasing Power Parity terms is 422 CFA Francs per day per capita in 2009 prices.						

[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 Senegal, 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

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)[4]. The applicable poverty line for USAID tool development is the one that yields the higher household poverty rate for a given country.

In Senegal 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 422 CFA Francs per capita per day.[5] At these values, the $1.25/day poverty line identifies 11.5% of households as "very poor." Schreiner (2009) reports a household poverty rate of 22.1% using 2005/2006 ESPS data. [6]

There are a number of reasons that may explain the difference in observed poverty rates using the $1.25/day line compared to the most recent ESPS data set from Senegal. These include:

[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 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 Senegal is 298.24.

[5] The calculation for the $1.25/day poverty line is 1.25*(298.2448)*(127.1/112.4) where the final term is the CPI adjustment from average 2005 prices to average 2009 prices.

[6] http://www.microfinance.com/English/Papers/Scoring_Poverty_Senegal_EN_2005.pdf The World Bank's PovcalNet provides a poverty headcount of 33.5% using population rather than household weights and the 2005 DHS.

- The ESPS survey took place during the dry season in the seasonal calendar. In comparison, the USAID/IRIS Center survey was implemented during the harvest season.
- The USAID/IRIS survey was done in two sessions whereas the ESPS was done in one single session. Lengthy interviews could lead to under-reporting of expenditures.
- The food expenditures recall period for the ESPS survey was longer than the USAID/IRIS survey. The recall period was ESPS was 30 days and 14 days for the USAID/IRIS survey. 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).
- Expenditure shares differ between the USAID/IRIS survey and the ESPS, particularly for the education and health sectors. Based on the USAID/IRIS survey, education expenditures account for 1.3% compared with 0.8% for the ESPS. While health expenses account for 5.4% of expenditures for the USAID/IRIS survey and 2.4% for the ESPS.

We considered carefully, but ultimately did not use, the national poverty line that is used by some sources for Senegal. First, the poverty line is based on 2001 data from the ESAM survey, which is rather old. Second, the expenditure data used in the ESAM 2001 and 2006 ESPS data had significant shortcomings (as mentioned above). Third, calculating the value of the national poverty line directly from our 2009 data would require calculating the expenditures required to consume 2,400 calories per adult equivalent and also non-food expenditures of those near the food poverty line. These calculations require a very rich dataset with larger sample sizes than the 2009 data set. Finally, the latest evidence using previous data is that the $1.25/day line would be the binding line, albeit narrowly, over the median poverty line.[7]

Hence the decision rule for Senegal'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

[7] Figure 2 in Schreiner (2009).

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 In-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)	54 (6.4%)	45 (5.3%)
Number of "true" not very-poor households (as determined by benchmark survey)	46 (5.5%)	697 (82.8%)

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

.39 Quantile regression — Number of obs = 842
Min sum of deviations 225.2976 — Pseudo R2 = 0.5005

Variable	Coef.	Std. Err.	T	P>|t|	[95% Conf. Interval]	
Intercept	7.2604	0.1666	43.5700	0.0000	6.9333	7.5875
Household size	-0.0606	0.0072	-8.4700	0.0000	-0.0747	-0.0466
Household size squared	0.0006	0.0002	2.8400	0.0050	0.0002	0.0104
Household head age	-0.0025	0.0060	-0.4200	0.6710	-0.0143	0.0092
Household head age squared	0.0000	0.0000	0.3200	0.7500	-0.0001	0.0001
Household lives in rural area	-0.0905	0.0507	-1.7900	0.0740	-0.1900	0.0089
Household lives in Diourbel	-0.2834	0.0707	-4.0100	0.0000	-0.4222	-0.1445
Household lives in Fatick	-0.0344	0.0820	-0.4200	0.6740	-0.1953	0.1264
Household lives in Kaolack	-0.0060	0.0728	-0.0800	0.93400	-0.1489	0.1368
Household lives in Kolda	-0.4373	0.0861	-5.0800	0.0000	-0.6064	-0.2683
Household lives in Louga	-0.1378	0.0916	-1.5100	0.1330	-0.3176	0.4192
Household lives in Matam	-0.1902	0.0828	-2.3000	0.0220	-0.3527	-0.0276
Household lives in Saint Louis	-0.0032	0.0683	0.0500	0.9620	-0.1309	0.1374
Household lives in Tambacounda	-0.2071	0.0936	-2.2100	0.0270	-0.3909	-0.0233
Household lives in Thies	-0.1264	0.0592	-2.1300	0.0330	-0.2426	-0.0100
Household lives in Zinguinchor	-0.2452	0.0846	-2.9000	0.0040	-0.4113	0.0790
Dependency ratio[8]	-0.0914	0.0208	-4.4000	0.0000	-0.1321	-0.0506
Wall is made of banco bricks	-0.0872	0.0514	-1.7000	0.0900	-0.1972	0.0137
Roof is made of thatch/straw	-0.2172	0.0534	-4.0600	0.0000	0.3221	-0.1123
Roof is made of concrete/cement	0.2034	0.0495	4.1100	0.0000	0.1063	0.3005
Toilet type is toilet with sewage	0.1833	0.0712	2.5700	0.0100	0.0435	0.3231
Garbage disposed by burying it	-0.1296	0.0726	-1.7800	0.0750	-0.2721	0.0129
Number of chairs owned	0.0127	0.0074	1.7100	0.0880	-0.0019	0.0273
Number of computers owned	0.0998	0.0290	3.4300	0.0010	0.0427	0.1569
Number of artisanal machetes owned	0.0627	0.0225	2.7900	0.0050	0.0185	0.1069
Household (HH) owns one or more tables	0.1061	0.0352	3.0100	0.0030	0.0370	0.1753
HH owns one or more sofas	0.1589	0.0485	3.2800	0.0010	0.0637	0.2541
HH owns one or more fans	0.1720	0.0433	3.9800	0.0000	0.0871	0.2570
HH owns one or more refrigerators	0.1973	0.0486	4.0600	0.0000	0.1019	0.2927
HH owns one or more cars	0.2461	0.0501	3.5500	0.0000	0.1100	0.3822
HH owns one or more cattle	0.1692	0.0264	3.3800	0.0010	0.0709	0.2674

[8] The dependency ratio is the ratio of household members assumed to be of non-working age, under 15 or over 65, to the number of working household members, ages 16-65.

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, as discussed in Section 5. 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 Senegal 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.

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 Senegal, 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 gender, household head's education by gender, household size, and housing conditions. Again, the additional information provided is for indicative purposes rather than statistical inference.

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

Senegal $2.50/day Line Share of Poor: 47.7%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step method						
Quantile regression (estimation point: 54)	83.25	82.84	17.16	17.91	0.36	82.09

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 In-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)	333 (39.6%)	69 (8.2%)
Number of "true" not poor households (as determined by benchmark survey)	72 (8.6%)	368 (43.6%)

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

.54 Quantile regression — Number of obs = 842
Min sum of deviations 233.3535 — Pseudo R2 = 0.5129

Variable	Coef.	Std. Err.	T	P>|t|	[95% Conf. Interval]	
Intercept	7.259	0.1918	37.8600	0.0000	6.8826	7.6353
Household size	-0.0589	0.0087	-6.8000	0.0000	-0.0759	-0.0419
Household size squared	0.0006	0.0003	2.3600	0.0190	0.0001	0.0012
Household head age	0.0033	0.0069	0.4800	0.6340	-0.0103	0.0169
Household head age squared	0.0000	0.0001	-0.4900	0.6250	-0.0002	0.0001
Household lives in rural area	-0.0816	0.0607	-1.3500	0.1790	-0.2008	0.0375
Household lives in Diourbel	-0.2436	0.0849	-2.8700	0.0040	-0.4102	-0.0769
Household lives in Fatick	-0.0278	0.0933	-0.3000	0.7650	-0.2110	0.1553
Household lives in Kaolack	-0.0079	0.0855	-0.0900	0.9260	-0.1757	0.1599
Household lives in Kolda	-0.4367	0.1018	-4.2900	0.0000	-0.6365	-0.2369
Household lives in Louga	-0.1283	0.1060	-1.2100	0.2270	-0.0002	0.0798
Household lives in Matam	-0.2483	0.0962	-2.5800	0.0100	-0.4372	-0.0594
Household lives in Saint Louis	-0.0245	0.0813	-0.3000	0.7620	-0.1842	0.1350
Household lives in Tambacounda	-0.2010	0.1091	-1.8400	0.0660	-0.4152	0.0131
Household lives in Thies	-0.1159	0.0677	-1.7100	0.0870	-0.2487	0.0170
Household lives in Zinguinchor	-0.1648	0.0967	-1.7000	0.0890	-0.3547	0.0251
Dependency ratio	-0.0962	0.0251	-3.8400	0.0000	-0.1455	-0.0506
Wall is made of banco bricks	-0.1280	0.0602	-2.1300	0.0340	-0.1972	-0.0098
Roof is made of thatch/straw	-0.2397	0.0633	-3.7900	0.0000	-0.3639	-0.1155
Roof is made of concrete/cement	0.2767	0.0558	4.9600	0.0000	0.1672	0.3862
Toilet type is toilet with sewage	0.0723	0.0861	0.8400	0.4020	-0.0968	0.2414
Garbage disposed by burying it	-0.0855	0.0846	-1.0100	0.3120	-0.2515	0.0805
Number of chairs owned	0.0119	0.0084	1.4100	0.1600	-0.0047	0.0284
Number of computers owned	0.1022	0.0413	2.4700	0.0140	0.0210	0.1834
Number of artisanal machetes owned	0.0508	0.0265	1.9200	0.0560	-0.0013	0.1029
HH owns one or more tables	0.0346	0.0423	0.8200	0.4140	-0.0485	0.1176
HH owns one or more sofas	0.2000	0.0558	3.5800	0.0000	0.0905	0.3097
HH owns one or more fans	0.1377	0.0510	2.7000	0.0070	0.0375	0.2379
HH owns one or more refrigerators	0.1964	0.0574	3.4200	0.0010	0.0837	0.3090
HH owns one or more cars	0.1983	0.0826	2.4000	0.0170	0.0362	0.3604
HH owns one or more cattle	0.1383	0.0572	2.4200	0.0160	0.0260	0.2507