

**Certification Criteria: Accuracy**  
**USAID/IRIS Tool for Vietnam**  
**Submitted: November 6, 2006**  
**Revised: November 9, 2006**

**1. Please describe the overall approach to the tool development**

The USAID/IRIS project on Developing Poverty Assessment Tools collected new data in four countries—Bangladesh, Peru, Kazakhstan, and Peru—to assess a selected set of indicators against the task of identifying “very poor” households based on household per capita expenditures. A composite survey questionnaire, compiled from several practitioner tools, was administered to a sample of 800 households. A benchmark for assessing measurement accuracy was developed using the expenditure module of the World Bank’s *Living Standards Measurement Survey* (LSMS). Administered to the same set of households exactly fourteen days later, this benchmark provided the best available quantitative information on the “true” poverty status of each sample household. Multiple statistical methods were then used to identify the 15 indicators (for each “step” if a 2-step model) within this composite survey that most accurately reflect the “true” poverty status of each household – that is, that most closely track the benchmark expenditure results. Identification of the most accurate set of indicator and the weights attached to them were done on the basis of criteria developed especially for this project. In addition to the four countries already listed, a comparative analysis drew on existing LSMS data sets from an additional eight countries to identify the 15 best poverty predictors (using a similar methodology and set of variables), to facilitate generalization of findings over a larger number of countries. The eight LSMS countries are: Albania, Ghana, Guatemala, India (Bihar and Uttar Pradesh), Jamaica, Madagascar, Tajikistan, and Vietnam. Thus, statistical testing for accuracy was carried out for twelve countries in total. The 110 indicators that appeared in the ‘best 15’ from at least one of the twelve countries were included in the next part of the project: testing for practicality.

The 110 indicators were divided into six surveys to be tested for practicality. Seventeen microenterprise organizations were selected by USAID to conduct the 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 brought in after the best 15 poverty indicators were determined for each country. If a best 15 indicator caused difficulties in testing, the indicator was dropped for the list and the next best indicator replaced it.

The end result of this development process was a country-specific poverty assessment tool for each of the twelve countries that predicts—rather than directly measures—household per capita expenditure based on a short set of indicators. Each country tool is incorporated into a data entry template that allows microenterprise practitioner to easily enter and store the responses of its sampled clients to indicator questions and will also calculate the percentage of that practitioner’s clients that are predicted to be very poor.

## **2. Please describe the data source used to develop or calibrate the tool.**

Eight of the twelve country tools for this project were developed from existing LSMS data. For the other four countries, original survey data was collected, using both a composite survey consisting of poverty indicators from multiple sources and a benchmark expenditure survey based on the LSMS expenditure module. The sample was selected to be nationally representative.

For Vietnam, the poverty assessment tool was developed from a 1998 LSMS survey on a nationally-representative sample.<sup>1</sup> Economic growth (GDP per capita) in Vietnam from 1998 to present has averaged 5.4 percent per year, while inflation has averaged 4.0 percent per year.<sup>2</sup> While the currency values in the data set have been adjusted for inflation, the pace of economic change in Vietnam raises concerns about the suitability of the IRIS poverty assessment tool. The choice is between a somewhat-outdated tool and a loan size tool with (most likely) strongly inaccurate predictions.<sup>3</sup> Although IRIS cannot definitively prove this without testing the two alternative tools on new data, we would expect that the accuracy performance of a loan size tool is likely so poor that it would be worth certifying the Vietnam country-specific tool. The sample size used for the statistical testing of the tools was 5935 households.

## **3. Please describe the process used to select the indicators included in the tool.**

LSMS data sets typically include variables related to education, housing characteristics, consumer durables, agricultural assets, financial assets, illness and disability, and employment. Roughly ninety indicators from all these 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.<sup>4</sup> This selection of the best model was based on the BPAC and PIE accuracy criteria.

---

<sup>1</sup> Additional information on the data set is available here:  
<http://www.worldbank.org/lsmc/country/vn98/vn98docs.html>

<sup>2</sup> Data available from: <http://www.imf.org/external/pubs/ft/weo/2006/01/data/dbginim.cfm>

<sup>3</sup> When a loan size tool was tested in Bangladesh, its BPAC was strongly negative (-49.5). Please see:  
[http://www.povertytools.org/Project\\_Documents/Bangladesh%20Accuracy%20Report%20Final.pdf](http://www.povertytools.org/Project_Documents/Bangladesh%20Accuracy%20Report%20Final.pdf)

<sup>4</sup> The set of indicators and their weights also depended on the selection of a 1-step or 2-step statistical model.

4. Please describe the estimation methods used to identify final indicators and their weights/coefficients.

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

<b>Vietnam (median)</b> Poverty Line: 1,393,582 Vietn. dong per person per year <sup>5</sup> Poverty Rate: 14.52%	<b>Total Accuracy</b>	<b>Poverty Accuracy</b>	<b>Under-coverage</b>	<b>Leakage</b>	<b>PIE</b>	<b>BPAC</b>
<b>Single-step methods</b>						
OLS	86.31	46.18	53.82	15.97	-7.42	8.33
Quantile regression (estimation point: 38)	<b>87.96</b>	<b>61.83</b>	<b>38.17</b>	<b>38.30</b>	<b>0.02</b>	<b>61.69</b>
Linear Probability	85.95	38.40	61.59	10.05	-10.1	-13.14
Probit	86.49	49.75	50.25	18.59	-6.21	18.09
<b>Two-step methods -- MAXR variable selection</b>						
OLS – 30 percentile cutoff	86.88	53.05	46.95	19.95	-5.29	26.05
Quantile (estimation points: 38, 12.3) 30 percentile cutoff	87.98	63.13	36.87	39.50	0.42	60.49
LP – 30 percentile cutoff	87.26	57.85	42.15	22.83	-3.79	38.53
Probit – 30 percentile cutoff	87.28	54.86	45.14	19.73	-4.98	29.45

**Table 2: Poverty Status of Sample Households, as Estimated by Model and Revealed by the Benchmark Survey**

	<b>Number of households<sup>6</sup> predicted as very poor by the tool</b>	<b>Number of households predicted as not very-poor by the tool</b>
<b>Number of “true” very poor households (as determined by benchmark survey)</b>	590 (9.94%)	345 (5.81%)
<b>Number of “true” not very-poor households (as determined by benchmark survey)</b>	369 (6.22%)	4631 (78.03%)

<sup>5</sup> The poverty line reported above is expressed in the prevailing prices at the time of data collection (December 1998). The poverty line and other monetary variables used in the accuracy results, however, are inflation-adjusted through the time of tool construction (May 2006), to ensure that the resulting poverty predictions are as current as possible. The poverty line for Vietnam, expressed in the most current available prices, is 21,569,181 Vietnamese dong per person per year.

<sup>6</sup> Values in the matrix are weighted for national representation.

**5. Please describe how coefficients and weights are used to compute prediction of poverty status or estimate of household expenditures.**

The weights attached to the indicators in the tool in each country are simply the regression coefficients for the statistical model exhibiting the highest statistical accuracy (according to the BPAC criterion). For Vietnam, the weights are from a 1-step Quantile model.

The weights are located in the “backend” analysis program of the EPI template as part of the extreme poverty rate calculation. While a skilled EPI user would be able to locate these values and potentially manipulate them, they would not be seen by the client or the interviewer during the normal course of interviewing, entering the data, or in calculating the extreme poverty rate.

**6. Please describe the decision rule used to classify households as very poor or not very-poor.**

The extreme poverty line for each of the twelve countries in the project was the higher of the two potential poverty lines specified in the legislation: \$1.08 a day (in PPP terms) OR the bottom half of households living below the poverty line (termed the ‘median’ poverty line). Two of statistical models (OLS and Quantile) used by the IRIS team predict the per capita consumption expenditures for each household, which is then compared to the binding poverty line to decide whether the household is very poor.<sup>7</sup> The other two statistical models (Linear Probability and Probit) predict the probability that a household is very poor (according to the binding, absolute poverty line). If this probability exceeds 0.5, the household is predicted to be very poor.<sup>8</sup>

For Vietnam, the binding poverty line was the ‘median’ line of 1,393,582 Vietnamese dong per person per month. Because the tool is based on a Quantile model, those households whose predicted expenditures according to the tool fall below this line will be considered very poor.

---

<sup>7</sup> For a 2-step OLS or Quantile model, the decision rule in the 1<sup>st</sup>-step compares the expenditures predicted for each household to a certain expenditure cutoff.

<sup>8</sup> For a 2-step Linear Probability or Probit model, the decision rule in the 1<sup>st</sup>-step compares the predicted probability that the households’ expenditures exceed a certain cutoff to the 0.5 value.

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

.38 Quantile regression  
Min sum of deviations 1374.281

Number of obs = 5935  
Pseudo R2 = 0.4571

Variable	Coef.	Std. Err.	t	P> t	[95% Confidence Interval]	
Intercept	8.2812	0.0715	115.7500	0.0000	8.1409	8.4214
Household head age	0.0145	0.0028	5.2700	0.0000	0.0091	0.0200
Household head age squared	-0.0001	0.0000	-5.2400	0.0000	-0.0002	0.0000
Household size	-0.1809	0.0097	-18.6500	0.0000	-0.1999	-0.1619
Household size squared	0.0070	0.0008	8.3000	0.0000	0.0053	0.0086
Household lives in rural area	-0.1428	0.0165	-8.6700	0.0000	-0.1750	-0.1105
Household lives in Red River Delta	-0.1466	0.0187	-7.8300	0.0000	-0.1833	-0.1099
Household lives in Northeast	-0.2536	0.0201	-12.6100	0.0000	-0.2930	-0.2142
Household lives in Northwest	-0.4876	0.0381	-12.7900	0.0000	-0.5623	-0.4129
Household lives in North Central Coast	-0.1676	0.0200	-8.3600	0.0000	-0.2068	-0.1283
Household lives in South Central Coast	-0.1469	0.0203	-7.2300	0.0000	-0.1867	-0.1070
Household lives in Central Highlands	-0.2548	0.0273	-9.3300	0.0000	-0.3083	-0.2013
Household lives in Mekong River Delta	-0.0525	0.0182	-2.8900	0.0040	-0.0882	-0.0168
Number of members with no education	-0.3267	0.0344	-9.4900	0.0000	-0.3942	-0.2593
Number of rooms in dwelling	0.0415	0.0054	7.6700	0.0000	0.03088	0.05209
Roof is made of thatch	-0.0863	0.0170	-5.0900	0.0000	-0.1195	-0.0530
Main cooking fuel is gas	0.2086	0.0257	8.1200	0.0000	0.1582	0.2589
Main cooking fuel is other	-0.0889	0.0158	-5.6200	0.0000	-0.1199	-0.0579
Main source of lighting is kerosene	-0.1063	0.0164	-6.4800	0.0000	-0.1385	-0.0741
Toilet is no toilet	-0.0741	0.0142	-5.2200	0.0000	-0.1019	-0.0462
Household owns one or more radios or radio receivers	0.0453	0.0110	4.1200	0.0000	0.0238	0.0668
Household owns one or more televisions	0.1844	0.0125	14.7700	0.0000	0.1599	0.2089
Household owns one or more refrigerators or freezers	0.2169	0.0226	9.5900	0.0000	0.1726	0.2612
Household owns one or more gas stoves, electric stoves, rice cookers, or pressure cookers	0.1890	0.0170	11.1500	0.0000	0.1558	0.2222
Household owns one or more motorbikes	0.3284	0.0150	21.8800	0.0000	0.2990	0.3578
Number of chickens owned	0.0023	0.0005	4.6600	0.0000	0.0013	0.0033
Total area of land owned	9.17E-06	1.10E-06	8.3600	0.0000	7.02E-06	1.13E-05
Household managed agricultural or forestry land	-0.0965	0.0141	-6.8500	0.0000	-0.1241	-0.0689