Poverty Assessment Tool Accuracy Submission
USAID/IRIS Tool for West Bank
Submitted: September 28, 2009
Revised to correct median household issue: September 30, 2009

The following report is divided into five sections. Section 1 describes the data set used to create the Poverty Assessment Tool for West Bank. 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 West Bank, existing data from the 2007 Palestine Expenditure and Consumption Survey (PECS) integrated survey were used to construct the poverty assessment tool. The full sample of 1,231 households is statistically representative of the population of West Bank and Gaza. However, the prospect of continued unrest in Gaza, coupled with the absence of USAID microenterprise programs there, led IRIS to select only the households residing in the West Bank. This choice was made in consultation with USAID. The final sample used for tool construction comprises 835 households. Given the limited sample size, out-of-sample accuracy tests were not conducted.

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, illness and disability, and employment. For West Bank, more than 70 indicators from the categories of education, housing, consumer durables, and agricultural assets were considered. Indicators relating to educational attainment were ultimately dropped for West Bank because the questions from which they were derived were confusing and the resulting data lacked consistency.

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 West Bank is the "median poverty line" – the level of expenditure that divides the poorest half of those living below the official poverty line from the less-poor half of the officially poor. Table 1 summarizes the accuracy results achieved by each of the eight estimation methods in predicting household poverty relative to this poverty line. For West Bank, the most accurate method, on the basis of BPAC, is the 2-step Quantile regression.

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

WEST BANK Median line***** Share of "very poor": 12.9%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	88.33	24.48	75.52	16.53	-7.48	-34.52
Quantile regression (estimation point: 32)	86.10	43.85	56.15	51.46	-0.61	39.17
Linear Probability	88.27	11.19	88.81	3.73	-10.79	-73.90
Probit	89.03	29.12	70.88	15.63	-7.00	-26.13
Two-step methods						
OLS – 36 percentile cutoff	90.07	37.59	62.41	15.92	-5.89	-8.90
Quantile (estimation points: 32, 12) 36 percentile cutoff	**87.70**	**52.24**	**47.76**	**47.45**	**-0.04**	**51.93**
LP – 32 percentile cutoff	89.91	41.46	58.54	21.08	-4.75	3.99
Probit – 32 percentile cutoff	90.47	45.39	54.61	20.54	-4.32	11.32
*Median poverty line is 96.22 Jordanian Dinars per adult equivalent per month. See Section 5 for details.						

For West Bank, the functionality of predicting the poverty rate at another poverty line—in this case, the official poverty line—has been added. When running the analysis routine with the Epi Info template, the user is presented the option to predict the extreme poverty rate (using the median line), the poverty rate (official line), or both. The methodology and the accuracy results for this prediction are discussed in Annex 1.

[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.

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 Epi Info 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 back-end, 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.

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. However, Palestine does not have a PPP value (for 2005 or 1993) and therefore the median poverty line was the default choice.

The median poverty line is the household per capita expenditure value of the 50th percentile below the official poverty line, at the level of prices prevailing when the household survey data were collected. In West Bank, the official poverty line for 2007 was 2,375 Israel New Shekels (NIS) per month for a household of two adults and four children. [4] Using the adult equivalence equation for Palestine, this household is equivalent to 3.31 adults.[5] Therefore, the official poverty line is 717.15 NIS per adult equivalent, or 123.73 Jordanian Dinars.[6] Finally, the median poverty line isolating the poorest half of the 25.7% of households living below the official poverty line is 96.22 Jordanian Dinars per adult equivalent per month. At this value, the median poverty line identifies 12.9% of 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.

[4] http://www.pcbs.gov.ps/Portals/_pcbs/PressRelease/population_dE.pdf.

[5] The adult equivalent equation is (numadult+0.46*numchild)^0.89, with numadult the number of adults in the household and numchild the number of children.

[6] This currency conversion is done to match the expenditure values, using the average 2007 exchange rate. According to the National Statistical Office, the expenditures values in the data set were not corrected for temporal and spatial price differences during survey collection, so we assume average 2007 prices.

Hence the decision rule for West Bank's USAID poverty assessment tool in classifying the "very poor" (and the "not very-poor") is whether that predicted per capita monthly expenditures of a household fall below (or above) the median 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 median poverty line is identified as "very poor," and each household whose estimated per capita consumption expenditures exceeds the median poverty line is identified as "not very-poor."[7]

An additional requirement for using the median poverty line is that the official poverty line on which it depends is actively used by the local government. This appears to be the case in West Bank, where the government uses the official poverty line for poverty monitoring.[8]

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)	56 (6.7%)	52 (6.2%)
Number of "true" not very-poor households (as determined by benchmark survey)	51 (6.1%)	676 (81.0%)

[7] For the 1st step of the 2-step quantile model, the decision rule compares the expenditures predicted for each household to a certain expenditure cutoff (here, the 36th percentile). Households whose expenditures are predicted to be above that cutoff are predicted to be not very-poor and are excluded from the 2nd step regression.
[8] See http://www.pcbs.gov.ps/Portals/_pcbs/PressRelease/population_dE.pdf.

Table 3: Regression Estimates using 2-step Quantile Method for Prediction at the Median Poverty Line

1st Step

.32 Quantile regression Number of obs = 835

Min sum of deviations 252.1093 Pseudo R2 = 0.2661

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Interecept	4.4317	0.3198	13.8600	0.0000	3.8040	5.0593
Household size	-0.1429	0.0242	-5.8900	0.0000	-0.1905	-0.0953
Household size squared	0.0042	0.0014	2.9300	0.0030	0.0014	0.0070
Household head age	0.0313	0.0103	3.0300	0.0030	0.0110	0.0516
Household head age squared	-0.0003	0.0001	-3.1000	0.0020	-0.0005	-0.0001
Household lives in refugee camp	-0.0454	0.0875	-0.5200	0.6040	-0.2172	0.1264
Household lives in rural area	-0.0269	0.0566	-0.4800	0.6350	-0.1379	0.0842
Number of rooms in dwelling	0.0826	0.0198	4.1800	0.0000	0.0438	0.1214
Walls of dwelling are made of cleaned stone	0.2373	0.0562	4.2200	0.0000	0.1269	0.3477
Walls of dwelling are made of stone and cement	0.1317	0.0708	1.8600	0.0630	-0.0071	0.2706
Water network is public water network	0.8014	0.2954	2.7100	0.0070	0.2217	1.3812
Water network is private water system	0.8666	0.3015	2.8700	0.0040	0.2748	1.4585
Sewage system is public sewage system	-1.0168	0.2347	-4.3300	0.0000	-1.4775	-0.5561
Sewage system is cesspit	-1.1877	0.2227	-5.3300	0.0000	-1.6249	-0.7506
Dwelling has no heating source	-0.0704	0.1114	-0.6300	0.5270	-0.2891	0.1482
Main source of energy for water heater is electricity	0.2287	0.0561	4.0700	0.0000	0.1185	0.3389
Household owns one or more cars	0.1555	0.0517	3.0100	0.0030	0.0541	0.2569
Household owns one or more solar boilers	0.1491	0.0537	2.7800	0.0060	0.0438	0.2545
Household owns one or more washing machines	0.0473	0.0931	0.5100	0.6120	-0.1354	0.2300
Household owns one or more bookcases	0.1753	0.0546	3.2100	0.0010	0.0682	0.2825
Household owns one or more televisions	0.1206	0.1248	0.9700	0.3340	-0.1243	0.3655
Household ones one or more videos (VCRs)	0.2097	0.0555	3.7800	0.0000	0.1008	0.3185

2nd STEP (Cutoff: 36 percentile)

.34 Quantile regression Number of obs = 439

Min sum of deviations 119.4187 Pseudo R2 = 0.1712

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	4.2543	0.2485	17.1200	0.0000	3.7659	4.7427
Household size	-0.0980	0.0336	-2.9200	0.0040	-0.1641	-0.0319
Household size squared	0.0018	0.0015	1.2100	0.2280	-0.0011	0.0047
Household head age	0.0053	0.0106	0.5000	0.6190	-0.0156	0.0262
Household head age squared	0.0000	0.0001	-0.4600	0.6460	-0.0002	0.0002
Household lives in refugee camp	0.0520	0.0822	0.6300	0.5270	-0.1095	0.2136
Household lives in rural area	0.1399	0.0499	2.8100	0.0050	0.0419	0.2379
Household head is widowed	-0.1971	0.0971	-2.0300	0.0430	-0.3879	-0.0062
One or more household members are refugees	-0.1779	0.0499	-3.5600	0.0000	-0.2761	-0.0798
Share of household members under than 15 years of age	0.2983	0.2284	1.3100	0.1920	-0.1508	0.7473
Share of household members between 16 and 64 years of age	0.4192	0.2143	1.9600	0.0510	-0.0019	0.8404
Number of rooms in dwelling	0.0826	0.0198	4.1800	0.0000	0.0438	0.1214
Dwelling is owned or rented with furniture	0.1945	0.0948	2.0500	0.0410	0.0082	0.3808
Walls of dwelling are made of stone and cement	0.2262	0.0630	3.5900	0.0000	0.1023	0.3501
Main source of energy for oven is gas	-0.0123	0.0516	-0.2400	0.8120	-0.1138	0.0892
Main source of energy for heating is kerosene	0.2175	0.0994	2.1900	0.0290	0.0221	0.4130
Main source of energy for water heater is electricity	0.0411	0.0713	0.5800	0.5640	-0.0990	0.1812
Household owns one or more cars	0.1674	0.0616	2.7200	0.0070	0.0464	0.2884
Household owns one or more solar boilers	0.1555	0.0500	3.1100	0.0020	0.0571	0.2539
Household owns one or more vacuum cleaners	0.2342	0.0617	3.7900	0.0000	0.1128	0.3555
Household owns one or more bookcases	0.1158	0.0772	1.5000	0.1340	-0.0359	0.2675
Household owns one or more televisions	0.2043	0.0896	2.2800	0.0230	0.0281	0.3805

Annex 1: Poverty Prediction at the Official 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 official poverty line. Following this logic, then, the "poor" ("not poor") in West Bank are defined as those whose predicted expenditures fall below (above) the official poverty line.

Table 4 summarizes the predictive accuracy results for the official poverty line using the Quantile model specification from the median poverty line. The indicators are the same as those in the model for the median 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 official poverty line.

Table 4: Accuracy Results Obtained for Prediction at the Official Poverty Line

West Bank Official Line Share of Poor: 25.7%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Two-step method						
Quantile (estimation points: 37, 20) 46 percentile cutoff	79.23	59.77	40.23	41.86	0.41	58.14

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 Official 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)	126 (15.1%)	85 (10.2%)
Number of "true" not poor households (as determined by benchmark survey)	88 (10.5%)	536 (64.2%)

Table 6: Regression Estimates using 2-step Quantile Method for Prediction at the Official Poverty Line

1st Step

.37 Quantile regression Number of obs = 835

Min sum of deviations 266.0274 Pseudo R2 = 0.2723

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	4.3534	0.3544	12.2800	0.0000	3.6577	5.0491
Household size	-0.1399	0.0292	-4.7900	0.0000	-0.1972	-0.0825
Household size squared	0.0044	0.0019	2.3700	0.0180	0.0008	0.0081
Household head age	0.0338	0.0108	3.1300	0.0020	0.0126	0.0549
Household head age squared	-0.0003	0.0001	-3.1700	0.0020	-0.0005	-0.0001
Household lives in refugee camp	-0.1104	0.0916	-1.2100	0.2280	-0.2902	0.0694
Household lives in rural area	-0.0356	0.0589	-0.6000	0.5460	-0.1512	0.0801
Number of rooms in dwelling	0.0680	0.0202	3.3600	0.0010	0.0283	0.1076
Walls of dwelling are made of cleaned stone	0.2483	0.0582	4.2700	0.0000	0.1341	0.3626
Walls of dwelling are made of stone and cement	0.1554	0.0741	2.1000	0.0360	0.0100	0.3008
Water network is public water network	0.8269	0.3386	2.4400	0.0150	0.1622	1.4916
Water network is private water system	0.8668	0.3446	2.5200	0.0120	0.1905	1.5431
Sewage system is public sewage system	-1.0076	0.2613	-3.8600	0.0000	-1.5206	-0.4946
Sewage system is cesspit	-1.1885	0.2506	-4.7400	0.0000	-1.6803	-0.6966
Dwelling has no heating source	-0.1154	0.1121	-1.0300	0.3040	-0.3355	0.1047
Main source of energy for water heater is electricity	0.2411	0.0586	4.1100	0.0000	0.1260	0.3561
Household owns one or more cars	0.1455	0.0539	2.7000	0.0070	0.0398	0.2512
Household owns one or more solar boilers	0.1950	0.0564	3.4600	0.0010	0.0843	0.3056
Household owns one or more washing machines	0.0483	0.0979	0.4900	0.6220	-0.1439	0.2405
Household owns one or more bookcases	0.1717	0.0565	3.0400	0.0020	0.0608	0.2826
Household owns one or more televisions	0.1797	0.1296	1.3900	0.1660	-0.0747	0.4341
Household ones one or more videos (VCRs)	0.1787	0.0574	3.1100	0.0020	0.0660	0.2913

2nd STEP (Cutoff: 46 percentile)

.43 Quantile regression Number of obs = 482

Min sum of deviations 141.6862 Pseudo R2 = 0.1547

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Intercept	4.5271	0.2068	21.8900	0.0000	4.1207	4.9336
Household size	-0.1018	0.0306	-3.3300	0.0010	-0.1619	-0.0417
Household size squared	0.0028	0.0015	1.8500	0.0640	-0.0002	0.0058
Household head age	0.0025	0.0087	0.2900	0.7700	-0.0145	0.0196
Household head age squared	0.0000	0.0001	-0.4100	0.6810	-0.0002	0.0001
Household lives in refugee camp	0.0967	0.0639	1.5100	0.1310	-0.0288	0.2222
Household lives in rural area	0.1015	0.0395	2.5700	0.0100	0.0239	0.1791
Household head is widowed	-0.1924	0.0793	-2.4300	0.0160	-0.3482	-0.0367
One or more household members are refugees	-0.1721	0.0397	-4.3400	0.0000	-0.2500	-0.0942
Share of household members under than 15 years of age	0.2636	0.1830	1.4400	0.1500	-0.0960	0.6232
Share of household members between 16 and 64 years of age	0.3873	0.1675	2.3100	0.0210	0.0582	0.7164
Number of rooms in dwelling	0.0517	0.0163	3.1800	0.0020	0.0197	0.0836
Dwelling is owned or rented with furniture	0.1562	0.0757	2.0600	0.0400	0.0074	0.3049
Walls of dwelling are made of stone and cement	0.2078	0.0510	4.0800	0.0000	0.1077	0.3080
Main source of energy for oven is gas	0.0380	0.0416	0.9100	0.3610	-0.0437	0.1197
Main source of energy for heating is kerosene	0.1748	0.0739	2.3700	0.0180	0.0296	0.3200
Main source of energy for water heater is electricity	-0.0117	0.0561	-0.2100	0.8350	-0.1220	0.0986
Household owns one or more cars	0.2025	0.0457	4.4300	0.0000	0.1127	0.2924
Household owns one or more solar boilers	0.1276	0.0387	3.2900	0.0010	0.0515	0.2037
Household owns one or more vacuum cleaners	0.2861	0.0468	6.1200	0.0000	0.1942	0.3780
Household owns one or more bookcases	0.0872	0.0566	1.5400	0.1240	-0.0240	0.1984
Household owns one or more televisions	0.1934	0.0756	2.5600	0.0110	0.0450	0.3419