VOL. 7 NUMBER 1 FIRST SEMESTER 2011

12

Well-being and Social Policy

Well-being and Social Policy

Edwin van Gameren — THE HEALTH INSURANCE REFORM IN THE NETHERLANDS AND ITS RELEVANCE FOR MEXICO

Nelly Aguilera, John Scott — FEASIBILITY ANALYSIS OF AN INTEGRATED HEALTH SYSTEM: FINANCIAL AND DISTRIBUTIVE IMPLICATIONS

Mark Schreiner — A SIMPLE POVERTY SCORECARD FOR MEXICO

Cristina Gutiérrez — REVIEW OF THE 2010 WORLD HEALTH REPORT. HEALTH SYSTEM FINANCING

VOL. 7 NUMBER 1 FIRST SEMESTER 2011

CISS

WELL-BEING AND SOCIAL POLICY
VOL. 7 NUMBER 1, pp. 47-85

A SIMPLE POVERTY SCORECARD FOR MEXICO

Mark Schreiner*
Microfinance Risk Management, L.L.C.
Center for Social Development, Washington University in Saint Louis
mark@microfinance.com

Abstract

This study uses Mexico's 2008 National Household Survey of Income and Expenditure to construct an easy-to-use scorecard that estimates the likelihood that a household has income below a given poverty line. The scorecard uses ten simple indicators that field workers can quickly collect and verify. Poverty scores can be computed on paper in the field in about five to ten minutes. The scorecard's accuracy and precision are reported for a range of poverty lines. The poverty scorecard is a practical way for local pro-poor programs in Mexico to monitor poverty rates, track changes in poverty rates over time, and target services.

Key Words: Poverty measurement, proxy-means tests, targeting, scoring, expenditure surveys.
JEL Classification: C53, D31, I32.

Introduction

This paper presents an easy-to-use poverty scorecard that local pro-poor programs in Mexico can use to estimate the likelihood that a household has income below a given poverty line, to monitor groups' poverty rates at a point in time, to track changes in groups' poverty rates between two points in time, and to target services to households.

The direct approach to poverty measurement via surveys is difficult and costly, asking households about a lengthy list of items. As a case in point, Mexico's 2008 National Household Survey of Income and Expenditure (ENIGH, for its initials in Spanish) runs 212 pages.

In contrast, the indirect approach via poverty scoring is simple, quick, and inexpensive (Figure 1). It uses ten verifiable indicators (such as "What fuel do you usually use to cook or heat food?" and "How many televisions does the household have?") to get a score that is highly correlated with poverty status as measured by income from the exhaustive survey.

---

* Mark Schreiner is the Director of Microfinance Risk Management, L.L.C., mark@microfinance.com. He is also Senior Scholar, Center for Social Development, Washington University in Saint Louis.
Acknowledgements: This paper is based on work funded by the Grameen Foundation with a grant from the Ford Foundation. Data come from Mexico's *Instituto Nacional de Estadística, Geografía e Informática*. Thanks go to Nigel Biggar, Frank DeGiovanni, Pilar Quintanilla, Tony Sheldon, and Jeff Toohig. More extensive documentation and scorecard-management tools are available at http://www.microfinance.com/#Mexico.

The poverty scorecard here differs from "proxy means tests" (Coady, Grosh, and Hoddinott 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for these local organizations are typically subjective and relative (such as participatory wealth ranking by skilled field workers) or blunt (such as rules based on land-ownership or housing quality). These approaches may be costly, their results are not comparable across organizations nor across countries, and their accuracy and precision are unknown.

The poverty scorecard here differs from that used by Mexico's *Oportunidades* program in that it is uses data and income-based poverty lines, its formula is public knowledge, and its accuracy is known. While the scorecard here is not meant to replace that of *Oportunidades*, its design makes it a practical tool for local pro-poor organizations who want to improve their social performance management.

The approach here aims to be simple, inexpensive, and understandable for non-specialists. After all, if program managers are to adopt poverty scoring on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Thanks to the predictive-modeling phenomenon known as the "flat max", simple scorecards are about accurate as complex ones.

The technical approach here is innovative in how it associates scores with poverty likelihoods, in the extent of its accuracy tests, and in how it derives formulas for standard errors.

The scorecard is based on the 2008 ENIGH conducted by Mexico's *Instituto Nacional de Estadística, Geografía e Informática.* Indicators are selected to be inexpensive to collect, easy to answer quickly, simple to verify, strongly correlated with poverty, and liable to change over time as poverty status changes.

All points in the scorecard are non-negative integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialists can collect data and tally scores on paper in the field in five to ten minutes.

Poverty scoring can be used to estimate three basic quantities. First, it can estimate a particular household's "poverty likelihood", that is, the probability that the household has per-capita income below a given poverty line.

Second, poverty scoring can estimate the poverty rate of a group of households at a point in time. This is simply the average poverty likelihood among the households in the group.

Third, poverty scoring can estimate changes in the poverty rate for a given group of households (or for two independent representative samples of households from the same population) between two points in time. This estimate is the change in the average poverty likelihood of the group(s) of households over time.

Poverty scoring can also be used for targeting services to poorer households. To help program managers choose a targeting cut-off, this paper reports several measures of targeting accuracy for a range of possible cut-offs.

This paper presents a single scorecard (Figure 1) whose indicators and points are derived from household income data and Mexico's national asset poverty line. Scores from this single scorecard are calibrated to poverty likelihoods for eight poverty lines.

Figure 1
Simple Poverty Scorecard for Mexico

Entity	**Name**	**ID**	**Date (DD/MM/YY)**	
Member:			**Joined:**	
Loan officer:			**Today:**	
Branch:			**Household size:**	

Indicator	Value	Points
1. How many household members are ages 0 to 17?	A. Four or more	0
	B. Three	7
	C. Two	11
	D. One	20
	E. None	28
2. What is the highest level that the female head/ spouse has passed in school?	A. None	0
	B. Up to third grade	5
	C. Fourth grade through high school	7
	D. College preparatory 1–3	10
	E. Normal/technical/commercial/no female head/spouse	14
	F. Professional, master's, or doctorate	20
3. How many household members have a written employment contract for a salary or for an indefinite period?	A. None	0
	B. One	6
	C. Two or more	16
4. What is the main material of the floor of this residence?	A. Dirt	0
	B. Cement/concrete	2
	C. Other	7
5. How is water supplied to the residence's toilet for flushing?	A. No toilet, or no water supply	0
	B. Carried by bucket	1
	C. Piped	3
6. Does the residence have a medium sink for washing dishes?	A. No	0
	B. Yes	4
7. What fuel do you usually use to cook or heat food?	A. Firewood	0
	B. Other	2
8. Does the household have a blender?	A. No	0
	B. Yes	4
9. Does the household have an electric iron?	A. No	0
	B. Yes	4
10. How many televisions does the household have?	A. One or none	0
	B. Two	5
	C. Three or more	12

The scorecard is constructed and calibrated using a sub-sample of the data from the 2008 ENIGH. Its accuracy is then validated on a different sub-sample from the 2008 ENIGH as well as on the entire 2004, 2005, and 2006 ENIGH surveys.[1] While all three scoring estimators are unbiased when applied to the population from which they were derived (that is, they match the true value on average in repeated samples from the same population from which the scorecard was built), they are—like all predictive models—biased to some extent when applied to a different population.[2]

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased in practice. (The direct survey approach is unbiased by definition.) There is bias because scoring must assume that the future relationships between indicators and poverty will be the same as in the data used to build the scorecard. It must also assume that these relationships will be the same in all sub-groups as in the population as a whole.[3] Of course, these assumptions—ubiquitous and inevitable in predictive modeling—hold only partly.

When applied to the 2008 validation sample for Mexico with n =16,384 and using the national capacity poverty line, the difference between scorecard estimates of groups' poverty rates and the true rates at a point in time is –0.9 percentage points.[4] For n =16,384, the 90% confidence intervals for these estimates are +/– 0.6 percentage points or less

When the scorecard built from the 2008 construction and calibration samples is applied both to the 2008 validation sample and to the entire 2006, 2005, and 2004 ENIGH with n =16,384 to measure change between two points in time, the difference between scorecard estimates and true values for changes in groups' poverty rates is much larger. This is because the scorecard assumes, for example, that August 2006 (before the recent economic crisis) is the same as August 2008 (during the crisis). This is an inherent property of scorecards; they must assume no structural change through time. Thus, when measuring change from now on, the accuracy of the scorecard will depend on how closely the time period resembles August 2008.

Section 1 below documents data and poverty lines. Section 2 places the new scorecard in the context of existing poverty maps for Mexico. Section 3 describes scorecard construction. Section 4 discusses practical considerations, and Sections 5 and 6 detail the estimation of households' poverty likelihoods and of groups' poverty rates at a point in time. Section 7 discusses estimating changes in poverty rates, and Section 8 covers targeting. The final section is a summary.

---

[1] Accuracy is not tested with the 2002 ENIGH because its question about educational attainment is incompatible with later surveys.

[2] Examples of "different populations" include nationally representative samples at another point in time or a non-representative sub-group (Tarozzi and Deaton 2007).

[3] Bias may also result from changes in the quality of data collection over time, from changes over time in the real value of poverty lines, from imperfect adjustment of poverty lines to account for differences in cost-of-living across time or geographic regions, or from sampling variation across surveys.

[4] Because the 2008 validation sample is representative of the same population as the data that was used to construct the scorecard, and because all the data come from the same time frame, the scorecard estimators are unbiased and these observed differences are due to sampling variation; the average difference would be zero if the whole 2008 ENIGH were to be repeatedly redrawn and divided into sub-samples before repeating the entire scorecard-building and accuracy-testing process.

1. Data and Poverty Lines

The scorecard is based on data from the 2008 ENIGH.[5] Households are randomly divided into three sub-samples (Table 1):

- Construction for selecting indicators and points
- Calibration for associating scores with poverty likelihoods
- Validation for measuring accuracy on data not used in construction or calibration

In addition, the 2004, 2005, and 2006 ENIGH are used to validate estimates of changes in poverty rates between two points in time.

The poverty scorecard is constructed using household-level weights. In turn, scores are calibrated to household-level poverty likelihoods, and accuracy is measured for household-level rates. Organizations can estimate person-level poverty rates by taking a household-size-weighted average of the household-level poverty likelihoods.

Table 2 reports poverty rates and poverty lines by urban/rural and for all-Mexico at both the household level and the person level based on the complete 2002, 2004, 2005, 2006, and 2008 ENIGH.

Mexico's three national poverty lines are defined in terms of income (*Comité Técnico para la Medición de la Pobreza 2002*). Usually, expenditure—not income—is preferred for measuring poverty (Deaton and Zaidi 2002). In Mexico, however, income and expenditure both give about the same poverty rates and about the same changes in poverty rates over time (de la Torre 2005). Furthermore, de la Torre points out that income tracks expenditure closely for households in the poorest four deciles. Finally, the measure of income in Mexico includes the value of self-produced/self-consumed goods as well as the rental value of owner-occupied housing, two values that are usually omitted from income and whose omission from income accounts for much of the typical preference for expenditure. All this suggests that for poverty measurement in Mexico, income is more or less equivalent to expenditure.

The first national poverty line represents the income required for basic nutrition. In 2008, this "food" line is MXN31.21/person/day in urban areas and MXN23.23 in rural areas, implying household-level poverty rates of 8.2% and 26.3% (Table 2).

[5] In 2008, the average surveyed household represented about 900 households. To prevent the breakdown of bootstrap estimates (see Singh, 1998), 222 households who each represented more than 5,000 households were omitted for 2008, 108 for 2006, 180 for 2005, and 216 for 2004. Furthermore, before random assignment to sub-samples, remaining households representing more than 4,000 households were duplicated and their weights divided by two. The newly replicated pair together represent the same number of households as the original heavily weighted household. Replication helps spread heavily weighted households across the construction, calibration, and validation sub-samples, which in turn reduces the influence of any single household on scorecard construction, calibration, and validation. This does not affect the unbiasedness of scoring estimators, but it does increase precision and thus decreases the difference between estimates and true values in a given sample.

Table 1
Sample Sizes and Household Poverty Rates by Sub-sample, Survey Round, and Poverty Line

			% with income below a poverty line							
									International 2005 PPP	
Sub-sample	**Round**	**Households**	**National Food**	**National Capacity**	**National Asset**	**125% Natl. Asset**	**150% Natl. Asset**	**USAID 'Extreme'**	**$1.25/day**	**$2.50/day**
All Mexico	2008	29,403	14.4	20.3	40.7	51.6	60.4	19.2	1.6	8.5
	2006	20,480	10.4	15.8	34.9	46.3	55.3	16.8	1.2	6.7
	2005	22,894	14.0	19.5	39.5	50.6	60.0	18.9	2.2	9.4
	2004	22,130	13.9	20.3	40.7	51.6	60.4	20.0	2.6	9.5
Construction										
Selecting indicators and weights	2008	9,768	14.2	20.1	40.6	51.5	60.1	18.9	1.5	8.2
Calibration										
Associating scores with likelihoods	2008	9,785	14.0	20.0	40.5	51.7	60.5	18.9	1.5	8.1
Validation										
Measuring accuracy	2008	9,850	15.0	21.0	40.9	51.8	60.6	19.9	1.8	9.2
Change in poverty rate (percentage points)										
From 2008 construction/ calibration to 2008 validation			–0.9	–0.9	–0.4	–0.2	–0.3	–1.0	–0.3	–1.0
From 2008 validation to 2006 for all Mexico			+4.7	+5.2	+6.1	+5.5	+5.3	+3.1	+0.6	+2.5
From 2008 validation to 2005 for all Mexico			+1.0	+1.4	+1.4	+1.2	+0.6	+0.9	–0.4	–0.2
From 2008 validation to 2004 for all Mexico			+1.1	+0.7	+0.2	+0.1	+0.2	–0.1	–0.8	–0.3

Source: ENIGH, after removing most heavily weighted cases and breaking up other heavily weighted cases.

The second national poverty line (the *capacidades* or capacity line) is the food line plus the income required for basic education and health care. This is the relevant line for *Oportunidades*. In 2008, the capacity line is MXN38.28/person/day in urban areas and MXN27.47 in rural areas, giving household-level poverty rates of 13.6% and 32.7%.

Finally, the third national poverty line (the *patrimonio* or asset line) is the capacity line plus the income required for clothing, shoes, housing, and transportation. The scorecard in this paper is constructed using this asset line. In 2008, this line is MXN62.62/person/day in urban areas and MXN42.16 in rural areas, giving household-level poverty rates of 33.3% and 53.6%.

For Mexico as a whole, poverty rates for these three lines increased an average of about 3.7 percentage points from 2006 to 2008, decreased an average of about 3.3 percentage points from 2005 to 2006, and decreased an average of about 0.6 percentage points from 2004 to 2005 (Table 1). The economic crisis in 2008 put poverty rates in Mexico more or less back to 2004.

Because local pro-poor organizations may want to use different or various poverty lines, this paper calibrates scores from its single scorecard to poverty likelihoods for eight lines:[6]

- National food
- National capacity
- National asset
- 125% of national asset
- 150% of national asset
- USAID (United States Agency for International Development) "extreme"
- USD1.25/day 2005 PPP (Purchasing Parity Power)
- USD2.50/day 2005 PPP

The 125% asset line and the 150% asset line are self-explanatory. The USAID "extreme" line is defined as the median income of people (not households) below the national line (U.S. Congress, 2002). The USD1.25/day line (2005 PPP) is derived from:

- 2005 PPP exchange rate for "individual consumption expenditure by households" (International Comparison Project, 2008): MXN7.65 per $1.00
- Price indices from *Banco de México*:[7] 100.3315 for August 2002, 111.4331 for August 2004, 115.2967 for August 2005, 120.1828 for August 2006, and 134.7458 for August 2008, along with 116.3710 for 2005 on average

[6] This paper focuses on the capacity line. Schreiner (2009a) gives details for all lines.
[7] http://www.banxico.org.mx/polmoneinflacion/estadisticas/indicesPrecios/indicesPreciosConsumidor.html, accessed 31 July 2009, basic bundle.

Using the formula in Sillers (2006), the USD1.25/day 2005 PPP line for Mexico as a whole in August 2008 is MXN11.14 per person per day, and the USD2.50/day line is twice that. These all-Mexico lines are adjusted for urban/rural differences in cost-of-living as implicitly reflected in the national food poverty lines, using population weights.

2. The Context of Poverty Scorecards for Mexico

This section discusses existing Mexico scorecards in terms of their goals, methods, poverty lines, indicators, accuracy, and precision. All the scorecards use indicators that are simple, inexpensive to collect, and verifiable. The relative strengths of the new scorecard are:

- Its estimates are tested out-of-sample, and accuracy, precision, and formulas for standard errors are reported
- It is based on the largest sample and on the latest nationally representative data
- Its accuracy is good enough for most of its likely purposes

2.1 López-Calva et al.

López-Calva et al. (2005) use poverty scorecards to construct a "poverty map" (Elbers, Lanjouw, and Lanjouw 2003) of average income at the level of Mexico's municipalities. Their goal is to improve the policy process through detailed information on the geographic distribution of poverty.

López-Calva et al. build ten scorecards (urban and rural for five groupings of states) using stepwise ordinary least squares on the logarithm of per-capita income for households in the 2000 ENIGH, using only indicators also in the 2000 census.

They then apply the resulting scorecards to households in the 2000 census to estimate average income levels by municipality and state. These estimates are more precise at these levels than would be possible with only ENIGH data. López-Calva et al. then make "poverty maps" that quickly show how average income varies across municipalities in a way that makes sense to lay people.

Poverty mapping and poverty scoring are similar in that they both:

- Build scorecards with nationally representative survey data and then apply them to other data on sub-groups that may not be nationally representative
- Use simple, verifiable indicators that are quick and inexpensive to collect
- Provide unbiased estimates when their assumptions hold
- Can be used to estimate poverty rates for groups
- Seek to be useful in practice and so aim to be understood by non-specialists

Table 2
Poverty Lines and Poverty Rates by Survey Round and by Urban/Rural/All Mexico

			Poverty line (MXN/person/day) and poverty rate (%) at the household level							
		Line or Rate	National Food	National Capacity	National Asset	125% Natl. Asset	150% Natl. Asset	USAID 'Extreme'	International 2005 PPP $1.25/day	International 2005 PPP $2.50/day
Urban	2002	Line	22.10	27.11	44.35	55.43	66.52	29.68	9.17	18.34
		Rate	8.5	13.2	34.5	46.7	55.9	16.1	0.4	4.8
	2004	Line	24.32	29.82	48.79	60.98	73.18	32.35	10.14	20.28
		Rate	8.7	14.2	34.3	45.8	55.0	16.5	1.0	5.3
	2005	Line	26.00	31.89	52.16	65.20	78.24	34.56	10.49	20.98
		Rate	7.7	12.4	32.0	43.8	53.7	15.2	0.7	4.1
	2006	Line	26.63	32.66	53.42	66.78	80.13	36.53	10.92	21.85
		Rate	5.9	10.6	29.3	40.7	50.0	14.0	0.5	3.3
	2008	Line	31.21	38.28	62.62	78.28	93.94	41.40	12.20	24.39
		Rate	8.2	13.6	33.3	44.3	53.9	15.9	0.7	4.2
Rural	2002	Line	26.62	19.23	29.52	36.90	44.28	15.71	6.75	13.50
		Rate	27.8	35.4	56.0	65.8	73.7	26.5	4.0	20.6
	2004	Line	18.02	21.31	32.70	40.88	49.05	18.20	7.52	15.03
		Rate	22.9	29.9	49.3	61.9	69.3	23.3	5.0	16.6
	2005	Line	19.21	22.71	34.86	43.58	52.29	18.70	7.75	15.50
		Rate	26.1	32.9	53.9	64.6	73.0	24.9	4.9	18.5
	2006	Line	19.68	23.27	35.72	44.65	53.58	20.94	8.08	16.15
		Rate	19.5	26.5	47.2	59.1	68.1	22.1	2.7	13.7
	2008	Line	23.23	27.47	42.16	52.70	63.24	22.54	9.08	18.16
		Rate	26.3	32.7	53.6	64.5	71.6	24.8	3.3	16.8
All Mexico	2002	Line	19.95	24.20	38.88	48.60	58.32	24.53	8.28	16.55
		Rate	15.6	21.4	42.4	53.7	62.5	19.9	1.7	10.6
	2004	Line	22.04	26.74	42.97	53.71	64.46	27.24	9.19	18.38
		Rate	13.8	19.9	39.7	51.6	60.2	19.0	2.4	9.4
	2005	Line	23.65	28.71	46.18	57.72	69.26	29.07	9.54	19.08
		Rate	14.1	19.4	39.6	51.0	60.4	18.5	2.2	9.1
	2006	Line	24.23	29.42	47.32	59.15	70.97	31.15	9.94	19.88
		Rate	10.6	16.1	35.5	47.1	56.2	16.8	1.2	6.9
	2008	Line	28.52	34.63	55.71	69.64	83.56	35.02	11.14	22.28
		Rate	14.3	20.1	40.2	51.2	59.9	18.9	1.6	8.5

Table 2 (continued)

		Line or Rate	Poverty line (MXN/person/day) and poverty rate (%) at the person level							
			National Food	National Capacity	National Asset	125% Natl. Asset	150% Natl. Asset	USAID 'Extreme'	International 2005 PPP $1.25/day	International 2005 PPP $2.50/day
Urban	2002	Line	22.10	27.11	44.35	55.43	66.52	29.68	9.17	18.34
		Rate	11.3	17.2	41.1	53.8	63.0	20.6	0.5	6.4
	2004	Line	24.32	29.82	48.79	60.98	73.18	32.35	10.14	20.28
		Rate	11.0	17.8	41.1	53.2	62.5	20.6	1.0	6.5
	2005	Line	26.00	31.89	52.16	65.20	78.24	34.56	10.49	20.98
		Rate	9.9	15.8	38.3	51.1	61.2	19.1	0.7	5.4
	2006	Line	26.63	32.66	53.42	66.78	80.13	36.53	10.92	21.85
		Rate	7.5	13.6	35.6	48.4	57.9	17.8	0.5	4.2
	2008	Line	31.21	38.28	62.62	78.28	93.94	41.40	12.20	24.39
		Rate	10.6	17.2	39.8	51.8	61.6	19.9	0.8	5.3
Rural	2002	Line	26.62	19.23	29.52	36.90	44.28	15.71	6.75	13.50
		Rate	34.0	42.6	64.3	73.6	80.5	32.1	5.0	25.5
	2004	Line	18.02	21.31	32.70	40.88	49.05	18.20	7.52	15.03
		Rate	28.0	36.2	57.4	69.5	75.8	28.7	6.6	20.9
	2005	Line	19.21	22.71	34.86	43.58	52.29	18.70	7.75	15.50
		Rate	32.3	39.8	61.8	72.1	79.3	30.9	6.6	23.5
	2006	Line	19.68	23.27	35.72	44.65	53.58	20.94	8.08	16.15
		Rate	24.5	32.7	54.6	66.6	75.6	27.3	3.7	17.6
	2008	Line	23.23	27.47	42.16	52.70	63.24	22.54	9.08	18.16
		Rate	31.8	39.1	60.8	70.9	77.2	30.4	4.6	20.9
All Mexico	2002	Line	19.87	24.10	38.68	48.36	58.03	24.35	8.24	16.49
		Rate	20.0	26.9	50.0	61.4	69.7	25.0	2.2	13.7
	2004	Line	21.96	26.63	42.76	53.45	64.14	27.05	9.16	18.31
		Rate	17.4	24.7	47.2	59.3	67.5	23.6	3.1	11.9
	2005	Line	23.48	28.48	45.74	57.18	68.61	28.68	9.47	18.95
		Rate	18.2	24.7	47.0	58.9	67.9	23.5	2.9	12.1
	2006	Line	24.07	29.21	46.92	58.64	70.37	30.80	9.88	19.75
		Rate	13.8	20.7	42.6	55.0	64.4	21.3	1.7	9.1
	2008	Line	28.34	34.39	55.25	69.06	82.88	34.60	11.07	22.14
		Rate	18.2	25.1	47.4	58.7	67.2	23.7	2.2	10.9

Source: ENIGH, complete sample. All-Mexico figures are population-weighted averages of urban and rural figures.

Strengths of the poverty mapping approach include that it:

- Has been published in top journals such as *Econometrica*
- Can be applied straightforwardly to measures of well-being beyond poverty rates
- Requires less data for scorecard construction and calibration
- Includes community-level indicators
- Uses only indicators that appear in a census

Strengths of poverty scoring include that it:

- Is simpler in terms of both construction and application
- Tests accuracy empirically
- Associates poverty likelihoods with scores non-parametrically
- Uses judgment and theory in scorecard construction to reduce overfitting
- Estimates poverty likelihoods for individual households
- Reports simple formulas for standard errors

The basic difference between the two approaches is that poverty mapping seeks to help governments design pro-poor policies, while poverty scoring seeks to help small, local pro-poor organizations to manage their outreach when implementing policies.[8]

Because the census measure of income differs from that in ENIGH, López-Calva et al. cannot test accuracy out-of-sample, that is, using different data than was used to construct the scorecard. Also, even though a central strength of poverty mapping (like poverty scoring) is that provides estimates of standard errors, López-Calva et al. do not report them.

The poverty maps in López-Calva et al. stand out because they are actually being used, informing federal budget distributions to local governments, leading to targeted interventions in the 50 poorest municipalities, and generally increasing awareness and improving the quality of the public debate on poverty and policy (López-Calva, Rodríguez-Chamussy, and Székely 2007). Still, the complexity of the methods has hindered understanding and thus slowed acceptance.[9]

[8] Another difference is that the developers of poverty mapping say that it is too inaccurate for targeting individual households or persons, while Schreiner (2008a) supports targeting as a legitimate, potentially useful application of poverty scoring.

[9] CONEVAL (2007) produces poverty maps combining the 2005 ENIGH with the 2005 *II Conteo de Población y Vivienda*, a mid-decade mini-census, but the documentation does not provide enough detail to permit an analysis here.

2.2 Tarozzi and Deaton and Demombynes et al.

A broader debate on the general accuracy of poverty mapping (and by extension, of poverty scoring) has played out against the background of Mexico in Tarozzi and Deaton (2007) and Demombynes et al. (2007).

2.2.1 Tarozzi and Deaton

Tarozzi and Deaton point out that sub-groups in a population (such as a given municipality, or the clients of a given pro-poor organization) may differ from the population in ways that are both linked with poverty and not fully captured by a scorecard. These differences cause estimates based on poverty mapping/scoring to differ from true values, so that reports of accuracy should include not only standard errors but also differences from true values.

Tarozzi and Deaton use Monte Carlo tests to show that sub-group differences can matter. To show that their concern is not merely theoretical, they also use the 2000 Mexico census (the same data source as in López-Calva et al.) to create synthetic household surveys of rural households in Chiapas, Oaxaca, and Veracruz. Using a poverty line of MXN6.57/day/person in 2000 prices (about $1.08/day 1993 PPP, according to Tarozzi and Deaton), they apply poverty mapping to these surveys, generate estimates of poverty rates, and compare the estimates out-of-sample to census data.[10] Deaton and Tarozzi were the first to do this for poverty mapping, although it has always been a standard feature of the poverty-scoring approach.

As in the present paper, Tarozzi and Deaton use logit to estimate the likelihood that a household has income below a given poverty line. They use 35 indicators and report bias, standard errors, and a formula for the standard error of their estimates.

2.2.2 Demombynes et al.

Demombynes et al. (2007) defend the poverty-mapping approach against the critique of Tarozzi and Deaton. They use expenditure data from a census of 20,544 households in some communities served by *Oportunidades* (formerly *Progresa*).

Demombynes et al. draw a series of synthetic surveys, construct a scorecard using a poverty line of MXN5.23/person/day in November 1997 prices, and then apply it to out-of-sample households, comparing estimates with true values. To get large enough "small areas" to test whether sub-groups effects matter, Demombynes et al. join 50 localities at random.[11] They build 10 scorecards with stepwise ordinary least-squares on the log of per-capita expenditure.

[10] Tarozzi (2008) further shows that the poverty-mapping approach, when applied to literacy rates in the 2000 Mexico census, leads to inaccuracies for sub-groups, suggesting that there would probably also be inaccuracies when applying the approach to income or poverty rates.

[11] Tarozzi and Deaton point out that this way of forming "small areas" wipes out most sub-group differences, invalidating tests for such differences.

From Table 4 in Demombynes et al. the average difference between estimated poverty rates and true values across 20 "small areas" with an average sample size of about 1,010 is + 0.7 percentage points, and the average 90% confidence interval for this difference is +/– 0.7 percentage points. For the scorecard here and 150% of the national asset line (the line that gives a poverty rate closest to the 61% in Demombynes et al.'s Table 5), the 2008 scorecard applied to the 2008 validation sample with n =1,024 gives a difference –1.2 percentage points and a 90% confidence interval of +/– 2.3 percentage points (Schreiner 2009a). Thus, Demombynes et al. is more accurate and more precise than the scorecard here, perhaps because they use about 17 indicators instead of 10, about half of which are at the community level (versus none in the scorecard here). Indeed, Demombynes et al. find that community-level indicators reduce standard errors by 41%.

In the end, Demombynes et al. concede that accuracy is reduced when a sub-group is not representative of the population from which the scorecard was built. They also contend, however, that community-level indicators mitigate such inaccuracies. After all, if a sub-group is different, then group-level indicators should help control for these differences, at least to some extent. While acknowledging Tarozzi and Deaton's point, Demombynes et al. conclude that "bias is low" (p. 18) and that community-level indicators "can go a long way" (p. 19) toward mitigating sub-group differences.

Similar conclusions come out of another paper that shares two authors with Demombynes et al. Using data from a census that asked about income in Brazil's state of Minas Gerais, Elbers, Lanjouw, and Leite (2008) state that the poverty-mapping approach "performs reasonably well" (p. 30). They find that differences between estimates and true values are small and that confidence intervals have about the right width. While acknowledging that Deaton and Tarozzi have a point, they conclude that in practice "a hypothetical policy maker, presented with [a poverty map] and its accompanying standard errors, would not come away with a wildly unrealistic picture of the spatial distribution of poverty" (p. 30).

2.2.3 Significance for poverty scoring

In essence, poverty scoring is a simpler version of poverty mapping, designed to be accurate enough to be useful, inexpensive enough to be used by local pro-poor organizations, and straightforward enough for non-specialists to understand and accept. Given this, does the critique of poverty mapping by Deaton—a possible future Nobel Prize winner in economics—mean that poverty scoring is not useful?

In their concluding remarks, Tarozzi and Deaton say (pp. 24–25):

> Overall, we believe that efforts to calculate welfare estimates for small areas . . . are certainly worthwhile, but we also believe that the current literature has not emphasized enough the limitations of the current methodologies and the very strong assumptions that they require in order to allow for meaningful inference. Such limitations must be stressed, and the precision of the estimates should be judged accordingly . . . [Users] should be aware that such maps may be subject to much more uncertainty and error than previously thought.

Tarozzi and Deaton want poverty maps to document not only standard errors but also differences between estimates and true values, as well as broader limitations. This is reasonable; users of any tool need to know what it can and cannot do, and in what contexts. This type of reporting has been standard for poverty scoring since the beginning of 2008. In particular, reporting includes both bias and standard errors and explicitly points out that reported accuracy holds only for sub-groups that are representative of a given country's population at a particular point in time. Thus, poverty scoring and poverty mapping is still potentially useful, even though poverty mapping's limitations were not fully reported.

But is poverty mapping/scoring accurate enough? And for what purpose? Consider, for example, for-profit lenders around the world with billions of dollars at risk in loans underwritten largely via credit-risk scorecards. Not only are these credit-risk scorecards much less accurate for their purposes than poverty maps/scorecards, but they are also subject to the same sub-group critiques. But even though credit-risk scorecards have limitations, they are nevertheless more useful—from a benefit/cost perspective—than alternatives.

The next question is then, what is the benefit of improved decisions versus the cost of improved decision-making? If national governments are targeting funds at the state-level, then the cost of poverty mapping/scoring is probably not worth the benefit; after all, governments probably can already rank states by poverty. If, however, federal governments are targeting funds at lower levels, then they may not know what the poorest entities are (although governments at lower levels should know). In any case, poverty mapping provides an objectivity that will likely favor poorer entities in the budget process, raising awareness among the polity and allowing politicians to deflect accusations of political bias by referring to the poverty map.

In the case of local, pro-poor organizations, no alternative for targeting households compares well with poverty scoring's combination of inexpensiveness, accuracy, and objectivity. Other targeting tools may be more accurate, but they cost more, are less objective, and their accuracy is unquantified. For managing an organization's social performance, scoring's measures of poverty rates are also valuable, showing managers which branches and which field agents serve poorer people, and whether the pro-poor organization as a whole is indeed pro-poor.

In short, no tool is a silver bullet, and poverty mapping and poverty scoring are accurate enough for some uses and not accurate enough for others. A key strength of the two approaches is their ability to report quantitative measures of accuracy.

In most cases, the inaccuracies poverty scoring are likely to be small, relative to the benefit/cost of additional accuracy. Given the alternatives for their purposes, poverty scorecards are usually "good enough for government work".

Also, Tarozzi and Deaton apply their critique only to estimates of poverty rates. Their critique may not to apply as strongly to estimates of changes in poverty rates (which sweep out sub-group fixed effects) or to rankings used for targeting (which tend to be invariant to macroeconomic fluctuations). For example, Schreiner (2006a) finds little degradation for targeting when a single all-Mexico scorecard is applied to urban/rural sub-groups. Of course, on the continuum of sub-groups between urban/rural down to a single household, at some point rankings may become too inaccurate for a given purpose. Still, even relatively inaccurate credit-risk scorecards have proven useful for targeting individual households, and, depending on the context and alternatives, poverty scorecards may likewise turn out to be the best choice for targeting and/or other uses.

2.3 CIMMYT

Bellon et al. (2004, "CIMMYT", the Spanish acronym for the "International Maize and Wheat Improvement Center", the authors' institution) make a poverty map for Mexico using the 2002 ENIGH and the 2000 census which they then test out-of-sample on the 2000 ENIGH. At the time, household-level data from these sources was not being released, so CIMMYT uses only community-level indicators at the level of the municipality. CIMMYT also differ from López-Calva et al. and Deaton and Tarozzi in that they consider expenditure rather than income, consider only rural areas rather than all areas, and use poverty rates with the national food line instead of income levels or a $1.08/day 1993 PPP line.

The scorecard aggregates 16 indicators from the 2002 ENIGH over municipalities to predict the logarithm of the ratio of per-capita household income to the national food line. CIMMYT do not report standard errors, but they do report correlations in levels and ranks based on the out-of-sample test. Their estimated poverty rate is 41.5% while the true rate is 32.4%, so bias is –9.1 percentage points.

Despite the limits on its data, CIMMYT stands out for the relevance of its application. The authors use the poverty map to inform CIMMYT policy as it relates to its mandate by comparing the map to the placement of agricultural test plots (finding that test plots tend to be in flat, fertile areas, while the poor tend to live in sloped, infertile areas), to the distribution of the variety of corn germplasm (finding that the poor are not necessarily the caretakers of genetic diversity), and to the distribution of farm production (finding that the poor grow corn and beans rather than wheat).

2.4 *Oportunidades'* targeting tool

In Mexico, the most important poverty scorecard is that used to verify households' eligibility for *Oportunidades*. While the indicators collected from applicants are public knowledge, and while it is known that the scorecard is built with discriminant analysis, the scorecard formula—for understandable reasons—is secret.

Thus, while it would be highly policy-relevant to compare targeting accuracy for the scorecard here versus that of *Oportunidades*, it is also impossible. Indeed, two papers that purport to evaluate the targeting effectiveness of *Oportunidades* (Medina, Hubert, and Soto 2000; and Skoufias, Davis, and Behrman 1999) in fact only test different statistical approaches to targeting in Mexico in general, because even these authors do not have access to the full *Oportunidades* scorecard.

To be clear, the poverty scorecard here is not a replacement or competitor of *Oportunidades'* scorecard. Rather, it is a feasible alternative for local pro-poor organizations who would like to use targeting and poverty-monitoring tools similar to that which the government uses with *Oportunidades*.

3. Scorecard Construction

For the Mexico scorecard, about 120 potential indicators (Schreiner 2009a) are initially prepared from the 2008 ENIGH in the broad areas of:

- Family composition (such as household size)
- Education (such as school attendance of children)
- Employment (such as number of household members working in agriculture)
- Housing (such as the main construction material of the floors, walls, and roof)
- Ownership of durable goods (such as televisions and refrigerators)

One use of the scorecard is to measure changes in poverty through time. This means that, when selecting indicators and holding other considerations constant, preference is given to more sensitive indicators. For example, ownership of a television is probably more likely to change in response to changes in poverty than is the age of the male head/spouse.

The scorecard itself is built using the national asset poverty line and Logit regression on the construction sub-sample (Table 1).[12] Indicator selection uses both judgment and statistics (forward stepwise, based on "*c*"). The first step is to use Logit to build one scorecard for each candidate indicator. Each scorecard's accuracy is taken as "*c*", a measure of ability to rank by poverty status (SAS Institute Inc. 2004).

One of these one-indicator scorecards is then selected based on several factors, including improvement in accuracy, likelihood of acceptance by users (determined by simplicity, cost of collection, and "face validity" in terms of experience, theory, and common sense), sensitivity to changes in poverty status, variety among indicators, and verifiability.

A series of two-indicator scorecards are then built, each based on the one-indicator scorecard selected from the first step, with a second candidate indicator added. The best two-indicator scorecard is then selected, again based on "*c*" and judgment. These steps are repeated until the scorecard has 10 indicators.[13] The selection of indicators was also informed by feedback on practical considerations from pro-poor organizations in Mexico.

The final step is to linearly transform and round the Logit coefficients into non-negative integers such that total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line).

This algorithm is the Logit analogue to the familiar R^2-based stepwise with least-squares. It differs from naïve stepwise in that the criteria for selecting indicators is acknowledged to include not only statistical accuracy but also judgment and non-statistical factors. The use of non-statistical criteria can improve robustness and helps ensure that indicators are simple and make sense to users.

[12] The estimates here depend on households' estimated poverty ranks, and ranks change little if the scorecard is constructed using, for example, the capacity line.
[13] Logit regression results are available on request.

The single poverty scorecard here applies to all of Mexico. Evidence from Mexico (Schreiner, 2006a), India (Schreiner 2006b), Sri Lanka (Narayan and Yoshida 2005), and Jamaica (Grosh and Baker 1995) suggests that segmenting scorecards by urban/rural does not improve targeting accuracy much, although—as pointed out by Tarozzi and Deaton—it may improve the accuracy of estimated poverty rates.

4. Notes on Scorecard Use

The main challenge of scorecard design is not to squeeze out the last drops of accuracy but rather to improve the chances that scoring is actually used. When scoring projects fail, the reason is not usually technical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to learn to use it properly (Schreiner 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the "flat max" (Hand 2006; Baesens et al. 2003; Kolesar and Showers 1985; Dawes 1979; Wainer 1976). The bottleneck is less technical and more human, not statistics but organizational change management. Accuracy is easier to achieve than adoption.

The scorecard here is designed to encourage understanding and trust so that users will adopt it and use it properly. Of course, accuracy is important, but so are simplicity, ease-of-use, and "face validity". Programs are more likely to collect data, compute scores, and pay attention to the results if, in their view, scoring does not make a lot of "extra" work and if the whole process generally seems to make sense.

To this end, the scorecard here in Figure 1 fits on a single page. The construction process, indicators, and points are simple and transparent. "Extra" work is minimized; non-specialists can compute scores by hand in the field because the scorecard has only 10 indicators (all of them categorical) and simple points (non-negative integers, and no arithmetic beyond addition).

A field worker using the paper scorecard would record participant identifiers, read each question from the scorecard, circle each response and its points, write the points in the far-right column, add up the points to get the total score, implement targeting policy (if any), and deliver the paper scorecard to a central office for data entry and filing.

Of course, field workers must be trained. High-quality outputs require high-quality inputs. If organizations or field workers gather their own data and if they believe that they have an incentive to exaggerate poverty rates (for example, if funders reward them for higher poverty rates), then it is wise to do on-going quality control via data review and random audits (Matul and Kline 2003).[14] IRIS Center (2007) and Toohig (2008) are useful nuts-and-bolts guides for planning, budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and controlling quality.

[14] If a program does not want field workers to know the scorecard's points, then it can erase the points in Figure 1 and compute scores later at the central office.

In particular, while collecting scorecard indicators is relatively easier than alternatives, it is still absolutely difficult. Training and explicit definitions of terms and concepts in the scorecard are essential.[15] For the example of Nigeria, one study finds distressingly low inter-rater and test-retest correlations for indicators as seemingly simple and obvious as whether the household owns an automobile (Onwujekwe, Hanson, and Fox-Rushby 2006). At the same time, Grosh and Baker (1995) find that gross underreporting of assets does not affect targeting.

For the first stage of targeting in Mexico's *Oportunidades*, Martinelli and Parker (2007) find that "underreporting [of asset ownership] is widespread but not overwhelming, except for a few goods . . . [and] overreporting is common for a few goods, which implies that self-reporting may lead to the exclusion of deserving households" (pp. 24–25). Still, most over- or underreports can be corrected by field agents who verify responses in a home visit. In fact, this is what *Oportunidades* does in the second stage of their targeting process, and this is the suggested procedure for the poverty scoring approach here.

5. Estimates of Household Poverty Likelihoods

The sum of scorecard points for a household is called the score. For Mexico, scores range from 0 to 100. While higher scores indicate less likelihood of being below a poverty line, the scores themselves have only relative units.

To get absolute units, scores must be converted to poverty likelihoods, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the example of the national capacity line with the 2008 ENIGH, scores of 25–29 have a poverty likelihood of 49.4% (Table 3). That is, among Mexican households who score 25–29, 49.4% can be expected to have per-capita household income below the national capacity line. Of course, the poverty likelihood associated with a score varies by poverty line.

5.1 Calibrating scores with poverty likelihoods

A given score is non-parametrically associated ("calibrated") with a poverty likelihood by defining the poverty likelihood as the share of households in the calibration sub-sample who have the score and who are below a given poverty line.

For the example of the capacity line (Table 4), there are 6,720 (normalized) households in the calibration sub-sample with a score of 25–29, of whom 3,322 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 25–29 is then 3,322 ÷ 6,720 = 49.4%.

[15] Appendix A in Schreiner (2009a) gives help for interpreting the indicators in Mexico's poverty scorecard based on ENIGH's *Manual de Encuestador*.

Table 3
Estimated Poverty Likelihoods Associated with Scores, All Eight Poverty Lines

	Poverty likelihood (Probability per-capita household income is below poverty line)							
	National	National	National	125% Natl.	150% Natl.	USAID	International 2005 PPP	
Score	Food	Capacity	Asset	Asset	Asset	'Extreme'	$1.25/day	$2.50/day
0–4	83.9	89.6	98.8	98.8	100.0	83.9	24.9	64.5
5–9	80.7	88.9	97.1	98.5	100.0	78.9	21.3	66.5
10–14	68.0	76.4	94.2	97.5	98.1	68.9	13.4	51.3
15–19	51.4	67.8	92.2	96.1	97.7	52.3	8.5	25.2
20–24	46.9	61.4	86.8	92.5	96.1	53.6	4.5	33.4
25–29	35.7	49.4	81.1	89.7	94.4	45.4	2.8	18.7
30–34	27.8	40.6	71.6	84.4	90.8	37.8	1.9	14.6
35–39	15.7	25.2	60.2	74.9	84.4	24.7	0.9	7.6
40–44	9.9	15.0	50.6	64.4	75.4	15.3	0.6	4.5
45–49	7.5	13.9	41.6	59.9	70.9	14.4	0.1	3.8
50–54	4.6	8.1	26.4	42.6	56.8	8.9	0.7	2.8
55–59	2.2	4.9	17.8	31.4	45.4	6.1	0.4	1.1
60–64	1.1	2.4	10.7	19.2	29.0	3.9	0.1	0.6
65–69	0.9	1.4	6.5	13.3	19.9	1.6	0.0	0.8
70–74	0.2	0.4	2.9	6.5	12.4	0.5	0.0	0.0
75–79	0.0	0.0	0.1	3.2	8.2	0.0	0.0	0.0
80–84	0.0	0.5	2.0	2.9	5.3	0.5	0.0	0.0
85–89	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4
Derivation of Estimated Poverty Likelihoods Associated with Scores, National Capacity Line

Score	Households below poverty line		All households at score		Poverty likelihood (estimated %)
0–4	404	÷	451	=	89.6
5–9	760	÷	855	=	88.9
10–14	982	÷	1,285	=	76.4
15–19	1,913	÷	2,822	=	67.8
20–24	2,715	÷	4,424	=	61.4
25–29	3,322	÷	6,720	=	49.4
30–34	3,416	÷	8,425	=	40.6
35–39	2,077	÷	8,231	=	25.2
40–44	1,435	÷	9,587	=	15.0
45–49	1,483	÷	10,668	=	13.9
50–54	783	÷	9,639	=	8.1
55–59	440	÷	9,054	=	4.9
60–64	199	÷	8,489	=	2.4
65–69	91	÷	6,439	=	1.4
70–74	21	÷	4,668	=	0.4
75–79	0	÷	3,995	=	0.0
80–84	11	÷	2,455	=	0.5
85–89	0	÷	915	=	0.0
90–94	0	÷	776	=	0.0
95–100	0	÷	101	=	0.0

Note: Number of all households normalized to sum to 100,000.

The same method is used to calibrate scores with estimated poverty likelihoods for the other seven poverty lines. Even though the scorecard is constructed partly based on judgment, this calibration process produces poverty likelihoods that are objective, that is, derived from survey data on income and quantitative poverty lines. The poverty likelihoods would be objective even if indicators and/or points were selected without any data at all. In fact, objective scorecards of proven accuracy are often based only on judgment (Fuller 2006; Caire 2004). Of course, the scorecard here is constructed with both data and judgment. The fact that this paper acknowledges that some choices in scorecard construction—as in any statistical analysis—are informed by judgment in no way impugns the objectivity of the poverty likelihoods, as this depends on using data in score calibration, not on using data (and nothing else) in scorecard construction.

Although the points in Mexico's poverty scorecard are transformed coefficients from a Logit regression, scores are not converted to poverty likelihoods via the Logit formula of $2.718281828^{score} \times (1 + 2.718281828^{score})^{-1}$. This is because the Logit formula is esoteric and difficult to compute by hand. Non-specialists find it more intuitive to define the poverty likelihood as the share of households with a given score in the calibration sample who are below a poverty line. In the field, converting scores to poverty likelihoods in this way requires no arithmetic at all, just a look-up table. This non-parametric calibration can also improve accuracy, especially when the assumptions of Logit do not hold and/or when calibration samples are large.

5.2 Accuracy of estimates of households' poverty likelihoods

As long as the relationship between indicators and poverty does not change and as long as the scorecard is applied to households who are representative of the same population from which the scorecard was constructed, this calibration process produces unbiased estimates of poverty likelihoods. The scorecard also produces unbiased estimates of poverty rates at a point in time, as well as unbiased estimates of changes in poverty rates between two points in time.[16]

The relationship between indicators and poverty does change with time and also—as Tarozzi and Deaton point out—across sub-groups, so the scorecard will generally be biased when applied after the August 2008 end date of ENIGH fieldwork (as it must be in practice) or when applied with non-nationally representative groups (as it probably would be by local, pro-poor organizations).

How accurate are estimates of households' poverty likelihoods when the assumption of representativeness holds? To check, the scorecard is applied to 1,000 bootstrap samples of size $n = 16{,}384$ from the 2008 validation sub-sample and the capacity line. For each score range, Table 6 shows the average difference between estimated and true poverty likelihoods as well as confidence intervals for the differences.

For example, the average poverty likelihood across bootstrap samples for scores of 25–29 is too high by 0.9 percentage points, and for scores of 30–34, the estimate is too low by 0.3 percentage points.[17] Table 5 also shows 90-, 95-, and 99% confidence intervals for the estimated differences.

For almost all scores less than 55, Table 5 shows differences—some of them large—between estimates and true values. This is because sampling variation causes the 2008 validation sub-sample to differ in distribution from the construction/calibration sub-samples and from Mexico's population. When the 2008 scorecard is applied to the 2006, 2005, and 2004 ENIGH, differences are due mostly to changes over time in the relationships between indicators and poverty.

[16] This follows because these estimates are linear functions of the unbiased estimates of households' poverty likelihoods.

[17] These differences are not zero, in spite of the estimator's unbiasedness, because the scorecard comes from a single sample. The average difference by score would be zero if samples were repeatedly drawn from the population and split into sub-samples before repeating the entire construction and calibration process.

For targeting, what matters is less the differences across all score ranges and more the differences in score ranges just above and below the targeting cut-off. This mitigates the effects of bias and sampling variation on targeting (Friedman 1997). Section 9 below looks at targeting accuracy in detail.

Of course, if estimates of groups' poverty rates are to be usefully accurate, then errors for individual households must largely cancel out. This is generally the case, especially for the 2008 validation sub-sample, as discussed in the next section.

Another possible source of bias is overfitting. By construction, the scorecard here is unbiased, but it may still be overfit when applied after August 2008 (the end date of ENIGH field work). That is, the scorecard may fit the 2008 ENIGH data so closely that it captures not only some real patterns but also some random patterns that, due to sampling variation, show up only in the 2008 ENIGH. Or the scorecard may be overfit in the sense that it becomes biased as the relationships between indicators and poverty change through time (for example, due to the economic crisis that started in 2008). Finally, the scorecard could also be overfit—as Tarozzi and Deaton highlight—when it is applied to samples from non-nationally representative sub-groups.

Overfitting can be mitigated by simplifying the scorecard and by not relying only on data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Bootstrapping scorecard construction—which is not done here—can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. Combining scorecards can also help, at the cost of complexity.

In any case, most errors in individual households' likelihoods do cancel out in the estimates of groups' poverty rates (see later sections). Furthermore, much of the differences between scorecard estimates and true values may come from non-scorecard sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, inconsistencies in data quality across time, and inconsistencies/imperfections in cost-of-living adjustments across time and space. These factors can be addressed only by improving data quantity and quality (which is beyond the scope of the scorecard), by updating data, or by reducing overfitting (which likely has limited returns, given the scorecard's current parsimony).

6. Estimates of a Group's Poverty Rate at a Point in Time

A group's estimated poverty rate at a point in time is the average of the estimated poverty likelihoods of the individual households in the group.

To illustrate, suppose a program samples three households on Jan. 1, 2009 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 61.4, 40.6, and 15.0% (capacity line, Table 4). The group's estimated poverty rate is the households' average poverty likelihood of $(61.4 + 40.6 + 15.0) \div 3 = 39.0\%$.[18]

[18] The group's poverty rate is not the poverty likelihood associated with the average score. Here, the average score is $(20 + 30 + 40) \div 3 = 30$, and the poverty likelihood associated with the average score is 40.6%. This is not the 39.0% found as the average of the three poverty likelihoods associated with each of the three scores.

6.1 Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Table 6 reports average differences between estimated and true poverty rates as well as precision (confidence intervals for the differences) for the Mexico scorecard applied to 1,000 bootstrap samples from the 2008 validation sample and the capacity line.

Summarizing across poverty lines for $n = 16{,}384$, Table 7 shows that the differences between the estimated poverty rate and the true rate for the 2008 validation sample are −1.2 percentage points or less. The average difference across the eight lines for the 2008 validation sample is −0.8 percentage points. For the capacity line, the difference is −0.9 percentage points.

Table 5
Bootstrapped Differences between Estimated and True Household Poverty Likelihoods with Confidence Intervals in a Large Sample (n = 16,384), 2008 Scorecard Applied to the 2008 Validation Sample, National Capacity Line

Difference between estimate and true value				
		Confidence interval (+/− percentage points)		
Score	Diff.	90 percent	95 percent	99 percent
0–4	−6.2	4.8	4.9	6.1
5–9	−3.4	3.8	4.6	6.4
10–14	+2.3	7.0	8.3	11.0
15–19	−8.1	5.9	6.3	7.2
20–24	+4.2	3.6	4.2	5.7
25–29	+0.9	3.1	3.6	4.8
30–34	−0.3	2.7	3.2	4.3
35–39	−0.8	2.4	2.8	3.8
40–44	−6.6	4.3	4.4	4.7
45–49	−1.0	1.6	1.9	2.3
50–54	−2.0	1.8	1.9	2.4
55–59	+0.1	1.1	1.3	1.7
60–64	+2.1	0.2	0.2	0.3
65–69	+0.1	0.8	1.0	1.3
70–74	+0.4	0.0	0.0	0.1
75–79	−3.1	2.5	2.7	3.1
80–84	+0.5	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

In terms of precision, Table 7 shows that the 90% confidence interval for a group's estimated poverty rate at a point in time in 2004 – 2008 with $n = 16{,}384$ and for all poverty lines is +/–0.8 percentage points or less).

The differences between estimates and true values are much larger for the 2008 scorecard applied to the 2006, 2005, and 2004 ENIGH, ranging from –0.7 to +4.5 percentage points, with an average absolute difference across lines and years of 2.5 percentage points. Part of these differences is due to sampling variation across survey rounds and in the division of the 2008 ENIGH into three sub-samples, as well as small design differences across ENIGH rounds. Mostly, however, the differences are due to changes in the relationships between indicators and poverty over time,[19] as the 2008 ENIGH took place during an economic crisis when poverty was increasing, while the previous rounds took place in non-crisis periods when poverty was decreasing. Future accuracy will depend on whether the future is like 2008 or more like previous years.

Table 6
Differences and Precision of Differences for Bootstrapped Estimates of Poverty Rates for Groups of Households at a Point in Time, by Sample Size, 2008 Scorecard Applied to the 2008 Validation Sample, National Capacity Line

Sample Size n	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
	Diff.	90 percent	95 percent	99 percent
1	–2.8	67.2	73.7	85.8
4	–1.1	32.5	37.9	52.7
8	–1.8	23.6	28.2	37.9
16	–1.2	16.8	20.0	25.5
32	–1.2	11.7	13.9	18.5
64	–1.2	8.5	10.2	12.8
128	–1.1	6.4	7.3	9.1
256	–1.0	4.4	5.1	6.7
512	–1.0	3.1	3.6	4.8
1,024	–0.9	2.2	2.6	3.2
2,048	–0.9	1.5	1.8	2.3
4,096	–0.9	1.0	1.3	1.7
8,192	–0.9	0.7	0.9	1.2
16,384	–0.9	0.5	0.6	0.9

[19] This is known because similar poverty-scoring exercises in Mexico and other countries for periods that do not include sudden crises consistently show much smaller differences.

Table 7
Differences, Precision of Differences, and the α Factor for Bootstrapped Estimates of Poverty Rates for Groups of Households at a Point in Time for all Poverty Lines, 2008 Scorecard Applied to the 2008 Validation Sample and to the 2006, 2005, and 2004 ENIGH

	Poverty line							
	National Food	National Capacity	National Asset	125% Natl. Asset	150% Natl. Asset	USAID 'Extreme'	International 2005 PPP $1.25/day	$2.50/day
Estimate minus true value								
2008 scorecard applied to 2008 validation	−0.9	−0.9	−0.4	−0.9	−1.2	−0.8	−0.1	−0.8
2008 scorecard applied to all 2006	+3.7	+4.8	+4.5	+3.7	+3.0	+1.9	+0.4	+1.8
2008 scorecard applied to all 2005	+1.6	+3.5	+4.5	+3.5	+1.3	+3.6	−0.4	+0.8
2008 scorecard applied to all 2004	+2.8	+2.6	+3.2	+3.8	+3.4	+1.8	−0.7	+0.5
Precision of difference								
2008 scorecard applied to 2008 validation	0.5	0.5	0.6	0.6	0.6	0.5	0.2	0.4
2008 scorecard applied to all 2006	0.3	0.4	0.5	0.6	0.6	0.4	0.1	0.3
2008 scorecard applied to all 2005	0.5	0.6	0.8	0.8	0.8	0.6	0.2	0.4
2008 scorecard applied to all 2004	0.5	0.6	0.6	0.6	0.6	0.6	0.3	0.4
α for sample size								
2008 scorecard applied to 2008 validation	1.03	1.02	0.96	0.95	0.92	1.04	0.99	1.07
2008 scorecard applied to all 2006	0.76	0.81	0.87	0.93	0.92	0.88	0.71	0.77
2008 scorecard applied to all 2005	1.14	1.14	1.22	1.24	1.21	1.14	1.14	1.11
2008 scorecard applied to all 2004	1.08	1.07	0.99	0.93	0.91	1.09	1.48	1.15

Notes: Precision is measured as 90-percent confidence intervals in units of +/− percentage points. Differences and precision estimated from 500 bootstraps of size n = 16,384. α is estimated from 1,000 bootstrap samples of n = 256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.

6.2 Standard-error formula for estimates of poverty rates at a point in time

How precise are the point-in-time estimates? Because they are averages, the estimates have a Normal distribution and can be characterized by their average difference vis-à-vis true values, along with the standard error of the average difference.

To derive a formula for the standard errors of estimated poverty rates at a point in time for indirect measurement via poverty scorecards (Schreiner 2008b), note that the textbook formula (Cochran 1977) that relates confidence intervals with standard errors in the case of direct measurement of poverty rates is $c = +/- z \cdot \sigma$, where:

c is the confidence interval as a proportion (e.g., 0.2 for +/–2 percentage points),

z is from the Normal distribution and is $\begin{cases} 1.64 \text{ for confidence levels of 90 percent} \\ 1.96 \text{ for confidence levels of 95 percent} \\ 2.58 \text{ for confidence levels of 99 percent} \end{cases}$

σ is the standard error of the estimated poverty rate, that is, $\sqrt{\frac{p \cdot (1-p)}{n}}$,

p is the proportion of households below the poverty line in the sample, and n is the sample size.

For example, with a sample n = 16,384, 90% confidence (z = 1.64), and a poverty rate p of 21.0% (the true rate in the 2008 validation sample for the capacity line, Table 2), the confidence interval c is $+/-1.64 \cdot \sqrt{\frac{0.210 \cdot (1-0.210)}{16{,}384}}$ = +/–0.522 percentage points.

Poverty scorecards, however, do not measure poverty directly, so this formula is not immediately applicable. To derive a formula for the Mexico scorecard, consider Table 6, which reports empirical confidence intervals c for the scorecard applied to 1,000 bootstrap samples of various sample sizes from the 2008 validation sample and the capacity line. For n = 16,384, the 90% confidence interval is 0.525 percentage points.[20] Thus, the ratio of confidence intervals with poverty scoring and with direct measurement is $0.525 \div 0.522 = 1.01$.

Now consider the same case, but with n = 8,192. The confidence interval under direct measurement is $+/-1.64 \cdot \sqrt{\frac{0.210 \cdot (1-0.210)}{8{,}192}}$ = +/–0.738 percentage points. The empirical

[20] Due to rounding, Table 6 displays 0.5, not 0.525.

confidence interval with the Mexico scorecard for the national capacity line (Table 6) is 0.720 percentage points. Thus for $n = 8{,}192$, the ratio is $0.720 \div 0.738{=}0.98$.

This ratio of 0.98 for $n = 8{,}182$ is not far from the ratio of 1.01 for $n = 16{,}384$. Indeed, across all sample sizes of 256 or more in Table 6, the average ratio turns out to be 1.02, implying that confidence intervals for indirect estimates of poverty rates via the Mexico scorecard and this poverty line are slightly wider than those for direct estimates. This 1.02 appears in Table 7 as the "α factor" because if $\alpha = 1.02$, then the formula relating confidence intervals c and standard errors σ for the Mexico scorecard is $c = +/- z \cdot \alpha \cdot \sigma$, and the standard error σ for point-in-time scoring estimates of poverty rates is $\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$.[21]

In general, σ could be more or less than 1.00. When σ is less than 1.00, it means that the scorecard is more precise than direct measurement. This occurs in about half the cases in Table 7.

7. Estimates of Changes in Group Poverty Rates Over Time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the households in the group.

7.1 Warning: Change is not impact

Scoring can estimate change. Of course, change could be for the better or for the worse, and scoring does not indicate what caused change. This point is often forgotten, confused, or ignored, so it bears repeating: poverty scoring simply estimates change, and it does not, in and of itself, indicate the reason for the change. In particular, estimating the impact of program participation requires knowing what would have happened to participants if they had not been participants. Knowing this requires either strong assumptions or a control group that resembles participants in all ways except participation. To belabor the point, poverty scoring can help estimate program impact only if there is some way to know what would have happened in the absence of the program. And that information must come from somewhere beyond poverty scoring.

7.2 Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section in which a program sampled three households on Jan. 1, 2009 with an average poverty likelihood (capacity line) of 39.0%. After this baseline, two sampling approaches are possible for the follow-up round:

[21] Schreiner (2009a) further presents sample-size formula and examples of their use.

- Score a new, independent sample, measuring change by cohort across samples
- Score the same sample at follow-up as at baseline

By way of illustration, suppose that a year later on Jan. 1, 2010, the program samples three additional households who are in the same cohort as the three households originally sampled (or suppose that the program scores the same three original households a second time) and finds that their scores are now 25, 35, and 45 (poverty likelihoods of 49.4, 25.2, and 13.9%, capacity line, Table 4). Their average poverty likelihood at follow-up is then (49.4 + 25.2 + 13.9) ÷ 3 = 29.5%, an improvement of 39.0 – 29.5 = 9.5 percentage points.[22]

This suggests that about one of 10 participants crossed the poverty line in 2009. (This is a net figure; some people start above the line and end below it, and vice versa.) Among those who started below the line, about one in four (9.5 ÷ 39.0 = 24.3%) ended up above the line. Again, poverty scoring does not reveal the reasons for this change.

7.3 Estimated changes in poverty rates in Mexico

Given the Mexico poverty scorecard built from the construction and calibration samples from the 2008 ENIGH, an estimate of the change in the poverty rate is the difference between the estimated poverty rate in the 2008 validation sample and the estimated poverty rate in the 2006, 2005, and 2004 ENIGH.

In Table 8, the difference between this estimate and the true value for the capacity line between 2008 and 2006 with n = 16,384 is +4.8 percentage points. The scorecard overstates the change in poverty because it assumes that August 2006 (when there was no economic crisis) is like August 2008 (when there was an economic crisis). Across all eight lines for 2008 and 2006, the average difference is 3.6 percentage points (Table 9), while the average true change is +4.1 percentage points (Table 1). Thus, the scorecard estimate is about twice as high as it should be. Results for 2005 and 2004 are a little better (Table 9), but still not good. In terms of precision, the 90% confidence interval across all lines and years is +/–0.8 percentage points or less.

Because the scorecard estimate is unbiased, these differences are due to sampling variation, changes in poverty lines and/or data collection, and—especially—changes over time in the relationship between indicators and poverty. The magnitude of the differences here are far greater than those in other tests (Schreiner 2009b, 2009c, 2009d, and 2008b; Chen and Schreiner 2009a and 2009b; Mathiassen 2008), suggesting that the differences are related to the economic crisis captured in the 2008 data.

[22] Of course, such a huge reduction in poverty is unlikely in a year's time, but this is just an example to show how poverty scoring can be used to estimate change.

Table 8
Differences and Precision of Differences for Bootstrapped Estimates of Changes in Group's Poverty Rates between Two Points in Time, 2008 Scorecard Applied to the 2008 Validation Sample and the 2006 ENIGH, National Capacity Line

Sample Size *n*	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
	Diff.	90 percent	95 percent	99 percent
1	+6.6	100.0	100.0	106.8
4	+5.1	42.3	50.6	68.8
8	+5.9	30.9	36.9	48.3
16	+5.2	21.4	25.5	33.2
32	+5.0	15.4	18.4	23.0
64	+5.2	11.1	13.4	17.1
128	+5.0	7.9	9.7	12.6
256	+4.9	5.6	6.7	8.6
512	+4.8	3.9	4.6	6.7
1,024	+4.7	2.8	3.3	4.1
2,048	+4.7	1.9	2.3	3.2
4,096	+4.7	1.5	1.7	2.3
8,192	+4.8	1.0	1.2	1.5
16,384	+4.8	0.7	0.8	1.1

Table 9
Differences, Precision of Differences, and the α Factor for Bootstrapped Estimates of Changes in Group's Poverty Rates between Two Points in Time for all Poverty Lines, 2008 Scorecard Applied to the 2008 Validation Sample and to the 2006, 2005, and 2004 ENIGH

	Poverty line							
	National Food	National Capacity	National Asset	125% Natl. Asset	150% Natl. Asset	USAID 'Extreme'	International 2005 PPP $1.25/day	$2.50/day
Estimated change minus true change								
2008 scorecard applied to 2008 validation and all 2006	+4.6	+4.8	+4.9	+4.6	+4.2	+2.6	+0.5	+2.6
2008 scorecard applied to 2008 validation and all 2005	+1.7	+2.4	+1.1	+1.6	+0.5	+2.9	+0.1	+0.9
2008 scorecard applied to 2008 validation and all 2004	+3.7	+3.4	+3.7	+4.7	+4.6	+2.6	–0.6	+1.3
Precision of estimated change minus true change								
2008 scorecard applied to 2008 validation and all 2006	0.6	0.7	0.8	0.8	0.8	0.7	0.2	0.5
2008 scorecard applied to 2008 validation and all 2005	0.7	0.8	1.0	1.1	1.0	0.8	0.3	0.6
2008 scorecard applied to 2008 validation and all 2004	0.7	0.8	0.9	0.9	0.8	0.8	0.3	0.6
α for sample size								
2008 scorecard applied to 2008 validation and all 2006	1.33	1.35	1.29	1.28	1.27	1.41	1.30	1.39
2008 scorecard applied to 2008 validation and all 2005	1.59	1.53	1.59	1.65	1.59	1.50	1.73	1.60
2008 scorecard applied to 2008 validation and all 2004	1.54	1.49	1.36	1.33	1.28	1.54	1.86	1.65

Notes: Precision is measured as 90-percent confidence intervals in units of +/– percentage points. α is estimated from 1,000 bootstrap samples of n = 256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.

7.4 Accuracy for estimated change in two independent samples

For two equal-sized independent samples, the same logic as in the previous section can be used to derive a formula relating the confidence interval c with the standard error σ of a poverty scorecard's estimate of the change in poverty rates over time:

$$c = +/- z \cdot \sigma = +/- z \cdot \alpha \cdot \sqrt{\frac{2 \cdot p \cdot (1-p)}{n}}.$$

z, c, and p are defined as above, n is the sample size at both baseline and follow-up,[23] and α is the average (across a range of bootstrapped sample sizes) of the ratio of the observed confidence intervals from a poverty scorecard and the theoretical confidence intervals from the textbook formula. All the α factors for Mexico exceed 1.00 (Table 9), so scoring in this case is less precise than direct measurement.

7.5 Accuracy for estimated change for one sample, scored twice

The general formula relating the confidence interval c to the standard error σ when using scoring to estimate change for a single group of households, all of whom are scored at two points in time, is (McNemar 1947):

$$c = +/- z \cdot \sigma = +/- z \cdot \alpha \cdot \sqrt{\frac{p_{12} \cdot (1-p_{12}) + p_{21} \cdot (1-p_{21}) + 2 \cdot p_{12} \cdot p_{21}}{n}}.$$

z, c, and α are defined as before, p_{12} is the share of all sampled households that move from below the poverty line to above it, and p_{21} is the share of all sampled households that move from above the line to below it. Schreiner (2009a) gives a formula for choosing n before sampling.

8. Targeting

If a local, pro-poor organization chooses to use poverty scoring for targeting, then households with scores at or below a cut-off are labeled *targeted* and treated—for program purposes—as if they are below a given poverty line. Households with scores above a cut-off are labeled *non-targeted* and treated—for program purposes—as if they are above a given poverty line.

Targeting is successful when households truly below a poverty line are targeted (*inclusion*) and when households truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when households truly below a poverty line are not targeted (*undercoverage*) or when households truly above a poverty line are targeted (*leakage*).

[23] This means that, for a given precision and with direct measurement, estimating the change in a poverty rate over time requires four times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

Table 10 depicts these four possible targeting outcomes. Targeting accuracy varies by score cut-off; a higher cut-off has better inclusion (but greater leakage), while a lower cut-off has better exclusion (but higher undercoverage).

A program should weigh these trade-offs when setting a cut-off. A formal way to do this is to assign net benefits—based on a program's values and mission—to each of the four possible targeting outcomes and then to choose the cut-off that maximizes total net benefits.

Table 11 shows the distribution of households by cut-off and targeting outcome for the capacity line and the 2008 validation sample. For an example cut-off of 29 or less, outcomes are:

- Inclusion: 14.3% are below the line and correctly targeted
- Undercoverage: 26.6% are below the line and mistakenly not targeted
- Leakage: 2.2% are above the line and mistakenly targeted
- Exclusion: 56.8% are above the line and correctly not targeted

Increasing the cut-off to 34 or less would improve inclusion and undercoverage but worsen leakage and exclusion. Which cut-off is preferred depends on total net benefit. If each targeting outcome has a per-household benefit or cost, then total net benefit for a given cut-off is:

Benefit per household correctly included	x	Households correctly included	–
Cost per household mistakenly not covered	x	Households mistakenly not covered	–
Cost per household mistakenly leaked	x	Households mistakenly leaked	+
Benefit per household correctly excluded	x	Households correctly excluded.	

Table 10
(All Poverty Lines): Possible Types of Outcomes from Targeting by Poverty Score

		Targeting segment	
		Targeted	**Non-targeted**
True poverty status	**Below poverty line**	**Inclusion** **Under poverty line** **Correctly** **Targeted**	**Undercoverage** **Under poverty line** **Mistakenly** **Non-targeted**
	Above poverty line	**Leakage** **Above poverty line** **Mistakenly** **Targeted**	**Exclusion** **Above poverty line** **Correctly** **Non-targeted**

To set an optimal cut-off, a program would assign—based on its values and mission—benefits and costs to possible outcomes, tally total net benefits for each cut-off using Table 11 for a given poverty line (see Schreiner 2009a), and finally select the cut-off with the highest total net benefit.

The most difficult step is assigning benefits and costs to targeting outcomes. Any program that uses targeting—with or without scoring—should thoughtfully consider how it values successful inclusion or exclusion versus errors of undercoverage and leakage. It is healthy to go through a process of thinking explicitly and intentionally about how possible targeting outcomes are valued.

Table 11
Households by Targeting Classification and Score, 2008 Scorecard Applied to the 2008 Validation Sample, National Capacity Line

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly not targeted
0–4	0.4	40.5	0.0	59.0
5–9	1.3	39.7	0.0	59.0
10–14	2.5	38.5	0.1	58.9
15–19	5.1	35.9	0.3	58.7
20–24	8.9	32.1	1.0	58.1
25–29	14.3	26.6	2.2	56.8
30–34	20.6	20.4	4.4	54.7
35–39	25.7	15.3	7.5	51.5
40–44	30.5	10.5	12.3	46.7
45–49	34.6	6.3	18.8	40.2
50–54	37.5	3.5	25.6	33.4
55–59	39.3	1.6	32.8	26.2
60–64	40.2	0.8	40.5	18.6
65–69	40.6	0.3	46.4	12.6
70–74	40.8	0.1	50.9	8.1
75–79	40.9	0.0	54.8	4.2
80–84	40.9	0.0	57.3	1.8
85–89	40.9	0.0	58.2	0.9
90–94	40.9	0.0	59.0	0.1
95–100	40.9	0.0	59.1	0.0

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefit, a program could set a cut-off to achieve a desired poverty rate among targeted households. The third column of Table 12 ("% targeted who are poor") shows the expected poverty rate among Mexican households who score at or below a given cut-off for the capacity line and the 2008 validation sample. Targeting households who score 29 or less would target 16.6% of all households (second column) and produce a poverty rate among those targeted of 61.8% (third column).

Table 12 also reports two other measures of targeting accuracy. The first is a version of inclusion ("% of poor who are targeted"). With a cut-off of 29, about half (48.8%) of all poor households are covered.

Table 12
Households Below the Poverty Line and All Households at a Given Score or at or Below a Given Score Cut-off, 2008 Scorecard Applied to the 2008 Validation Sample, National Capacity Line

Targeting cut-off	% All households who are targeted	% Targeted who are poor	% Of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.5	93.7	2.0	14.9:1
5–9	1.3	91.3	5.7	10.5:1
10–14	2.6	85.3	10.5	5.8:1
15–19	5.4	79.8	20.6	3.9:1
20–24	9.8	70.4	33.0	2.4:1
25–29	16.6	61.8	48.8	1.6:1
30–34	25.0	54.6	65.0	1.2:1
35–39	33.2	47.7	75.6	0.9:1
40–44	42.8	41.9	85.5	0.7:1
45–49	53.5	36.4	92.8	0.6:1
50–54	63.1	32.3	97.3	0.5:1
55–59	72.2	28.8	99.2	0.4:1
60–64	80.7	25.9	99.5	0.3:1
65–69	87.1	24.0	99.8	0.3:1
70–74	91.8	22.8	99.8	0.3:1
75–79	95.8	21.9	100.0	0.3:1
80–84	98.2	21.4	100.0	0.3:1
85–89	99.1	21.2	100.0	0.3:1
90–94	99.9	21.0	100.0	0.3:1
95–100	100.0	21.0	100.0	0.3:1

The final targeting measure in Table 12 is the number of successfully targeted poor households for each non-poor household mistakenly targeted (right-most column). For a cut-off of 29, covering 1.6 poor households means leaking to 1 non-poor household.

8. Conclusion

This paper presents a simple poverty scorecard for Mexico that can be used to estimate the likelihood that a household has income below a given poverty line, to estimate the poverty rate of a group of households at a point in time, and to estimate changes in the poverty rate of a group of households between two points in time. The scorecard can also be used for targeting.

The scorecard is inexpensive to use and can be understood by non-specialists. It is designed to be practical for local pro-poor organizations who want to improve how they monitor and manage their social performance in order to speed up their participants' progress out of poverty.

The scorecard is built with a sub-sample of data from the 2008 ENIGH, tested on a different sub-sample from the 2008 ENIGH and on the 2006, 2005, and 2004 ENIGH, and calibrated to eight poverty lines.

Accuracy is reported for estimates of households' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time. Targeting accuracy and formula for standard errors are also reported.

When the scorecard is applied to the 2008 validation sample with $n = 16{,}384$, the absolute difference between estimates and true poverty rates at a point in time is always less than 1.2 percentage points and averages—across the eight poverty lines—about –0.8 percentage points. With 90% confidence, the precision of these differences is +/– 0.6 percentage points or less. In this case, the scorecard is usually more precise than direct measurement.

When used to measure change across independent samples of $n = 16{,}384$ between the 2008 validation sample and the 2006, 2005, or 2004 ENIGH, the average absolute difference between estimates and true changes across poverty lines and years is large (about 2.5 percentage points), with a 90% confidence interval of +/– 0.8 percentage points or less. The scorecard overestimates changes mainly because the 2008 ENIGH data from which the scorecard was built reflect a time of economic crisis, while earlier survey rounds reflect non-crisis periods. Future accuracy will depend on how closely the economic situation in Mexico resembles that of August 2008.

For targeting, programs can use the results reported here to select a cut-off that fits their values and mission.

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard here focuses on transparency and ease-of-use. After all, a perfectly accurate scorecard is worthless if programs feel so daunted by its complexity or its cost that they do not even try to use it. For this reason, the poverty scorecard is kept simple, using ten indicators that are inexpensive to collect and that are straightforward to verify. Points are all zeros or positive integers, and scores range from 0 to 100. Scores are related to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise simple to apply. The design attempts to facilitate adoption by helping managers understand and trust scoring and by allowing non-specialists to generate scores quickly in the field.

In sum, the poverty scorecard is a practical, objective way for pro-poor programs in Mexico to monitor poverty rates, track changes in poverty rates over time, and target services. Estimates of poverty rates and changes in rates are more accurate for periods similar to August 2008, the point in time when the data used to construct the scorecard was collected. This same poverty scoring approach can be applied to any country with similar data from a national income or expenditure survey.

References

Baesens, Bart, Van Gestel Tony, Viaene Stijn, Stepanova Maria, Suykens Johan A. K., and Vanthienen Jan. "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring", *Journal of the Operational Research Society*, Vol. 54, (2003): 627–635.

Bellon, Mauricio R., Hodson Dave P., Martínez-Romero Eduardo, Montoya Yinha, Becceril Javier, and White Jeffrey W. "Geospatial Dimensions of Poverty and Food Security—A Case Study for Mexico", Mexico City: International Maize and Wheat Improvement Center, (2004), http://www.cimmyt.org/GIS/povertymexico/poverty_mapping_final_draft.pdf, accessed 29 July 2009.

Caire, Dean. "Building Credit Scorecards for Small Business Lending in Developing Markets", Bannock Consulting, (2004), http://www.microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf, accessed 31 July 2009.

Chen, Shiyuan, and Schreiner Mark. "A Simple Poverty Scorecard for Bangladesh", (2009a), http://www.microfinance.com/English/Papers/Scoring_Poverty_Bangladesh.pdf, accessed 31 July 2009.

Chen, Shiyuan, and Schreiner Mark. "A Simple Poverty Scorecard for Vietnam", (2009b), http://www.microfinance.com/English/Papers/Scoring_Poverty_Vietnam_EN.pdf, accessed 31 July 2009.

Coady, David, Grosh Margaret, and Hoddinott John. "The Targeting of Transfers in Developing Countries: Review of Experience and Lessons" (2002), http://info.worldbank.org/etools/docs/library/79646/Dc%202003/courses/dc2003/readings/targeting.pdf, accessed 31 July 2009.

Cochran, William G. *Sampling Techniques, Third Edition*, Nueva York: Wiley, (1977).

Comité Técnico para la Medición de la Pobreza. "Medición de la Pobreza: Variantes Metodológicas y Estimación Preliminar", Secretaria de Desarrollo Social, México, D.F., (2002), http://www.sedesol.gob.mx/archivos/801588/file/Docu01.pdf, accessed 26 July 2009.

CONEVAL. "Los Mapas de Pobreza en México", México, D.F.: Consejo Nacional de Evaluación de la Política de Desarrollo Social (2007), http://www.coneval.gob.mx/mapas/mapas/Informe_Tecnico.zip, accessed 29 July 2009.

Dawes, Robyn M. "The Robust Beauty of Improper Linear Models in Decision Making", *American Psychologist*, Vol. 34, No. 7, (1979): 571–582.

de la Torre, Rodolfo. "Ingreso y Gasto en la Medición de la Pobreza", SEDESOL Documentos de Investigación No. 22, México, D.F., (2005), http://www.sedesol.gob.mx/archivos/801588/file/Docu_22_2003.pdf, accessed 26 July 2009.

Deaton, Angus, and Zaidi Salman. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis", Living Standards Measurement Study Working Paper No. 135, Washington, D.C. World Bank (2002), http://go.worldbank.org/8YRCR9ERJ0, accessed 26 July 2009.

Demombynes, Gabriel, Elbers Chris, and Lanjouw Peter. "How Good a Map? Putting Small-Area Estimation to the Test", World Bank Policy Research Working Paper No. 4155 (2007), Washington, D.C., http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2007/03/26/000016406_20070326150728/Rendered/PDF/wps4155.pdf, accessed 28 July 2009.

Elbers, Chris, Lanjouw Jean O., and Lanjouw Peter. "Micro-Level Estimation of Poverty and Inequality", *Econometrica*, Vol. 71, No. 1, (2003): 355–364.

Elbers, Chris, Lanjouw Peter, and Leite Phillippe George. "Brazil within Brazil: Testing the Poverty Map Methodology in Minas Gerais", World Bank Policy Research Working Paper No. 4513 (2008), Washington, D.C., http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2008/02/26/000158349_20080226134003/Rendered/PDF/wps4513.pdf, accessed 28 July 2009.

Friedman, Jerome H. "On Bias, Variance, 0–1 Loss, and the Curse-of-Dimensionality", *Data Mining and Knowledge Discovery*, Vol. 1, (1997): 55–77.

Fuller, Rob. "Measuring the Poverty of Microfinance Clients in Haiti", (2006), http://www.microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf, accessed 31 July 2009.

Grosh, Margaret, and Baker Judy L. "Proxy Means Tests for Targeting Social Programs: Simulations and Speculation", LSMS Working Paper No. 118 (1995), Washington, D.C.: World Bank, http://go.worldbank.org/W9OWN57PD0, accessed 31 July 2009.

Hand, David J. "Classifier Technology and the Illusion of Progress", *Statistical Science*, Vol. 22, No. 1, (2006): 1–15.

International Comparison Project. "Tables of Results", Washington, D.C.: World Bank, (2008), http://siteresources.worldbank.org/ICPINT/Resources/icp-final-tables.pdf, accessed 31 July 2009.

IRIS Center. "Manual for the Implementation of USAID Poverty Assessment Tools", (2007), http://www.povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf, accessed 31 July 2009.

Kolesar, Peter, and Showers Janet L. "A Robust Credit Screening Model Using Categorical Data", *Management Science*, Vol. 31, No. 2, (1985): 124–133.

López-Calva, Luís F., Meléndez Álvaro, Rascón Ericka G., Rodríguez-Chammusy Lourdes, and Székely Pardo Miguel. "Poniendo al Ingreso de los Hogares en el Mapa de México", Documento de Trabajo EGAP–2005–04, Tecnológico de Monterrey, Campus Ciudad de México, (2005), http://alejandria.ccm.itesm.mx/egap/documentos/EGAP-2005-04.pdf, accessed 27 July 2009.

Martinelli, César, and Parker Susan W. "Deception and Misreporting in a Social Program", Centro de Investigación Económica and Instituto Tecnológico Autónomo de México, (2007), http://ciep.itam.mx/~martinel/lies4.pdf, accessed 31 July 2009.

Mathiassen, Astrid. "The Predictive Ability of Poverty Models: Empirical Evidence from Uganda", Discussion Paper No. 560 (2008), Statistics Norway, Division for Development Cooperation, http://www.ssb.no/publikasjoner/DP/pdf/dp560.pdf, accessed 31 July 2009.

Matul, Michal, and Kline Sean. "Scoring Change: Prizma's Approach to Assessing Poverty", MFC Spotlight Note No. 4, Warsaw, Poland: Microfinance Centre for Central and Eastern Europe and the New Independent States, (2003), http://www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf, accessed 31 July 2009.

McNemar, Quinn. "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages", *Psychometrika*, Vol. 17, (1947): 153–157.

Medina, Cinthia, Hubert Celia, and Soto Humberto. "Comparación de Distintas Metodologías para la Identificación de Familias Beneficiarias", Secretaría de Desarrollo Social, México, D.F., (2000).

Narayan, Ambar, and Yoshida Nobuo. "Proxy Means Tests for Targeting Welfare Benefits in Sri Lanka", Report No. SASPR–7, (2005), Washington, D.C.: World Bank, http://siteresources.worldbank.org/EXTSAREGTOPPOVRED/Resources/493440-1102216396155/572861-1102221461685/Proxy+Means+Test+for+Targeting+Welfare+Benefits.pdf, accessed 31 July 2009.

Onwujekwe, Obinna, Hanson Kara, and Fox-Rushby Julia. "Some Indicators of Socio-Economic Status May Not Be Reliable and Use of Indices with These Data Could Worsen Equity", *Health Economics*, Vol. 15, (2006): 639–644.

Rodríguez-Chamussy, Lourdes, and Székely Miguel. "Poverty Maps and Public Policy in Mexico", in Tara Bedi, Aline Coudouel, and Kenneth Simler (eds) *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, Washington, D.C.: World Bank, (2007): 188–207, http://go.worldbank.org/P6S3FQPOU0, accessed 29 July 2009.

SAS Institute Inc. "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*, Cary, NC, (2004), http://support.sas.com/documentation/cdl/en/statug/59654/HTML/default/statug_logistic_sect035.htm, accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for Mexico", (2009a), http://www.microfinance.com/English/Papers/Scoring_Poverty_Mexico_2008.pdf., accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for Mexico", (2009b), http://www.microfinance.com/English/Papers/Scoring_Poverty_Mexico_2006.pdf. , accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for Peru", (2009c), http://www.microfinance.com/English/Papers/Scoring_Poverty_Peru.pdf, accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for the Philippines", (2009d), http://www.microfinance.com/English/Papers/Scoring_Poverty_Philippines.pdf, accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for Ecuador", (2008a), http://www.microfinance.com/English/Papers/Scoring_Poverty_Ecuador.pdf, accessed 31 July 2009.

Schreiner, Mark. "A Simple Poverty Scorecard for Peru", (2008b), http://www.microfinance.com/English/Papers/Scoring_Poverty_Peru_May_2008.pdf, accessed 31 July 2009.

Schreiner, Mark. "Un índice de pobreza para México", memo for Grameen Foundation, (2006a), http://www.microfinance.com/Castellano/Documentos/Scoring_Pobreza_Mexico_2002.pdf, accessed 31 July 2009.

Schreiner, Mark. "Is One Simple Poverty Scorecard Enough for India?", memo for Grameen Foundation, (2006b), http://www.microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, accessed 31 July 2009.

Schreiner, Mark. "Scoring: The Next Breakthrough in Microfinance?" Consultative Group to Assist the Poor Occasional Paper No. 7, (2002), Washington, D.C., http://pdf.usaid.gov/pdf_docs/PNACQ633.pdf, accessed 31 July 2009.

Sillers, Don. "National and International Poverty Lines: An Overview", Washington, D.C.: United States Agency for International Development, (2006) , http://www.microlinks.org/file_download.php/Poverty_lines___An_Overview_1_4_06.pdf?URL_ID=12247&filename=11549869641Poverty_lines___An_Overview_1_4_06.pdf&filetype=application%2Fpdf&filesize=108185&name=Poverty_lines___An_Overview_1_4_06.pdf&location=user-S/, accessed 31 July 2009.

Singh, Kesar. "Breakdown Theory for Bootstrap Quantiles", *Annals of Statistics*, Vol. 26, (1998): 1719–1732.

Skoufias, Emmanuel, Davis Benjamin, and Behrman Jere R. "An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico", (1999), http://www.ifpri.org/themes/progresa/pdf/skoufias_target.pdf, accessed 31 July 2009.

Tarozzi, Alesandro. "Can Census Data Alone Signal Heterogeneity in the Estimation of Poverty Maps?", (2008), http://www.econ.duke.edu/~taroz/TarozziHet2008.pdf, accessed 31 July 2009.

Tarozzi, Alessandro, and Deaton Angus. "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas", (2007), http://www.princeton.edu/~deaton/downloads/20080301SmallAreas_FINAL.pdf, accessed 31 July 2009.

Toohig, Jeff. "PPI Pilot Training Guide", Grameen Foundation, (2008), http://www.progressoutofpoverty.org/toolkit, accessed 31 July 2009.

United States Congress. "Amendments to the Microenterprise for Self-Reliance Act of 2000 (Public Law 106–309)", (2002), http://www.microlinks.org/file_download.phpAmendMicroenterpriseAct2000.pdf?URL_ID=7744&filename=11205460851AmendMicroenterpriseAct2000.pdf&filetype=application%2Fpdf&filesize=95834&name=AmendMicroenterpriseAct2000.pdf&location=user-S/, accessed 27 July 2009.

Wainer, Howard. "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin*, Vol. 83, (1976): 223–227.