An Expert-Based Poverty Scorecard for Rural China

Mark Schreiner

18 May 2012

This document is at microfinance.com/#China
中國的版本在 microfinance.com/Chinese

Abstract

This paper presents an expert-based poverty scorecard for rural China. The scorecard's indicators, response options, and points—and therefore its definition of poverty status—are based on expert judgment because consumption data from China's Rural Household Survey are not available. Field workers can quickly collect and check the scorecard's 16 indicators, and they can compute scores on paper. Until consumption data are available, the expert scorecard is a practical way for pro-poor programs in rural China to measure the distribution of poverty among their clients, to track changes in that distribution, and to segment clients. While the expert scorecard does not use a consumption-based definition of poverty status, research and experience suggests that it will still be useful in many common applications.

Author

Mark Schreiner is the Director of Microfinance Risk Management, L.L.C., microfinance.com. He is also Senior Scholar at the Center for Social Development in Washington University in Saint Louis.

Acknowledgements

This paper was commissioned by Grameen Foundation (GF) through the generous support of the Overseas Resources Foundation and You Change PuRong Information Advisory Company Limited. At GF's global office, Mary Jo Kochendorfer provided management, with support from Jeff Toohig and Steven Wright. In GF's Asia Region, Sonia Wong（王淑里）helped with management, with translation, and with the organization and facilitation of the meeting of experts in Beijing. She was supported by Cristopher Lomboy. You Change PuRong Information Advisory Company Limited convened the expert meeting, with thanks to Jun Li（李钧）and Xiang Wang（王翔）. Ronda Zhang（张若丹）and Jane Sun（孙屹青）provided translation. The Chinese Academy of Social Sciences hosted the meeting of experts on 10–11 April 2012. Special thanks go to the China-poverty experts for their contributions and their engagement:

Shaohua Chen	（陈少华）	World Bank
Enjiang Cheng	（程恩江）	International Poverty Reduction Center in China
Xiaoshan Du	（杜晓山）	Chinese Academy of Social Sciences and China Association of Microfinance
Yanhong Hao	（郝彦宏）	National Bureau of Statistics of China
Guangwen He	（何广文）	China Agricultural University
Guilin Huo	（霍桂林）	Chifeng Zhaowuda Women's Sustainable Dev. Assoc.
Zhipu Long	（龙治普）	Ningxia Huimin Microfinance Company, Limited

You Change PuRong Information Advisory Company, Limited, was represented at the expert meeting by Xiang Wang（王翔）.

The expert-based poverty scorecard differs from what GF calls the Progress out of Poverty Index® in that the PPI® is based on data and uses a consumption-based definition of poverty. GF promotes the PPI® and the expert-based poverty scorecard here as performance-management tools to help institutions achieve their social objectives more effectively.

Expert-Based Poverty Scorecard for Rural China

Entity	**Name**	**ID**	**Date (DD/MM/YY)**	
Member:			Date joined:	
Field agent:			Date scored:	
Service point:			Number of HH members:	

Indicator	Response	Points	Score
1. How many permanent members does the household have?	A. Six or more	0	
	B. Five	1	
	C. Four	2	
	D. Three	4	
	E. Two	6	
	F. One	7	
2. How many household members are less than six years-old or more than 60-years-old?	A. All are older than 60 or less than 6 (regardless of number)	0	
	B. Four or more	0	
	C. Three	1	
	D. Two	2	
	E. One	3	
	F. None	5	
3. Do all household members who are of age receive compulsory education?	A. No	0	
	B. Yes, and some are boarding	1	
	C. Yes, and none are boarding	2	
	D. None are of age	4	
4. What is the highest educational attainment of a household member who is in the labor force?	A. No one in labor force	0	
	B. Illiterate	0	
	C. Primary	1	
	D. Junior high	5	
	E. Senior high or higher	8	
5. How many household members work as migrants?	A. None	0	
	B. One	2	
	C. Two or more	5	
6. How many household members have stable income from employment or work outside of agriculture?	A. None	0	
	B. One	2	
	C. Two or more	5	
7. What are the main construction materials of the residence?	A. Others	0	
	B. Masonry and wood	3	
	C. Reinforced concrete	6	

Expert-Based Poverty Scorecard for Rural China (cont.)

8. What is the main fuel used for cooking?	A. Firewood, or others	0
	B. Coal, or biogas (methane)	2
	C. Natural gas, coal gas, liquefied petroleum gas, electricity, gasoline, diesel, or solar	5
9. What is the household's main source of drinking water?	A. Pond, or others	0
	B. River or lake	1
	C. Shallow well	2
	D. Deep well	3
	E. Tap, or bottled	6
10. Does the household have a color TV?	A. No	0
	B. Yes	2
11. Does the household have a refrigerator or freezer?	A. No	0
	B. Yes	4
12. Does the household have a washing machine?	A. No	0
	B. Semi-automatic	2
	C. Automatic	6
13. What is the best form of mechanized transport that the household has?	A. None	0
	B. Bicycle	1
	C. Motorcycle or motorized bicycle/scooter/moped	5
	D. Automobile, truck, etc.	10
14. What is the best form of agricultural traction that the household has?	A. None	0
	B. Draught animal	2
	C. Mini or walking tractor	4
	D. Motor vehicle, large or medium tractor, threshing machine, harvester, or motor tricycle	6
15. What type of insurance does the household have?	A. None	0
	B. Non-commercial medical	1
	C. Non-commercial old age	4
	D. Commercial (any type)	6
16. Does the household receive the Minimum Living Standard Subsidy?	A. Yes	0
	B. No	2

Household Worksheet

If convenient, complete this worksheet at the start of the interview using the "Guidelines for the Interpretation of Scorecard Indicators". Then use the results to mark the responses for "Number of HH members" and for the first six indicators in the scorecard.

Name of permanent household member	How old is [name] in years?	Does [name] receive compulsory education?	In labor force?	Migrant?	Highest educational attainment?	Stable income from employment, or non-ag. income from working?
1.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
2.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
3.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
4.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
5.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
6.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
7.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
8.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
9.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
10.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
11.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y
12.		Not of age No Yes (boarding) Yes (non-boarding)	N Y	N Y		N Y

- Record the number of permanent household members under "Number of household members"
- For Indicator 1, count the number of permanent household members and mark the corresponding response on the scorecard
- For Indicator 2, count the number of members younger than 6 or older than 60 and mark the response on the scorecard. If all members are older than 60 or younger than 6, then mark (A)
- For Indicator 3, mark (A) if there are any members of age who do not go to school, (B) if all members of age go to school, and if some go to boarding school, (C) if all members of age go to school but none go to boarding school, or (D) if no members are of age
- For Indicator 4, mark the response corresponding to the highest education level attained by a member in the labor force
- For Indicator 5, mark the response based on the number of members who work as migrants
- For Indicator 6, mark the response based on the number of members who have stable income from employment or who work outside of agriculture.

An Expert-Based Poverty Scorecard for Rural China

1. Introduction

How poor are participants in pro-poor programs in China? The first step towards an answer is to define *poverty*. For most governments, people are *poor* if their consumption (the financial value of the resources they use up) is less than a poverty line.[1] Head-count rates of consumption-based poverty are estimated via national sample surveys that ask households about the value all of the goods and services that they consume. These consumption surveys are long, costly, and complex, so they are done only by governments, only periodically, and only on a sample basis.

Poverty scoring is an alternative approach (Schreiner, 2006). Constructed with data from national consumption surveys, poverty scorecards use a few simple indicators to estimate the likelihood that a household has consumption below a poverty line. Field workers can record responses and compute scores in about 10 minutes. Scoring's estimates are unbiased and have known standard errors. More important, poverty scoring is designed to be quick, transparent, and low-cost so as to spark voluntary uptake by non-specialists in local, pro-poor organizations who seek to manage progress towards their social mission more quantitatively and intentionally by:

- Estimating poverty rates for participants at a point in time
- Estimating changes in participants' poverty rates over time
- Segmenting participants based on their relative poverty scores

Among microfinance organizations, scoring's estimates of poverty rates are increasingly the preferred indicator of depth and breadth of outreach. But nothing about poverty scoring is specific to microfinance, and it is also being used by governments and by other non-microfinance organizations with a social mission.

[1] Instead of consumption, some countries use income (the value of financial inflows).

There are consumption-based poverty scorecards for about 45 countries,[2] covering about 77 percent of people in the world who fall under $1.25/day 2005 PPP. China—with 13 percent of the world's people under $1.25/day—is the biggest omission.[3] There is no consumption-based poverty scorecard for China because data from its national consumption surveys is available only inside China's government.

This paper proposes an expert-based poverty scorecard for rural China.[4] Chinese poverty experts and microfinance leaders selected the scorecard's indicators, response options, and points, and their choices implicitly define *poverty*. The expert approach has a long and distinguished history in scoring research and practice, and tests here with data from Bosnia-Herzegovina, Cambodia, and Mali suggest that poverty as measured by expert-based scores is largely concordant with poverty as defined by consumption.

Unlike data-based scorecards, the expert-based scorecard produces only scores and not also poverty likelihoods. Likelihoods are cardinal; they can be added and averaged to give single-figure summary measures such as poverty rates. This is because the distance between poverty likelihoods is known and constant; for example, 20 percent is as far from 30 percent as 40 percent is from 50 percent.

In contrast, scores are ordinal. Even though scores are represented by numbers, the numbers are just ordered symbols. Like letters of the alphabet, scores cannot be added or averaged because the distance between them is not known nor constant; the distance between scores of 20 and 30 is not the same as between scores of 40 and 50.

Nevertheless, expert-based scorecards can still be useful because managers can analyze score *distributions*. This is not as simple and clear-cut as analyzing single-figure poverty rates, but it is usually better than nothing. The main lesson so far from data-based poverty scoring is that the mere act of measuring poverty can help to refocus an organization on its social mission and to promote a more quantitative, intentional culture of management. The use of expert-based scorecards will likely have a similar effect.

Section 2 below discusses the definition of *poverty*. Section 3 tells how the scorecard was constructed. Section 4 shows how to analyze distributions of scores. Section 5 looks at how expert scorecards and data-based scorecards for Bosnia-Herzegovina, Cambodia, and Mali rank households vis-à-vis consumption-based poverty status. Section 6 discusses how the "flat maximum" phenomenon supports the probable usefulness of expert scoring. The final section is a summary.

---

[2] `progressoutofpoverty.org/ppi-country` and `microfinance.com/#Poverty_Scoring`

[3] Thanks to Brian James for compiling these figures from country-level data for different years at `data.worldbank.org/indicator/SI.POV.DDAY` (poverty rate by $1.25/day 2005 PPP) and `data.worldbank.org/indicator/SP.POP.TOTL` (population).

[4] The focus is rural because most of China's poor people live in rural areas.

2. Definition of *poverty*

This section reviews two broad ways to define *poverty*.

2.1 Consumption

The consumption approach defines people as *poor* if they live in a household where the financial value of per-capita consumption is less than a poverty line. The main point in favor of this approach is that consumption is probably the single-best indicator of realized well-being; if only one thing can be measured and if all measures are equally costly, then measure consumption. Another strength is that defensible, quantitative poverty lines can be derived using economic logic, nutritional norms, and data on households' consumption of non-food goods and services (Ravallion, 1998). Also, single-figure consumption-based poverty rates are simple to communicate and understand. Finally, measures of income may stand in for measures of consumption, and income—at least in developed countries where most households have tax returns and pay stubs—may be simple to measure.

In developing countries, however, consumption is difficult and costly to measure, and income cannot serve as a stand-in. For example, farmers often eat some of their own produce, generating no cash income nor a paper trail. It is difficult for a survey to place a value on a farmer's non-traded production or to account for the consumption of hundreds of possible goods and services. Furthermore, consumption varies over time. Chronic poverty is a greater concern than short-term poverty, but most consumption measures are annual or shorter. Consumption also depends on borrowing and saving, so more saving and less consumption in the short term may not signal less well-being in the long-term.

Sahn and Stifel (2003) highlight other issues with measuring consumption:

- It depends on error-prone recall ("How much rice did you eat last month?")
- It requires prices, often for non-tradables, and depreciation rates of consumer durables. For example, the value of owner-occupied housing is a big share of consumption but is difficult to estimate. (What would a farmer's hut rent for?)
- It relies on price indexes that may be low-quality or non-existent (across years between surveys, across months when a given survey is in the field, across regions within a country, and across countries)

In addition to the data-based poverty scorecards described in the introduction, consumption-based poverty is estimated for proxy-means tests (Coady, Grosh, and Hoddinott, 2002) and poverty maps (Elbers, Lanjouw, and Lanjouw, 2003). Governments target social programs to households with proxy-means tests (for example, Camacho and Conover, 2011; Martinelli and Parker, 2009) and to regions with poverty maps (Bedi, Coudouel, and Simler, 2007).

These two approaches estimate the level of per-capita consumption via regional regressions of indicators on the logarithm of consumption as measured in a national survey. They then convert estimated consumption into a poverty likelihood or into poor/non-poor status based on a poverty line.

In most ways, scorecards for proxy-means tests and poverty maps resemble data-based poverty scorecards, except that poverty scoring is:

- Simpler, quicker, and less costly to implement, as its indicators are selected not only for their links with poverty but also for their ease of collection
- More transparent, as it is aimed not at government experts but at managers in local, pro-poor organizations who will not use a tool unless they understand it
- More robust, as it loses less accuracy when applied later in time or to sub-groups

Wang, Wang, and Wang (2007) is a consumption-based proxy-means test for rural China. It uses the 2002 Rural Poverty Monitoring Survey and reports ranking accuracy for a single poverty-likelihood cut-off. While well-done and technically sound, it cannot be used as a scorecard by local, pro-poor organizations because it has 46 indicators, points with many decimal places and negative values, community-level indicators, and complex indicators such as "Square root of the per-capita amount of grain stored at home" and "Ratio of area sown to cash crops to total sown area".

In 2002/3, China's National Bureau of Statistics and the World Bank made a pilot poverty map for Yunnan Province with the 1997 Agricultural Census and consumption data from the 1997 RHS (Ahmad and Goh, 2007). The results were not published or actively disseminated, so the scorecard cannot be used by local, pro-poor organizations. The poverty map was used by the National Development and Reform Commission to review county-level project funding in Yunnan and by the World Bank and the U.K.'s Department of International Development to help target beneficiary areas for their Poor Rural Communities Development Project.

2.2 Assets

The asset approach defines people as *poor* if their capital (physical, human, financial, and social) falls below a threshold. Arguments for the asset-based view include Carter and Barrett (2006), Schreiner and Sherraden (2006), Sahn and Stifel (2003), and Sherraden (1991). The main points in its favor are that:

- Asset ownership is easier to measure accurately than consumption
- Access to resources in the long term—and thus capacity to produce income and to consume—depends on the control of assets
- Assets get at capability more directly, the difference between, say, "Does your income permit adequate sanitation?" versus "Do you have a flush toilet?"

In practice, governments and international donors acknowledge the importance of assets but focus mostly on consumption, probably because it seems more objective, it has a shorter horizon, and its single dimension makes it easier to communicate. In particular, there is no consensus on a single, theory-based, data-derived asset poverty line. Even if there were a single agreed-on asset poverty line, one government or organization could label a household as *poor* if its asset score falls below a chosen cut-off, while others could reasonably choose different cut-offs.

While the asset view and the income/consumption view are distinct, they are also tightly linked. After all, income/consumption are flows of resources received/consumed from the use of stocks of assets. Both views are low-dimensional simplifications—due to practical limits on definitions and measurement—of a higher-dimensional and more complete conception of the production of human well-being.

Whatever the strengths of a given view, the asset/consumption choice usually seems to hinge—as in this paper—on data availability and on whether cross-country comparisons are a central goal. For example, the Demographic and Health Surveys do not collect consumption data, so most DHS research defines *poverty* in terms of an asset index made up of all housing characteristics and consumer durables available in the DHS (Rutstein and Johnson, 2004).[5] Inequality in health is then measured as the variation in outcomes by quintiles of the asset index.

Point values for DHS-like indexes come from the standardized first principle component of the variance-covariance matrix of the indicators. This process does not explicitly model any particular definition of poverty; rather, it finds the linear combination that maximizes the explained variation among the indicators. Nevertheless, the resulting ranks are related to poverty, especially when this is seen as “permanent income” or “expected long-term control over resources”. DHS-like indexes turn out to be correlated in intuitive ways with outcomes such as fertility (Bollen, Glanville, and Stecklov, 2007), use of emergency obstetric care (Pitchforth *et al.*, 2007), maternal and child mortality (Knowles *et al.*, 2008), food security (Dekker, 2006), child health and nutrition (Sahn and Stifel, 2003), and education (Filmer and Pritchett, 2001). As presented, the DHS indexes are too complex for use by local pro-poor organizations; even though they usually have only about 15 indicators, they generally require marking more than 100 responses and adding more than 200 point values, all of which have four decimal places and half of which are negative.[6]

[5] Asset-based poverty definitions use indexes as a quantitative way to aggregate over diverse types of assets (such as housing, education, employment, and asset ownership).
[6] Schreiner (2010) shows that points in DHS-like indexes can be transformed to give the same ranks with only half the number of point values (all of which are zero or positive whole numbers) and with marking only one answer per indicator.

Ranks by DHS indexes are not comparable across countries because indicators and points vary by country. For cross-country (and cross-time) comparisons, Sahn and Stifel (2000) make an asset index from pooled data for eight African countries, all with a common set of nine indicators from DHS surveys. They then apply the index across the eight countries (and across consecutive DHS surveys in a given country), measuring differences in poverty rates as defined by the 25th and 40th percentiles of the baseline index. Because assets are likely to be measured accurately, Sahn and Stifel provide a convincing ranking of the eight countries by poverty and robust evidence that their poverty decreased over about 1987–97.[7]

Going beyond Sahn and Stifle, Alkire and Santos (2010) build a world-wide poverty index. Unlike the DHS-like indices above but like the poverty scorecard for rural China, the Multi-dimensional Poverty Index is completely expert-based, as its indicators and points are selected by hand. Also like the scorecard here, MPI points are all zero or positive whole numbers, and scores range from 0 to 100. Alkire and Santos apply the MPI to about 100 countries with data from the DHS or similar surveys that do not measure consumption. For each country, they report a single figure that combines a head-count poverty rate—based on a poverty line set at scores of 30 or less—and a measure of the distance between poor households and the poverty line.

Not surprisingly, MPI poverty is different from (but positively correlated with) consumption-based measures. Alkire and Santos explain how the MPI indicators square with the Millenium Development Goals, with Nobel Prize-winner Amartya Sen's thinking, and with lessons from participatory consultations. In the end, the common-sense indicators are similar to those in most other asset indexes and in poverty scorecards, except for indicators for child nutrition, schooling, and mortality that do not pertain to all households or that are not easy to measure.

Poverty can be seen in terms of consumption or assets. If there is no consumption data (as for China), then the only option is assets. Regardless of definitions, poverty tools tend to rank most households similarly. Of course, consumption-based poverty is not the gold-standard benchmark, and asset-based tools should not be judged on how well they replicate consumption-based ranks. Rather, both views are useful according to their own strengths and weaknesses.

[7] Booysen *et al.* (2008) reprise Sahn and Stifel (2000) for the following 10 years or so.

3. Scorecard construction

This section describes the construction of rural China's expert-based poverty scorecard. The scorecard was made with care by knowledgeable Chinese experts in an attempt to design a simple tool that is reasonably accurate and that is acceptable to non-specialist managers working in pro-poor organizations.

3.1 Experts

Experts were invited based on their:

- Knowledge of poverty in China
- Knowledge of data on poverty in China
- Interest in poverty measurement
- Interest in microfinance

The seven experts are listed in the Acknowledgments.

3.2 Candidate indicators

Before the two-day expert meeting, the author—based on experience with data-based poverty scorecards for about 45 countries—culled about 60 candidate indicators from items in the questionnaire for China's 2011 Rural Household Survey. When available, published tabulations from the RHS (National Bureau of Statistics, 2011) were used to help narrow the list (for example, by dropping from consideration rare consumer durables such as cameras) or to re-group response options.[8] Drawing indicators from the RHS helps to limit the scope of the task and saves time in the expert meeting. Also, the RHS questions have precedent and reflect years of thought and testing. Furthermore, the same questions are in China's Rural Poverty Monitoring Survey, so they should be relevant and easy to ask.[9]

[8] Indicators based on recall or opinion were also eliminated, as were indicators that are not usually acceptable to users.

[9] At first, it was hoped that using RHS indicators would permit calibrating expert-based scores to data-based poverty likelihoods, should there later be indirect access to RHS data. The Chinese experts, however, sometimes changed questions or response options, breaking the perfect correspondence with the RHS. In any case, if there turns out to be indirect access to the RHS data later, then it should be possible to construct a completely data-based scorecard from scratch.

The candidate indicators covered seven broad areas:

- Household demographics
- Education
- Employment
- Insurance coverage
- Housing
- Consumer durable assets
- Productive durable assets

The "unit-weight" scoring literature emphasizes the importance of variety in indicators, so the experts were asked to select at least one indicator from each group and no more than four from any single group. The goal was to end up with a scorecard with 15–20 indicators.

3.3 Criteria for selecting indicators and grouping responses

Much as in this paper, the experts were first introduced to the basic concepts of poverty scoring, including the distinctions between data- and expert-based scorecards and between consumption- and asset-based poverty. To demonstrate that small refinements to expert scorecards do not matter much, the group compared accuracy for a data-based scorecard and an expert, unit-weight scorecard for Cambodia that used common-sense indicators and responses. Given a consumption standard, the resulting expert scorecard turned out to be less accurate than data-based scorecard, but not hopelessly so (see Section 5).

After that, the experts—with the help of the author and GF's facilitator—together selected 1–4 indictors from each of the seven groups, resolving disagreements by voting. The experts were asked to base their judgments on the indicators':

- Strength of association with poverty
- Ease of collection
- Acceptability and concordance with common sense
- Validity and reliability
- Evenness of distribution of responses
- Resistance to misrepresentation
- Variety when combined with other indicators
- Relevance for low-consumption/low-asset households
- Relevance in rural areas
- Stability of relationships with poverty over time
- Possibility of change as a given household's poverty changes over time
- Relevance for women and for the Millennium Development Goals

In the debate, some experts consulted results from regressions that they had run with RHS data. The experts were engaged and often disagreed, but they worked together quickly and amicably toward reasonable compromises.

At the end of the first day, the experts had settled on a draft scorecard with about 18 indicators with corresponding response options and points.

3.4 Field test

The second day began with two teams applying the scorecard to about 7 households each in a rural village about two hours north of Beijing. The goals were to detect problematic indicators, note response options in need of modification, discover holes to be plugged in the "Guidelines to Interpretation", and in general to check how well the scores fit with the experts' subjective perceptions of the households' poverty. In particular, the field test highlighted:

- The usefulness of the insurance question, something that had been in doubt before
- The awkwardness a single, all-rural-China scorecard, given regional variation
- The need to revise some response options to be both exhaustive (covering all possible answers) and exclusive (matching any given answer to only one option)
- The need for the "Guidelines for Interpretation" to clarify a plethora of details

3.5 Finalizing the scorecard

The experts met for the last time in the afternoon after the field test. They modified a handful of indicators (dropping, adding, and combining), and they reviewed and adjusted the points to ensure that they were of a consistent scale and that they reflected their judgments about the strength of the relationships with poverty.

As on the first day, most disagreements and doubts on the second day related to the constraints to:

- Hew closely to the RHS indicators and response options
- Make a single scorecard for all of rural China, rather than many regional ones

After acquiescing to these constraints, the experts agreed on a final set of indicators, response options, and points, giving a scorecard with 16 indicators that produces scores that range from 0 (most poor) to 87 (least poor).

4. Analyzing the distribution of scores

Scores from the expert-based poverty scorecard for rural China are ordinal, not cardinal, so analysis can proceed only in terms of their distribution, not in terms of their average. This section shows how to do this.

4.1 Why analyze?

The poverty scorecard for rural China seeks to help inform decisions by managers in pro-poor organizations. The purpose of a decision is to cause an improvement in the fulfillment of a mission. Thus, a decision rests on a belief in the ability to cause a change. For example, a manager might decide to offer savings services—and not only loans—because she believes that savings will deepen and broaden poverty outreach. If the manager does not believe that a decision will probably cause positive change, then the decision is—at best—senseless or random.

For their part, poverty scorecards—be they expert-based or data-based—only measure change. Just as a scale reports weight but not why weight has changed, scorecards do not reveal what causes observed changes in poverty. This presents a challenge: managers use scoring to inform decisions, and decisions follow from beliefs in causes, but scoring by itself does not identify causes.

Fortunately, scoring is only one of many inputs in a decision. In particular, managers also use other data and their intelligence, knowledge, experience, and judgment. It makes sense for managers to seek more information and expend more energy in decision-making roughly up the point where the cost of doing so equals the expected improvement in the decision. At that point, managers will have guesses as to the causes of effects, judgments about the drivers under their control that can affect the causes, and rough estimates of the costs and benefits of it all, along with a lot of uncertainty. That is how real-life choices go.

Poverty scoring can add value in this messy process because it is low-cost and unusually quantitative, contributing information of a different nature than other sources or information that managers might overlook (because social accounting is less developed and more difficult than financial accounting).

If scoring is to point managers towards causes of changes, then scores need to be compared with something, that is, put in context. Juxtaposing scores with other factors can help managers divine the *why* behind observed relationships. This is what *analysis* means: to put in context to get hints about where to look deeper for causes.

For example, suppose that comparing distributions of scores suggests that new participants in 2012 were less poor than new participants in 2011. This might spur managers to think about possible causes of this change and to search for drivers at their disposal to nudge the causes in the desired direction. For example, managers could ask whether policy changes in 2012 can explain lower poverty among new participants. Or managers could compare the distribution of poverty among male and female participants. If male participants are less poor, and if the composition of new clients shifted towards males in 2012, then managers could ask whether this caused the fall in poverty among incoming participants and perhaps whether a possible driver to tilt the composition back toward females would be a new product tailored to female-specific savings demands (*e.g.*, Vonderlack and Schreiner, 2001).

In this process, data are incomplete, and managers do the heavy lifting; the analysis of scores just points at a few promising stones to look under.

4.2 Analyzing distributions

Scores from an expert-based scorecard are ordinal; lower scores mean more poverty, but—for example—halving a score does not double the implied poverty. Numbers represent scores, but those numbers are merely ordered symbols (like letters in the alphabet or colors in the spectrum) rather than markers of uniform intervals on a line. Because the ordered symbols are not all defined in the same units, they cannot be added or averaged. Analysis must therefore proceed in terms of distributions, that is, the relative frequencies of each score.

So what do scores represent? And why can points be added up to get scores, but scores cannot be added with other scores?

Each score is the sum of points corresponding to one or more combinations of response values for the scorecard's 16 indicators. For example, score 0 corresponds to having the "poorest" response to all 16 indicators.[10] Likewise, score 1 corresponds to having the "poorest" response to all 16 indicators, except for having the "second-poorest" response for one of indicators 1, 2, 3, 4, 9, 13, or 15. So score 1 maps seven sets of responses to a single level of poverty that is less than that of score 0.

The "poverty units" of a given score are defined by the sets of responses that add to that score. These sets are unique for each score, so every pair of scores has different differences between their sets of responses. For example, the "poverty unit" difference between scores 0 (1 combination) and 1 (7 combinations) is not the same as the difference between scores 1 and 2 (32 combinations). With scores, $2 - 1 \neq 1 - 0$, so the poverty of a group cannot be summarized in anything briefer than a distribution.[11]

4.3 Example analysis of distributions

Figure 1 shows hypothetical data on the distribution of poverty scores for incoming participants in 2011 and incoming participants in 2012. A *distribution* is the cross-tab frequency of participants by score. For example, in 2011 no incoming participants had score 0, 13 had score 1, 37 had score 2, and so on.

---

[10] This treats the two options with zero points in indicators 2 and 4 like a single option.

[11] This is true for scores from any scorecard. If there are data on a definition of poverty measured without reference to the scorecard itself (such as consumption), then scores can be calibrated to cardinal numbers which can be added and averaged to give single-figure summaries.

The distribution can also be expressed as the percentage of participants by score. This facilitates comparisons when—as here—the number of participants differs by group (10,000 in 2011 versus 13,000 in 2012). In 2011, 0 percent of incoming participants had score 0, 0.13 percent had score 1, 0.37 percent had score 2, and so on.

Comparing percentage distributions for 2011 and 2012, the share of new participants is higher in 2012 for scores 0–18. After that crossing point, the share of new participants is lower in 2012.

So were new participants poorer in 2011 or 2012? Figure 2—depicting the percentage distributions in Figure 1—shows that their poverty was greater in 2011. This is because lower scores (representing greater poverty) were a higher share of scores in 2011, and higher scores (representing less poverty) were a lower share in 2011. The two distributions cross once.

As a rule, if two percentage distributions cross only once, then the one more "to the left" (like 2011 in Figure 2) represents greater poverty. If two distributions do not cross at all, then again the one more "to the left" represents greater poverty (Figure 3). Finally, if two distributions cross more than once (Figure 4), then there is no clear-cut implication, as each distribution is sometimes more "to the left".[12]

Another way to look at these scores is via cumulative distributions. Whereas a percentage distribution gives the share of participants with a given score, a cumulative distribution gives the share of participants with a given score or lower. For the 2011 example in Figure 1, the cumulative distribution at score 5 is the sum of the frequency distributions for scores 0, 1, 2, 3, 4, and 5, or $0 + 0.13 + 0.37 + 0.61 + 0.82 + 1.01 = 2.94$. That is, 2.94 percent of new participants in 2011 have score 5 or lower.

Figure 4 shows the cumulative distributions that correspond to the percentage distributions in Figures 1 and 2. The 2011 cumulative distribution is "to the left" of the 2012 cumulative distribution; for all scores, a higher share of 2011 participants have a given score or lower. For example, Figure 4 shows that about 69 percent of new participants in 2011 had scores of 33 or lower, versus about 63 percent in 2012.

Figures 5 and 6 are the cumulative distributions that correspond to the percentage distributions in Figure 3 (no crossing) and Figure 4 (double-crossing). In Figure 5, 2011 is "to the left" and thus poorer than 2012. In Figure 6, the cumulative distributions cross, so the comparison is indeterminate; 2012 is "to the left" for scores of 44 or less, but 2011 is "too the left" for scores of 45 or less.

These examples show that the analysis of score distributions has one strength and two weaknesses. The strength is that it can sometimes reveal the direction of

[12] In Figure 4, 2012 is leftmost for scores 0–57, but 2011 is leftmost for scores 33–87. Regardless of the overlap, the net difference is unknown because scores are ordinal.

change. The weaknesses are that it does not put a single figure on the size of the change and that it sometimes cannot reveal even the direction of change.

In practice, score distributions will differ from the examples here by being:

- More jagged, due to sampling variation and other errors
- Closer together, because poverty changes slowly

Thus, real-world comparisons of score distributions will not always be clear-cut. If there are small differences between distributions or multiple crossings with irregular weaving, then it probably suggests little or no change, or at least that samples are too small—or that time frames are too short—to detect change with much confidence.

The above discussion of cumulative distributions also suggests that organizations could set cut-offs—in the example of Figure 5, at score 33—to segment participants into two groups:

- Those at or below the cut-off
- Those above the cut-off

Slicing the cumulative distribution in this way can be useful because it produces single-figure measures. In some cases, it can also produce unambiguous comparisons even if the difference between the entire distributions is indeterminate. For the case in Figures 4 and 7, the change in poverty between 2011 and 2012 is known if the user only cares about the share of new participants who score, say, 33 or less.

Using cut-offs can help to set benchmarks, track change, and compare poverty across groups or organizations. After an initial application of the expert scorecard, for example, an organization might find that 25 percent of new participants in the past year scored 33 or less, and they might then set a goal of having 28 percent of new participants in the next year score 33 or less. The exact meaning of the poverty levels implied by a cut-off is difficult to make explicit, as it corresponds to perhaps millions of sets of responses to scorecard indicators. Still, the cut-offs allow an organization to be explicit and quantitative about how it wants poverty by this measure to change over time, or about how poverty by this measure has in fact changed over time, or about how poverty differs between groups or organizations.

Of course, score cut-offs are not poverty lines, so users of the expert scorecard should say “those at or below score *x*” and “those above score *x*”, not *poor* and *non-poor*. This is because scores rank participants by relative poverty; everyone is poor (or non-poor), just to different degrees. Using *poor*/*non-poor* is thus misleading. It is also confusing; different users can legitimately use different cut-offs, and even a single user may have reason to use different cut-offs in different analyses.

5. Accuracy vis-à-vis consumption-based poverty

The expert scorecard for rural China defines poverty in terms of the assets captured by its indicators, so it is, by definition, 100-percent accurate. Nonetheless, it is useful to check the concordance between its household rankings and those of a data-based scorecard with a consumption definition of poverty. Of course, this is not an endorsement of the consumption view as the gold-standard benchmark; rather, consumption is merely the most relevant alternative for which comparisons are possible.

Comparisons of ranks are reported for three countries—Bosnia-Herzegovina, Cambodia, and Mali—that have expert-based poverty scorecards whose indicators and response options are drawn from items in a national consumption survey and that also have a data-based scorecard based on consumption poverty from that same national survey. The results suggest that there is reason to believe that the expert poverty scorecard for China will be usefully accurate.

5.1 Bosnia-Herzegovina

The approach to poverty scoring developed by Schreiner (2006) grew out of an expert scorecard for Bosnia-Herzegovina (BiH) by Matul and Kline (2003). They use a 0/1/2 "unit-weight" point scheme and seven indicators loosely based on items in BiH's 2002 Living Standard Measurement Survey (LSMS). This linkage allows Schreiner *et al.* (2004) to compare consumption-based ranks from the expert scorecard (Figure 8) and a "hybrid" (Caire, 2003) expert-and-data-based scorecard that uses the same seven indicators and the same response options but that selects points based on data.[13]

Figure 9 compares ranks from the two scorecards with consumption-based poverty in the 2002 BiH LSMS. The horizontal axis is the share of households below a given percentile based on scores from a given scorecard. The vertical axis is the share of all poor households with scores at or below a given percentile.[14] Accuracy increases as a scorecard's curve approaches closer to the upper-left corner.

The hybrid scorecard is more accurate because it concentrates more consumption-poor households in lower scores. But the difference is not great. For example, about 63 percent of consumption-poor households are in the bottom quintile of scores from the hybrid card, versus about 59 percent for the bottom quintile of scores from the expert card. If this hybrid scorecard for BiH is usefully accurate, then the expert scorecard probably is too.

[13] Instead of keeping Matul and Kline's seven indicators, the scorecard could have been reconstructed from scratch to be completely data-based, but this was not done.
[14] The national poverty line in BiH for 2002 is 2,200 Convertible Marks, giving a poverty rate of about 19 percent.

5.2 Cambodia

On the first day of the expert meeting for rural China's poverty scorecard, the author compared accuracy vis-à-vis consumption poverty by Cambodia's national poverty line[15] for two scorecards:

- Expert-based with 30 indicators and an asset view of poverty
- Data-based with 10 indicators and a consumption view

Both scorecards used indicators from Cambodia's 2004 Socio-Economic Survey (Schreiner, 2009), and the data-based scorecard also used data-based points. As for BiH, the Cambodia expert card had a "unit-weight" point scheme. Its indicators and response options were pre-selected by the author, but they were all common-sense and would not differ much if the Chinese experts—even without knowing what country they were dealing with—had done the selection.

Figure 10 compares the two scorecards' ranking power. The data-based scorecard is more accurate, but again the difference is not great. For example, about 46 percent of consumption-poor households are in the bottom quintile of scores from the data-based card, versus about 41 percent for the expert card. In Cambodia as in BiH, if the data-based card is usefully accurate, then the expert card probably is too.

5.3 Mali

Schreiner (2007) made an expert-based poverty scorecard for pro-poor organizations in Mali. An expert card was done because data from Mali's 2001 Poverty Evaluation Survey was unavailable. Even so, all 20 indicators in the expert card were drawn from items in the national survey's questionnaire in the hope that the expert scores could later be calibrated to data.

Soon after, the data became available. Instead of just calibrating poverty likelihoods to the expert-based scores, however, a new data-based scorecard for consumption-based poverty was also constructed from scratch (Schreiner, 2008).[16]

Figure 11 compares the two scorecards' ranking power. As usual, the data-based card is more accurate, but only slightly. For example, about 32 percent of consumption-poor households are in the bottom quintile of data-based scores, versus about 30 percent by expert-based scores. These results are like those of BiH and Cambodia.

---

[15] In average 2004 prices, the national line is KHR1,825 per person per day, giving a poverty rate of about 30 percent.

[16] This used Mali's national poverty line of FCFA395 per person per day with an associated poverty rate of about 57 percent.

In the three tests here, expert-based poverty scorecards rank households by consumption-based poverty status almost as well as data-based scorecards. This bodes well for the potential usefulness of the expert-based poverty scorecard for rural China. This may seem surprising, but it fits well with decades of scoring research.

6. Why expert scorecards work

Data-based scorecards beat expert-based scorecards, but not by a lot. Expert cards fare well as long as they have reasonable indicators whose response options are ordered in line with their correlation with the outcome of interest. These conditions are usually met for social-science outcomes such as poverty because:

- Most good indicators are intuitive
- Response options have common-sense correlations with outcomes
- Most good indicators are highly correlated with each other

Statistical fireworks may help model physical processes, but not to model people.

Could this be true? If simple, low-cost approaches are almost as powerful as complex, high-cost ones, then why do many peer-reviewed scoring articles focus on new, complex techniques? And why do many scoring firms tout neural nets or support-vector machines instead of expert scorecards?

As usual, incentives are why. The academy rewards novelty and more math, not usefulness and less math; innovation, not adoption of innovations; technical progress towards an ideal, not finding solutions that are "good enough for government work"; higher benefits, not lower costs; being right, not being helpful. There is always an obstruse statistical approach that performs better than others with given data. In most cases, however, the improvement is trivial and may be due to overfitting, that is, tailoring the scorecard so closely to a particular sample that it captures not only real patterns but also false ones that, due to chance, appear in only this sample.[17]

For their part, scoring firms need to justify their claimed value by leading their clients to believe that they possess some secret algorithmic sauce. But black-box approaches are just a way to say, "Pay no attention to the man behind the curtain". Falkenstein (2009) calls this "alpha deception", pretending to have special, valuable skills. For example, credit-scoring firms often claim that a process called "reject inference" can improve prediction by estimating whether rejected loan applicants would have repaid on time. But the complex models of reject inference do not add predictive power (Crook and Banasik, 2004; Kiefer and Larson, 2006; Hand and Henley, 1993); they are mere distractions, a magician's trick.

The previous section showed three cases of low returns to greater complexity in scorecard construction and in point schemes.[18] In the scoring literature, sharply decreasing returns to complexity is a stylized fact with a name, the *flat maximum.*

---

[17] Hand (2006) calls this the "illusion of progress" in scoring technology.

[18] Caire and Schreiner (2012) present similar results for credit scoring in three countries with cross-tab weighting, an extremely simple data-based point scheme.

The flat maximum is good news; expert cards and simple data-based cards need not sacrifice much power in return for the transparency and lower costs that promote buy-in, integration in front-line operations and in software/database systems, and the deep organizational change that ultimately drives successful adoption. Simple scorecards are not inferior, “dumbed-down” consolation prizes for the weak in math; rather, they focus on greater risks to success than technical accuracy.

According to Hand (2006, pp. 1, 3, and 12):

> A large number of comparative studies have been conducted in attempts to establish the relative superiority of [scoring] methods. . . . These comparisons often fail to take into account important aspects of real problems, so that the apparent superiority of more sophisticated methods may be something of an illusion. In particular, simple methods typically yield performance almost as good as more sophisticated methods. . . . The improvements attributed to the more advanced and recent developments are small, and aspects of real practical problems often render such small differences irrelevant, or even unreal. . . . When building predictive models of increasing complexity, the marginal gain from complicated models is typically small compared to the predictive power of the simple models.

The flat maximum shows up repeatedly in credit scoring and medical diagnosis (Breiman, 2001; Holte, 1993; Lovie and Lovie, 1986; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Myers and Forgy, 1963). It explains why, when different approaches are compared head-to-head, they usually end in a dead heat (for example, Baesens *et al.*, 2003).

One reason for the flat maximum is that, when it comes to point schemes, “It don’t make no nevermind” (Wainer, 1976). Benjamin Franklin extolled the virtue of expert cards with “unit-weight” (0/1/2) point schemes (Dawes and Corrigan, 1974), and there are mathematical reasons why greater sophistication has decreasing returns (Grove, 2003; Bloch and Moses, 1988; Einhorn and Hogarth, 1975; Tukey, 1948).

The key to scoring—be it expert-based or data-based—is picking indicators whose response options can be ordered in line with their correlation with the outcome to be estimated.[19] Having more indicators and a greater variety of indicators also helps (Einhorn and Hogarth, 1975). Deciding whether to use an indicator and how to order its response options should be easy; if it is not, then it probably should be left out.

Besides the flat maximum, the power of 0/1/2 point schemes, and the tests here for BiH, Cambodia, and Mali, other research (Dana and Dawes, 2004; Cohen, 1990; Kolesar and Showers, 1985) supports the potential usefulness of expert scorecards. Of course, expert scorecards do not give something for nothing. Their value comes from their robustness against overfitting and—more important—their use of expert knowledge/judgment that a data sample—if available—might not capture as well. The ideal approach would combine data and experts, although the flat maximum suggests that the extra effort may not be worth it.

Expert scorecards can be useful because it is the nature of things (at least when measuring social-science outcomes) that "the dramatic steps in improvement in classifier accuracy are made in the simple first steps" (Hand, 2006, p. 5).

[19] For Dawes and Corrigan (1974, p. 105), "The whole trick is to decide what variables to look at and then know how to add." Or in Einhorn and Hogarth's (1975, pp. 172–173, 187) more academic words: "The weighting problem is subsidiary, to a large degree, to specifying the relevant variables to put into the model . . . . Provided that one can state the sign of the zero-order correlations between the independent and the dependent variables, one can confidently use a unit-weighting scheme".

7. Conclusion

This paper presents an expert-based poverty scorecard for rural China. Its 16 indicators, their response options, and their points were selected based on judgment and common sense by Chinese poverty experts and microfinance leaders. The scorecard's point scheme is simple, and its definition of poverty is asset-based. Field workers can apply the scorecard in about 10 minutes, producing ordinal scores. In addition to segmenting participants, the scorecard can measure the distribution of poverty among participants at a point in time, and it can often provide an indication of the direction of differences between two groups or for a single group over time. The analysis of score distributions will not always be as clear-cut as it would be with a consumption-based definition of poverty, but, until there is indirect access to China's consumption data, it is the best alternative.

The expert scorecard is designed to be simple, quick, and low-cost. The goal is to spark voluntary up-take by non-specialist managers in pro-poor organizations who want to improve their decision-making in the service of their social mission. Because "you manage what you measure", the use of poverty scoring should help to put social accounting more on par with financial accounting. It can also nudge the organizational culture towards greater transparency and intentionality.

Despite their simplicity and asset-based view of poverty, expert scorecards in Bosnia-Herzegovina, Cambodia, and Mali are almost as accurate as data-based scorecards when ranking households by consumption-poverty status. This is unsurprising, given scoring's flat maximum and the power of simple point schemes.

All in all, rural China's expert poverty card is not just "better than nothing" or "good enough for government work"; it probably has 70–90 percent of the power of a data-based scorecard tailored to consumption poverty. The empirical tests here—as well as decades of academic research and practical experience—bode well for its likely usefulness. What matters is data and common sense, not statistical sophistication, and this offers the hope that poverty scoring can reach the masses in China and elsewhere.

Guidelines for the Interpretation of Scorecard Indicators

These Guidelines are an integral part of the expert-based poverty scorecard for rural China. They define—in more detail than is feasible in the scorecard itself—the meaning and interpretation of the scorecard indicators and response options.

Quoted passages are from the National Bureau of Statistics, 2010, "Explanation of Indicators", Rural Household Survey, Beijing.

Non-quoted passages are the author's instructions (sometimes based on advice from the Chinese experts) meant to fill gaps in the NBS' formal guidelines, clarify ambiguities, or generally improve data quality and the data-collection process.

Data-based poverty scorecards derive their indicators, response options, and guidelines to interpretation directly from the data, questionnaire, and enumerator manual associated with a national expenditure survey, and it is critical that data-based scorecards replicate the wording and guidelines used in the national survey. Expert-based scorecards, however, do not need to hew so closely to the national survey. Thus, the guidelines for interpretation here go beyond those provided by China's NBS. This is done to increase consistency of interpretation across scorecard users.

All users of the expert-based poverty scorecard for rural China should follow all the guidelines here strictly and consistently. In cases where these guidelines are vague or incomplete, interpretation should be left to the judgment of the enumerator and the respondent. That is, local pro-poor organizations should *not* tell its enumerators how to interpret cases not covered here. Leaving non-covered cases to the judgment of the enumerator and respondent permits a natural variation among interpretations in vague/incomplete cases, and—when done across organizations—helps to reduce non-comparability due to different organizations' following different top-down policies.

For example, suppose—contrary to fact—that these Guidelines did not tell whether broken washing machines are to be counted. In this hypothetical case, organizations should not establish nor communicate policies about whether to count broken washing machines. If field agents ask for guidance, they should be told "Use your best judgment." This will help avoid systematic differences among how organizations treat the case of broken washing machines.

Guidelines for conducting the interview[20]

Be prepared
Carry enough copies of the scorecard with you, and at least two pencils. Study the scorecard and these guidelines carefully and repeatedly so that you can appear to glide through the interview effortlessly. You cannot expect the respondent to understand the questions and the response options unless you do.

Conduct the survey at the homestead
The survey should be done at the homestead, as this facilitates checking responses. Doing the survey in the lobby of a field office, at a participant meeting, by telephone, in a bus, out in the fields, etc. are unsanctioned, off-label uses. If the scorecard is applied off the homestead, then this should be made clear when results are reported.

Introduce yourself politely
Observe local customs when approaching the residence. Introduce yourself and greet the household, observing the appropriate formalities before explaining the purpose of your visit. Tell them that you come on behalf of your organization. Advise the household that participation is voluntary, and—before asking any scorecard questions—ask for permission to do the survey. Explain that it will take less than 15 minutes and that the household's answers will be combined with those of many other households and used to help your organization to understand its clientele better. If the survey is being done on a sample basis, explain that not all households are interviewed, just those lucky enough to be selected. Emphasize that the household's answers will not affect its relationship with the organization nor will they be shared with the government or anyone else.

Speak with a knowledgeable respondent
Address the survey to a single respondent, preferably the household head or his/her spouse. While other household members may be present and may contribute, always direct the questions to a single respondent, and give that respondent the final say. If the head or his/her spouse are not available, then address the survey to another knowledgeable adult or teenager. The respondent need not be a participant with the organization doing the survey. Do not address the survey to young children, servants, or others who are not household members. If the respondent refuses to respond, cannot respond, or seems to lie, then finish the interview normally and then discreetly discard the survey. If the issue is the respondent's lack of knowledge, then come back later to try again with a different respondent.

---

[20] Parts of this section are loosely inspired by Institut National de la Statistique et de l'Analyse Economique, 2009, *Enquête de Suivi, Enquête Modulaire Intégrée sur les Conditions de Vie au Bénin (EMICoV), Manuel de l'Enquêteur*, Cotonou.

Try to avoid doing the survey in the presence of people who are not household members
The presence non-household members may affect the respondent's answers or produce anxiety or embarrassment. When possible, politely ask non-members to retire until the survey is finished. They will not always comply, but do your best.

Complete the identifying information at the top of the scorecard first
Ask for the name of the participant in your organization who lives in the household, then write it in the space provided. Ask for (and record) the participant's national identification number. (If the participant also has an internal identification number for your organization, it could also be recorded.) The respondent may not be (and does not need to be) the person who participates in your organization.

Next, ask for the date that the participant joined your organization, and record it. If the respondent does not know the date, then probe for an approximation (for example, just the year, without a day or month). You may have to ask a series of questions about whether the person joined before or after specific memorable historical events. For example, you could ask, "Did you join before or after the current president was elected?" If the respondent provides only a month and year, then record the day as "1", as if it were the first day of the month. If the respondent provides only a year, then record the day and month as "1/1" as if it were January 1.

In the spaces identifying the "field agent", record your own name and your own national identification number. (If your organization uses an internal identification number for you, then record that as well.)

Finally, in the spaces identifying the "service point", record the name of the service point (such as a health clinic, if your organization runs health clinics, or a branch office, if you organization makes microloans) and the organization's internal identification number for that service point.

Complete the "Household Worksheet" second
Before turning to the scorecard's 16 indicators, you may find it helpful to complete the optional "Household Worksheet". The worksheet covers a series of items for each permanent household member and will supply the responses for the first six scorecard indicators. Detailed guidelines for each worksheet item are discussed below under the rubric of the indicator related to them.

Be neutral throughout the interview
People may to try to give the responses that they think that you would like to hear. To avoid revealing any expectations, keep your tone of voice and body language neutral. Do not react to responses by saying "Good" or "Correct", and do not smile or frown. Just say nothing, or "Thank you", or repeat back the response ("OK, the household has a television. The next question is . . .") Do not suggest answers. For example, if a respondent says something that does not answer the question, do not say "I think what

you mean is . . . Is that right?" Instead, probe in a way that invites an improved response, saying things like "Could you explain a little more?", "I do not understand; could you please repeat?", or "Oh, there is no hurry. Take your time to think about it." Unless the respondent asks for it, do not read the list of response options aloud. Just keep probing, politely.

Read the question exactly as written and in the order presented
Reading the question exactly as written helps to preserve neutrality. Reading the questions in the order presented helps keep responses comparable across households.

Ask the question unless you know the response with certainty without asking
In some cases, you can know a response without asking the question. For example, if the question is "Does the household own a television?", and you can see children watching television in the next room, then there is no need to ask the question.

If, however, you do not see a television, you cannot assume that the household does not have one. After all, it could be out-of-sight in another room or out for repairs. If you cannot know the response with certainty without asking, then you should ask the question of the respondent.

The same holds if the question is "How many televisions does the household own?" Even if you can see one television, there may be others that you cannot see, so you need to ask the question of the respondent.

Be patient and tactful with non-cooperative respondents
Sometimes a respondent will not answer, most often by shrugging, chuckling, or making some other non-committal sounds. Or a respondent may say, "I don't know", say something irrelevant, act bored, contradict an earlier response, tell a long anecdote, or just refuse to answer. In such cases, take your time and stay calm. If it helps, repeat the assurances given at the start of the interview, or take a break for a moment to chat about some safe subject unrelated to the survey. Always be cordial, and try to act and speak so that the respondent feels at ease. In almost all cases, the respondent is participating in the survey as a personal favor to you, the person sitting in front of him or her, and could not care less about how the survey might help your organization. Thus, maintain a cordial tone and professional demeanor, gently bringing the survey questions back to the respondent.

Get responses for all indicators
If any questions are left unanswered, then the poverty score cannot be computed. In particular, the point value of an unanswered question is not zero; indeed, it is not anything. If it is not possible to get responses for all indicators, then interview should be discarded. Of course, you should never make up answers. It is better to discard a survey as incomplete than to submit one with false information.

Be redundant and take notes
It is easy to make errors when recording responses and their point values. To reduce errors, you should circle both the verbal description of the response and its point value, and you should also write the point value in the column marked "Score". This redundancy will help you mark the response correctly in the field as well as to help the data-entry person to understand what you marked in the office.

In some cases, you may be unsure how to interpret a response, perhaps because you do not recall perfectly these guidelines or perhaps because these guidelines are vague or incomplete. When this happens, draw a square around the response and its point value, and write in the margin a detailed description of the issue. When you are back at the office, check these guidelines again. If they do not help, then you must judge—on your own—which response option fits best. In this process, you should give extra weight to the option that you marked in the field, as your first instinct is often best. Do not think about it too much, and do not consult any source other than yourself and these guidelines to resolve the issue.

Conduct the survey in Chinese, or, if necessary, in the local language
If some respondents will not understand Chinese easily, then the organization should hire a professional translator to translate the scorecard and these guidelines to the local language(s) before the survey goes to the field. To check accuracy, the organization should translate the translation back into Chinese as well as having it checked by a second translator. The final translation in the local language should be sent to GF, which will then share it with other organizations who are known to be applying the scorecard, thus ensuring a single, consistent translation across organizations.

If you yourself are not fluent in the language in which the scorecard will be administered, then you should work in tandem with a translator. The translator should study the scorecard and guidelines in the local language and know them as well as you do. During the survey, the translator should translate the responses back to you to be marked on the scorecard. Be alert and try to make sure that the translator is doing as good a job as you would do.

Do not take advantage of the household's good nature
Because you are their guest, the household may give you special treatment. Do not take advantage of this, other than to complete the survey. You may accept offers of small snacks or drinks, per local custom, but do not ask for food or drink. Do not bring anyone with you who is not employed by your organization to help you. Keep all information confidential, and do not gossip about what other households may have told you. Do not do things—such as trying to sell things to the respondent—unrelated to the task at hand.

Turn your mobile telephone off
It is not polite to ask respondents for their time and then to put them on hold while you attend to your personal business.

Thank the household for participating
When the survey is over, double-check that all questions have a response marked. Tell the household that the interview is over, repeat that their answers will be kept confidential, note that you may return later if something needs to be clarified, and profusely thank them for their time and willingness to participate.

Guidelines for specific scorecard indicators

1. How many permanent members does the household have?

According to "Explanation of Indicators", *permanent household members* are those who "live at home all year or for more than 6 months and who depend on the household for their economic livelihood. People who work away from home [migrants] who send most of their income back home should also be counted as permanent household members even though they may live away from home for more than 6 months out of the year. The same holds for national workmen and retirees who live elsewhere but who depend on the household for their livelihood. In contrast, active-duty soldiers, students attending technical secondary school or higher (excluding day students), and persons working away from home all year (except perhaps for going home for family visits or to see a doctor) who do not have stable employment or residence are not to be counted as permanent household members."

If using the "Household Worksheet", make sure that each person listed there fulfills at least one of the following criteria:

- Lives with the household at least six months of the year and depends on the household for their economic livelihood
- Works as a migrant and remits most income back to the household
- Lives elsewhere but depends on the household for their economic livelihood

Write the total number of permanent household members in the scorecard header next to "Number of HH members". Then circle the response option for Indicator 1 that corresponds to the number of permanent household members.

2. <u>How many household members are less than six years-old or more than 60-years-old?</u>

In other words, this indicator asks about the number of permanent household members who are five-years-old or younger or 61-years-old or older.

If all household members are less than six-years-old or older than 60-years-old, then circle response option (A), regardless of the number of household members. Otherwise, circle the appropriate response option according to the number of members who fall into one of these age ranges.

3. Do all household members who are of age receive compulsory education?

If you are using the "Household Worksheet", record each members' age and then circle "Not of age" for any member who is not of age to attend compulsory education.

For each member who is of age (that is, for whom "Not of age" is not circled), ask whether they attend compulsory education. If they do not, then circle "No". If they do, then ask whether the school is boarding or non-boarding. Then circle "Yes (boarding)" or "Yes (non-boarding)" as appropriate.

If a household member is not of the age to attend compulsory education but nevertheless attends compulsory education, then mark "Yes (boarding)" or "No (non-boarding)" according to the type of school, as if the person were of age.

If all household members have "Not of age" circled, then mark response (D).

If some household members are of age but have "No" circled for attending compulsory education, then mark response (A).

If some members have "Yes (boarding)" circled and none have "No" circled for attending compulsory education, then mark response (B). In other words, if all children who are of age attend school, and if some or all attend boarding school (even if some others attend non-boarding school), then circle (B).

Otherwise, mark response (C). That is, mark (C) if there are children who are of age for compulsory education and if all of them attend non-boarding schools.

4. What is the highest educational attainment of a household member who is in the labor force?

A permanent household member is considered to be in the labor force if he or she is currently self-employed or is working as a farmer, casual (day) laborer, wage- or salary-earner, or an unpaid helper/apprentice on a family or non-family farm or business. Unemployed people who are looking for work are not counted as being in the labor force. People—such as housewives—who work only on non-farming household production (such as caring for children, cooking for the household, cleaning house, shopping, caring for elderly parents, etc.) are not considered to be in the labor force.

If you are using the "Household Worksheet", circle "Y" for each household member in the labor force, otherwise circle "N". Then ask about the highest education attainment of members who are in the labor force.

Mark the response option for Indicator 4 that corresponds with the highest educational attainment of a household member who is in the labor force. If no household members are in the labor force, then mark response option (A).

If the household member in the labor force with the highest educational attainment is illiterate, then mark response option (B), even if the person has completed primary school or higher.

5. How many household members work as migrants?

According to “Explanation of Indicators”, people are considered to be *working as migrants* if they “work outside the administrative division of this township, even if they come back home frequently or even every day. Such people are considered to be migrants as long as the duration of the work exceeds a week.”

If you are working through the “Household Worksheet”, circle “Y” for each household member who works as a migrant, otherwise circle “N”. Then mark the response option for Indicator 5 based on the number of members marked as migrants.

Refer to the definition of *permanent household member* for Indicator 1 to determine whether migrants are to be considered as permanent household members. In general, migrants are household members if they send most of their income home, regardless of how long or often they live with the household.

6. <u>How many household members have stable income from employment or work outside of agriculture?</u>

Stable income from employment means wage or salary income from a job. It does not include non-employment sources of income such as savings and investments, insurance payments, remittances, or government transfers. *Work outside of agriculture* means employment or self-employment outside of farming, forestry, animal husbandry, hunting, and fishing.

If you are using the "Household Worksheet", the answer to this question is "No" for those outside the labor force, for casual laborers (also known as "day laborers", that is, those who sell their manual labor by the day or hour for short periods and for specific, once-off tasks to different employers), for farmers, and for unpaid helpers/apprentices in any type of enterprise or farm.

The answer to this question is "Yes" for those in the labor force who sell their labor to a single employer for a long period or on a permanent basis for a salary or wage, for those who are self-employed outside of agriculture, and for paid helpers/apprentices in non-agricultural enterprises.

The answer to this question is "Yes" for farm workers who receive a salary or wage, but it is "No" for self-employed farmers, even though they may employ salary or wage workers or casual (day) laborers.

In summary, this question does not count those outside the labor force, those who do not receive a stable income, and those who are self-employed farmers. It does count all others.

7. What are the main construction materials of the residence?

According to “Explanations of Indicators”, this refers to materials “used for weight-bearing structures such as beams, posts, and walls. *Reinforced concrete* is concrete or cement strengthened with rebar. *Masonry and wood* is bricks, stones, and timber (it excludes adobe houses with only bricks and stones without wood). *Other* refers to everything except reinforced concrete and masonry and wood.”

Main means the material that accounts for the largest share. For example, if 55 percent of the residence’s weight is borne by structures made of masonry and wood, and 45 percent is borne reinforced concrete, then the main construction material is masonry and wood. As another example, if 20 percent of a residence’s weight is borne by reinforced concrete, 35 percent is borne by masonry and wood, and 45 percent is borne by other materials, then the main construction material is “other”.

If a respondent insists that two materials each bear exactly half the weight, or that three materials each bear exactly one-third the weight, then mark the worst-quality material as being the most common.

If the respondent cannot determine the main material, then take a quick look around the residence and record your best guess.

8. What is the main fuel used for cooking?

Record the fuel used to cook the largest share of the household's meals in the past year. Because households may use different fuels in different seasons, the main fuel may not be the fuel in use on the day of the interview. Only consider fuels used for cooking, not fuels for lighting or heating.

If the household says that it does not know which is its "main" cooking fuel, then ask the household about what cooking fuels it uses throughout the year and then make your own best judgment about which is the main one.

9. What is the household's main source of drinking water?

This asks for the source of drinking water used in the largest share of days in the past year. Because households may use different water sources in different seasons, the main source of drinking water may not be the source used on the day of the interview.

The question refers to drinking water; different sources of water may be used for washing, irrigation, and other non-drinking uses.

"Explanations of Indicators" defines the response options as follows:

- *Pond water* comes from a non-flowing source such as a natural or artificial pond, dam, water storage tank, etc.
- *Other water* refers to any source not covered in another response option
- *River or lake water* comes from a flowing source such as a river, lake, reservoir, or spring
- *Shallow well water* comes from a well with a well head less than 3 meters above the water's surface
- *Deep well water* comes from a well with a well head more than 3 meters above the water's surface
- *Tap water* is transported by pipe to a household's homestead after having been purified and sterilized in public facilities

Bottled water is purified water that is delivered in bottles rather than via pipes.

Water that is transported to the homestead by pipe but that has not been purified and sterilized in public facilities does not count as "tap water".

Water delivered via truck counts as tap water if it has been purified. Otherwise, it counts as pond water.

If the household says that it does not know its "main" source of drinking water, then ask the household about what sources it uses throughout the year and then judge for yourself which is the main one.

10. Does the household have a color TV?

According to “Explanation of Indicators”, assets (such as color TVs) “that have been damaged and cannot be used (or have not been used for a long time) and that are not planned to be used again should not be counted.”

To nail down the meaning of “a long time”, a color TV in disrepair counts if it has been broken for less than 12 months and if the household plans to have it repaired before its total time in disrepair exceeds 12 months.

If a household has a working color TV but does not use it—whether due to choice or because, for example, it does not have electricity—it still counts as having a color TV.

A household is considered to “have” a color TV if it owns one (even if it is not on the homestead) or if there is a non-owned color TV on the homestead. Thus, a household that has lent out its color TV temporarily is counted as still having it, as is a household with a color TV on the homestead that is borrowed, rented, or being purchased on installments. A household is not considered to “have” a color TV if it is lent out and is not expected to be returned.

11. Does the household have a refrigerator or freezer?

According to “Explanation of Indicators”, assets (such as refrigerators or freezers) “that have been damaged and cannot be used (or have not been used for a long time) and that are not planned to be used again should not be counted.”

To nail down the meaning of “a long time”, a refrigerator or freezer in disrepair counts if it has been broken for less than 12 months and if the household plans to have it repaired before its total time in disrepair exceeds 12 months.

The answer to this question is “Yes” if the household owns either a refrigerator or a freezer (or both) that works or that will not remain in disrepair for more than a total of 12 months.

If a household has a working refrigerator or freezer but does not use it—whether due to choice or because, for example, it does not have electricity—it still counts as having a refrigerator or freezer.

A household is considered to “have” a refrigerator or freezer if it owns one (even if it is not on the homestead) or if there is a non-owned refrigerator or freezer on the homestead. Thus, a household that has lent out its refrigerator or freezer temporarily is counted as still having it, as is a household with a refrigerator or freezer on the homestead that is borrowed, rented, or being purchased on installments. A household is not considered to “have” a refrigerator or freezer if it is lent out and is not expected to be returned.

12. Does the household have a washing machine?

An *automatic washing machine* requires operator intervention only at the beginning of the process to put the clothes in, to add detergent/bleach/other chemicals, to set any dials, and to press Start. All by itself, the machine then adds water from a piped-in source, washes and rinses the clothes, spins most of the water out so that the clothes are not dripping wet when removed, and drains the water.

A *semi-automatic washing machine* is a washing machine that is not automatic. It requires operator intervention not only at the beginning of the process but also again later, or it may require the operator to do something manually to add water at the beginning, to drain water at the end, or to move the clothes to a drying tub while they are still dripping wet. A semi-automatic washing machine may also require hand-cranking by the operator. Also, if the clothes are still dripping wet at the end of the process, then the washing machine is semi-automatic because the operator must then do something to remove this water before starting the final drying process.

According to "Explanation of Indicators", assets (such as washing machines) "that have been damaged and cannot be used (or have not been used for a long time) and that are not planned to be used again should not be counted."

To nail down the meaning of "a long time", a washing machine in disrepair counts if it has been broken for less than 12 months and if the household plans to have it repaired before its total time in disrepair exceeds 12 months.

If a household has a working washing machine but does not use it—whether due to choice or because, for example, it does not have electricity or easy access to water—it still counts as having a washing machine.

If the household owns both a semi-automatic washing machine and an automatic washing machine, then mark response option (C).

A household is considered to "have" a washing machine if it owns one (even if it is not on the homestead) or if there is a non-owned washing machine on the homestead. Thus, a household that has lent out its washing machine temporarily is counted as still having it, as is a household with a washing machine on the homestead that is borrowed, rented, or being purchased on installments. A household is not considered to "have" a washing machine if it is lent out and is not expected to be returned.

13. What is the best form of mechanized transport that the household has?

Mechanized transport is defined as a machine used for transportation. Walking, running, etc. is not mechanized transport, nor is riding an animal.

Bicycles are two-wheeled human-powered machines.

Motorcycles, motorized bicycles/scooters/mopeds are two-wheeled machines powered by a non-human source of energy such as electricity, gasoline, natural gas, etc. Machines that can run on both human power and non-human power are counted here as well.

Automobile, truck, etc. are machines with four or more wheels powered by non-human sources of energy. They include buses, vans, jeeps, etc.

Wheelchairs count as "None". Animal-pulled carts or wagons count as "Bicycle". Non-motorized tricycles count as bicycles, and motorized tricycles count as "Motorcycle or motorized bicycle/scooter/moped".

The *best form of mechanized transport* is defined as the one with the highest point value in the scorecard. Thus, if a household has both a bicycle and a motorcycle, mark response option (C). Likewise, if the household has a bicycle, a motorcycle, and a car, mark response option (D).

According to "Explanation of Indicators", assets (such as forms of mechanized transport) "that have been damaged and cannot be used (or have not been used for a long time) and that are not planned to be used again should not be counted."

To nail down the meaning of "a long time", a form of mechanized transport in disrepair counts if it has been broken for less than 12 months and if the household plans to have it repaired before its total time in disrepair exceeds 12 months.

If a household has a working form of mechanized transport but does not use it—whether due to choice or because, for example, it cannot afford to pay for fuel—it still counts as having a form of mechanized transport.

A household is considered to "have" a form of mechanized transport if it owns one (even if it is not on the homestead) or if there is a non-owned form of mechanized transport on the homestead. Thus, a household that has lent out its mechanized transport temporarily is counted as still having it, as is a household with a form of mechanized transport on the homestead that is borrowed, rented, or being purchased on installments. A household is not considered to "have" a form of mechanized transport if it is lent out and is not expected to be returned.

14. What is the best form of agricultural traction that the household has?

The *best form of agricultural traction* is the one with the highest point value in the scorecard. Thus, draught animals are the worst form (but better than nothing). Mini or walking tractors are better than draught animals. Motor vehicles, large or medium tractors, threshing machines, harvesters, or motor tricycles are the best of all.

Draught animals pull or carry things. They include horses, oxen, donkeys, mules, yaks, llamas, camels, water buffalo, and elephants. If the household has a draught animal, it counts for the purposes of this question even if the animal—for whatever reason, including illness, injury, or old age—is not used for agricultural traction.

Mini or walking tractors are agricultural machines powered by non-human, non-animal energy sources (such as diesel, coal, or gasoline) which move over land and which are guided by an operator who walks behind them.

Motor vehicles, large or medium tractors, threshing machines, harvesters, or motor tricycles are machines powered by non-human, non-animal energy sources (such as diesel, coal, or gasoline) which move over land and which are guided by an operator who rides on the machine.

For the purposes of this question, threshing machines and harvesters are considered as forms of agricultural traction even though they are not used for plowing and even though they may not move over land or may not be guided by an operator who rides on them.

Motor vehicles such as automobiles, trucks, etc. are not counted unless they are used to pull plows or other things through fields. In particular, motor vehicles used only to haul agricultural products are not counted.

According to "Explanation of Indicators", assets (such as mechanized forms of agricultural traction) "that have been damaged and cannot be used (or have not been used for a long time) and that are not planned to be used again should not be counted."

To nail down the meaning of "a long time", a form of mechanized agricultural traction in disrepair counts if it has been broken for less than 12 months and if the household plans to have it repaired before its total time in disrepair exceeds 12 months.

A household is considered to "have" a form of agricultural traction if it owns one (even if it is not on the homestead) or if there is a non-owned form of agricultural traction on the homestead. Thus, a household that has lent out its form of agricultural traction temporarily is counted as still having it, as is a household with a form of agricultural traction on the homestead that is borrowed, rented, or being purchased on installments. A household is not considered to "have" a form of agricultural traction if it is lent out and is not expected to be returned.

15. <u>What type of insurance does the household have?</u>

Insurance is a financial arrangement in which a group of households make regular payments into a pool and then receive pay-outs from the pool if a specified event occurs. Example events that trigger insurance pay-outs are illness, old age, and death.

Medical insurance has pay-outs triggered by specific types of illness.

Old-age insurance has pay-outs triggered by reaching a certain age.

Commercial insurance is an insurance arrangement run by a private, for-profit firm. *Non-commercial insurance* is run by a non-profit organization (such as a cooperative) or by the government.

A household is considered to "have" insurance of a given type and from a given source if any of its permanent members are covered by an insurance policy of that type and from that source. It does not matter who pays for the insurance; it could be the household itself, or it could be some other entity (such as an employer). What matters is that someone in the household is insured.

In principle, the response option corresponding to the most costly type of insurance should be marked. In practice for the purposes of the scorecard, the rule is that old-age insurance from a non-commercial source is always counted as more costly than medical insurance from a non-commercial source, and insurance (regardless of its type) from a commercial source is always counted as more costly than insurance (regardless of its type) from a non-commercial source.

If a household has no insurance coverage, or if it is covered only by non-medical, non-old-age insurance from a non-commercial source (say, life insurance from a non-commercial source), then mark response option (A), "None".

If a household has only medical insurance from non-commercial sources, then mark response option (B).

If a household has only old-age insurance from non-commercial sources, or it is has both medical and old-age insurance and both are from non-commercial sources, then mark response option (C).

If a household has any type of insurance from a commercial source (including medical or old-age insurance, but also including other types of insurance such as life or auto), then mark response option (D).

16. Does the household receive the Minimum Living Standard Subsidy?

The Minimum Living Standard Subsidy is a cash transfer from the government to households that have been determined to be very poor.

References

Ahmad, Yusuf; and Chor-Ching Goh. (2007) "Poverty Maps of Yunnan Province, China: Uses and Lessons for Scaling Up", pp. 143–152 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.

Alkire, Sabina; and María Emma Santos. (2010) "Acute Multidimensional Poverty: A New Index for Developing Countries", Oxford Poverty and Human Development Initiative Working Paper No. 38, `ophi.org.uk/wp-content/uploads/ophi-wp38.pdf`, retrieved 4 May 2012.

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

Bedi, Tara; Coudouel, Aline; and Kenneth Simler. (2007) *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, Washington, D.C.: World Bank.

Bloch, Daniel A.; Lincoln E. Moses. (1988) "Non-Optimally Weighted Least Squares", *American Statistician*, Vol. 42, No. 1, pp. 50–53.

Bollen, Kenneth A.; Glanville, Jennifer L.; and Guy Stecklov. (2007) "Socio-Economic Status, Permanent Income, and Fertility: A Latent-Variable Approach", *Population Studies*, Vol. 61, No. 1, pp. 15–34.

Breiman, Leo. (2001) "Statistical Modeling: Two Cultures", *Statistical Science*, Vol. 16, No. 3, pp. 199–231.

Caire, Dean. (2004) "Building Credit Scorecards for Small-Business Lending in Developing Markets", `http://www.microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, retrieved 4 May 2012.

_____.; and Mark Schreiner (2012) "Cross-Tab Weighting for Retail and Small-Business Scorecards in Developing Markets".

Camacho, Adriana; and Emily Conover. (2011) "Manipulation of Social-Program Eligibility", *American Economic Journal: Economic Policy*, Vol. 3, No. 2, pp. 41–65.

Coady, David; Grosh, Margaret; and John Hoddinott. (2002) "The Targeting of Transfers in Developing Countries: Review of Experience and Lessons", `info.worldbank.org/etools/docs/library/79646/Dc%202003/courses/dc2003/readings/targeting.pdf`, retrieved 4 May 2012.

Cohen, Jacob. (1990) "Things I Have Learned (So Far)", *American Psychologist*, Vol. 45, No. 12, pp. 1304–1213.

Crook, Jonathan; and John Banasik. (2004) "Does Reject Inference Really Improve the Performance of Application Scoring Models?" *Journal of Banking and Finance*, Vol. 28, No. 4, pp. 857–874.

Dana, Jason; and Robyn M. Dawes. (2004) "The Superiority of Simple Alternatives to Regression in Social-Science Prediction", *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 3, pp. 317–331.

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

_____; and Bernard Corrigan. (1974) "Linear Models in Decision Making", *Psychological Bulletin*, Vol. 81, No. 2, pp. 95–106.

Dekker, Marleen. (2006) "Estimating Wealth Effects without Expenditure Data: Evidence from Rural Ethiopia", *Ethiopian Journal of Economics*, Vol. 15, No. 1, pp. 35–54.

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

Einhorn, Hillel J.; and Robin M. Hogarth. (1975) "Unit-Weighting Schemes for Decision-Making", *Organizational Behavior and Human Performance*, Vol. 13, pp. 171–192.

Falkenstein, Eric. (2009) *Finding Alpha: The Search for Alpha when Risk and Return Break Down*, Hoboken: Wiley.

Filmer, Deon; and Lant Pritchett. (2001) “Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India”, *Demography*, Vol. 38, No. 1, pp. 115–132.

Grove, William M. (2003) “Correction and Extension of Wainer’s ‘Estimating Coefficients in Linear Models: It Don’t Make No Nevermind’”, psych.umn.edu/faculty/grove/127correctionandextensionof.pdf, retrieved 4 May 2012.

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

_____; and William E. Henley. (1993) “Can Reject Inference Ever Work?”, *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 5, pp. 45–55.

Holte, Robert C. (1993) “Very Simple Classification Rules Perform Well on Most Commonly Used Datasets”, *Machine Learning*, Vol. 11, pp. 63–91.

Kiefer, Micholas M.; and C. Erik Larson. (2006) “Specification and Informational Issues in Credit Scoring”, *International Journal of Statistics and Management Systems*, Vol. 1, pp. 152–178.

Knowles, James C.; Bales, Sarah; Cuong, Le Quang; Oanh, Tran Thi Mai; and Duong Huy Luong. (2008) “Health Equity in Viet Nam: A Situational Analysis Focused on Maternal and Child Mortality”, background paper prepared for UNICEF Consultancy on “Equity in Access to Quality Healthcare for Women and Children” (April 8-10, Ha Long City, Vietnam).

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

Lovie, Alexander D.; and Patricia Lovie. (1986) “The Flat-Maximum Effect and Linear Scoring Models for Prediction”, *Journal of Forecasting*, Vol. 5, pp. 159–168.

Martinelli, César; and Susan W. Parker. (2007) “Deception and Misreporting in a Social Program”, *Journal of the European Economic Association*, Vol. 4, No. 6, pp. 886–908.

Matul, Michal; and Sean Kline. (2003) “Scoring Change: Prizma’s Approach to Assessing Poverty”, Spotlight Note No. 4, Warsaw, Poland: Microfinance Centre for Central and Eastern Europe and the New Independent States, imp-act.org/sites/default/files/mfc_sn4.pdf, retrieved 4 May 2012.

Myers, James H.; and Edward W. Forgy. (1963) "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol. 58, No. 303, pp. 779–806.

National Bureau of Statistics. (2011) *China Yearbook, 2010: Rural Household Survey*, Department of Rural Surveys, Beijing.

Pitchforth, Emma; van Teijlingen, Edwin; Graham, Wendy; and Ann Fitzmaurice. (2007) "Development of a Proxy Wealth Index for Women Utilizing Emergency Obstetric Care in Bangladesh", *Health Policy and Planning*, Vol. 22, No. 3, pp. 311–319.

Rutstein, Shea Oscar; and Kiersten Johnson. (2004) "The DHS Wealth Index", DHS Comparative Reports No. 6, Calverton, MD: ORC Macro, `measuredhs.com/pubs/pdf/CR6/CR6.pdf`, retrieved 4 May 2012.

Sahn, David E.; and David C. Stifel. (2003) "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data", *Review of Income and Wealth*, Series 49, No. 4, pp. 463–489.

_____. (2000) "Poverty Comparisons over Time and across Countries in Africa", *World Development*, Vol. 28, No. 12, pp. 2123–2155.

Schreiner, Mark. (2010) "Expenditure-Based Poverty: Estimates from the Ethiopia DHS", presentation for Futures Group International, L.L.C., Washington, D.C., 6 July.

_____. (2009) "A Simple Poverty Scorecard for Cambodia", `microfinance.com/English/Papers/Scoring_Poverty_Cambodia_EN_2004.pdf`, retrieved 4 May 2012.

_____. (2008) "A Simple Poverty Scorecard for Mali", `microfinance.com/English/Papers/Scoring_Poverty_Mali_2005_PPP_EN.pdf`, retrieved 4 May 2012.

_____. (2007) "A Judgmental Poverty Scorecard for Mali", report to Trickle Up and FIDES.

_____. (2006) "Un Índice de Pobreza para México", `microfinance.com/Castellano/Documentos/Scoring_Pobreza_Mexico_2002.pdf`, retrieved 4 May 2012.

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2004) "Poverty Scorecards: Lessons from a Microlender in Bosnia-Herzegovina", `microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf`, retrieved 4 May 2012.

Stillwell, William G.; Barron, F. Hutton; and Ward Edwards. (1983) "Evaluating Credit Applications: A Validation of Multi-Attribute Utility-Weight Elicitation Techniques", *Organizational Behavior and Human Performance*, Vol. 32, pp. 87–108.

Tukey, John W. (1948) "Approximate Weights", *Annals of Mathematical Statistics*, Vol. 19, pp. 91–92.

Vonderlack, Rebecca M.; and Mark Schreiner. (2002) "Women, Microfinance, and Savings: Lessons and Proposals", *Development in Practice*, Vol. 12, No. 5, pp. 602–612.

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

Wang, Sangui; Wang, Pingping; and Heng Wang. (2007) "Identifying Poverty Predictors Using China's Rural Poverty Monitoring Survey", pp. 91–116 in Guntur Sugiyarto (ed.) *Poverty Impact Analysis: Selected Tools and Applications*, Manila: Asian Development Bank.

Figure 1: Data on hypothetical distribution of scores for new clients in 2011 and 2012

	# Participants		Distribution (%)	
Score	**2011**	**2012**	**2011**	**2012**
0	0	0	0.00	0.00
1	13	86	0.13	0.66
2	37	152	0.37	1.17
3	61	191	0.61	1.47
4	82	220	0.82	1.69
5	101	243	1.01	1.87
6	119	260	1.19	2.00
7	136	274	1.36	2.11
8	152	285	1.52	2.19
9	164	294	1.64	2.26
10	178	301	1.78	2.32
11	188	307	1.88	2.36
12	199	311	1.99	2.39
13	207	313	2.07	2.41
14	215	314	2.15	2.42
15	222	315	2.22	2.42
16	227	315	2.27	2.42
17	232	313	2.32	2.41
18	236	312	2.36	2.40
19	239	309	2.39	2.38
20	240	305	2.40	2.35
21	242	302	2.42	2.32
22	242	298	2.42	2.29
23	243	293	2.43	2.25
24	241	288	2.41	2.22
25	240	283	2.40	2.18
26	238	277	2.38	2.13
27	236	271	2.36	2.08
28	232	264	2.32	2.03
29	229	259	2.29	1.99
30	225	252	2.25	1.94
31	221	245	2.21	1.88
32	216	238	2.16	1.83
33	211	231	2.11	1.78
34	206	224	2.06	1.72
35	200	217	2.00	1.67
36	195	209	1.95	1.61
37	188	203	1.88	1.56
38	183	195	1.83	1.50
39	176	188	1.76	1.45
40	170	181	1.70	1.39
41	163	173	1.63	1.33
42	157	167	1.57	1.28
43	150	159	1.50	1.22
44	144	152	1.44	1.17
45	137	145	1.37	1.12
46	131	139	1.31	1.07
47	124	131	1.24	1.01
48	117	125	1.17	0.96
49	111	119	1.11	0.92
50	105	112	1.05	0.86
51	98	106	0.98	0.82
52	93	99	0.93	0.76
53	86	94	0.86	0.72
54	81	88	0.81	0.68
55	75	82	0.75	0.63
56	70	77	0.70	0.59
57	64	71	0.64	0.55
58	59	66	0.59	0.51
59	55	62	0.55	0.48
60	49	56	0.49	0.43
61	46	52	0.46	0.40
62	40	48	0.40	0.37
63	37	43	0.37	0.33
64	33	40	0.33	0.31
65	30	36	0.30	0.28
66	26	32	0.26	0.25
67	23	28	0.23	0.22
68	20	26	0.20	0.20
69	18	22	0.18	0.17
70	15	20	0.15	0.15
71	12	17	0.12	0.13
72	11	15	0.11	0.12
73	9	13	0.09	0.10
74	7	10	0.07	0.08
75	6	9	0.06	0.07
76	5	7	0.05	0.05
77	3	6	0.03	0.05
78	3	5	0.03	0.04
79	2	3	0.02	0.02
80	1	3	0.01	0.02
81	1	1	0.01	0.01
82	0	2	0.00	0.02
83	1	0	0.01	0.00
84	0	1	0.00	0.01
85	0	0	0.00	0.00
86	0	0	0.00	0.00
87	0	0	0.00	0.00
Total:	10,000	13,000	100	100

Figure 2: Depiction of comparison of distribution of scores for new clients in 2011 compared with 2012 (single crossing)

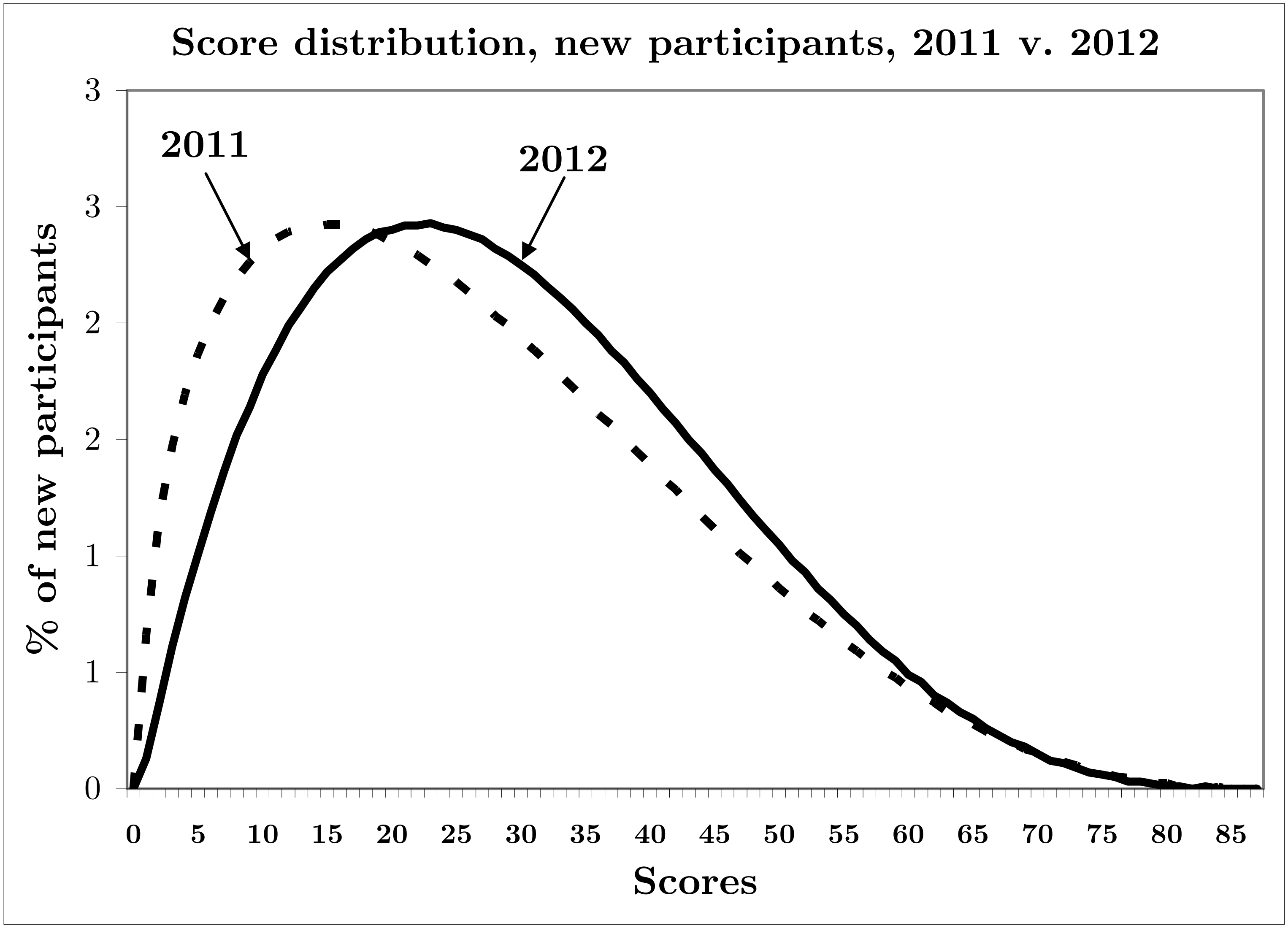

Figure 3: Example of non-crossing distributions of scores

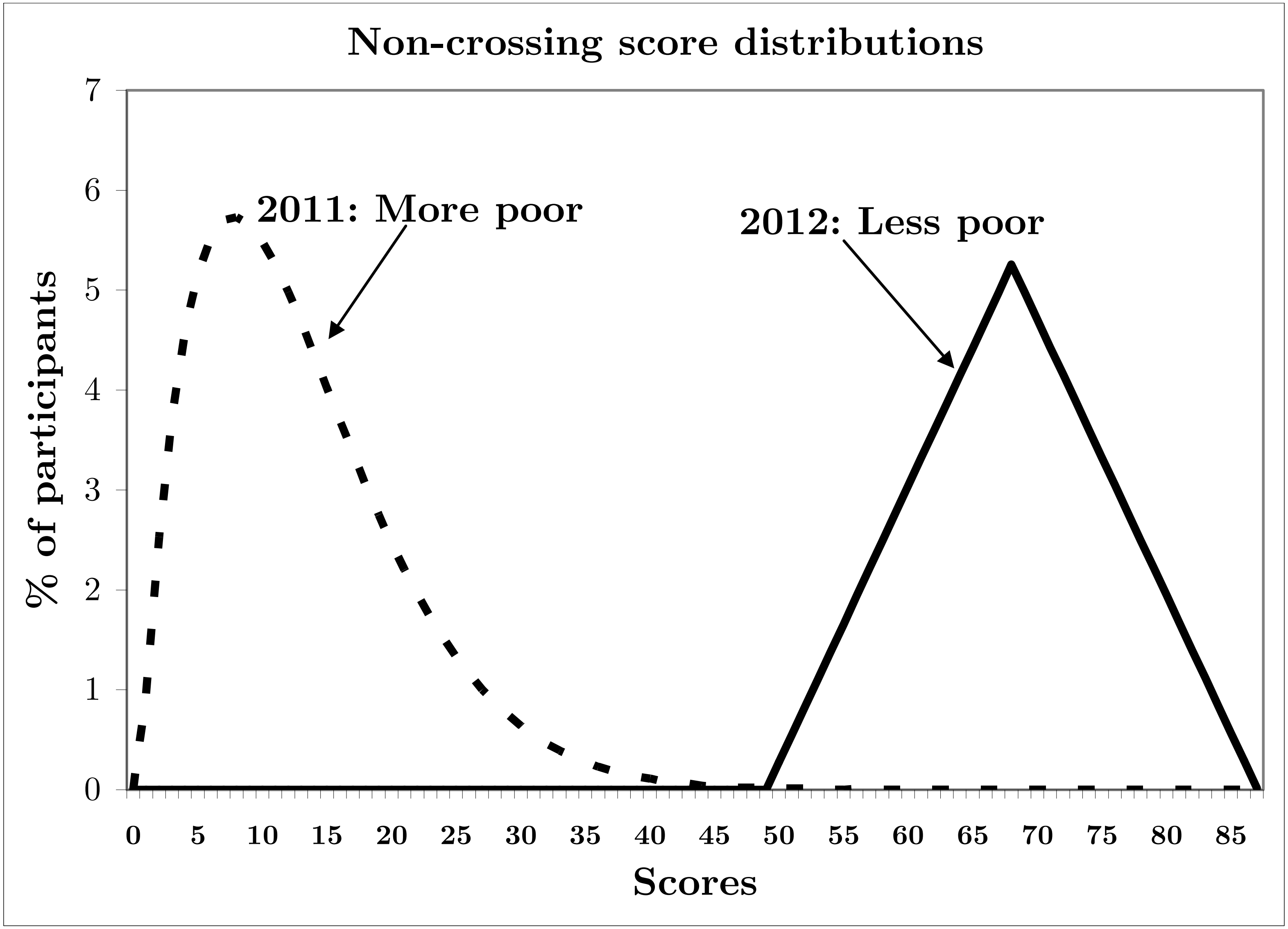

Figure 4: Example of twice-crossing distributions of scores

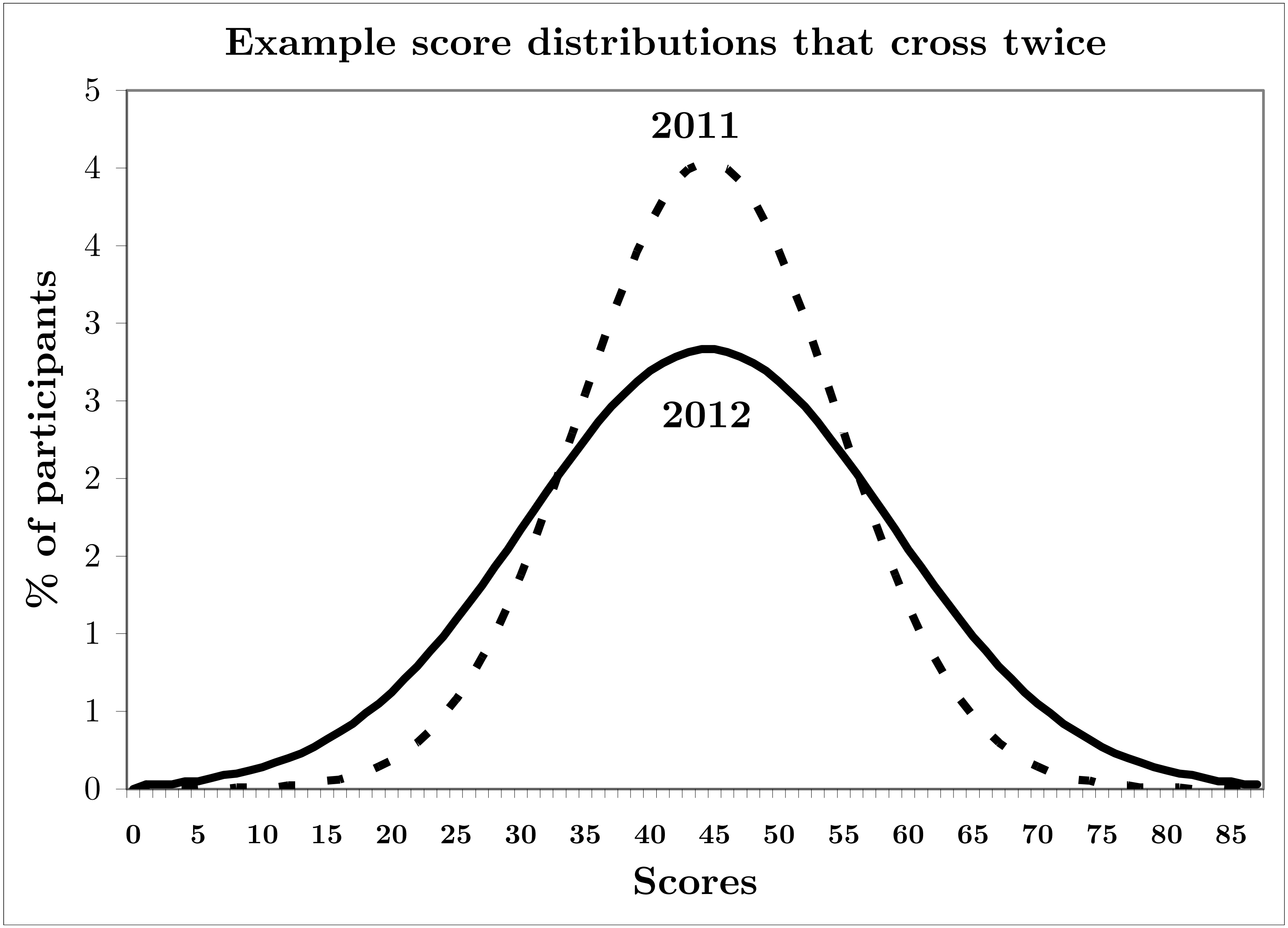

Figure 5: Cumulative distributions of scores for new clients in 2011 compared with 2012 (corresponding to percentage distributions in Figure 2)

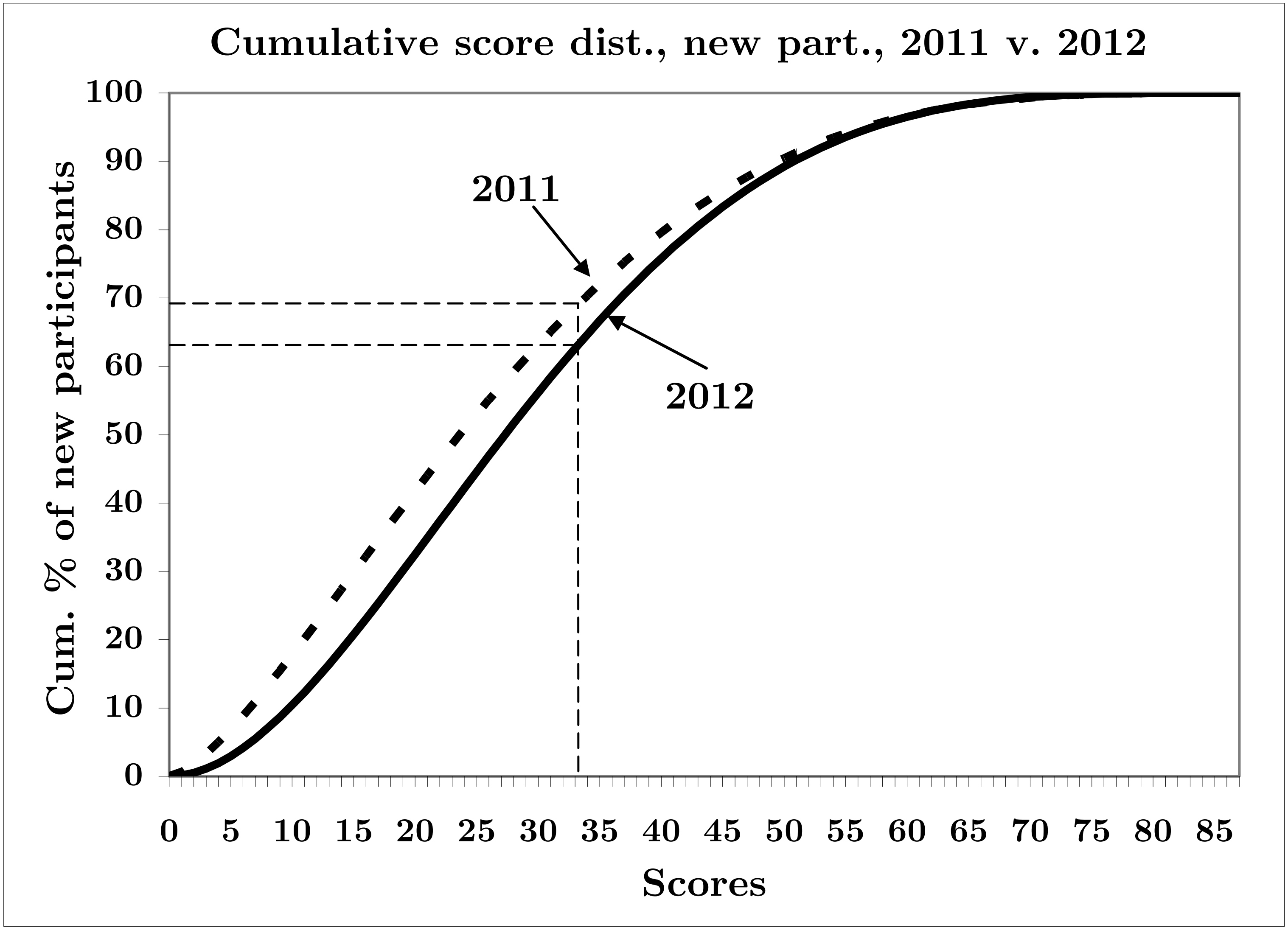

Figure 6: Example of non-crossing cumulative distributions of scores (corresponding to percentage distributions in Figure 3)

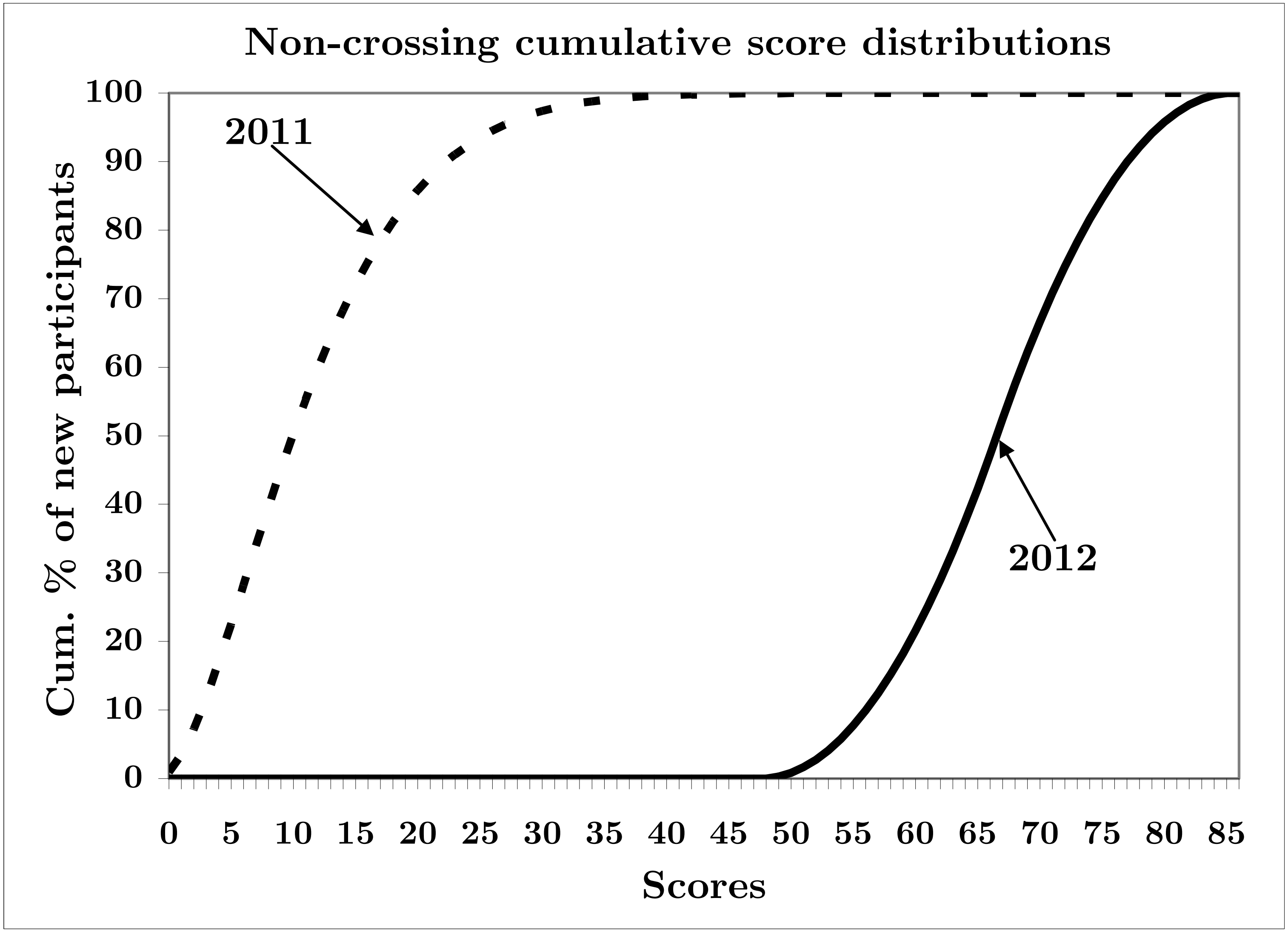

Figure 7: Example of crossing cumulative distributions of scores (corresponding to percentage distributions in Figure 4)

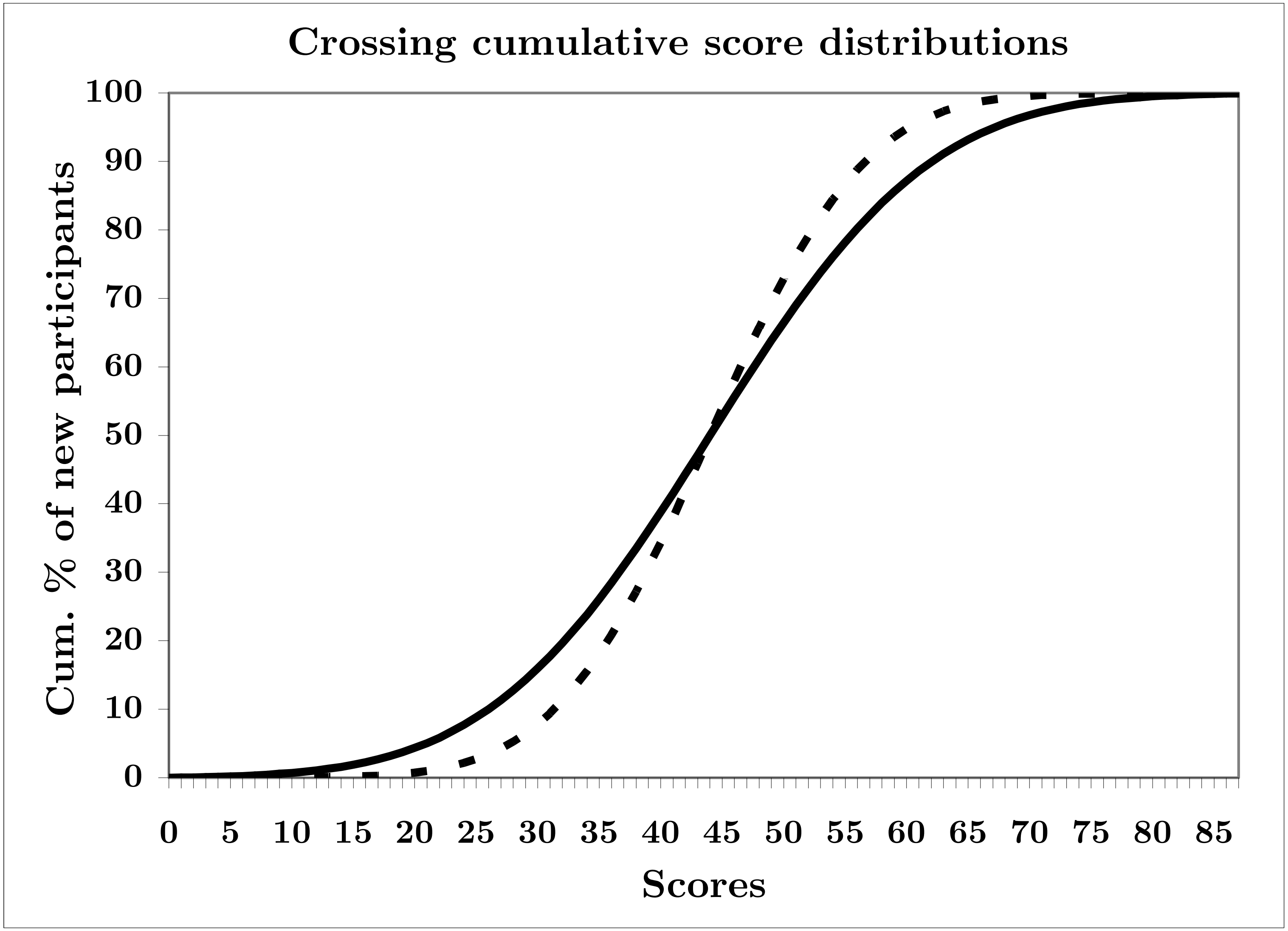

Figure 8: Expert poverty scorecard and expert-plus-data hybrid poverty scorecard for Bosnia-Herzegovina

		Points	
Indicator	**Value**	**Expert**	**Statistical**
1. Ownership of car	No	0	0
	Yes	1	12
2. Education level of female household head/partner/spouse	≤ Primary	0	0
	> Primary	1	4
3. Number of household members	6 or more	0	0
	5	1	8
	4	1	11
	3	1	19
	2	1	27
	1	1	34
4. Ownership of stereo CD player	No	0	0
	Yes	1	8
5. Location of residence	Rural or peri-urban	0	0
	Urban	1	6
6. Average times eats meat each week with main meal	Rarely (0-2)	0	0
	Sometimes (3-5)	1	8
	Often (6 or more)	2	20
7. Average times eats sweets each week with main meal	Rarely (0-2)	0	0
	Sometimes (3-5)	1	8
	Often (6 or more)	2	16
Minimum possible score (most-likely poor)		0	0
Maximum possible score (least-likely poor)		9	100

Source: Schreiner *et al.* (2004)

Figure 9: Ranking power for expenditure-based poverty status, expert poverty scorecard versus expert-plus-data hybrid poverty scorecard for Bosnia-Herzegovina

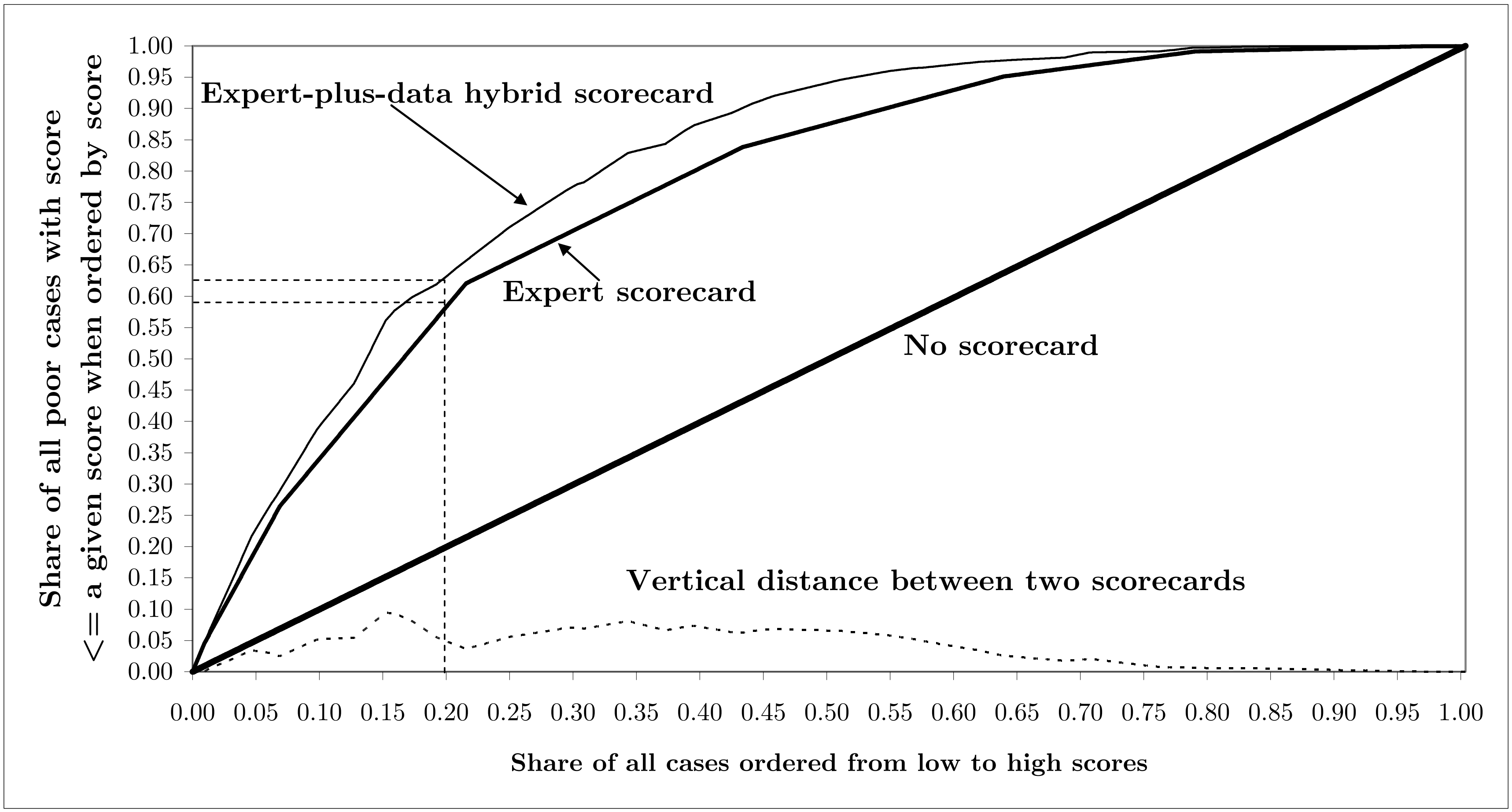

Figure 10: Ranking power for expenditure-based poverty status, expert poverty scorecard versus data-based poverty scorecard for Cambodia

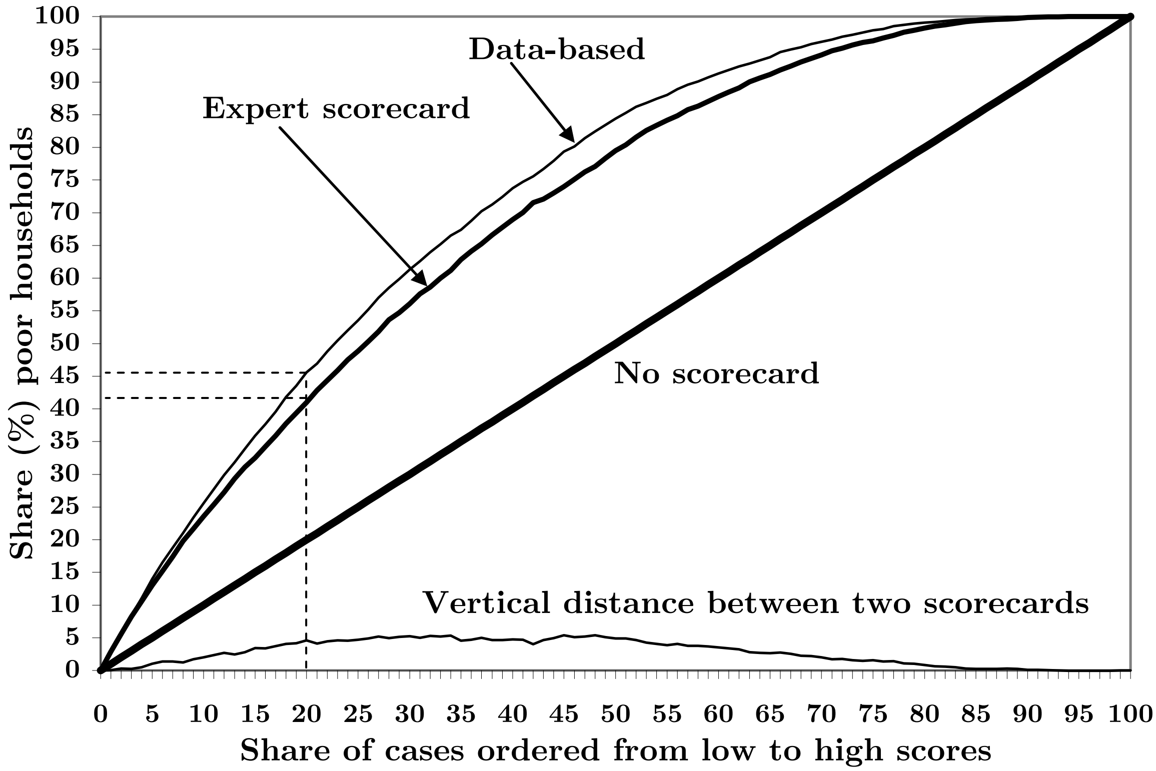

Figure 11: Ranking power for expenditure-based poverty status, expert poverty scorecard versus data-based poverty scorecard for Mali