The Power of Prizma's Poverty Scorecard: Lessons for Microfinance

Mark Schreiner, Michal Matul, Ewa Pawlak, and Sean Kline

January 6, 2005

Microfinance Risk Management
6970 Chippewa St. #1W, Saint Louis, MO 63109-3060, U.S.A.
Telephone: (314) 481-9788, http://www.microfinance.com

Abstract

How well does Prizma's poverty scorecard identify poor clients? The scorecard applies 0/1 weights to seven simple and inexpensive-to-collect indicators to give a score from 0 (most likely poor) to 9 (least likely poor). The scorecard has good power, not only for ranking clients by relative poverty but also for identifying the likelihood that a client is poor by an absolute standard. While poverty outreach varied across branches and products, the scorecard suggests that Prizma's overall poverty rate among new clients is about 18 percent, close to the rate for Bosnia-Herzegovina as a whole. Technical refinements improve the scorecard's power, though not by much, as most power comes from a single indicator, how often the client eats meat. Still, one-indicator scorecards overstate Prizma's overall poverty rate. Loan size is an indicator of poverty, but the scorecard is more powerful. Overall, poverty scoring can help microfinance organizations target the poor, track changes in clients' poverty over time, manage depth of outreach, and report on clients' absolute poverty.

Acknowledgements

Prizma's (http://www.prizma.ba) technical advisor and staff, in particular Sean Kline (sean@prizma.ba) and Maja Gizdić (maja@prizma.ba), initiatated the development of Prizma's poverty scorecard in 2002. The scorecard was built in 2003–4 with significant inputs (indicator selection, statistical testing, and pilot testing) by Michal Matul (michal@mfc.org.pl) and Ewa Pawlak (ewa@mfc.org.pl) of the Microfinance Centre for Central and Eastern Europe and the New Independent States (http://ww.mfc.org.pl). The scorecard was developed under Imp-Act (http://www.imp-act.org) and was funded by the Ford Foundation. The scorecard was based on data from a 2001 Living Standards Measurement Survey (http://www.worldbank.org/lsms/lsmshome.html) by the World Bank, the State Agency for Statistics in Bosnia-Herzegovina, the Federation Statistical Institute, and the Republika Srpska Statistical Institute. Data on Prizma's clients came from the Kredits Loan Tracking System (http://www.kreditslts.net) and was provided by Don Eggert, Mirza Čehajić, and Kenan Crnkić. This paper was funded by the Consultative Group to Assist the Poor (http://www.cgap.org).

The Power of Prizma's Poverty Scorecard: Lessons for Microfinance

1. Introduction

Development assistance in general—and microfinance in particular—aims to improve the lives of the poor. But what share of microfinance clients are poor? Current measures of poverty outreach rely on rough indicators such as lending method (group borrowers are assumed to be poorer than individual borrowers), client gender (women are assumed to be poorer than men), branch location (rural dwellers are assumed to be poorer than urban dwellers), and loan size (poor people are assumed to use small loans more than large loans). More and more microlenders are also using poverty scorecards that add up a few weighted client-level indicators to produce a score that is assumed to be associated with a poverty status (Zeller, 2004; Hatch and Frederick, 1998).

While these indicators and scores are correlated with poverty, the strength of the correlations is unknown. This paper examines the power of a simple poverty scorecard that estimates the likelihood that a given client is poor.[1] The average of each client's poverty likelihood is an estimate of the overall share of clients who are poor. The scorecard examined here was built by Prizma (a microlender in Bosnia-Herzegovina, BiH) and the Microfinance Centre for Central and Eastern Europe and the New Independent States (Matul and Kline, 2003).

Prizma's poverty scorecard is powerful; that is, it does a good job of identifying clients who are likely to be poor. Because the scorecard was derived from a Living Standard Measurement Survey (World Bank *et al.*, 2002), it also estimates absolute, expenditure-based poverty rates. About 17.9 percent of new borrowers between December 2003 and September 2004 were poor (versus 19.3 percent for all BiH).

These measures of power have two caveats. First, Prizma does not collect all the scorecard indicators exactly as in the national survey, so the estimates are biased to some unknown degree. Second, the scorecard does not completely control for the fact that Prizma's clients are not a random sample of the population of BiH.

Prizma's scorecard is simple, with seven indicators and 0/1 weights. Using statistically optimal weights improves power, but only a little, probably because one indicator—how often the client eats meat—is dominant. Very simple one-indicator scorecards overstate the likelihood that a given Prizma client is poor. In particular, loan size does not identify poor clients as well as the poverty scorecard.

[1] While this paper was in its final stages of revision, Zeller, Alcaraz, and Johannsen (2004) appeared with an analysis of the power of poverty scorecards in Bangladesh.

This paper describes Prizma's poverty scorecard and how it was built. It then measures the scorecard's power and compares it with an "expanded" scorecard with statistically optimal weights, a "meat-only" scorecard with only one indicator, and a scorecard based on loan size. Estimates of overall poverty rates from Prizma's scorecard are compared with those from one-indicator scorecards (including one based on loan size) and from a "benchmarkable" scorecard with indicators directly linked to the national survey. The conclusion discusses lessons for poverty scoring in microfinance.

2. Prizma's poverty scorecard

Constructing and applying Prizma's poverty scorecard involved:

- Measuring the absolute, expenditure-based poverty status of households in a national random sample
- Selecting non-expenditure indicators that were not only simple and inexpensive to collect but also correlated with absolute, expenditure-based poverty status
- Constructing a scorecard by assigning weights to the non-expenditure indicators to reflect their correlation with expenditure-based poverty status
- Adding up the weighted non-expenditure indicators to produce poverty scores for the surveyed households
- Collecting from Prizma's clients the non-expenditure indicators used in the scorecard and using them to produce poverty scores
- Defining the poverty likelihood of a Prizma client with a given poverty score as the observed poverty rate among surveyed households with the same score
- Defining Prizma's overall poverty rate as its clients' average poverty likelihood
- Checking that poverty scores for Prizma clients made sense for different branches, products, and geographic areas

This process rests on two basic assumptions. The first is that Prizma's clients—like the surveyed households—are a random sample from the population of BiH. The second is that the relationship between non-expediture indicators and expenditure-based poverty status does not change through time. Consequences of violating these assumptions are discussed later.

2.1 An expenditure-based measure of absolute poverty

Absolute poverty status was derived from the 2001 Living Standards Measurement Survey for BiH that recorded expenditure and a wide range of other data for a national random sample. A household was *poor* if annual per capita consumption (adjusted for the local cost of living) was less than 2,200 Convertible Marks (World Bank *et al.*, 2002). At purchasing-power parity, this poverty line was about $14 per person per day.[2] The overall poverty rate in BiH was 19.3 percent.

---

[2] The poverty line in Convertible Marks per year was changed to Purchasing Power Parity dollars per day as follows. First, the ratio of PPP dollars per dollar (5.08) was derived as 2001 GDP per capita in PPP dollars (5,970) divided by GDP per capita in nominal dollars (1,175) as reported in the *2003 Human Development Report* (`http://www.undp.org/hdr2003/indicator/cty_f_BiH.html`). Second, the December 31, 2001 exchange rate of 2.22 Convertible Marks per dollar (`http://www.cbbh.gov.ba/kursne/211201.html`) was used to convert 2,200 Convertible Marks to 991 dollars. Third, multiplying 991 dollars by the ratio of PPP

2.2 Indicators in the poverty scorecard

Prizma worked with the Microfinance Centre for Central and Eastern Europe and the New Independent States to select non-expenditure indicators that (Matul and Kline, 2003):

- Correlated strongly with poverty status, both in the past and future
- Appeared in the national survey, enabling linkage to an absolute poverty line
- Kept data-collection costs low:
    - Already collected as part of the loan evaluation, or easy to start to collect
    - Did not embarrass clients or loan officers or make them uncomfortable
- Elicited truthful reports that an internal auditor could verify
- Took different values across clients
- Took different values for a given client as poverty changes over time

Analysts first brainstormed a long list of candidate indicators, drawing on their country knowledge, poverty studies in BiH (Dunn and Tvrtkovic, 2003; Prism Research, 2003; World Bank *et al.*, 2002), and input from managers, front-line staff, and client focus groups. The list was then narrowed using the criteria above.

2.2.1 Benchmarkable indicators

For example, analysts expected that owners of cars were less likely to be poor than non-owners. In the national survey, car ownership was indeed strongly correlated with expenditure-based poverty: 11 percent of car owners were poor, versus 26 percent of non-owners (Figure 1). Car ownership also varied across households (55 percent were owners, 45 percent non-owners). In a pilot test with a prototype scorecard in one branch, Prizma found that clients were comfortable answering the question truthfully. Car ownership also promised to be a useful indicator because changes in car ownership by a given client over time are probably correlated with changes in poverty.

---

dollars to dollars (5.08) gives a poverty line of 5,034 PPP dollars per year. Finally, converting from years to days gives a poverty line of 13.79 PPP dollars per day.

Following this same process, the scorecard incorporated indicators for:

- *Education of the female household head/spouse/partner.* In the national survey, female education was highly correlated with overall household education. Also, until quite recently, all Prizma clients were women, so asking only about female's education simplified data collection. Among the 64 percent of surveyed households whose female head had only a primary education, 24 percent were poor. Among the other 36 percent of households, 11 percent were poor (Figure 1)
- *Household size.* Among the 17 percent of households with 6 or more members, 40 percent were poor. Among the other 83 percent, 15 percent were poor
- *Stereo CD ownership.* Among the 78 percent of households who did not own stereo CD players, 23 percent were poor. Among the other 22 percent, 8 percent were poor

Prizma collected these four indicators—car ownership, female education, household size, and stereo CD ownership—exactly as in the national survey. Thus, a scorecard using only these four indicators could be directly benchmarked to the national survey's absolute, expenditure-based measure of poverty status. (In principle, a four-indicator scorecard could also be benchmarked to the international $1-per-day "ultra-poor" poverty line, except that such extreme poverty was so rare in BiH that no extremely poor households were sampled in the World Bank survey.)

2.2.2 Non-benchmarkable indicators

Prizma's scorecard includes three additional indicators—location of residence, frequency of eating meat, and frequency of eating sweets—that were collected differently than in the national survey. In strict terms, scores using these indicators cannot be linked directly to the survey's poverty measure. Still, these three indicators were highly correlated with poverty, so even if they break the scorecard's direct link with an absolute benchmark, they increase power to rank clients by relative poverty.

For location of residence, Prizma recorded whether the client lives in an urban area (more than 10,000 residents) or a rural/peri-urban area. The national survey, however, assigned location status by municipality, even though many municipalities have both rural and urban areas. In the survey, about 21 percent of people in rural municipalities were poor versus 13 percent in urban municipalities (Figure 1). This does not, however, necessarily imply anything about the poverty of clients whose location of residence is defined differently. The estimates of overall poverty rates in this paper assume that Prizma's definition of rural/urban matches the national survey definition.

A CGAP Poverty Assessment Survey (Henry *et al.*, 2003) for Prizma found that the frequency of eating meats and sweets was highly correlated with poverty (Prism Research, 2003). The national survey used to build Prizma's poverty scorecard recorded spending on meat in Convertible Marks. Prizma, however, found it impractical to ask clients to report spending. Instead, Prizma asked about frequency: the times per week the household eats meat and the times per week the household eats sweets (usually cakes) with the main meal. If all households were the same size and if all people eat the same amount per meal, then frequency (measured by Prizma) is perfectly correlated with spending (measured by the survey). In fact, larger households can spend more on meat (or sweets) even if they eat less frequently, and not everyone eats the same amount. Thus, Prizma's indicator is not equivalent to the survey indicator, breaking the direct link between the score and the absolute, expenditure-based poverty benchmark.

Knowing this, the scorecard builders divided the survey distribution of spending on meat in three classes in such a way that the distribution of surveyed households matched the distribution of a sample of Prizma's clients across three frequency classes ("rarely" for 0–2 times per week, "sometimes" for 3–5, and "often" for 6 or more). Figure 1 shows that spending on meat was highly correlated with poverty in the survey: the lowest spenders had a poverty likelihood of 42 percent, versus 19 percent for those in the middle and 4 percent for the highest spenders. (Spending on sweets was also correlated with poverty, but not as strongly.) Still, the correlation between spending (measured in the survey) and frequency (measured by Prizma) is unknown. Except for one scorecard that uses only the four benchmarkable indicators, the estimates here of Prizma's overall poverty rate assume that spending and frequency are equivalent.

2.2.3 Excluded indicators

The scorecard does not include all indicators that appear in both the survey and Prizma's data base. For example, an indicator for single mothers was left out because female-headed households with children had about the same poverty rate as male-headed household with children.

Scorecard builders also considered—but ultimately rejected—some survey indicators that were strongly correlated with poverty but that fell short on other criteria. For example, refugee status in 2001 was strongly correlated with poverty (among the 16 percent of households in which the female head was a refugee, 37 percent were poor, while among non-refugees, 17 percent were poor). Prizma's managers believe, however, that this correlation is fading. In this case, including the indicator would cause the scorecard to overestimate the poverty likelihood of refugees. Also, data collection is difficult because the definition of refugee status is constantly changing among government, aid agencies, and among refugees themselves.

Likewise, the survey found that the unemployed were more likely to be poor. The survey's aggregate unemployment figure, however, was rather high, and scorecard builders suspected that many part-time or unregistered workers were counted as unemployed. Because Prizma's loan officers would (correctly) report part-time or unregistered clients as employed, a scorecard that included employment would underestimate poverty likelihood. Furthermore, Prizma's "enterprise" loan product is limited to the self-employed, so an employment indicator—regardless of its correlation with poverty—would not help rank "enterprise" clients by poverty likelihood. For these reasons, Prizma's scorecard does not include employment.

Finally, while television ownership was highly correlated with poverty, about 96 percent of Prizma's clients were owners. With so little variation across clients (and less variation through time), the indicator would not help rank clients by poverty likelihood.

2.2.4 Lessons for the selection of poverty indicators

Poverty scorecards use indicators to estimate a client's poverty likelihood, but power is not the only criteria for choosing indicators. Even indicators that are strongly correlated with poverty in a national survey might produce misleading estimates if they do not really measure what they appear to measure, if the relationship between the indicators and poverty will change, or if the lender and the survey do not record the indicators in the same way. Other indicators might offend clients, embarrass loan officers, or just be too difficult to collect accurately. Like credit scoring (Schreiner, 2002), poverty scoring depends more on data quality than technical wizardry.

Building poverty scorecards requires "domain expertise", that is, knowledge of microfinance and of how a specific lender works in its local context. Feedback from front-line staff is also key, as are pilot tests and generous doses of care and good sense. Building a poverty scorecard is not rocket science, but it is not a cake walk either. The power of Prizma's poverty scorecard comes less from the specific weights assigned to the indicators than from its someone's knowing that the cultural standard is to eat cake with the main meal and that the culture's love of music makes the lack of a stereo CD player an indicator of poverty.

2.3 Weights for indicators in the poverty scorecard

Five of the seven indicators in Prizma's poverty scorecard had "Yes/No" answers. A client either did or did not own a car, have more than a primary education, have 6 or more household members, own a stereo CD, or live in an urban area. Weights of zero (0) were given to values correlated with greater poverty in the survey, and weights of one (1) were given to values correlated with less poverty. The two remaining indicators (frequency of eating meat and sweets) had values of "rarely", "sometimes", or "often" with weights of 0, 1, or 2, again reflecting the survey correlations with poverty.

Figure 2 lists indicators, values, and weights for the "original" scorecard just described as well as for an "expanded" scorecard with statistically optimal weights (discussed below). In the original scorecard, scores range from 0 (most likely poor) to 9 (least likely poor). The 0/1 weights in the original scorecard assume, for example, that owning a stereo CD has the same link with poverty as does owning a car. Can such a simple weighting scheme accurately identify the likelihood that a client is poor?

Both practice and theory lend support to 0/1 weights. Such scorecards have been used by banks to predict creditworthiness, hospitals to identify at-risk pregnancies, phone companies to predict bill-payment, and colleges to screen potential matriculants (Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Myers and Forgy, 1963). Wainer (1976) shows mathematically why 0/1 scorecards can work. Such scorecards are also robust to dirty data and—important in some contexts—let front-line workers compute scores on paper.

Of course, statistically optimal weights cannot do worse than 0/1 weights. The next section examines how well the original and expanded scorecards identify the poor.

3. Power to identify the poor

Any scorecard can add up weighted indicators. But is the result related to poverty? Prizma's poverty scorecard was derived from a national survey with an absolute, expenditure-based measure of poverty, so its power can be checked. This section measures how well Prizma's scorecard identifies the poor in the national survey.

3.1 Prizma's original scorecard

Given a score from Prizma's original scorecard, Figure 3 shows the number of poor households surveyed, the number of all households, and the poverty likelihood. For example, 46.2 households had a score of 0 (households were weighted for national representativeness), and 44 of these were poor. Thus, a score of 0 was associated with a poverty likelihood of 95.3 percent (44) 46.2). Households with a score of 0 represented 0.9 percent of all households but 4.5 percent of poor households.

Among households with a score of 1, 71.4 percent (214 of 300) were poor. These represented 6 percent of all households and 22 percent of the poor. Among households with a score of 2, 47 percent were poor. Given that the overall poverty rate in BiH was 19.3 percent, Prizma's original scorecard effectively assigned lower scores (and higher poverty likelihoods) to poor households.

The scorecard also assigned higher scores (and lower poverty likelihoods) to non-poor households. For example, 1,069.8 households had scores in the range from 6 to 9, and 8.3 of them (0.7 percent) were poor.

Whereas Figure 3 shows the number of households with a given score, Figure 4 shows the number of households with a given score or less. For example, 1,082.6 households had a score of 2 or less, of which 604.7 (55.9 percent) were poor. These represented 21.5 percent of all households and 62.1 percent of the poor.

These cumulative measures are useful for formal measures of power. For example, "lift" in Figure 4 is the share of the poor with a given score or less, relative to the overall poverty rate. For example, 55.9 percent of households with a score of 2 or less were poor, giving a lift of 2.9 times above the overall share of 19.3 percent (55.9) 19.3 = 2.9). More lift means more power.

Another formal measure of scorecard accuracy is the "Power Curve" (Figure 5). The line curving toward the northwest corner shows power to identify the poor, plotting the share of poor households with a given score or less (vertical axis) against the share of all households with that score or less (horizontal axis). For example, 26.6 percent of poor households (and 6.9 percent of all households) had a score of 1 or less. The greater the power to identify the poor, the more the northwest curve will approach the left and top borders. For example, the curve almost touches the top border for the 20 percent of households with the highest scores, as the scorecard correctly identifies almost all of these households as non-poor.

In Figure 5, the line curving toward the southeast corner shows the power to identify non-poor households. The closer this curve is to the bottom and right borders, the greater the power. (For reference, the diagonal line in the middle of Figure 5 shows the share of poor and non-poor that would be identified in the absence of a scorecard.)

The greater the area between the curves, the greater the power. One indicator of this area is the Kolmogorov-Smirnov distance, the vertical distance between the poor and non-poor curves.[3] For Prizma's poverty scorecard, the maximum KS is 0.50.[4]

Overall, the original scorecard does well, assigning a high proportion of low scores to the poor and a high proportion of high scores to the non-poor.

3.2 Expanded scorecard

To refine Prizma's original scorecard, an expanded scorecard with statistically optimal (not 0/1) weights was constructed. It used the same indicators as the original scorecard, except household size had six classes (1, 2, 3, 4, 5, or 6 or more) rather than two (5 or less, or 6 or more). Scores ranged from 0 (most likely to be poor) to 100 (least likely to be poor). Although there is more arithmetic, expanded scores can still be computed on paper by front-line workers.

Compared with the original scorecard, the expanded scorecard places less weight on education and location of residence (Figure 2). The relative weights on ownership (car and stereo CD) and food consumption (meats and sweets) do not differ much between scorecards. The main difference is in household size, with the expanded scorecard assigning larger, more finely distinguished weights to this indicator.

Given a score from the expanded scorecard, Figure 6 shows the number of poor households surveyed, the number of all households, and the poverty likelihood. (Figure 7 shows this for households with a given score or less.) As in the original scorecard, 44 of 46.2 households scoring 0 were poor (poverty likelihood 95.3 percent).

In the expanded scorecard, 5.7 percent of households had scores from 1 to 15, with 80.2 percent of them poor. In the original scorecard, 6 percent of households had a score of 1, with 71.4 percent of them poor. Thus, the expanded scorecard more accurately identified the poor.

The expanded scorecard also identified the non-poor better. About 1,837 households had scores from 36 to 100, and 22.5 (1.2 percent) were poor. In the original scorecard, 1,827 households had scores from 4 to 9, and 47.2 (2.6 percent) were poor.

Figure 8 compares lift for the two scorecards. For all scores (except 0), the expanded scorecard has more lift. For the lowest-scoring 10 percent of households, the concentration of the poor (relative to their overall concentration of 19.3 percent) is

---

[3] Statistical measures are inferior to measures based on the benefits or costs of correctly or incorrectly identifying a poor client (Granger and Pesaran, 2000). If benefits and costs are known, power curves give the information needed to evaluate scorecards.

[4] Mays (2000) says a maximum KS from 0.41 to 0.50 is "good", 0.51 to 0.60 is "very good", and 0.61 to 0.70 is "excellent".

about 0.5 units greater with the expanded scorecard (3.9, poverty likelihood of 3.9 x 19.3 percent = 75.3 percent) versus the original (3.4, poverty likelihood of 3.4 x 19.3 percent = 65.6 percent). The original scorecard has good lift, but the expanded scorecard does better.

The expanded scorecard also has a better power curve (Figure 9), more closely approaching the northwest and southeast corners. At 0.63, it also has a greater maximum Kolmogorov-Smirnov distance.

Should the expanded scorecard replace the original one? It depends. If the goal is to identify clients likely to be non-poor (perhaps to disqualify them from participation), then the KS distances at high scores in Figure 10 show that both scorecards work about equally well. To identify clients likely to be poor (or to estimate the overall share of poor clients), however, then the KS distances for low scores in Figure 10 show that the expanded scorecard works better. Still, the original scorecard is remarkably powerful, especially given its simple weighting scheme.

3.3 Indicator importance

How important is each indicator for power? For the expanded scorecard, importance is measured as the reduction in the log-likelihood (a statistical measure of power) caused by removing a given indicator while keeping all others (Brieman, 2001). Figure 11 shows the results, normalized on a scale from 0 to 100.

In the expanded scorecard, household size is the most powerful indicator, followed by frequency of eating meat, frequency of eating sweets, and car ownership. The final three indicators—ownership of a stereo CD player, location of residence, and education—contribute little.

Except for household size, the most-important indicators are liable to change in the mid-term if poverty changes. This suggests that the expanded scorecard can track changes in poverty over time. Location of residence and education are not likely to change with poverty status, but they are not powerful indicators anyway.

Because the original scorecard was not derived statistically, importance is measured as the change in the maximum KS distance caused by removing a given indicator while keeping all others. Two results stand out in Figure 12. First, the frequency of eating meat is by far the most powerful indicator. Second, removing location of residence, education, or frequency of eating sweets *increases* the maximum KS distance.[5,6]

[5] While adding indicators to a statistical (regression) scorecard cannot harm power in terms of R^2, this scorecard is not statistical, and the maximum KS distance is not R^2.
[6] This does not mean that removing all three indicators at once would improve power. Furthermore, while removing one indicator at a time increases the maximum KS distance, it also decreases the KS distance at some points of the score distribution.

How powerful is a one-indicator, meat-only scorecard? The power curve in Figure 13 and the lift chart in Figure 14 show that while a meat-only scorecard is better than no scorecard at all and while it identifies the non-poor almost as well as the original scorecard, a meat-only scorecard still sacrifices a lot of power for identifying the poor. Collecting the six additional indicators in the original scorecard is probably worthwhile.

In sum, meat consumption is a powerful indicator of poverty status. After all, 42 percent of households who eat meat "rarely" are poor (Table 1), versus 5 percent for those who eat meat "frequently". Accurate estimates of poverty likelihood require accurate data on meat consumption. Household size is also a powerful indicator, especially when split in six classes. In BiH, a two-indicator scorecard with meat consumption and household size would likely be both simple and powerful.

The original and expanded scorecards are simple, inexpensive, and powerful; they effectively assign lower scores to clients who are more likely to be poor. For a given Prizma client, the poverty likelihood is defined as the poverty likelihood associated with that client's score. As discussed next, Prizma's overall poverty rate can then be computed as its clients' average poverty likelihood.

4. Overall poverty rate

Prizma's poverty scorecard ranks clients by relative poverty. These ranks can help managers improve targeting, track changes in poverty status over time, and manage depth of outreach. Ranking clients requires that lower scores be associated with higher poverty likelihoods, but it does not require knowing the exact likelihoods.

Donors, however, want measures of absolute poverty, and that requires exact likelihoods. Because Prizma's poverty scorecard was benchmarked to an expenditure survey, these likelihoods are known (at least to the extent that the scorecard indicators match those in the survey). This enables measuring Prizma's overall poverty rate.

4.1 Why measure rates of absolute poverty (and how)

For managers, poverty ranks may be enough. Ranks allow targeting services to clients with low scores, inferring the direction of changes in poverty over time for individuals, and using average client scores to manage poverty outreach.

For donors, however, ranks are not enough. Absolute measures are required to compare apples with apples when alloting funds across organizations. Boards and managers—as in the case of Prizma—may also seek absolute measures to provide a clearer sence of poverty outreach. Absolute measures also act as a check on claims of poverty outreach by microfinance advocates and create incentives for managers to innovate to reach more and poorer clients (Dunford, 2002a). For example, publishing comparable (that is, absolute) measures of poverty (for example, in the MIX, `http://www.themix.org`) would increase pressure to improve depth of outreach.

In addition, all recipients of microenterprise assistance from the U.S. Agency for International Development must—as of October 2005—report the share of their clients who are "very poor", defined as those living on less than a dollar per day (Sillers, 2003) or those among the poorest half of people below the country-specific poverty line. The U.S. Congress (Public Law 108-31) requires that these measures be objective (linked with an income- or expenditure-based poverty line), quantitative (not "more or less poor" but "above or below the poverty line") and low-cost (Zeller, 2004).

There are three broad approaches to meeting these goals. The first uses Lot Quality Assurance Sampling (Davis, 2002; MkNelly *et al.*, 2002) and an expenditure survey with a small sample of clients. It then estimates the probability that at least 50 percent of all clients are poor. Lot Quality Assurance Sampling has high per-client costs because of the survey but low total costs because very few clients are surveyed.

The second approach is that taken by Prizma and described in this paper. It produces objective, quantitative poverty measures without doing additional surveys. Furthermore, it scores all clients (not just a sample), so it can be used for targeting.[7] This approach assumes that clients are selected at random. If they are not, then it assumes that there are enough scorecard indicators to control for non-random differences between clients and non-clients that affect both poverty status and the

[7] Matul and Kline (2003) discuss the variety of reasons behind Prizma's approach.

probability of being selected as a client. This assumption about "selection effects" will be revisited in detail below.

The third approach—used by IRIS (`http://www.povertytools.org`) to help microfinance organizations meet Congress's mandate—is like Prizma's approach except the scorecard is based not on an existing expenditure survey but rather a new special-purpose expenditure survey on a national random sample. Doing a new survey allows IRIS to include non-expenditure indicators not in existing surveys. Of course, doing a new survey is also costly, and sample size may be smaller than in existing surveys.

Only Lot Quality Assurance Sampling (because it measures poverty directly) avoids bias in the link between scores and poverty likelihoods due to "selection effects". These effects occur because clients are self-selected (they choose to apply to programs) and program-selected (programs choose which applicants to accept). Both self-selection by clients and program-selection by lenders are partly based on client characteristics (for example, "work ethic", "good looks", or "business sector") that are correlated with poverty but that are omitted from the scorecard. Because the indicators in the scorecard are perfect indicators of poverty, a client and a non-client can have the same poverty score—and even the same values for all scorecard indicators—but different poverty statuses. Clients could be more or less likely to be poor than non-clients even though—in terms of indicators in the scorecard—they look identical.

If a special-purpose expenditure survey (such as in IRIS' approach) included both clients and non-clients, then it could measure selection bias as the difference in poverty likelihood between clients and non-clients with identical poverty scores. Organizations could then adjust their poverty rates for "selection effects".

Yet another alternative is to include in the expenditure survey an indicator for the presence of formal loans. Some surveyed households would have formal loans and some would not, but all clients (of a microlender) would. This will increase clients' scores (assuming omitted indicators are positively correlated with selection as a client and negatively correlated with poverty), building an adjustment for selection bias into the scorecard. (This works best if weights are statistically optimal rather than 0/1.)

Overall, the three approaches reflect trade-offs between different goals. Lot Quality Assurance Sampling checks whether a given standard of poverty outreach is met, but it is probably less accurate for estimating an overall poverty rate, and it cannot track changes in poverty status for a large number of clients over time. The approaches of Prizma and IRIS fulfill all three goals, and they can also help to target services. Compared with Lot Quality Assurance Sampling, however, they may be more costly. Overall, Prizma and IRIS are quite similar and have similar on-going costs, but IRIS has greater up-front costs (because its conducts a survey) and offers greater accuracy (because it provides indicators absent from existing surveys).

4.2 Prizma's overall poverty rate

Prizma's overall poverty rate is its clients' average poverty likelihood. Loan officers collected the required scorecard indicators for 5,177 first-time borrowers from December 2003 to September 2004. Given the resulting scores, the poverty likelihood of each client was defined as the poverty likelihood of households in the national survey with that same score. For example, among surveyed households with a score of 1, 71.4 percent were poor (Figure 3), so Prizma clients with a score of 1 were assigned a poverty likelihood of 71.4 percent.

For the original scorecard, Figure 15 shows the distribution of Prizma's new clients by score. The average poverty likelihood is the share of cases with a given score multiplied by the associated poverty likelihood, summed for all scores. For Prizma, this was 14.6 percent.[8] The estimate from the expanded scorecard (Figure 16) is 17.9 percent. This figure is probably closer to the true poverty rate, at it accounts better for fine graduations in poverty status for households of different sizes.

Prizma's poverty rate is quite close to the national poverty rate of 19.3 percent. Is this poverty outreach high or low? There is no simple answer, and the national average is not necessarily an appropriate benchmark. After all, the distribution of creditworthy borrowers in BiH might not be uniform over the distribution of poverty. At the same time, Prizma's poverty outreach may be high compared with the (unknown) poverty outreach of other microlenders in BiH or compared with the (unknown) poverty outreach that is sustainable. In any case, Prizma has an explicit mission to serve the poor, and measuring poverty outreach helps the board to monitor the fulfillment of the mission as it helps managers look for new ways to improve. By measuring poverty, however, Prizma risks "looking bad" vis-à-vis competitors who lack such measurements and who thus can claim (because there is no evidence to the contrary) that they have greater poverty outreach (Pritchett, 2002).

While external stakeholders focus on Prizma's overall poverty rate, managers are also interested in poverty rates by loan product and by branch.[9] Disaggregating poverty rates can help pinpoint products and branches with greater or lesser poverty outreach, possibly suggesting ways to deepen outreach.

The poverty rate at Prizma's Sarajevo branch using the original scorecard is 23.9 percent, five times the 4.7-percent rate for Banja Luka (Figure 17). The Zenica branch had a poverty rate of 18.9 percent, versus 8.0 and 12.0 percent for the Mostar and Bihać branches. The reasons for these differences are not immediately clear. Banja Luka (the branch with the smallest concentration of poor clients) is entirely in the

---

[8] From the third and fourth columns of Figure 15, this is (0.003 x 0.953) + (0.044 x 0.714) + (0.090 x 0.471) + (0.130 x 0.194) + (0.170 x 0.106) + (0.191 x 0.051) + (0.165 x 0.009) + (0.120 x 0.009) + (0.061 x 0.000) + (0.025 x 0.000) = 0.132, or 13.3 percent to within rounding error.

[9] Prizma also seeks to disaggregate poverty rates by drop-out status and by a host of other variables to help it manage social performance.

Republic of Srpska, which is generally poorer than the rest of the the Federation of Bosnia and Herzegovina. Banja Luka, however, is in one of the least-poor municipalities in the Republic of Srpska, and service has been limited so far to urban areas. A law in the Republic of Srpska limits Prizma to enterprise loans, and the staff at this newer branch may be more risk averse as they compare themselves with older branches.

The branches in Sarajevo, Mostar, and Zenica are all in the the Federation of Bosnia and Herzegovina but also serve some clients from the Republic of Srpska. Each of these branches serves some very poor communities and some not-very-poor communities. The branch in Mostar, for example, serves the city of Mostar (one of the least-poor areas in the country) as well as the eastern Republic of Srpska (one of the poorest areas). Likewise, Zenica serves a mix of communities. The branches in Mostar and Zenica also serve larger numbers of Croats, the ethnic group least represented among the poor.

The Sarajevo branch—with the highest concentration of poor clients at 23.9 percent—operates in the area with the most competition and thus may have the deepest outreach because it serves those whom competitors will not or cannot. The Sarajevo branch also reaches underserved, low-income suburbs in the Republic of Srpska that include some of the poorest areas in the country. Most clients at the Sarajevo branch are Bosniaks, the second-poorest ethnic group (after Serbs). Given the geography of poverty in the country, outreach to Serbs in the rural areas of the central and eastern Republic of Srpska might be an opportunity for Prizma to reach higher concentrations of poor clients and increase its overall poverty rate.

Figure 18 breaks down poverty rates by loan product. Most new borrowers were "enterprise" borrowers who received group loans for business use or "basic needs" borrowers who received individual, small, short, unrestricted loans based on the guarantee of a household member with a salaried job. Basic-needs loans are often used for emergencies, and basic-needs borrowers were more likely to be poor (16.4 percent) than enterprise borrowers (13.2 percent). This difference might result from the group-individual distinction or the enterprise/emergency distinction. Either way, managers could investigate the reasons and perhaps take advantage of them to improve outreach. For example, Prizma's trimesterly appraisal process encourages supervisors to explore social and financial performance. Disaggregating the poverty rate by loan officers, the municipalities they serve, and by clients' drop-out status might offer insights about who is reaching and retaining poor clients and spark discussions about possible explanations.

Overall, poverty outreach seems to vary more by branch than by loan product, perhaps highlighting the importance of branch placement and branch managers' outreach within their service areas. Also, newer/smaller/non-growing branches (those that had fewer "new" clients between December 2003 and September 2004) had lower concentrations of poverty, perhaps because older/larger/growing branches face more pressure (or are more able, due to their experience) to go beyond less-poor clients. Finally, the current law only allows Prizma to make loans, but allowing savings services, money-transfer services, and insurance might help improve poverty outreach.

4.3 Poverty rates from a fully benchmarked scorecard

The poverty estimates above assume that all the indicators collected by Prizma are linked directly to the national survey. As discussed earlier, however, location of residence and frequency of consumption of meats and sweets were not directly linked. Does scoring work with only the four fully benchmarked indicators (ownership of cars and stereo CDs, education, and household size)?

The power curve and lift chart in Figures 19 and 20 show that although the four-indicator, fully benchmarked scorecard identifies those most-likely and least-likely to be poor almost as well the seven-indicator scorecard, it does sacrifice accuracy for "middle" scores (assuming—perhaps incorrectly—that the seven-indicator scorecard is accurate).

Still, the benchmarked scorecard may accurately estimate overall poverty rates if errors for individual clients, on average, cancel each other out. The overall poverty rate as estimated by the four-indicator scorecard is 14.4 percent (Figure 21), almost equal to the original scorecard's 14.6 percent.

Is this a coincidence? Or, for overall poverty rates, are four indicators just as good as seven? And if they are, why stop at four? Why not use just one?

4.4 Simple benchmarkable scorecards and selection effects

Suppose Prizma estimated its poverty rate with a one-indicator scorecard. Figure 1 gives seven such scorecards, one for each indicator in the original scorecard. For the example of car ownership, non-owners have a 26-percent poverty likelihood versus 11 percent for owners. For meat consumption, poverty likelihood is 42 percent for those who eat meat "rarely", 19 percent for "sometimes", and 4 percent for "often".

Applied to the surveyed households, all seven one-indicator scorecards give the (correct) poverty rate of 19.3 percent. For example, 55 percent of surveyed households were not car owners, and 26 percent were poor. Furthermore, 45 percent were car owners, and 11 percent of these were poor. The average poverty likelihood for all households was then (0.55 x 0.26) + (0.45 x 0.11) = 0.1925, or 19.3 percent. Looking at meat consumption, the poverty rate was (0.47 x 0.28) + (0.31 x 0.17) + (0.22 x 0.05) = 0.1953, again (within rounding error) 19.3 percent.

Applied to Prizma's clients, each of the seven one-indicator scorecards gives a different poverty rate (Figure 22). For example, 54 percent of Prizma's clients owned cars, and 46 percent did not. Using the poverty likelihoods from the national survey, Prizma's poverty rate based on car ownership was (0.46 x 0.26) + (0.54 x 0.11) = 0.179, or 18.1 percent (within rounding). Looking at meat consumption, Prizma's poverty rate was (0.19 x 0.42) + (0.50 x 0.19) + (0.30 x 0.04) = 0.1868, or 18.7 percent. The estimated poverty rates from the one-indicator scorecards are all larger than the original scorecard's 14.6 percent but smaller than BiH's overall rate of 19.3 percent.

This happens because indicators for Prizma's clients had values associated with lower poverty likelihoods. For example, 45 percent of surveyed households owned a car, versus 54 percent of Prizma's clients. Likewise, 25 percent of surveyed households ate meat “rarely”, versus 19 percent for Prizma’s clients. In short, Prizma’s clients were not selected at random. Rather, they were self-selected and program-selected, with the result that—compared with the average BiH household—their indicators usually had values associated with lower poverty likelihoods.

Prizma clients were also more likely to have multiple indicators with values associated with lower poverty likelihoods. For example, 15.0 percent of surveyed households owned a car and ate meat “often”, versus 20.0 percent for Prizma’s clients. In the survey, these households had a poverty likelihood of 2.2 percent.

Given these results for indicators included in the scorecard, Prizma’s clients probably had values for indicators omitted from the scorecard that were positively correlated with the probability of selection as a Prizma client (such as “work ethic”, “good looks”, and “business sector”) and negatively correlated with poverty. As more of these indicators are omitted, the more a scorecard will overstate poverty rates.

While all scorecards necessarily omit some indicators, including more indicators reduces bias. First, more included indicators means fewer omitted indicators, weakening selection effects. Second, included indicators are correlated with omitted indicators and so partially “represent” omitted indicators. Including more indicators strengthens this representation until, at some point, the omitted indicators are effectively included.[10] Exactly how many indicators are optimal, however, is unknown.

At some point, the cost of more indicators outweighs the benefits. In principle, the strongest indicators should be included first. Sooner or later, the gain from additional indicators is so small that the process can stop. For Prizma, one indicator overstated poverty, while four gave about the same poverty rate as seven. While this does not imply that four indicators are always sufficient, it does suggest that small and large scorecards can at least sometimes produce similar poverty rates.

Of course, the strongest indicators are not known beforehand. Also, the strongest indicators might not be very strong, and the strength of a given indicator depends partly on what other indicators are included. Prizma’s estimated poverty rate could fall (or rise) if an eighth indicator was discovered that was correlated with poverty and not highly correlated with the other seven indicators.

A practical approach to scorecard-building is to identify indicators highly correlated with poverty and then put the strongest indicators in small scorecards, expanding until the power curve, lift chart, and/or estimated poverty rate stop changing. The results for Prizma suggest that accurate estimates of overall poverty rates do not require as large of scorecards as do accurate estimates of the poverty likelihood of individual clients. Thus, scorecards for targeting, tracking, and managing poverty outreach will be larger than those for reporting poverty rates to donors.

---

[10] Although this method of controlling for selection is simple and obvious, it requires more and better indicators rather than fancy technique, so it has been largely ignored (Schreiner and Sherraden, forthcoming; Benjamin, 2003).

4.5 Loan size and poverty

Does poverty correlate with loan size? In the absence of alternatives, the amount disbursed is a common measure of poverty, although the true correlation between loan size and depth of outreach is unknown (Dunford, 2002b; Schreiner, 2001).

There is no direct, expenditure-based measure of the poverty status of Prizma clients, only the indirect scorecard-based measure. Thus, the data at hand do not provide a direct way to link poverty with loan size. It is possible to test, however, the correlation between loan size and poverty likelihood as derived from Prizma's scorecard.

Figure 23 shows a loan-size-only scorecard and the poverty likelihoods associated with each score for Prizma's clients. For example, a client with a 300KM loan has a poverty likelihood of 18.3 percent, while the figure for a client with a 1,200KM loan is 11.5 percent. By definition, the overall poverty likelihood is 14.6 percent.

Estimates from the loan-size-only scorecard are not highly correlated with those from the original scorecard (Figure 24). For example, 35 percent of all poor clients had scores in the lowest decile in the original scorecard, while 13 percent of all poor clients had scores in the lowest decile in the loan-size-only scorecard. The highest quartile contained 2 percent of all poor clients for the original scorecard but 20 percent of poor clients for the loan-size-only scorecard.

Of course, the loan-size-only scorecard does not directly estimate poverty likelihood; rather, it tries to reproduce the estimates from the original scorecard. Still, the results suggest that the loan-size-only scorecard is not very powerful.

5. Lessons for microfinance

5.1 Summary

Prizma's poverty scorecard powerfully identifies poor clients without the cost of measuring expenditure. By ranking clients by relative poverty, it helps managers target the poor, track changes in poverty, and manage depth of outreach. Because the scorecard is based on an expenditure survey, it can also report clients' absolute poverty.

The scorecard uses 0/1 weights and seven inexpensive-to-collect indicators. Statistically optimal weights improve power, but only a little. A single indicator—how often the client eats meat—supplies much of the power. One-indicator scorecards, however, overstate the overall poverty rate. Loan size is correlated with poverty, but not nearly as strongly as the poverty score.

5.2 Lessons

Based on the analysis of Prizma's poverty scorecard, this final section discusses ten broad lessons for microfinance.

First, poverty scoring can work, and it need not be complex or costly. Ranking clients by relative poverty requires finding yes/no (or low/average/high) indicators correlated with poverty. Most microfinance organizations probably already collect several such indicators, and—if desired—they might be able to collect a few more without adding too much to the workload of loan officers and clients.

Second, if a scorecard is derived from indicators that appear in expenditure surveys, then it can estimate poverty rates based on national or international benchmarks. Because clients are self-selected and program-selected, however, such estimates are biased. Reducing bias requires using a scorecard that includes many indicators and/or surveying both clients and non-clients.

Third, both theory and experience provide support for 0/1 weights. In general, data quantity and quality matter more than statistical sophistication. After all, no amount of manipulation can substitute for an unrecorded indicator or squeeze meaning from carelessly recorded data. Collecting good data and monitoring its quality is difficult, but the long-term reward will only increase as scoring—for poverty, repayment behavior, drop-out, and other uncertain future outcomes—becomes more widespread.

Fourth, "domain knowledge" (of the specific country, and of microfinance in the specific organization) is key. For example, home ownership may not be linked with poverty if almost all clients own a home. Likewise, religion or ethnicity might be highly correlated with poverty but difficult to record without undermining client rapport. Even within a given organization, a single scorecard might not work for all regions, branches, or products. In short, common sense and knowledge of the local context matter and may lead to customized scorecards for different client segments.

Fifth, scorecard power is not the only performance criterion. If scorecards lie unused, the culprit is usually not weak power but rather a mismanaged change process. Staff must see scoring as simple and worthwhile. As always, training helps, especially for improving data quality (Matul and Kline, 2003).

Sixth, organizations might opt for two scorecards, the first with more indicators (not all benchmarked to an absolute poverty line) that managers can use to rank individual clients by relative poverty, and a second with fewer indicators (all benchmarked) that donors can use to estimate absolute poverty rates for all clients. Including non-benchmarked indicators gives the larger scorecard greater power and thus greater usefulness for targeting and for tracking change over time. While there are diminishing returns to adding indicators, clients are not selected at random, so very small scorecards are severely biased. The number of indicators required to control for differences between clients and the average person ("selection effects") is unknown and varies by case, so—unless bias has been measured with an expenditure survey of clients and non-clients—scorecards should probably include at least 5 to 10 indicators that are strongly correlated with poverty.

Seventh, except for organizations with very strong pro-poor targeting, small benchmarked scorecards will probably overstate poverty rates. Unfortunately, it is tempting to use small scorecards, mostly because they are simpler and less costly. Indeed, unless an organization plans to use poverty scoring for managing depth of outreach, it will have weak incentives to collect quality data and build a powerful scorecard. Thus, if donors want accurate reporting of poverty rates, they should support poverty scorecards that managers find useful for their own purposes.

Eighth, loan size is correlated with poverty, but—at least in the case of Prizma—not as strongly as the poverty score.

Ninth, there is nothing about the scorecard scorecard that is specific or unique to microfinance. If a scorecard includes enough relevant indicators, then it might serve all the poverty-measurement purposes of a given microlender or country. Indeed, the benefit-cost ratio would be very large if the World Bank and national statistical agencies (when they do LSMS or other expenditure surveys) would assign a few person-weeks to building a poverty scorecard based on their expenditure data.

Tenth and most important, poverty scoring can promote a culture of intentional, explicit management of depth of outreach. Equipped with poverty scores, managers no longer must guess clients' poverty status nor how it changes over time. Instead, they can use a standard yardstick to reward branches and loan officers who improve depth of outreach. Lack of evidence about poverty outreach no longer supports business-as-usual complacency, and greater knowledge may spur innovation. Managers cannot hide behind ignorance when they report subjective (and perhaps sanguine, see Dunford, 2002a) estimates of overall poverty rates. Measurement feeds management, and managers with poverty scores may well wonder how they made decisions before, and boards may likewise wonder how they governed an organization explicitly committed to serving poor people when they had no data or metric to monitor progress in this realm.

References

Benjamin, Daniel J. (2003) "Does 401(k) Eligibility Increase Saving? Evidence from Propensity-Score Sub-Classification", *Journal of Public Economics*, Vol. 87, pp. 1259–1290.

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

Davis, Robb. (2002) "Lot Quality Assurance Sampling (LQAS) for Microfinance Institutions: A Management Tool to Efficiently Assess Poverty Outreach", Davis, CA: Freedom from Hunger.

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

Dunford, Chris. (2002a) "Why Set a Threshold for Service Orientation to the Very Poor?" Davis, CA: Freedom from Hunger, `http://www.ffhtechnical.org/publications/pdfs/tresholdsapr02.pdf`.

______. (2002b) "What's Wrong with Loan Size?", Davis, CA: Freedom from Hunger, `http://www.ffhtechnical.org/publications/pdfs/loansize0302.pdf`.

Dunn, Elizabeth; and Josip Tvrtkovic. (2003) "Microfinance Clients in Bosnia and Herzegovina: Report on Baseline Survey", Foundation for Sustainable Development of the Federation of Bosnia and Herzegovina, `http://www.odraz.ba/Documents/LIPII_Report_on_BaselineFindings.pdf`.

Granger, Clive W.J.; and M. Hashem Pesaran. (2000) "Economic and Statistical Measures of Forecast Accuracy", *Journal of Forecasting*, Vol. 19, pp. 537–560.

Hatch, John K.; and Laura Frederick. (1998) "Poverty Assessment by Microfinance Institutions: A Review of Current Practice", Microenterprise Best Practices, `http://www.povertytools.org/documents/PAbyMFIs.pdf`.

Henry, Carla; Sharma, Manohar; Lapenu, Cecile; and Manfred Zeller. (2003) "Microfinance Poverty Assessment Tool", Technical Tool No. 5, Washington, D.C.: Consultative Group to Assist the Poorest, `http://www.cgap.org/docs/TechnicalTool_05.pdf`.

Kolesar, Peter; and Janet L. Showers. (1985) "A Robust Credit Screening Model Using Categorical Data", *Management Science*, Vol. 31, No. 2, pp. 123–133.

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

Matul, Michal; and Sean Kline. (2003) “Scoring Change: Prizma’s Approach to Assessing Poverty”, Spotlight Note No. 4, Warsaw: Microfinance Centre for Central and Eastern Europe and the New Independent States, http://www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf.

Mays, Elizabeth. (2000) “The Basics of Scorecard Development and Validation”, pp. 89–105 in Elizabeth Mays (ed.) *Handbook of Credit Scoring*, Chicago: Glenlake, ISBN 0–8144–0619–X.

MkNelly, Barbara; Valadez, Joseph; Treiber, Jeanette; and Robb Davis. (2001) “Considering the Applicability of Lot Quality Assurance Sampling (LQAS) to Credit with Education Progress Tracking”, Davis, CA: Freedom from Hunger.

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.

Prism Research (2003) “Poverty Assessment of Prizma Clients Using the CGAP Index”, Sarajevo.

Pritchett, Lant. (2002) “It Pays to be Ignorant: A Simple Political Economy of Rigorous Program Evaluation”, *Journal of Policy Reform*, Vol. 5, No. 4, pp. 251–269.

Schreiner, Mark. (2002) “Scoring: The Next Breakthrough in Microfinance?” Occasional Paper No. 7, Washington, D.C.: Consultative Group to Assist the Poorest, http://www.cgap.org/html/p-occasional-papers07.html.

_____. (2001) “Seven Aspects of Loan Size”, *Journal of Microfinance*, Vol. 3, No. 2, pp. 27–47.

_____; and Michael Sherraden. (forthcoming) *Can the Poor Save? Saving and Asset Accumulation in Individual Development Accounts*, New York: Aldine de Gruyter.

Sillers, Don. (2003) “National and International Poverty Lines: An Overview”, Washington, D.C.: U.S. Agency for International Development, http://www.povertytools.org/documents/Poverty%20lines%20-%20An%20Overview%201_23_04.doc.

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.

Wainer, Howard. (1976) "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin*, Vol. 83, pp. 213–217.

World Bank; State Agency for Statistics in Bosnia-Herzegovina; Federation Statistical Institute; and the Republika Srpska Statistical Institute. (2002) "Welfare in Bosnia and Herzegovina, 2001: Measurement and Findings", October, `http://www.rzs.rs.ba/anketeLSMS/WelfareBiH.zip`.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", report to IRIS and USAID as part of the Developing Poverty Assessment Tools Project, `http://www.povertytools.org/documents/Review%20of%20Poverty%20Assessment%20Tools.pdf`.

Zeller, Manfred; Alcaraz V., Gabriela; and Julia Johannsen. (2004) "Developing and Testing Poverty-Assessment Tools: Results from Accuracy Tests in Bangladesh", Accelerated Microenterprise Advancement Project, IRIS, University of Maryland, `http://www.povertytools.org/documents/Bangladesh%20Accuracy%20Report.pdf`.

Figure 1: Correlation of indicators with poverty status, national survey

		National survey	
Indicator	**Value**	**% cases with value**	**% with value who are poor**
1. Ownership of car	No	55	26
	Yes	45	11
2. Education level of female household head/partner/spouse	≤ Primary	64	24
	> Primary	36	11
3. Number of household members	6 or more	17	40
	5 or less	83	15
4. Ownership of stereo CD player	No	78	23
	Yes	22	8
5. Location of residence	Rural or peri-urban	75	21
	Urban	25	13
6. Average times eats meat each week with main meal	Rarely (0-2)	25	42
	Sometimes (3-5)	40	19
	Often (6 or more)	35	4
7. Average times eats sweets each week with main meal	Rarely (0-2)	47	28
	Sometimes (3-5)	31	17
	Often (6 or more)	22	5

Note: In the national survey, 19.3 percent of all cases were poor.

Figure 2: Prizma's original and expanded scorecards

		Weight	
Indicator	**Value**	**Original**	**Expanded**
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

Figure 3: Distribution of surveyed households by score, original scorecard

	# of cases		Likelihood	% of cases	
Score	**Poor**	**All**	**poor (%)**	**Poor**	**All**
0	44.0	46.2	95.3	4.5	0.9
1	214.0	300.0	71.4	22.0	6.0
2	346.6	736.4	47.1	35.6	14.6
3	212.3	1,096.3	19.4	21.8	21.8
4	109.9	1,033.2	10.6	11.3	20.5
5	38.8	757.8	5.1	4.0	15.0
6	5.5	619.0	0.9	0.6	12.3
7	2.8	307.9	0.9	0.3	6.1
8	0.0	117.5	0.0	0.0	2.3
9	0.0	25.4	0.0	0.0	0.5
Total:	974.1	5,039.6	19.3	100	100

Figure 4: Distribution of surveyed households with a given score or less, original scorecard

	# of cases		Likelihood	% of cases		
Score	**Poor**	**All**	**poor (%)**	**Poor**	**All**	**Lift**
0	44.0	46.2	95.3	4.5	0.9	4.9
1	258.0	346.1	74.5	26.5	6.9	3.9
2	604.7	1,082.6	55.9	62.1	21.5	2.9
3	816.9	2,178.8	37.5	83.9	43.2	1.9
4	926.9	3,212.1	28.9	95.2	63.7	1.5
5	965.7	3,969.8	24.3	99.1	78.8	1.3
6	971.2	4,588.8	21.2	99.7	91.1	1.1
7	974.1	4,896.7	19.9	100.0	97.2	1.0
8	974.1	5,014.3	19.4	100.0	99.5	1.0
9	974.1	5,039.6	19.3	100	100	1.0

Figure 5: Power Curve, original scorecard

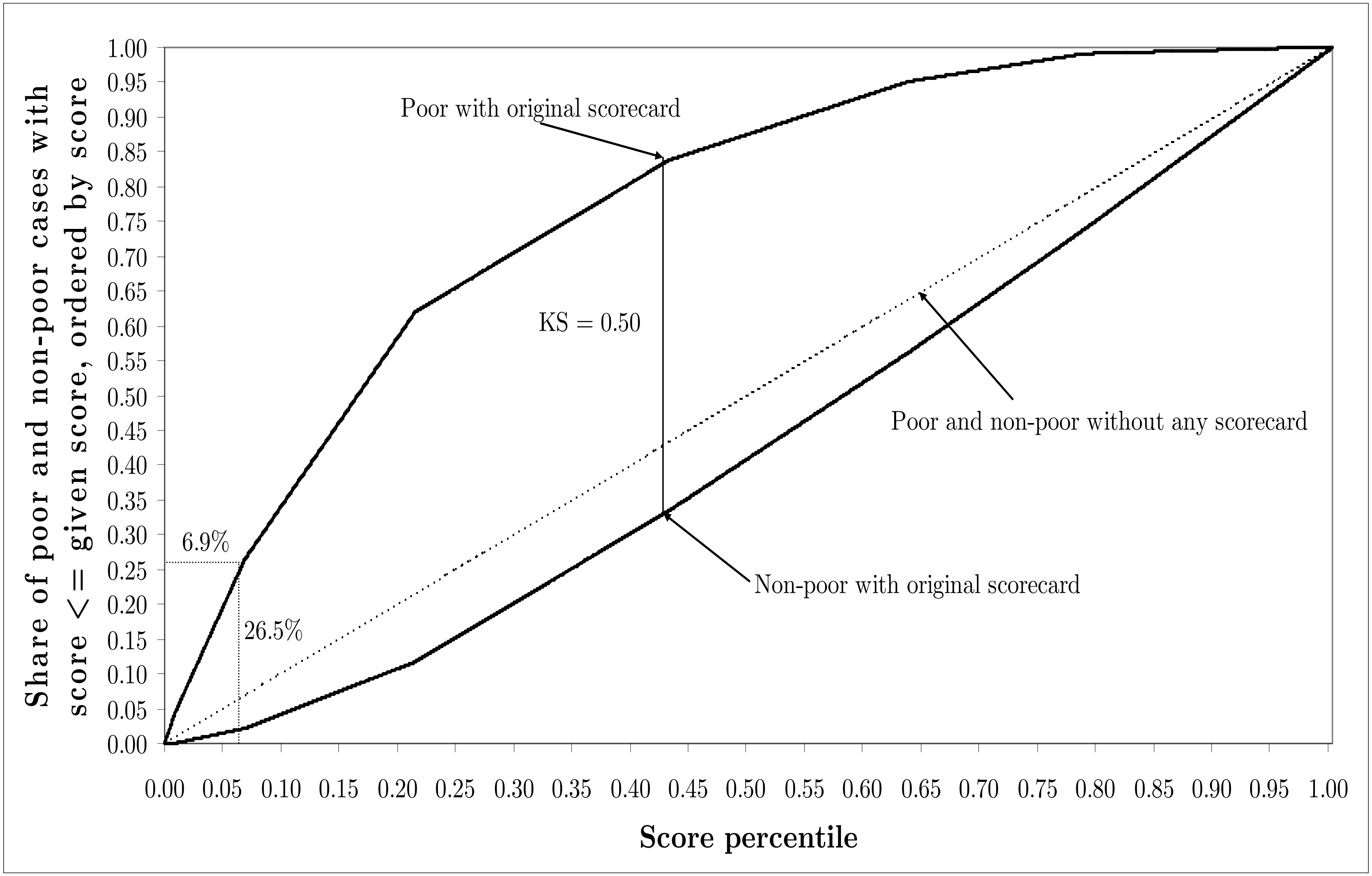

Figure 6: Distribution of surveyed households by score, expanded scorecard

	# of cases		Likelihood	% of cases	
Score	**Poor**	**All**	**poor (%)**	**Poor**	**All**
0	44.0	46.2	95.3	4.5	0.9
1–15	229.1	285.8	80.2	23.5	5.7
16–26	356.1	713.5	49.9	36.6	14.2
27–35	241.8	1,093.4	22.1	24.8	21.7
36–45	80.6	1,063.3	7.6	8.3	21.1
45–53	20.2	755.5	2.7	2.1	15.0
54–62	1.9	578.5	0.3	0.2	11.5
63–69	0.4	321.4	0.1	0.0	6.4
70–77	0.0	122.9	0.0	0.0	2.4
78–100	0.0	59.1	0.0	0.0	1.2
Total:	974.1	5,039.6	19.3	100	100

Figure 7: Distribution of surveyed households with a given score or less, expanded scorecard

	# of cases		Likelihood	% of cases		
Score	**Poor**	**All**	**poor (%)**	**Poor**	**All**	**Lift**
0	44.0	46.2	95.3	4.5	0.9	4.9
1–15	273.1	331.9	82.3	28.0	6.6	4.3
16–26	629.2	1,045.4	60.2	64.6	20.7	3.1
27–35	871.0	2,138.8	40.7	89.4	42.4	2.1
36–45	951.6	3,202.2	29.7	97.7	63.5	1.5
45–53	971.8	3,957.7	24.6	99.8	78.5	1.3
54–62	973.7	4,536.2	21.5	100.0	90.0	1.1
63–69	974.1	4,857.6	20.1	100.0	96.4	1.0
70–77	974.1	4,980.5	19.6	100.0	98.8	1.0
78–100	974.1	5,039.6	19.3	100.0	100.0	1.0

Figure 8: Lift, original versus expanded scorecards

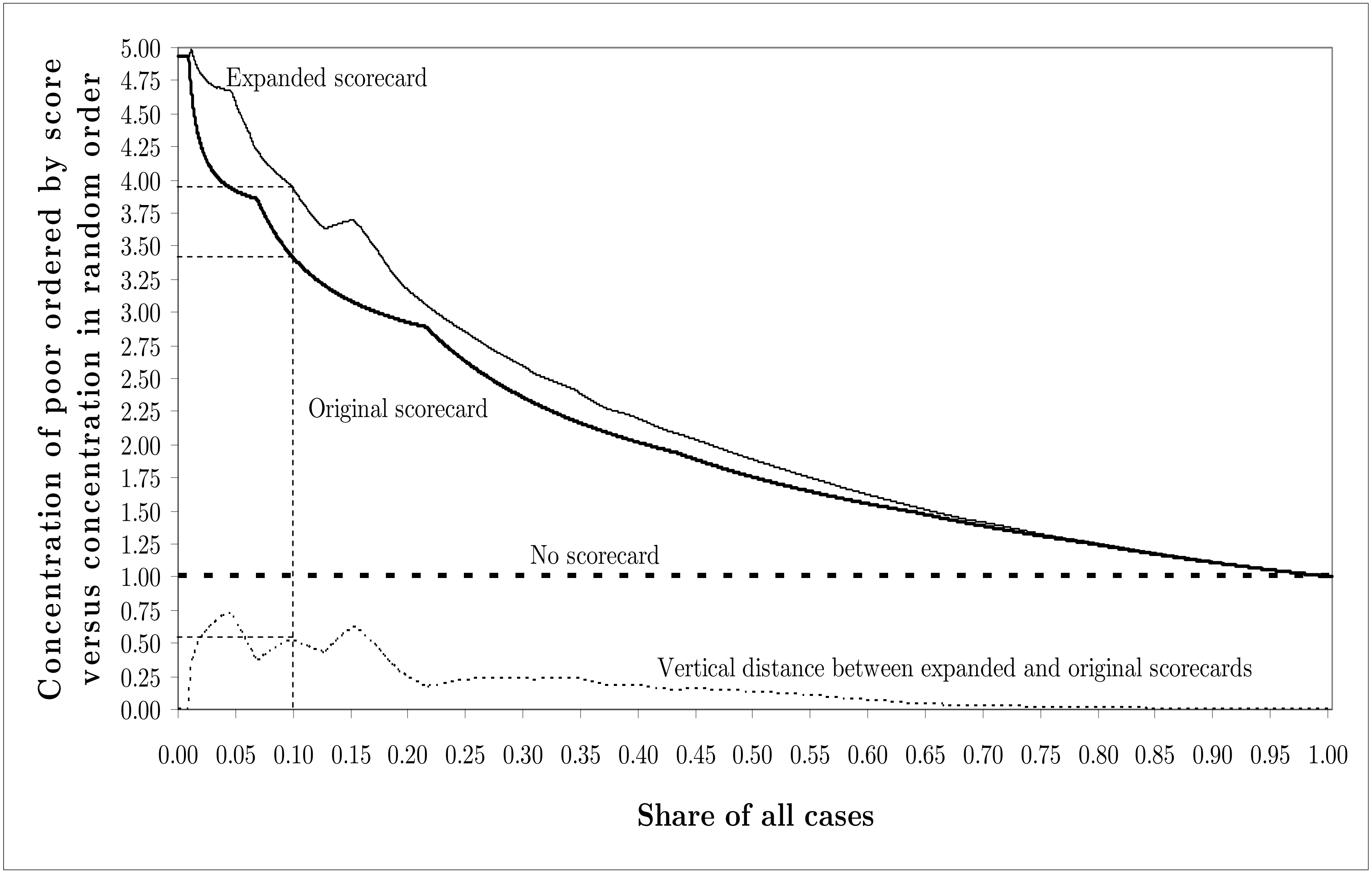

Figure 9: Power Curve, original versus expanded scorecards

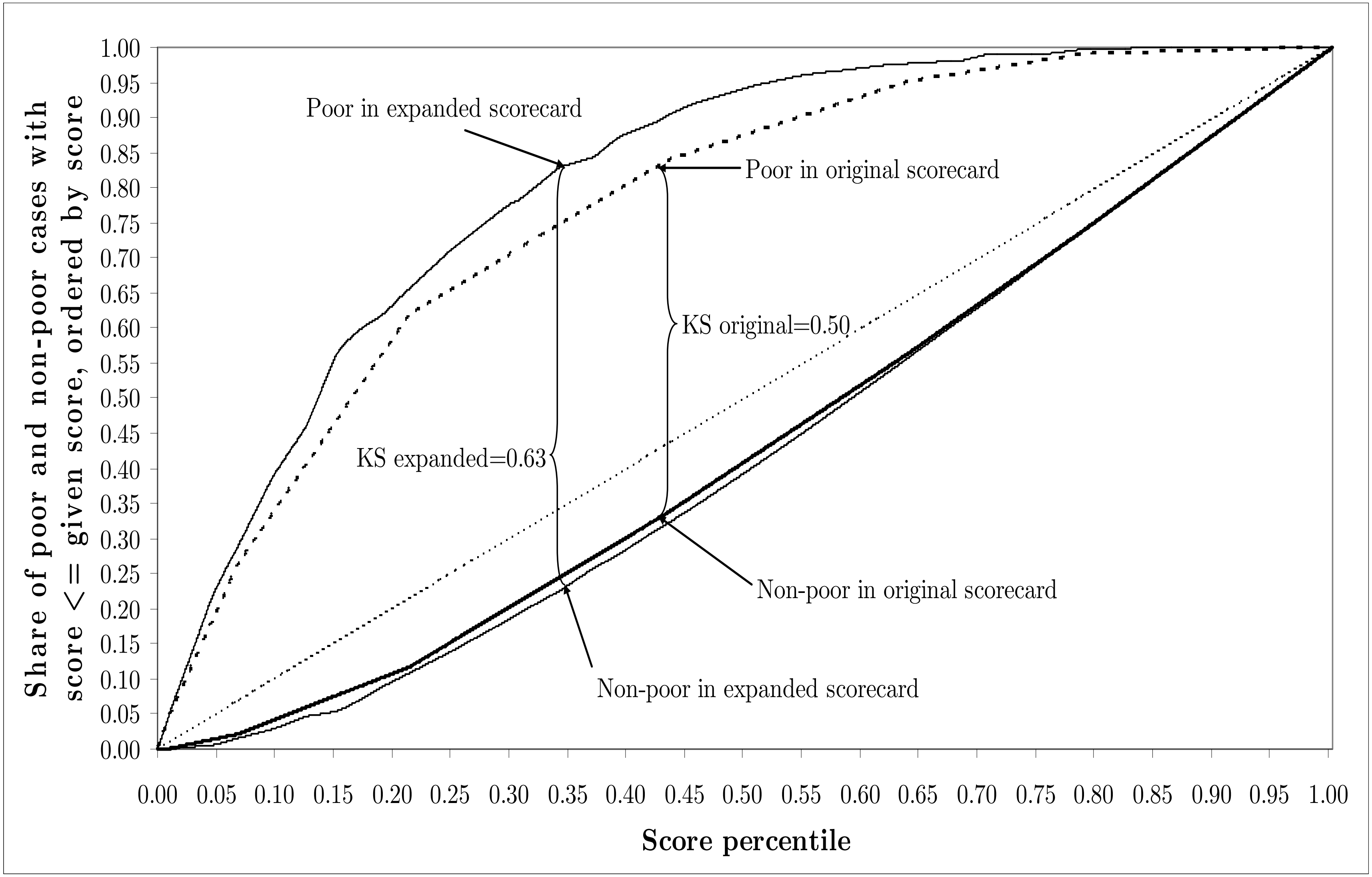

Figure 10: KS distances, original and expanded scorecards

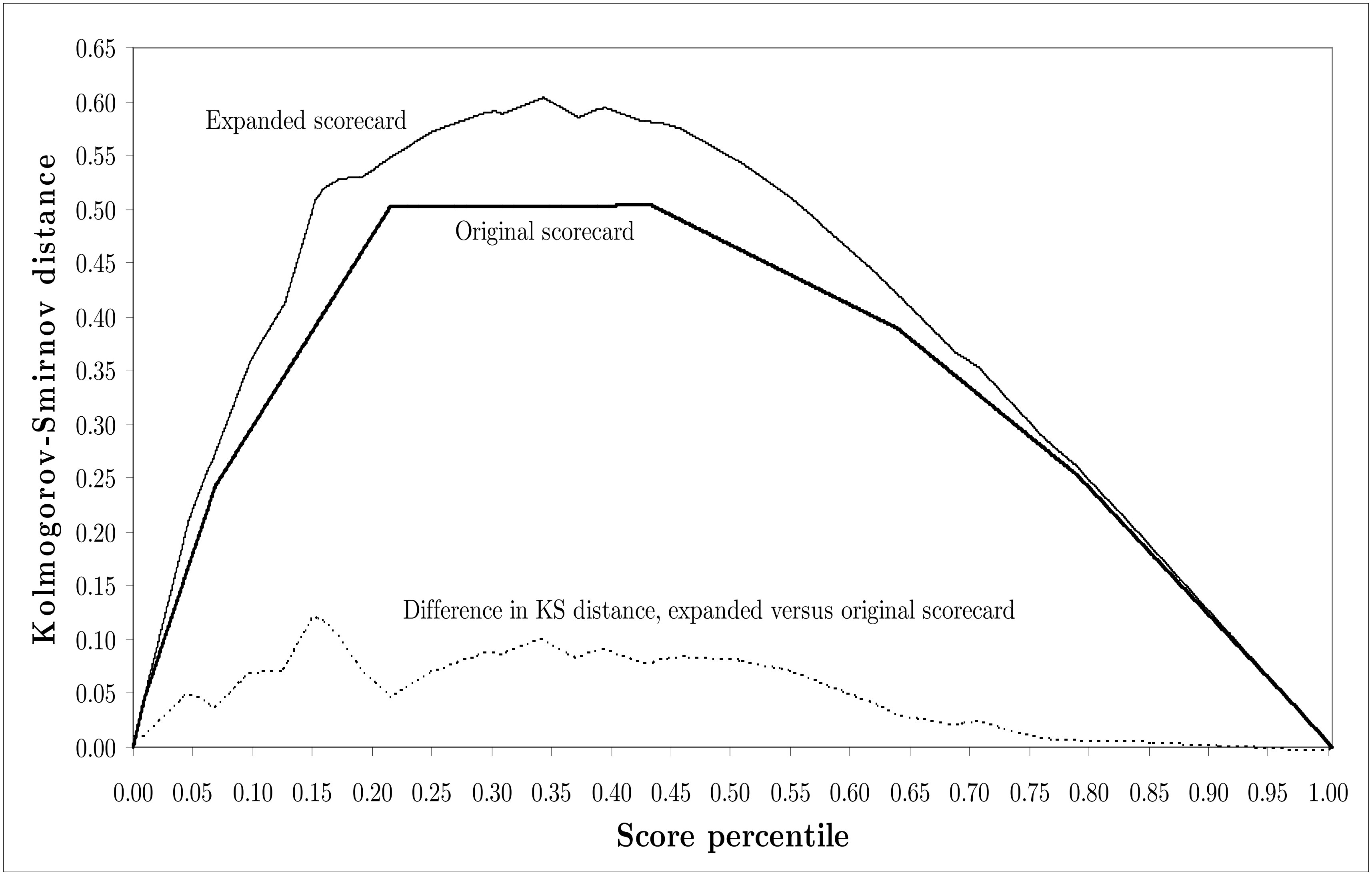

Figure 11: Importance of indicators, expanded scorecard

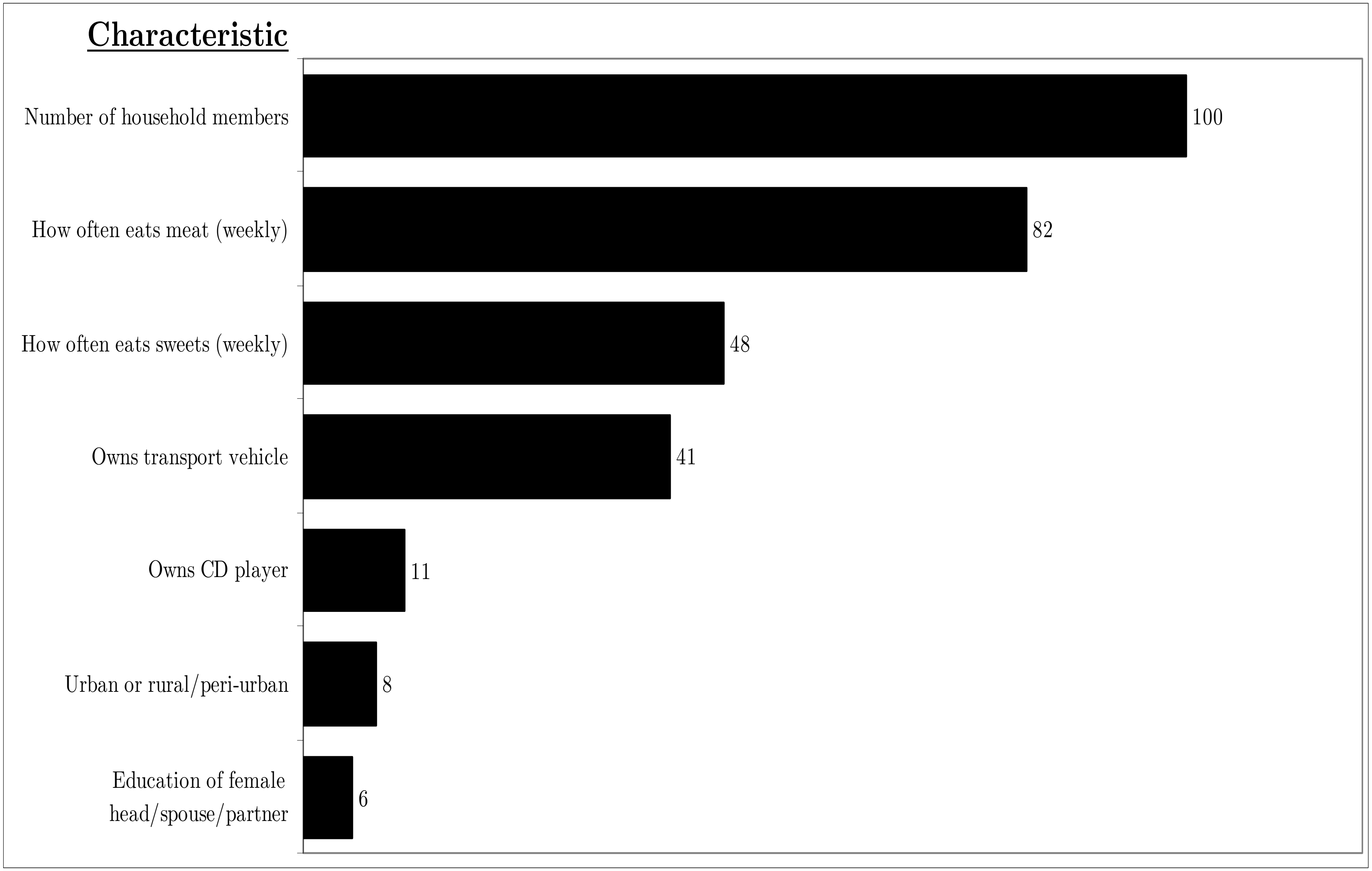

Figure 12: Importance of indicators, original scorecard

Figure 13: Power curve, meat-only and original scorecards

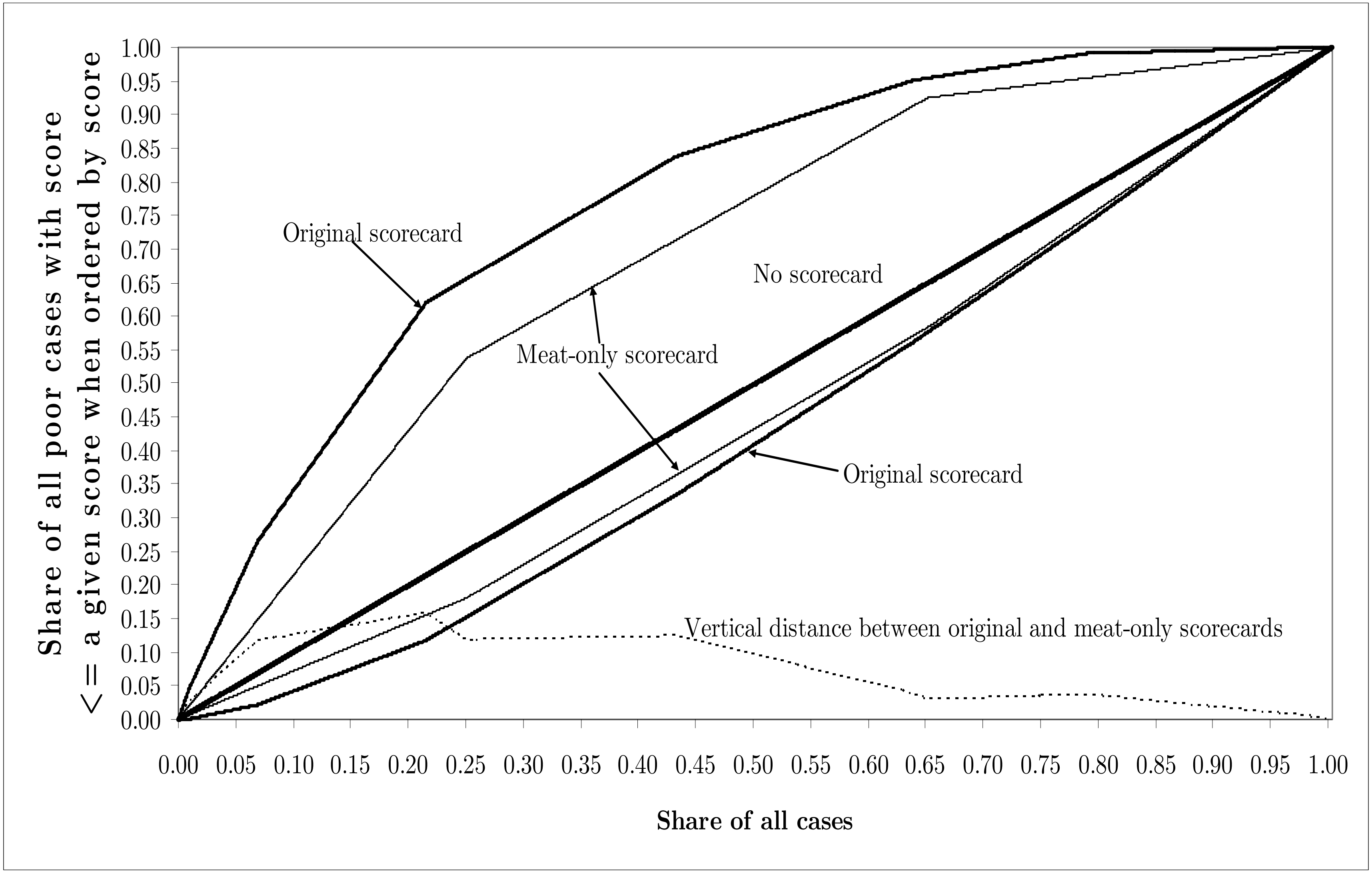

Figure 14: Lift, meat-only and original scorecards

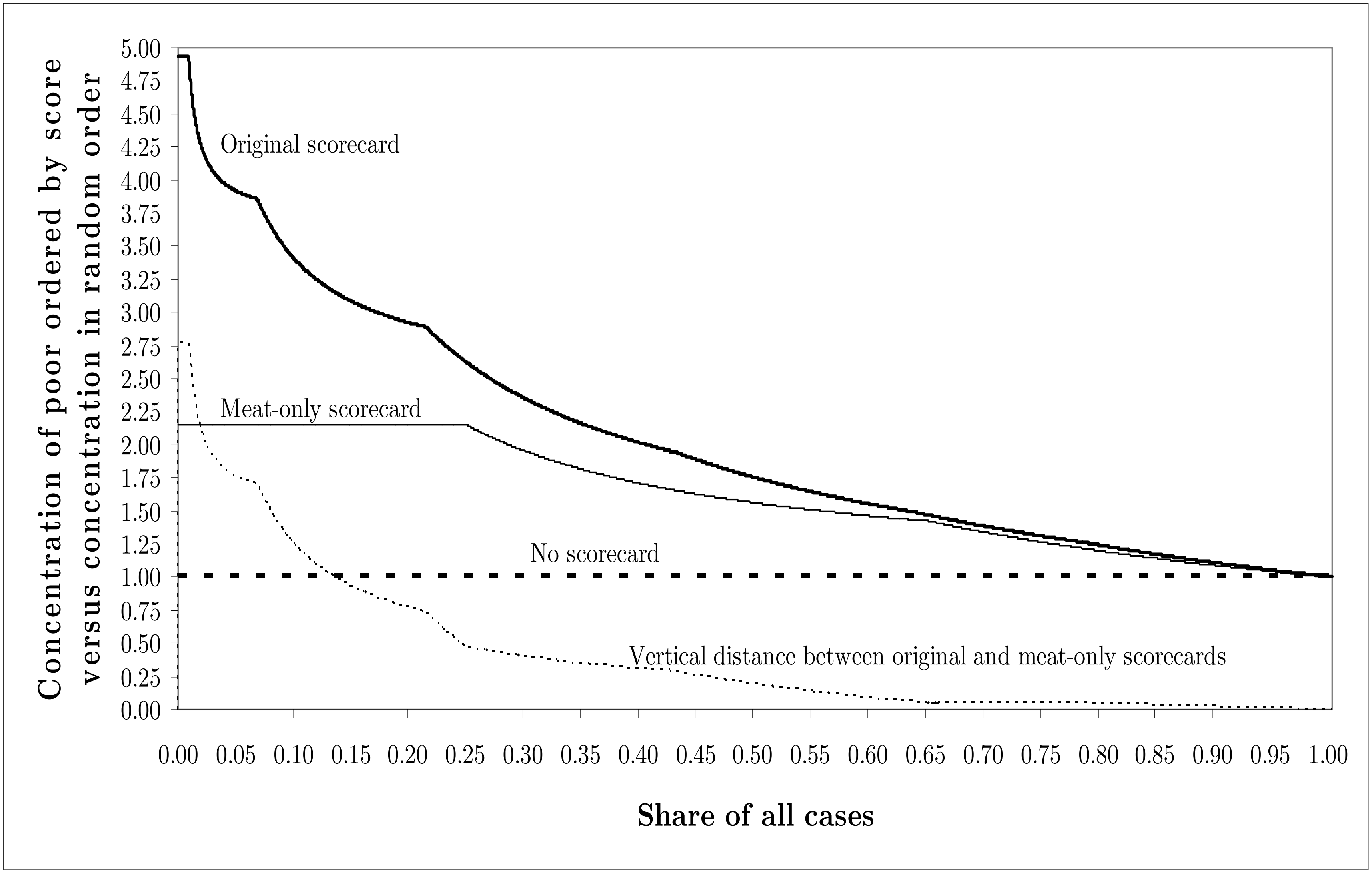

Figure 15: Prizma's overall poverty rate, original scorecard, new borrowers from December 2003 to September 2004

Score	Cases	% of cases	Assumed likelihood poor (%) (% with score poor in survey)
0	36	0.7	95.3
1	269	5.2	71.4
2	503	9.7	47.1
3	691	13.3	19.4
4	912	17.6	10.6
5	980	18.9	5.1
6	820	15.8	0.9
7	571	11.0	0.9
8	291	5.6	0.0
9	104	2.0	0.0
Total:	5,177	100	14.6

Figure 16: Prizma's overall poverty rate, expanded scorecard, new borrowers from December 2003 to September 2004

Score	Cases	% of cases	Likelihood poor (%) (% with score poor in survey)	# poor
0–3	36	0.7	95.3	34.3
4–5	18	0.3	100.0	18.0
6–7	5	0.1	100.0	5.0
8–9	101	2.0	87.8	88.6
10	2	0.0	100.0	2.0
11	46	0.9	88.4	40.7
12–13	96	1.9	66.3	63.7
14	3	0.1	64.5	1.9
15	45	0.9	52.2	23.5
16	131	2.5	65.0	85.2
17	4	0.1	63.3	2.5
18	10	0.2	52.5	5.3
19	72	1.4	48.4	34.8
20	153	3.0	78.3	119.8
21	2	0.0	41.8	0.8
22	10	0.2	45.6	4.6
23	92	1.8	29.7	27.3
24	152	2.9	20.8	31.6
25	13	0.3	30.4	4.0
26	29	0.6	33.2	9.6
27	140	2.7	30.3	42.4
28	212	4.1	25.5	54.2
29	17	0.3	22.0	3.7
30	21	0.4	10.9	2.3
31	172	3.3	26.3	45.2
32	195	3.8	9.6	18.7
33	16	0.3	26.9	4.3
34	56	1.1	23.0	12.9
35	132	2.5	12.9	17.1
36	210	4.1	18.0	37.8
37	43	0.8	14.1	6.1
38	45	0.9	10.3	4.6
39	178	3.4	9.5	16.9
40	220	4.2	7.0	15.4
41	30	0.6	4.6	1.4
42	43	0.8	2.4	1.0
43	169	3.3	4.1	6.9
44	117	2.3	3.9	4.6
45	63	1.2	1.9	1.2
46	88	1.7	3.4	3.0
47	137	2.6	1.8	2.5
48	188	3.6	8.0	15.0
49	55	1.1	0.7	0.4
50	24	0.5	0.0	0.0
51	161	3.1	0.5	0.9
52	109	2.1	3.9	4.3
53	38	0.7	4.6	1.7
54	75	1.4	0.4	0.3
55	83	1.6	0.9	0.8
56	144	2.8	0.0	0.0
57	66	1.3	0.2	0.1
58	30	0.6	0.0	0.0
59	106	2.0	0.0	0.0
60	140	2.7	0.0	0.0
61	35	0.7	0.0	0.0
62	45	0.9	0.0	0.0
63	68	1.3	0.1	0.1
64	24	0.5	0.0	0.0
65	44	0.8	0.3	0.2
66	64	1.2	0.0	0.0
67	36	0.7	0.0	0.0
68	77	1.5	0.0	0.0
69	21	0.4	0.0	0.0
70	1	0.0	0.0	0.0
71–72	57	1.1	0.0	0.0
73	20	0.4	0.0	0.0
74	50	1.0	0.0	0.0
75–76	10	0.2	0.0	0.0
77–78	52	1.0	0.0	0.0
79–80	13	0.3	0.0	0.0
81–84	5	0.1	0.0	0.0
85–86	9	0.2	0.0	0.0
87–92	1	0.0	0.0	0.0
93–99	2	0.0	0.0	0.0
100	0	0.0	0.0	0.0
Total:	5,177	100.0	17.9	929

Figure 17: Prizma's poverty rate by branch, original scorecard, new borrowers from December 2003 to September 2004

Branch	Cases	Poverty rate
Banja Luka	655	4.7
Mostar	745	8.0
Bihać	1,576	12.0
Zenica	998	18.9
Sarajevo	1,203	23.9
Total:	5,177	14.6

Figure 18: Prizma's poverty rate by loan product, original scorecard, new borrowers from December 2003 to September 2004

Product	Cases	Share (%) loans to individuals	Poverty rate
Farming	64	100	8.6
Enterprise	2,777	0	13.2
Basic needs	2,062	91	16.4
Small farm	211	48	16.9
Housing	63	100	17.8
Total:	5,177	41	14.6

Figure 19: Power curve, original and benchmarkable scorecards

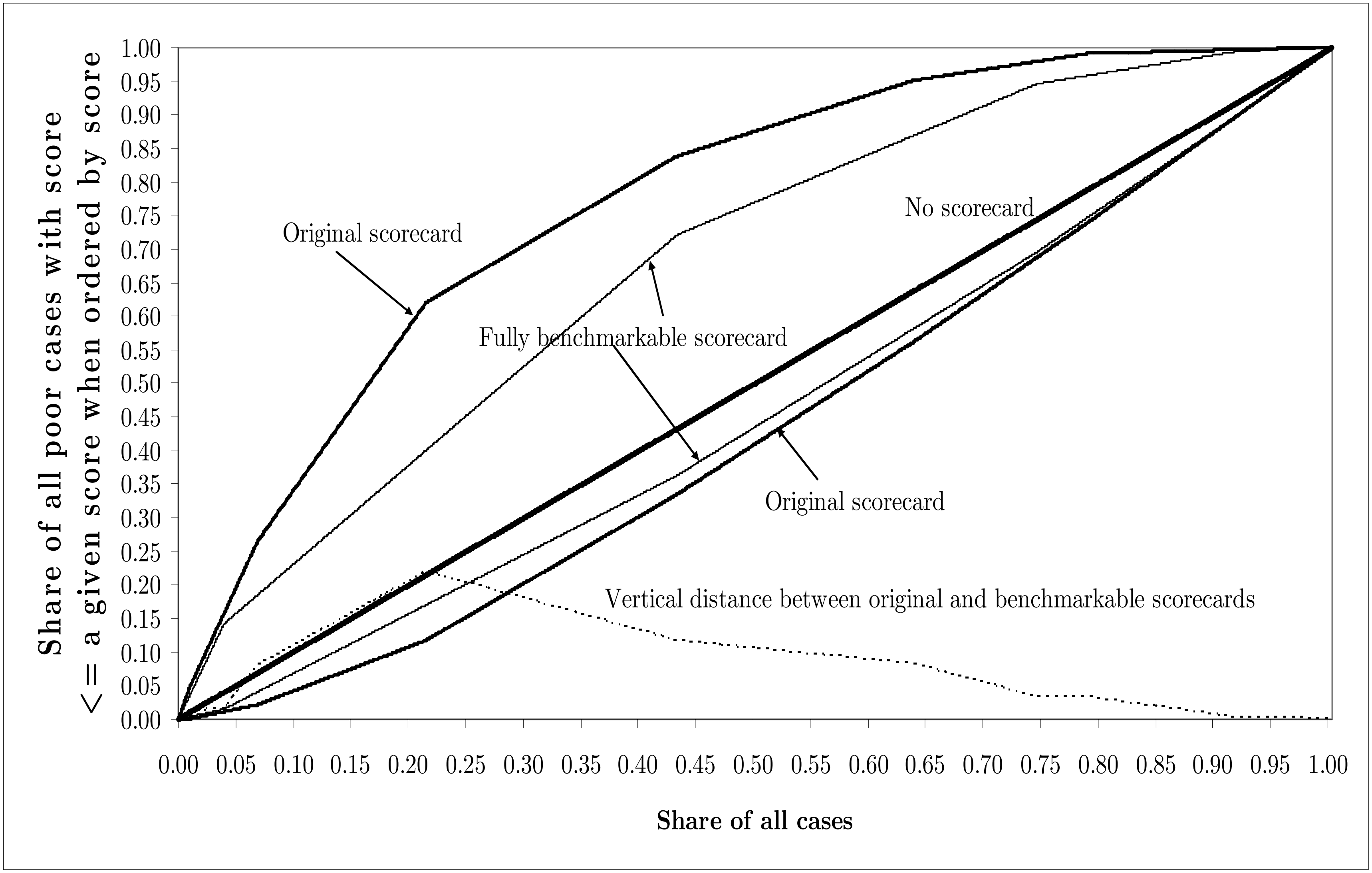

Figure 20: Lift, original and benchmarkable scorecards

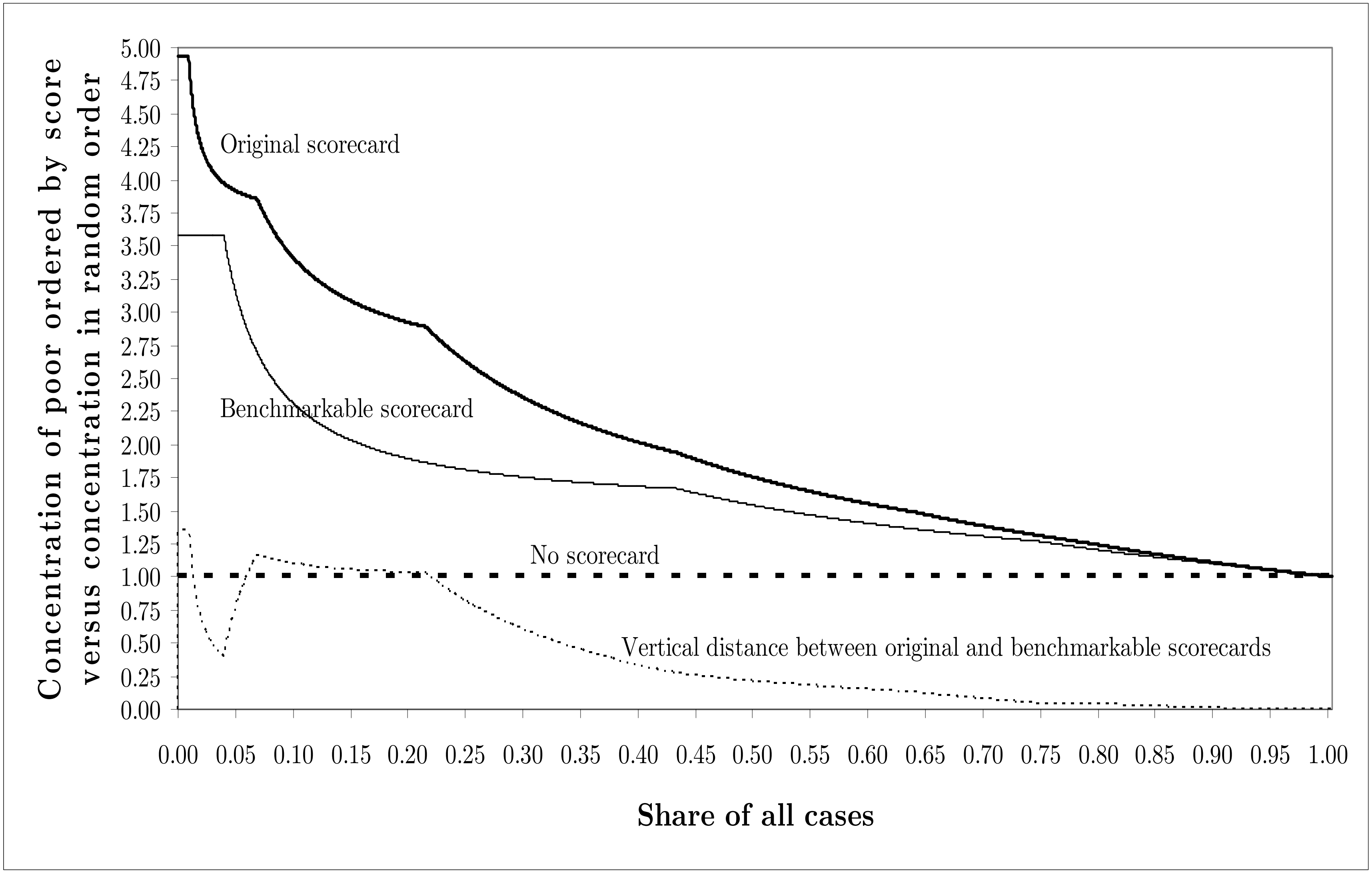

Figure 21: Prizma's overall poverty rate, benchmarkable scorecard, new borrowers from December 2003 to September 2004

Score	Cases	% of cases	Likelihood poor (%) (% with score poor in survey)
0	204	3.9	69.2
1	981	18.9	28.5
2	1,667	32.2	13.9
3	1,545	29.8	5.4
4	780	15.1	1.1
Total:	5,177	100	14.4

Figure 22: Prizma's overall poverty rate, single-indicator scorecards, new borrowers from December 2003 to September 2004

Indicator	**Value**	**% Prizma cases with value**	**% surveyed cases with value who are poor**	**Overall poverty rate for Prizma**
1. Ownership of car	No	46	26	
	Yes	54	11	18.1
2. Education level of female household head/partner/spouse	≤ Primary	34	24	
	> Primary	66	11	15.2
3. Number of household members	6 or more	23	40	
	5 or less	77	15	20.8
4. Ownership of stereo CD player	No	64	23	
	Yes	36	8	17.2
5. Location of residence	Rural or peri-urban	78	21	
	Urban	22	13	19.6
6. Average times eats meat each week with main meal	Rarely (0-2)	19	42	
	Sometimes (3-5)	50	19	
	Often (6 or more)	30	4	18.7
7. Average times eats sweets each week with main meal	Rarely (0-2)	34	28	
	Sometimes (3-5)	32	17	
	Often (6 or more)	33	5	16.7

Note: In the national survey, 19.3 percent of all cases were poor.

Figure 23: Loan-size-only scorecard based on Prizma's original poverty scorecard for new borrowers from December 2003 to September 2004

Amount disbursed	Score	% of cases	Likelihood poor
0 to 400 KM	0	4.3	18.3
401 to 599 KM	1	35.6	16.3
600 to 800 KM	2	13.8	15.9
801 to 1000 KM	3	18.2	14.3
1001 KM or more	4	28.1	11.5
	Total:	100.0	14.6

Figure 24: Power curve, original and loan-size-only scorecards