Scoring Arrears at a Microlender in Bolivia

Mark Schreiner

April 6, 2004

Center for Social Development
Washington University in St. Louis
Campus Box 1196, One Brookings Drive, St. Louis, MO 63130-4899, U.S.A.
schreiner@gwbmail.wustl.edu, http://www.microfinance.com

Abstract

Can scoring models help microlenders in poor countries as much as they have helped credit-card lenders in rich countries? This paper presents a scorecard that predicts the probability that loans from a microlender in Bolivia will have arrears of 15 days or more. Although arrears in microfinance depend on many factors difficult to include in scorecards, the paper shows that inexpensive, simple-to-collect data does have some predictive power. In microfinance, scoring will not replace loan officers, but it can flag high-risk cases and act as a cross-check on loan officer's judgment.

Author's Note

Mark Schreiner is a Senior Scholar in the Center for Social Development at Washington University in Saint Louis and a consultant with Microfinance Risk Management. He works to help the poor build assets via improved financial services.

Acknowledgments

I am grateful for support from the anonymous microfinance lender in Bolivia and for funding from the Division of Asset Building and Community Development of the Ford Foundation. I am also thankful for comments from the anonymous referees, John Adams, Jonathan Conning, Guillermo Rabiela, Loïc Sadoulet, Laura Viganò, and participants in the Third Annual Seminar on New Development Finance.

Scoring Arrears at a Microlender in Bolivia

1. Introduction

Microlenders make small, short, unsecured loans to the self-employed poor. Few of these borrowers have standard collateral, credit-bureau records, or formal wage jobs. Historically, lenders lacked low-cost ways to judge the risk of these borrowers. If lenders set interest rates to cover the high per-dollar costs of small loans, then they were accused of usury, but if they set lower rates, then they lost money.

Microfinance has been defined by new ways to cut the cost of judging the credit risk of the self-employed poor. For example, group lenders tap the knowledge of risk held as a sunk cost by neighbors of a potential borrower. Likewise, individual lenders control risk through detailed evaluations of the borrowers and their businesses, frequent repayments, stepped loan sizes, and chattel collateral (Bond and Rai, 2002).

Scoring is new to microfinance, but it may help reduce the costs of making loans to the self-employed poor. Scoring compares simple-to-observe quantified data about the borrower, loan, and lender with similar past cases. The share of similar past borrowers who had repayment problems is an estimate of the risk that a current potential borrower will also have repayment problems.

Credit-card lenders in rich countries make massive numbers of small, short, unsecured loans each year at very low costs because data-based scorecards inexpensively and accurately forecast the risk of potential borrowers (Lewis, 1990). In high-income countries, the microloan is often the credit card, and most home loans and car loans are made based almost entirely on scoring. Of course, microlenders also use a type of implicit, subjective "scoring" in that they evaluate borrowers based on their own historical experiences and prejudices. Data-based scoring differs only in that it is explicit and consistent (Longhofer, 2002). Most careful research (Berger, Frame, and Miller, 2002; Frame, Padhi, and Woosely, 2001; Martell, *et al.*, 1999) suggests that scoring—combined with credit bureaux—has improved the depth and breadth of formal loans in high-income countries. With better knowledge of risk, lenders can approve more poor-but-safe borrowers and reject more non-poor-but-risky borrowers. In this way, lenders save time that they would have spent dunning delinquent borrowers and can use the newfound time to find new borrowers (Schreiner, 2002). Can microlenders in poor countries likewise take advantage of scoring?

Scoring can help microlenders to judge risk, but it is unlikely to replace human loan officers anytime soon. For example, the most important factors in credit-card scorecards—employment and credit record—are often unavailable in low-income countries because credit bureaux are absent and potential borrowers are self-employed.

This paper tests whether a simple scorecard can predict the risk of costly arrears—spells of 15 days or more—for borrowers from a microlender in Bolivia. The scorecard shows how characteristics are linked with risk. In historical tests, the scorecard predicts better than naïve models but worse than credit-card models. Thus, scoring may help cut the costs of individual microloans (but probably not group joint-liability loans, see Schreiner, 2002), not as a replacement for loan officers' judgment, but rather as an additional filter for high-risk cases that would otherwise slip by.

Section 2 below gives the background for the scorecard. Section 3 reports how characteristics of the borrower, the loan, and the lender are linked with arrears, and Section 4 tests the scorecard's predictive power. Section 5 concludes.

2. Scoring for a Bolivian microlender

This section discusses the market for microfinance in Bolivia, reviews past work on scoring for microfinance, and presents the data and scorecard used in this paper.

2.1 Microfinance in Bolivia

Bolivia is the showcase of microfinance in Latin America. In spite of its sparse population and deep poverty, microfinance has a high rate of penetration. Most Latin American countries have, at most, one microfinance lender with more than 10,000 borrowers; Bolivia has a dozen. Several microlenders have converted from unregulated not-for-profits to regulated for-profits. Most borrowers are near the poverty line but are not among the poorest (Navajas, *et al.*, 2000).

As recounted in Rhyne (2001) and Mosley (2003), profits in microfinance attracted competition from Bolivian banks and Chilean consumer-finance companies, and, by 1996, the market started to saturate. Arrears skyrocketed, in part because the new entrants tolerated high arrears and weakened the culture of repayment for all borrowers, and in part because microlenders, in the battle for market share, made loans to people already indebted elsewhere. The crisis in Brazil in 1999 also hurt repayment from the women traders who make up the bulk of microfinance portfolios. As arrears rose to more than double their historical levels, interest in scoring heightened as microlenders sought to find new ways to control risk.

2.2 Past work on scoring for microfinance

Many statistical models have linked arrears to characteristics of the microlender, borrower, and loan (Nannyonga, 2000; Reinke, 1998; Zeller, 1998; Sharma and Zeller, 1997). In broad terms, these models have not been very useful as scorecards (nor were they intended for this purpose) for three reasons. First, most use small samples and so may lack robustness. Second, some use characteristics that most microlenders do not already collect or that are expensive to collect. Third and most importantly, they do not check predictive power. A historical test is needed to confirm that the scorecard truly can predict risk and—just as importantly—helps to convince loan officers and credit managers that the scorecard works. Most past academic statistical models aim to detect characteristics linked with risk, not to help lenders to score potential borrowers.

Viganò (1993) is the best scorecard for microfinance in the literature. It links default with 53 characteristics at a rural development bank in Burkina Faso. With a small sample ($n = 100$), prediction was checked with the jack-knife (Efron and Tibshirani, 1993). Unfortunately, the small sample also required that the 53 characteristics be condensed in 13 factors, obscuring the links between risk and specific characteristics. The scorecard also has the technical drawbacks common to discriminant analysis (Eisenbeis, 1981).

The scorecard in this paper is an improvement in three ways. First, the sample is large (39,956 loans repaid in 1988–96). Second, the focus is less on the statistical significance of the estimated coefficients and more on the power to predict arrears for

10,555 loans repaid in 1997. Statistical significance need not imply predictive power (Hand, 1994). Third, the scorecard uses only characteristics that most microlenders already collect.

2.3 Data and scorecard

The Bolivian microlender makes loans to urban individuals in trade and manufacturing. It bases risk evaluations almost exclusively with the personal judgment of loan officers; few loans are collateralized, and a credit committee discusses only very large or unusual loans. From August 1988 until the end of 1996, 1,987 loans out of 39,956 (5 percent) had costly arrears, defined as a spell of 15 days or more. Such long spells are costly to the lender because they require extra collection efforts. In the first nine months of 1997, 8.6 percent of loans went "bad" (913 of 10,555 loans).

The following characteristics are known for all loans disbursed and paid:

- Date of disbursement
- Amount disbursed
- Type of guarantee
- Branch
- Loan officer
- Gender of the borrower
- Sector of the firm
- Number of spells of arrears
- Length of the longest spell of arrears

This is an unusually short list; most scorecards for microlenders would also use the age, marital status, education, and length of residence of the borrower; ownership of a phone, house, or car; and measures of the size and financial strength of the household and firm. Thus, the test in this paper is conservative: if a scorecard with few characteristics works, then a scorecard with a full complement of characteristics on the borrower, loan, and lender would work even better.

Scorecard construction uses knowledge of the characteristics of past cases at the time of disbursement and of their subsequent repayment performance to infer future repayment risk for similar current cases. Because data exist only for cases that were approved under the lender's standard evaluation process, the scorecard applies only to current cases that also have been approved under this process. Otherwise, sample-selection bias can degrade scorecard performance (Crook and Banasik, 2004).

The scorecard predicts "costly arrears", defined as a dichotomous dependent variable that is 1 for "bad" loans with a spell of at least 15 days and 0 for all other "good" loans. The scorecard is derived from a logit model.

The characteristics linked with risk were chosen based on theory and experience. At the point that a loan has been provisionally approved under the standard evaluation process, these characteristics can be taken as given. Of course, the terms of the loan contract—such as amount disbursed, the term to maturity, and the guarantee requirements—do depend on the evaluation of risk by the lender. For loans provisionally approved under the lender's standard evaluation, however, the loan terms are fixed. Thus, the scorecard applies only to cases that, in the absence of scoring, would have been approved.

The test below thus checks how well the scorecard flags high-risk cases that the loan officers and the credit committee nevertheless believed to be low-risk.

3. Links between risk and characteristics

Microlenders want to predict the risk of arrears, and they also want to know which characteristics are linked with risk. This section discusses these linkages, and the next section discusses predictive power.

The scorecard is derived from a logit regression based on 39,956 loans repaid by the end of 1996. The χ^2 statistic for the scorecard as a whole had a p-value of 0.01, and 56 of 109 estimated coefficients had p-values of 0.10 or less.

3.1 Experience as a borrower

The experience of the borrower is measured as the number of previous loans and also as the number of months since the first disbursement. Table 1 shows scorecard weights (derived from estimated logit coefficients) that show how risk changes with the borrower's experience. Positive weights mark increased risk, and negative weights mark decreased risk. Table 1 also shows p-values and the means of the characteristics.

3.1.1 Number of previous loans

Looking at weights with p-values below 0.10, the risk of a loan's going "bad" decreases with the number of past loans. For example, risk is 5.4 percentage points less for a borrower with 7 previous loans than for a first-time borrower. All else constant, first-time borrowers (46 percent of cases) are the worst risks.

3.1.2 Months since the first loan

Experience in months since the first loan is a different measure of experience than the number of previous loans because, for example, a borrower could get three one-month loans or three one-year loans. Ranges of numbers of months are defined as a set of dummy variables (Table 1).

Although not all the weights have low p-values, the broad pattern suggests that risk increases with time as a borrower. The effect is large; a borrower whose first disbursement took place 54–147 months ago is 3.3 percentage points more likely to go "bad" than a new borrower.

This may reflect regression to the mean. Borrowers tend to ask for their first loan during uncommonly good times when their ability to repay is at a peak. If the first loan is repaid on time, then the lender may encourage the borrower to take larger and longer loans, even if the borrower is less able to repay such a loan than the first, smaller, shorter loan.

In any case, this is a new result. While it is common wisdom that risk decreases with more experience seen as numbers of loans, no one has discussed that risk increases with experience seen as months as a borrower.

3.2 Arrears in the previous loan

Experience with scoring for microfinance suggests that repayment history is the best predictor of future repayment performance. Most microlenders cannot check borrowers' histories in credit bureaux, but they use knowledge of past repayment performance for their own repeat borrowers. Table 2 shows how risk is linked with arrears in the previous loan. Spells of arrears were common, but most were short.

3.2.1 Length of spells of arrears in days

The weights on arrears seen as the longest spell in days in the previous loan are large; compared with no arrears (67 percent of cases), cases with one day of arrears had 2.4 percentage points less risk, and cases with 31 or more days of arrears in the previous loan had 1.6 percentage points more risk in the current loan.

This result is new: for this microlender, repeat borrowers with shorter spells in the previous loan were less likely to go "bad" than those with no arrears in the previous loan. This is surprising; common sense suggests that more past arrears would always mean more risk. Why would a short spell be better than no spell?

Perhaps some arrears are due to random shocks, and perhaps borrowers who have had some arrears but who worked to get back on track in just a few days are, on average, less likely to have long spells than those who have not yet fallen into arrears but who might not be so robust once they do hit an unlucky stretch.

3.2.2 Number of spells

The number of spells of arrears in the previous loan is strongly linked with risk (Table 2). Compared with 0–1 spells, risk increases for 2–4 spells and then starts to decrease. This may reflect traders who make frequent installments but who are often a day or two late, not from negligence but because they wait to combine the trip to pay the installment at the branch with other errands. For them, the number of spells of arrears reveals little about the risk of long spells of arrears.

3.3 Gender of the borrower

Probably the best-known stylized fact in microfinance is that women are safer than men. The Bolivian lender made most (58 percent) of its loans to women (Table 3). All else constant, women were indeed less risky, but only by 0.2 percentage points (the p-value is 0.34, so the difference in risk between men and women may very well be nil).

Does this disprove the stylized fact? Without controlling for other factors, women are almost half as risky as men; 3.6 percent of women in the sample went "bad" versus 6.9 percent of men. But after controlling for other factors—many of them correlated with gender—most of the gender gap in risk vanishes. At least for this lender, gender *per se* is not strongly linked with risk. Rather, gender is associated with other factors that do cause risk. For example, Bolivian women are more likely than Bolivian men to be traders than manufacturers, and Bolivian women are more likely to have smaller businesses and to take smaller, shorter loans. So if the lender observes only gender, then gender is strongly linked with risk, but if the lender observes many characteristics and accounts for their linkages with risk, then gender is much less predictive.

3.4 Sector

About 53 percent of loans went to traders, and their risk was 4 percentage points less than manufacturers. This is a large weight; average risk in the sample was 5 percent, so manufacturers were almost twice as risky as traders. (More finely grained class of sectors—for example, agriculture and services as well as manufacture and trade—would improve prediction, but the data for this lender lack this level of detail.)

Changing sectors between consecutive loans was associated with 0.5 percentage points more risk, but the p-value is high, and few borrowers switched sectors.

3.5 Amount disbursed

The link between risk and the amount disbursed is weak. In 1998 dollars, each $100 disbursed was linked with an increase in risk of 0.02 percentage points (Table 3).

In terms of changes in amount disbursed between consecutive loans, a $100 increase had virtually no link with risk, but a $100 decrease was associated with a decrease of 0.1 percentage points. Apparently, this lender successfully rations borrowers suspected as high risks.

In this case, the link between risk and the amount disbursed is so weak that the microlender has little scope to affect arrears via changes in amount disbursed. The average loan is already small ($680), and the average increases ($140) and decreases ($25) are even smaller. If the amount disbursed were reduced by $100, risk as predicted by the scorecard would change by 0.8 percentage points. In any case, the scorecard is properly used only after the microlender has provisionally approved the loan under its traditional evaluation process and has set the terms and conditions of the loan contract.

3.6 Guarantees

Of the four types of guarantees, the only one with a large effect and a small p-value is “no guarantee” (Table 3). Here, the lack of a guarantee does not cause risk, but it does reveal risk as judged by the loan officer. Most likely, only borrowers judged as very low risks in the normal evaluation process had the privilege of borrowing without a guarantee. Changes in the guarantee between consecutive loans are not linked with risk.

3.7 Branches

All branches are not equal (Table 4). Compared with “other” (the central office and four small branches), the safest branch was associated with 1.3 percentage points less risk. The few borrowers who switched branches were less risky by 0.8 percentage points. Such results are useful to microlenders because branch performance is susceptible to policy, for example through bonuses or training.

3.8 Loan officers

This microlender bases its normal evaluation on the subjective judgment of loan officers. Of course, officers differ in their ability to “smell” high-risk cases, and they may also take time to learn the ropes and to sharpen their “sixth sense”.

Perhaps surprisingly, risk increases as loan officers gain experience (Table 4). For example, cases handled by a new loan officer with 0–6 months of experience are 3.2 percentage points less risky than cases handled by an old hand with 148 months of experience. Although loan officers learn to work smarter with time, the amount of work that they must do also grows as their portfolios expand. Furthermore, the quality of new borrowers may degrade as loan officers get past the "cream" in the neighborhoods where they work and start to recruit more "typical" borrowers.

Beyond experience, loan officers differ in their ability to sense high-risk cases (Table 5). Compared with "other" officers (those with less than 300 loans paid-off) the safest officer was linked with 4.8 percentage points less risk, and the riskiest officer was linked with 2.1 percentage points more risk. Loan officers are not interchangeable parts; microfinance rests on personal relationships, so who the person is is important. This matters because lender policy influences loan officers more directly than borrowers.

The 12 percent of borrowers who changed loan officers—usually because the loan officer quit—were 0.5 percentage points more risky (Table 5). Thus, decreased staff turnover may lead to decreased arrears.

3.9 Date of disbursement

To control for seasonal or once-off changes in the market or lender policy, the scorecard controls for the year and month of disbursement. Loans disbursed in the months before Christmas when business is heaviest are more risky. Compared with 1988–91, risk increased in 1992–93 before falling in 1994–96.

In sum, risk depends on gender, sector, arrears in the previous loan, the experience of the borrower, the experience of the loan officer, the specific loan officer, and the specific branch. Seasonality and changes in policy and the market also matter.

4. Predictive power

Scoring uses what is known from the past to forecast what will take place in the future. This section checks how well the scorecard built on data from 1988–96 classifies loans repaid in the first nine months of 1997.

By most measures, the scorecard does indeed have some predictive power. Still, it is less powerful than most scorecards for credit cards. This reflects the challenge of microfinance to judge risk without reference to credit bureaux or formal wage jobs. Risk is correlated with inexpensive-to-observe characteristics, and lenders can use this to reduce arrears, but the link is too weak for scoring to replace loan officers completely.

In 1988–96, 5 percent of the Bolivian microlender's loans went "bad". In 1997, 8.6 percent went "bad". A naïve model would predict that 5 percent of loans would go "bad" in 1997, but the scoring model predicted 6.4 percent. Thus, about one-third of the increase in problematic loans was due to changes in characteristics that appear in the scorecard, while two-thirds of the increase was due to other factors such as changes in competition and in the macroeconomy.

Scoring also predicts the risk of each loan. For example, if the Bolivian lender had used the scorecard in 1997 with a rejection threshold of 0.10 (that is, if it had rejected all provisionally approved applicants with a risk of 10 percent or higher, and approved all others), then the share of "bad" loans would have decreased from 8.6 to 6.9 percent. With a threshold of 0.05, the share of "bads" would have fallen to 4.8 percent.

As the threshold approaches zero, fewer “bad” loans sneak through but more “good” loans are mistakenly rejected. Scoring gives estimates of risk, but lenders must choose how to balance risk against the cost to reduce it and against other goals.

If estimated risk exceeds a threshold, then—for the purposes of the historical test—a loan is rejected; otherwise, it is approved. Given knowledge of what would have happened had the case been approved (because, in reality, the cases in the historical test were approved), scoring has four possible outcomes:

- *“Good” approved*: a “good” with predicted risk below the threshold
- *“Bad” rejected*: a “bad” with predicted risk above the threshold
- *“Bad” approved*: a “bad” with predicted risk below the threshold
- *“Good” rejected*: a “good” with predicted risk above the threshold

For thresholds from 0–0.30 and for 1, the outcomes for the historical test with 1997 data for the Bolivian lender are in Table 7. In the test sample, 913 (8.6 percent) of loans were “bads”, and 9,642 (91.4 percent) were “goods”. As the threshold increases, “goods” approved increase and “goods” rejected decrease; however, “bads” rejected decrease, and “bads” approved increase. Lenders choose a threshold based on the trade-offs among the four outcomes, their goals, and the benefits and costs of each outcome.

The all-bad threshold is so low (0.00) that all loans are rejected. The all-good threshold is so high (1.00) that all loans are approved. The all-bad model is a straw person, but the all-good model is not; it is equivalent to policy that the Bolivian lender used once it has approved a borrower through its normal evaluation when it did not have a scorecard.

4.1 Good/bad separation

The most basic test of a scorecard is how well it separates goods from bads. The cumulative distributions of estimated risk for "goods" and "bads" (Figure 1) show that the scorecard achieves some separation. The distribution of "goods" (mean 0.062, median 0.042) is always left of the distribution of "bads" (mean 0.098, median 0.077).

4.2 "Hit" rates

To what extent does the scorecard separate goods from bads? The proper measure of the sharpness of separation depends on the goals of the lender (Hand, 1994; Kennedy, 1998). *"Hit" rates* are best if a lender wants to optimize the share of "goods" approved and/or "bads" rejected. Table 7 shows the share of "goods" approved, the share of "bads" rejected, and the total "hit" rate ("goods" approved and "bads" rejected as a share of all cases).

In terms of the share of "goods" approved, the all-good threshold of 100 percent beats the scorecard at all possible thresholds. On the other hand, the share of "bads" rejected for the scorecard beats the all-good model at all possible thresholds.

An all-bad model would have approved none of the "goods" and rejected all of the "bads". Thus, the scorecard always does better in terms of "goods" approved but always does worse in terms of "bads" rejected.

The total "hit" rate for the scorecard never beats the highest naïve "hit" rate (0.914) that is achieved by simply predicting that all cases will be "good". If the Bolivian lender only wanted to maximize the "hit" rate, then it would predict that all loans would be good. In practice, however, the loss from a "bad" approved exceeds the benefit from a "good" approved. Likewise, the cost avoided due to a "bad" rejected exceeds the benefit missed due to a "good" rejected. Because lenders do not weigh all outcomes the same, they generally will do better with a scorecard than with their current (implicit) naïve all-good threshold.

Figure 2 shows the trade-off between the share of "goods" approved and the share of "bads" rejected. The diagonal represents a policy that rejects loans at random. Scoring has more power as its curve bends away from the diagonal; a perfect scorecard would trace the upper border and then the right border, forming a mirror-image "L".

The scorecard is near the upper border for shares of "bads" rejected above 0.8 and near the right border for shares of "goods" approved near 0.8. This suggests that the scorecard would work well as a "super-pass" or "super-fail" filter. The lender could use the scorecard to approve very low risks ("super-passes") without further ado and to flag high risks ("super-fails") for more review.

In sum, the scorecard predicts risk well. It separates "goods" from "bads" imperfectly (no scorecard is perfect), but, on average, it assigns higher risk to "bads" than to "goods". If the lender puts more weight on successfully rejecting a "bad" than on successfully approving a "good", then scoring beats the all-good naïve model currently used once a borrower is approved by the traditional evaluation process.

5. Conclusion

Both credit-card lenders in high-income countries and microfinance lenders in low-income countries make massive numbers of small, short, unsecured loans. Unlike credit-card lenders, however, microfinance lenders do not use scorecards.

Can scoring help microfinance? A scoring model for arrears at a microlender in Bolivia suggests that it can. The model pinpoints characteristics that are associated with risk and, more importantly, it predicts risk better than the all-good "naïve" model currently used by the lender in which all loans approved by the traditional evaluation process are disbursed. Still, scoring for microfinance is less powerful than scoring for credit cards, so scoring and knowledge of quantitative characteristics will not replace loan officers and their knowledge of qualitative character anytime soon.

How should scoring be used? As usual, the math is the easy part. The difficult work is to collect the data and then to use the risk forecasts. The scorecard here is not powerful enough to accept or to reject applicants without a standard evaluation; risk is linked with the characteristics in the scorecard, but it still depends strongly on factors that only the loan officer can observe. Also, the scorecard starts from the premise that an applicant has already been provisionally approved under the normal evaluation.

The scorecard is probably most appropriately used as a "super-fail" filter that flags high-risk cases that deserve more careful review. Thus, scoring channels effort to borderline cases where rewards are greatest.

Even lenders who do not score each borrower can still use knowledge of the weights in scorecards to inform policy choices. For example, the Bolivian lender might try to attract more traders because they are safer than manufacturers. Likewise, the lender might refer to a special credit committee all loans to borrowers who had a spell of arrears in their most recent loan of more than 15 days. Finally, the scorecard isolates the risk associated with individual branches and individual loan officers. This allows the lender to target training and incentives to those who need it the most.

In the end, the greatest challenge to scoring for microfinance is not technical but managerial. After all, predictive power can be tested with historical data, so no microlender should have to use a scorecard that does not forecast risk well. But some microlenders—especially those whose mission focuses more on service to the poorest than on profitability—fear that scoring will overstate the risk of the poor or that knowing the risk of the poor will lead to mission drift.

Both fears are valid, and both may be addressed with proper management. Scoring will not overstate the risk of the poor as long as scorecards are based on historical data for loans commonly used by the poor. For example, suppose that a poor person applies for a microloan and is run through two scorecards, one constructed from data for mortgage loans to middle-class, salaried civil servants and one constructed from data for microloans for poor people. The civil-servant scorecard will likely overstate the risk of the self-employed poor (because they will appear to be very below-average civil servants), but the microloan scorecard will, on average, give an accurate estimate of the risk of the poor. Thus, the key to avoiding shortchanging the poor is to not blindly apply scorecards constructed with data from one type of loan and one type of population to a different product and population.

But what if the data show that the poor are indeed worse risks, even for loans tailored to their demands? This knowledge need not automatically lead to mission drift; what kind of mission-driven organization would abandon its mission just because it learns that success is more difficult than expected? No, knowledge is better than ignorance. If the poor are riskier, then managers can use that knowledge to make better decisions about how to make trade-offs between cost and depth of outreach. Even the most mission-driven poverty lender has a limit; a loan made to one poor person with an 80-percent risk of default is a loan not made to another less-risky poor person (or, through time, several loans not made to other less-risky poor people). So while poverty-focused lenders are willing to accept more risk, they (and their borrowers) benefit from greater knowledge of risk. After all, arrears harm borrowers at least as much as lenders, as borrowers suffer worry and humiliation and may end up selling assets to repay debts (Mosley, 2001). Scoring gives managers better knowledge of repayment risk; whether managers use that knowledge for good or ill is up to them. The best way to improve the odds of good use is not to suppress scoring but to educate managers about what scoring can and cannot do and when scoring is appropriate.

References

Berger, Allen N.; Frame, W. Scott; and Nathan H. Miller. (2002) "Credit Scoring and the Availability, Price, and Risk of Small-Business Credit", Federal Reserve Bank of Atlanta Working Paper No. 2002–06, `http://www.frbatlanta.org/filelegacydocs/wp0206.pdf`.

Bond, Phillip; and Ashok Rai. (2002) "Collateral Substitutes for Microfinance", manuscript, `http://finance.wharton.upenn.edu/~pbond/research/limits-jul2-02.pdf`.

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.

Efron, Bradley; and Robert J. Tibshirani. (1993) *An Introduction to the Bootstrap*, New York: Chapman and Hall, ISBN 0–412–04231–2.

Eisenbeis, Robert A. (1981) "Credit Scoring Applications", pp. 167–198 in Edward I. Altman, Robert B. Avery, Robert A. Eisenbeis, and Joseph F. Sinkey, Jr. (eds) *Application of Classification Techniques in Business, Banking and Finance*, Greenwich: JAI Press, ISBN 0–89232–044–4.

Frame, W. Scott; Padhi, Michael; and Lynn Woosley. (2001) "The Effect of Credit Scoring on Small Business Lending in Low- and Moderate-Income Areas", Federal Reserve Bank of Atlanta Working Paper No. 2001–6, `http://www.frbatlanta.org/filelegacydocs/wp0106.pdf`.

Hand, D.J. (1994) "Assessing Classification Rules", *Journal of Applied Statistics*, Vol. 21, No. 3, pp. 3–16.

Hand, D.J.; and W.E. Henley. (1997) "Statistical Classification Methods in Consumer Credit Scoring: A Review", *Journal of the Royal Statistical Association Series A*, Vol. 160, Part 3, pp. 523–541.

Kennedy, Peter. (1998) *A Guide to Econometrics, Fourth Edition*, Cambridge, MIT Press, ISBN 0–262–11235–3.

Lewis, Edward M. (1990) *An Introduction to Credit Scoring*, San Rafael: Athena Press, ISBN 99–956–4223–9.

Longhofer, Stanley D. (2002) "Statement of Stanley D. Longhofer", Perspectives on

Credit Scoring and Fair Mortgage Lending, Final Installment, Federal Reserve Bank of Saint Louis, `http://www.stls.frb.org/community/other_pubs.html`.

Martell, Javier; Panichelli, Paul; Strauch, Rich; and Sally Taylor-Shoff. (1999) "The Effectiveness of Scoring on Low-to-Moderate-Income and High-Minority Area Populations", San Rafael, CA: Fair, Isaac Company.

Mosley, Paul. (2003) "Bolivia during the Global Crisis 1998–2003: Towards a 'Macroeconomics of Microfinance'", manuscript.

Nannyonga, Harriet Linda. (2000) *Determinants of Repayment Behavior in the Centenary Rural Development Bank of Uganda*, unpublished Ph.D. dissertation, Department of Agricultural, Environmental, and Development Economics, The Ohio State University, `nannyonga.2@osu.edu`.

Navajas, Sergio; Schreiner, Mark; Meyer, Richard L.; Gonzalez-Vega, Claudio; and Jorge Rodriguez-Meza. (2000) "Microfinance and the Poorest of the Poor: Theory and Evidence From Bolivia", *World Development*, Vol. 28, No. 2, pp. 333–346.

Reinke, Jens. (1998) "How to Lend like Mad and Make a Profit: A Micro-credit Paradigm versus the Start-Up Fund in South Africa", *Journal of Development Studies*, Vol. 34, No. 3, pp. 44–61.

Rhyne, Elisabeth. (2001) *Mainstreaming Microfinance: How Lending to the Poor Began, Grew, and Came of Age in Bolivia*, Bloomfield: Kumarian, ISBN 1–56549–127–0.

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

Sharma, Manohar; and Manfred Zeller. (1997) "Repayment Performance in Group-Based Credit Programs in Bangladesh: An Empirical Analysis", *World Development*, Vol. 25, No. 10, pp. 1731–1742

Viganò, Laura. (1993) "A Credit Scoring Model For Development Banks: An African Case Study", *Savings and Development*, Vol. 17, No. 4, pp. 441–482.

Zeller, Manfred. (1998) "Determinants of Repayment Performance In Credit Groups: The Role of Program Design, Intra-Group Risk Pooling, and Social Cohesion", *Economic Development and Cultural Change*, Vol. 46, No. 3, pp. 599–620.

Table 1: Scorecard weights for borrower experience

Experience of borrower		Mean	Weight	p-value
Previous loans	0	0.460	0.000	N/A
	1	0.247	–0.012	0.50
	2	0.131	–0.023	0.21
	3	0.070	–0.028	0.12
	4	0.039	–0.032	0.09
	5	0.022	–0.040	0.05
	6	0.013	–0.034	0.10
	7	0.008	–0.054	0.04
	8	0.005	–0.026	0.28
	9 or more	0.006	–0.025	0.31
Months since first loan	0–6	0.466	0.000	N/A
	7–19	0.170	0.015	0.40
	20–53	0.233	0.021	0.23
	54–147	0.125	0.033	0.07
	148 or more	0.007	0.031	0.11

Table 2: Scorecard weights for arrears in the previous loan

Arrears in previous loan		Mean	Weight	p-value
Longest spell in days	0	0.674	0.000	N/A
	1	0.127	–0.024	0.01
	2	0.054	–0.018	0.01
	3	0.034	–0.018	0.01
	4	0.028	–0.013	0.03
	5	0.012	–0.008	0.28
	6	0.009	0.003	0.63
	7	0.016	0.001	0.85
	8	0.007	0.017	0.02
	9	0.004	0.016	0.05
	10–14	0.014	0.012	0.05
	15–23	0.009	0.028	0.01
	24–30	0.003	0.020	0.03
	31 or more	0.007	0.016	0.03
Number of spells	0-1	0.761	0.000	N/A
	2	0.062	0.009	0.08
	3	0.044	0.011	0.05
	4	0.032	0.015	0.01
	5 or 6	0.041	0.008	0.16
	7 or more	0.059	0.006	0.27

Table 3: Scorecard weights for gender, sector, amount disbursed, and guarantee

Characteristic		Mean	Weight	p-value
Gender	Male	0.422	0.000	N/A
	Female	0.578	–0.002	0.35
Sector	Manufacturing	0.473	0.000	N/A
	Trade	0.527	–0.040	0.01
Changed sector		0.006	0.005	0.52
Amount disbursed	Level	676	0.0000023	0.03
	Increase	140	–0.0000003	0.89
	Decrease	25	–0.0000123	0.01
Guarantee	Other	0.029	0.000	N/A
	Personal	0.475	0.002	0.55
	No guarantee	0.248	–0.009	0.01
	Multiple	0.248	–0.004	0.29
Changed guarantee		0.100	0.001	0.76

Table 4: Scorecard weights for the branch and the experience of the loan officer

Characteristic		Mean	Weight	p-value
Branch	Other	0.438	0.000	N/A
	1	0.114	–0.013	0.23
	2	0.072	–0.012	0.12
	3	0.161	–0.010	0.13
	4	0.044	–0.008	0.29
	5	0.053	–0.007	0.51
	6	0.078	–0.003	0.56
	7	0.040	0.000	0.98
Changed branch		0.024	–0.008	0.10
Experience of loan officer in months	0-6	0.062	0.000	N/A
	7-19	0.204	0.006	0.09
	20-53	0.322	0.009	0.02
	54-147	0.335	0.020	0.01
	148 or more	0.078	0.032	0.01

Table 5: Scorecard weights for specific loan officers

Loan officer	Mean	Weight	p-value
Other	0.116	0.000	N/A
1	0.008	–0.048	0.01
2	0.067	–0.038	0.01
3	0.019	–0.037	0.01
4	0.009	–0.037	0.01
5	0.037	–0.033	0.01
6	0.025	–0.025	0.01
7	0.038	–0.024	0.01
8	0.045	–0.024	0.01
9	0.059	–0.023	0.01
10	0.048	–0.020	0.01
11	0.016	–0.019	0.04
12	0.015	–0.018	0.09
13	0.017	–0.017	0.10
14	0.014	–0.016	0.17
15	0.031	–0.015	0.02
16	0.027	–0.014	0.04
17	0.035	–0.013	0.02
18	0.024	–0.012	0.03
19	0.010	–0.007	0.31
20	0.016	–0.006	0.30
21	0.019	–0.005	0.57
22	0.031	–0.004	0.71
23	0.019	–0.002	0.81
24	0.011	–0.001	0.95
25	0.016	0.002	0.88
26	0.022	0.002	0.84
27	0.016	0.002	0.79
28	0.015	0.004	0.58
29	0.010	0.005	0.54
30	0.010	0.005	0.49
31	0.035	0.007	0.55
32	0.010	0.007	0.31
33	0.009	0.008	0.52
34	0.041	0.009	0.40
35	0.016	0.009	0.44
36	0.014	0.021	0.01
37	0.011	0.021	0.01
Changed officer	0.116	0.005	0.05

Table 6: Scorecard weights for year and month of disbursement

Characteristics		Mean	Weight	p-value
Year of disbursement	1988-1991	0.083	0.000	N/A
	1992	0.086	0.040	0.01
	1993	0.131	0.068	0.01
	1994	0.198	0.059	0.01
	1995	0.353	0.056	0.01
	1996	0.150	0.050	0.01
Month of disbursement	January	0.056	0.000	N/A
	February	0.064	0.006	0.13
	March	0.088	0.004	0.28
	April	0.091	0.002	0.65
	May	0.102	0.003	0.48
	June	0.096	0.006	0.12
	July	0.081	0.006	0.16
	August	0.081	0.006	0.13
	September	0.087	0.009	0.04
	October	0.086	0.008	0.06
	November	0.089	0.009	0.03
	December	0.079	0.010	0.02

Table 7: Power to predict with historical data

		Threshold							
		All-bad							All-good
Measure	Formula	0.00	0.05	0.10	0.15	0.20	0.25	0.30	1.00
“Goods” approved	GA	0	5,343	7,791	8,976	9,330	9,491	9,561	9,642
“Bads” rejected	BR	913	646	335	173	98	52	27	0
“Bads” approved	BA	0	267	578	740	815	861	886	913
“Goods” rejected	GR	9,642	4,299	1,851	666	312	151	81	0
Share of “Goods” approved	GA/(GA+GR)	0.00	0.55	0.81	0.93	0.97	0.98	0.99	1.00
Share of “Bads” rejected	BR/(BR+BA)	1.00	0.71	0.37	0.19	0.11	0.06	0.03	0.00
Total “hit” rate	(GA+BR)/N	0.09	0.57	0.77	0.87	0.89	0.90	0.91	0.91

Note: There are 10,555 cases, 9,642 “goods” and 913 “bads”.
The “good” rate is 0.914, and the “bad” rate is 0.086.

Figure 1: Cumulative distributions of predicted risk for "bads" and "goods"

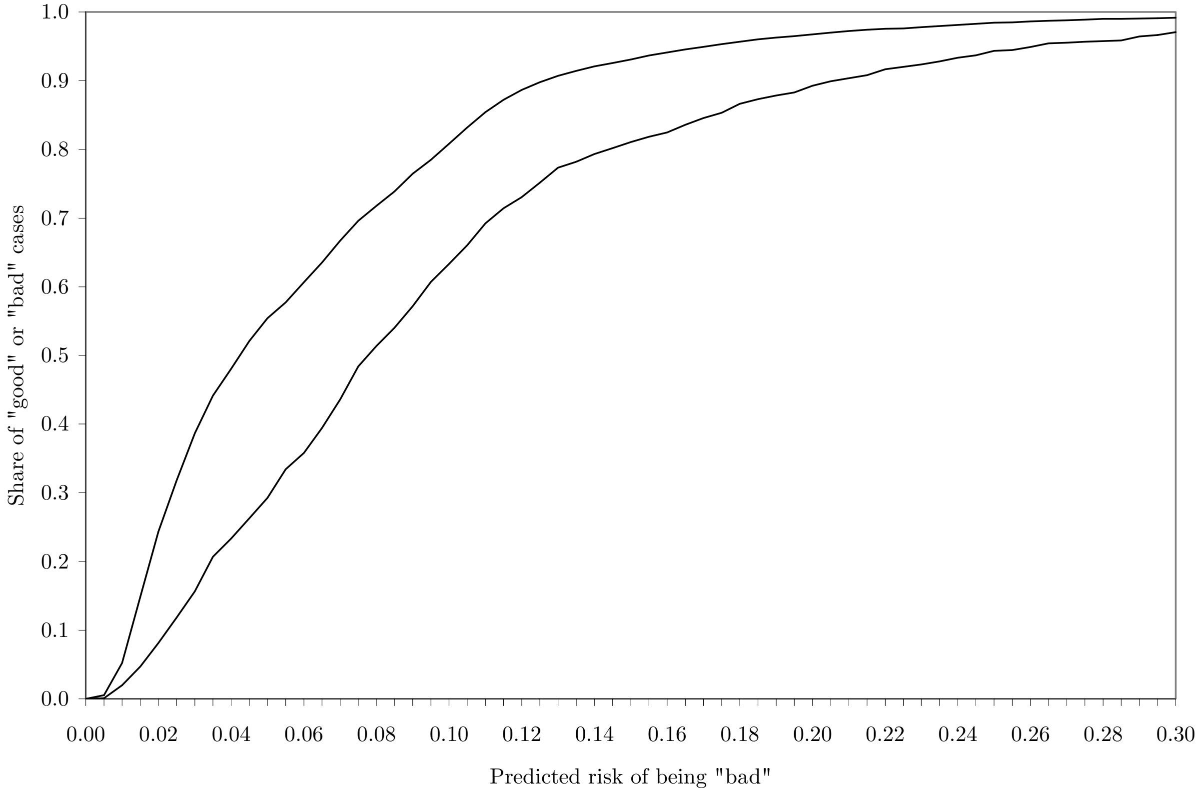

Figure 2: Trade-off between the share of "goods" approved and the share of "bads" rejected