Benefits and Pitfalls of Statistical Credit Scoring for Microfinance

Mark Schreiner

December 23, 2004

Microfinance Risk Management
6970 Chippewa St. #1W, St. Louis, MO 63109-3060, U.S.A.
Telephone: (314) 481-9788, http://www.microfinance.com
and
Center for Social Development
Washington University in St. Louis
Campus Box 1196, One Brookings Drive, St. Louis, MO 63130-4899, U.S.A.

Abstract

This paper discusses the benefits and pitfalls of credit scoring applied to microfinance. Although scoring will not replace joint-liability groups nor loan officers, it does have enough predictive power to significantly improve the evaluation of the risk of loans applicants. This paper discusses what scoring can and cannot do, describes the data that microlenders who plan to use scoring should start to collect from all loan applicants, and outlines the basic steps in a scoring project.

Author's Note

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

Acknowledgments

A later version of this paper appears in *Savings and Development*, 2004, Vol. 28, No. 1, pp. 63–86. Many people have helped me learn about credit scoring for microfinance. I am particularly grateful to Hans Dellien, Sean Kline, and Gary Woller.

Benefits and Pitfalls of Statistical Credit Scoring for Microfinance

1. What is scoring?

Scoring is the use of the knowledge of the performance and characteristics of past loans to predict the performance of future loans. For example, when a loan officer judges risk by mentally comparing a current applicant with her experience with other applicants, it is scoring, albeit implicit and subjective. Likewise, when a microlender adopts a policy not to renew loans to clients who had spells of arrears in excess of 30 days in their previous loan, it is scoring, albeit simple and one-dimensional. Thus, although the name *scoring* may be new to microfinance, scoring itself is old hat.

Statistical scoring is the use of quantitative knowledge of the performance and characteristics of past loans recorded in an electronic data base to predict the performance of future loans. The evaluation of the repayment risk of the self-employed poor is the central challenge of microfinance. The innovations of microfinance to date have been the use of joint-liability groups and detailed evaluations of individual applicants to judge risk; scoring promises to be the next jump in efficiency.

For example, a statistical scoring system might start with a base risk of 10 percent. It might then add a percentage point if the applicant is a man, add 8 basis points for each $50 that would be disbursed, add 50 basis points for each month the loan will be outstanding, and subtract 4.5 percentage points if the borrower owns a telephone. The scorecard might then subtract 3 percentage points if the borrower is a maize farmer but add 4 percentage points if the borrower is a goat breeder, subtract 5 percentage points for second-time borrowers but add 0.5 percentage points for each loan after the second, and add or subtract percentage points depending on the specific loan officer assigned to the application.

The sum of the weighted characteristics is a probability that the loan would, if disbursed, eventually go “bad”, where *bad* is defined by the lender. The weights of each characteristic in the scorecard are based on a statistical analysis of the relationship between the characteristic and repayment in the lender’s historical data base.

To use scoring in daily work, the microlender might specify four policy ranges: super-safe, regular, risky, and super-risky. Applicants with predicted risk in the “super-safe” range are quickly approved and may qualify for lines of credit or other special rewards. Applicants with predicted risk in the “regular” range are approved as they always have been. Applicants with predicted risk in the “risky” range receive extra attention from the loan officer and from the credit committee. The amount, term, or guarantee specified in the loan contract may be adjusted for “risky” cases in an attempt to control risk. Finally, applicants with risk in the “super-risky” range are summarily rejected. By testing various policy ranges on historical data, the lender can have a good idea of the trade-offs involved even before the policies go into effect.

2. Benefits of statistical scoring

Statistical scoring quantifies risk and has several important potential advantages when compared with implicit or subjective scoring.

2.1 Statistical scoring quantifies risk as a probability

For example, predicted risk under statistical scoring might be 4.5 percent that a loan will at some point have arrears of 30 days or more. In contrast, subjective scoring might merely express that a loan is of below-average risk, and this is based in large part on a qualitative feeling.

2.2 Statistical scoring is consistent

The scorecard treats identical applications identically. Two people with the same characteristics have the same predicted risk. Under subjective scoring, however, risk judgements might vary by loan officer or even with the mood of a given loan officer.

2.3 Statistical scoring is explicit

With statistical scoring, the exact process (the scorecard) used to predict risk is known and can be communicated. Subjective scoring, on the other hand, depends on a vague process that even its users sometimes cannot explain. The difficulty with the replication of subjective scoring explains the great time and energy spent in training and capacity-building for loan officers who must use subjective scoring.

2.4 Statistical scoring accounts for a wide range of risk factors

Evaluation guidelines may specify that an application must meet a few financial ratios and other simple policy rules, but, unlike statistical scoring, subjective scoring cannot consider more than a handful of characteristics. Furthermore, subjective scoring is usually limited to "death penalty" rules, such as the value of the guarantee must be at least 200 percent of the loan value, or no loan. In contrast, statistical scoring can quantify the risk trade-off due to guarantee coverage of only 180 percent, or only x percent. Compared with subjective scoring, statistical scoring enables more refined risk evaluation and more deliberate risk management.

2.5 Statistical scoring can be tested before use

For example, a newly constructed scorecard can be applied to currently outstanding loans, using only characteristics known to the lender at the moment of disbursement. This predicted risk can then be compared with observed risk to date. This reveals how scoring would have worked, had it been in place at the time of disbursement of currently outstanding loans. Subjective scoring could also be tested on historical data, but it is prohibitively costly and, as far as I know, has never been done.

Second only to the quantification of risk, the ability to test statistical scoring before use is its most important strength. A great mistake of those lenders who use "expert systems" (scorecards with weights derived from experience and assumptions, rather than historical relationships in the data base) is not the use of assumed weights but rather the failure to test the “expert system” on historical data. A stock-picker or horse-bettor would test his system on historical data before placing his own money at risk in the real world; microlenders should do the same with their scorecards.

2.6 Statistical scoring reveals trade-offs

By showing what the lender can expect from various policy choices, statistical scoring improves risk management. For example, the historical test might reveal to management that, among all historically approved loans that would have had a risk forecast in excess of 50 percent, about 62 percent in fact ended up with arrears in excess of 30 days. Also, the historical test might reveal that about 8.5 percent of all loans currently outstanding have a predicted risk in excess of 50 percent. Thus, scoring suggests that if the lender, for example, were to adopt a policy to reject all loans with a risk forecast in excess of 50 percent, then it would avoid about 6 “bads” for each four “goods” that it would lose, and that disbursements would decrease by about 8.5 percent.

Of course, the historical test of scoring cannot tell managers what policy to choose, but it can inform them about probable consequences of various choices. Subjective scoring works, but no one knows what would happen with different policies. Scoring reveals what would have happened, exactly what good management requires.

2.7 Scoring reveals the links between risk and the characteristics of the borrower, the loan, and the lender

For example, the received wisdom in microfinance is that women repay better than men. For a given lender, scoring not only confirms or denies this wisdom but also reveals precisely how much gender matters. Scoring can also reveal, for example, how risk is linked with past arrears, with the type of business, and with adjustments to the terms and conditions of the loan contract. Scoring can even tell management how loan officers would fare if they all managed an identical portfolio.

In contrast, subjective scoring links risks with characteristics based on beliefs derived from experience and/or handed-down wisdom, but the beliefs and wisdom may be incorrect, or at least imprecise. Scoring uses statistics to derive the historical links between risk and characteristics. In general, statistical scoring confirms the general direction of subjective judgement (for example, past arrears do signal greater risk of future arrears, and carpenters are indeed among the riskiest of borrowers), but statistical scoring—unlike subjective scoring—reveals the precise strength of the links.

2.8 Statistical scoring does not require changes in the current evaluation process before the credit committee

Scoring uses the current data base in its current form. Although the lender may start to collect more data with an eye toward a more-powerful scorecard in 1 to 3 years, the only characteristics that loan officers need to collect for scoring are those that they currently collect. Likewise, keypunch operators enter the same data as always.

Once data are collected and keypunched, the management-information system (MIS) computes the score and displays it in various reports, for example, in the daily list of cases to review in the credit committee, in the daily list of loans in arrears for each loan officer, and in the weekly list of outstanding loans for each loan officer. The MIS also automatically produces follow-up and monitoring reports that allow managers, on a monthly basis, to monitor the continued performance of scoring. In short, although scoring is technically complex, its use in practice is automated; front-line personnel do not need to compute risk forecasts, they "merely" need to decide how to use them.

2.9 Statistical scoring reduces time spent in collections

The main benefit of statistical scoring is that loan officers spend less time in collections. For a lender's first scorecard, a wise practice is to start with a simple "disbursement scorecard" that uses data known before disbursement to predict repayment behavior after disbursement. Disbursement scoring has three functions, all of which reduce time spent in collections. First, it reduces the number, value, and length of loans disbursed to high-risk applicants. This reduces the number of times loans fall into arrears and thus saves loan officers time in collections.

Second, once a loan has been disbursed, the score highlights outstanding cases at-risk of problems, even though they are still fine. Loan officers might be extra-aware of these borrowers. They might even pay such borrowers "courtesy visits" even before any arrears, just to reinforce the presence of the lender in the mind of the borrowers.

Third, once a loan has fallen into arrears, the disbursement score can help loan officers to prioritize collections efforts; they can visit first those loans with a high risk of staying in arrears for a long time. For example, loan officers receive each morning a report that lists all their borrowers in arrears. If this report includes the predicted risk of reaching, say, 30 days of arrears, and if cases in the report are ordered by their predicted risk, then loan officers might decide to visit high-risk borrowers the very first day after they fall into arrears, whereas they might not bother visiting low-risk borrowers—who might cure themselves anyway and who might feel offended or embarrassed by a collections visit—until they fall several days behind.

2.10 Statistical scoring affects profits, and the first-round effect can be estimated

Suppose a microlender knew the net cost of a "bad" loan and also the net profit of a "good" loan. (Knowledge of these costs and benefits is very useful, regardless of the use of scoring.) This is equivalent to knowing the net benefit of avoiding a bad loan and the net cost of not making a good loan. Given the performance of the scorecard in the historical test, a lender can estimate the first-round direct effect on profits of a given super-risky range. The change in profit is the number of bad loans that would be avoided multiplied by the net benefit per bad avoided, net of the number of good loans lost multiplied by the net cost per good lost.

2.11 Statistical scoring beats the "automatic" grade

Some lenders assign an "automatic" grade to each loan. For example, their MIS assigns a grade from 1 to 5 as a function of arrears in the previous loan. (This is a very simple "expert system" based on a single characteristic, previous arrears.) Some lenders also assign a "subjective" grade determined by the loan officer. For example, perhaps a bad spell of arrears was due to a freak home fire (in which case the subjective grade is better than the automatic grade). Or perhaps a borrower always paid one day late, but grumbled and complained the whole time (in which case the subjective grade is worse than the automatic grade).

Scoring is more useful than an automatic grade for three reasons. First, an automatic grade *assumes* a relationship between past arrears and future risk; scoring *derives* the actual historical relationship. Second, an automatic grade does not exist for new borrowers because they do not have a repayment record; scoring, however, still predicts risk for new borrowers (although predictive power is less than for repeat borrowers). Third, an automatic grade is based only on past repayment performance; scoring accounts for past performance, plus a host of other characteristics.

Of course, statistical scoring cannot account for the subjective factors that enter into the subjective grade, but, if the microlender currently finds that some measure that resembles the automatic grade is useful, then the lender should find that the predicted risk produced by scoring is much more useful.

3. Pitfalls of statistical scoring

Scoring has several pitfalls. The microlender who fails to respect these runs the risk of a failed (unused) scorecard or, worse, a misused scorecard, with a subsequent and rapid increase in arrears. Scoring is powerful, so its misuse can be very damaging.

3.1 Statistical scoring requires data on many loans

Scorecard construction requires, as a gross rule-of-thumb, an electronic data base with at least 1,000 paid-off loans classed as "bad". Few microlenders have been large enough for long enough to accumulate this much data, and, furthermore, some large, experienced lenders have failed to archive their data properly.

3.2 Statistical scoring requires a lot of data on each loan

For most types of loans in rich countries like the United States, scorecards based on 10-15 borrower characteristics have almost completely displaced subjective evaluation by loan officers (Mester, 1997). Of course, scoring will not replace loan officers in microfinance. Almost all U.S. borrowers have wage jobs and extensive credit histories in credit bureaus; in contrast, almost all microfinance borrowers are self-employed and have no formal credit record. Thus, the best predictors of risk are not available in microfinance; scoring must compensate for the lack of a few very strongly predictive characteristics with the use of a larger number of less-strongly predictive characteristics, not only of the borrower but also of the loan and of the lender.

3.3 Statistical scoring requires high-quality data

All data bases have some dirt and noise. As long as the inaccuracies or random values are not too much, statistical scoring can detect the links between the characteristics in the data base and credit risk.

But how much inaccuracy is too much? The answer is empirical. If, during scorecard construction, the noise in a characteristic completely drowns out the signal, the statistical results will show this, and the scorecard will omit the characteristic.

For example, a common trouble-spot is the type of business (for example, farmers, seamstresses, carpenters, petty traders). The microlender may have asked loan officers to record this data—and keypunch operators to type it into the MIS—for years, even though the lender, so far, has not actually used the data in any way. Through time, loan officers and keypunch operators learned that there was no reward to taking the time to carefully record the type of business. Thus, data quality deteriorated. Now, scoring uses this data, but it may not be as predictive as it could have been, had the lender monitored data quality from the start. Of course, the lender never had suspected that it would use scoring, but what the lesson is to design processes that ensure data quality from now on, not just for the current scorecard but also for possible renovations and expansions 1 to 3 years down the road.

3.4 Statistical scoring requires a consultant

The concept of scoring—and the use of scoring—are straightforward. Indeed, because scoring focuses attention on the judgement of risk and on the explicit consideration of trade-offs, a by-product of the scoring process is a general shift in organizational culture toward conscious, explicit, pro-active risk management. (Another hoped-for by-product is a management habit of asking business questions that can be addressed by quantitative analysis of the data base.)

Still, scorecard construction involves complex statistical work, and project management and monitoring is best done by someone with experience. Clearly, the use of a consultant reduces flexibility and increases dependency. The consultant should train lender personnel well enough that, even though they will never build a scorecard from scratch, they can still monitor scoring and make sensible changes to its use.

3.5 Statistical scoring depends on integration with the MIS

Once the scorecard is constructed, someone (preferably a full-time, in-house programmer) must integrate the scorecard in the MIS. If the programmer cannot make time for the extra work or if the programmer makes errors, then the scoring project can founder. To avoid this, the local programmer must first understand scoring and its usefulness. Second, the local programmer must be free to dedicate 3-5 months full-time to integration and implementation; scoring cannot be integrated into an MIS at night and on weekends. Third, the local programmer must be paid fairly for her efforts.

3.6 Statistical scoring seems to fix what ain't broke

The current system of subjective scoring works; many microlenders provide quality, efficient, low-cost loans to the self-employed poor. Statistical scoring recognizes this, and it also recognizes that subjective scoring is indispensable for judging those aspects of risk associated with characteristics and qualities that are not (or cannot be) quantified and/or have not been recorded in an electronic data base.

Thus, statistical scoring only aims to add an additional step to the traditional evaluation process. Historically, scoring projects fail not because of statistical flaws but because the people within the lender's organization reject scoring. After all, scoring represents a fundamental change; a computer purports to help the loan officer and the credit manager. Applications that would have been approved without modification are modified or even rejected. Risk management shifts to a new level of care and explicitness, and the information technology department ascends in power.

Some people within the microlender resist such change. They doubt (quite reasonably) that an outside consultant who has never met any of their clients can nevertheless identify high-risk (and low-risk) applicants. Furthermore, loan officers may take offense when scoring assigns a high risk to a loan that they judged good enough to bring to the credit committee. Many employees, in spite of repeated historical tests, will persist in their disbelief that scoring has any predictive power.

The best way to forestall resistence is repeated education, both for upper managers and, at the branch level, credit managers and loan officers. Front-line personnel must have confidence that scoring works, and they cannot have confidence in a magic box. Confidence requires understanding, and understanding requires training and long-term follow-up. Other keys to the acceptance and correct use of scoring are a knowledgeable champion among upper management and built-in monitoring reports in the MIS that allow loan officers to compare, for their own clients, predicted and observed risk. It may also be necessary to explain to loan officers what factors led to a given application having high or low risk. Finally, the lender should plan to ask for feedback at all levels regularly so that employees feel that their concerns are heard.

3.7 Statistical scoring can reject applications, but it cannot approve or modify them

Unless the microlender has data on all rejected applications, statistical scoring does not apply to the "through-the-door" population of applicants; rather, scoring applies only to applicants who would have been approved under traditional, subjective evaluation standards. This can come as a shock; many lenders expect scoring to substitute for the credit committee and the loan officer's detailed, personal visit to the client's residence and/or place of business. Unfortunately, statistical scoring will not reduce time spent in the evaluation stage.

Why is this? Scoring compares current applications with those past applications that are recorded in the data base. All applications in the data base—unless the lender keyed-in data on rejected applications—were approved, that is, judged creditworthy under traditional, subjective evaluation standards. (Even if rejected applications had been keyed in, the repayment performance of these never-disbursed loans remains unknown.) Applications from drunks, or liars, or new businesses were rejected and never entered the database. Thus, the only loans in the data base are those that have been screened for subjective risk factors. If scoring were applied to unscreened applicants, those applicants—judged only in terms of their quantified characteristics—would look much less risky than they really are. After all, an applicant can have all the objective hallmarks of a good borrower but yet still be a thief.

If a microlender wants to apply scoring to "through-the-door" applicants before they are visited by a loan officer, then it should start to key-in data from all applications, whether rejected or approved. This would eventually allow the construction of a scorecard that would predict not eventual repayment risk but rather the risk that a loan would be rejected after the loan officer's visit.

Scoring simply ignores all factors linked with risk that are not quantified and recorded in the data base. Thus, scoring cannot displace loan officers and the subjective, one-on-one risk evaluation that has been a chief innovation of microfinance.

So how can scoring improve efficiency? First, scoring reduces losses from default. Second—and much more importantly—scoring reduces arrears and thus reduces the time spent by loan officers in collections. In many microlenders, loan officers spend 40 percent or more of their time in collections; if scoring can cut this by 25 percent, then loan officers have another half-day per week to seek more and better customers.

3.8 Statistical scoring does not approve, it can only reject

Abuses of scoring usually take one of two polar forms. In the first, scoring is ignored. In the second, loan officers and credit managers abdicate to scoring their responsibility to screen borrowers. The first abuse is merely a waste of resources; the second can destroy a lender.

As just noted, the data base contains only approved loans. Even though loan officers and credit managers judged all these loans to be of low risk, some nevertheless went bad. The purpose of scoring is to detect these cases that slip through the traditional evaluation technology. Scoring, however, cannot compare a current application with both past approvals and past rejects, so it cannot accurately forecast risk for applications not subjected to the traditional screens.

In this same vein, scoring cannot reject loans; it merely highlights high-risk cases. Loan officers and credit managers must decide how to manage risk, whether by modifying the terms of the loan contract, by making preventive "courtesy" visits even before a loan has problems, or by rejecting the application outright. In fact, the first-round, direct effects of scoring will almost certainly lengthen the evaluation process (although not by much) and reduce the number and value of disbursements. The second-round, indirect effects, however, provide loan officers more time to seek out additional, low-risk clients. Time saved due to the indirect effects, after a few months, will almost certainly overwhelm time lost due to the direct effects.

3.9 Statistical scoring assumes that a large share of risk is linked with quantified characteristics in the data base

Statistical scoring assumes, for example, that risk is linked with gender, age, place of residence, past arrears, type of business, and the terms of the loan contract. This is a safe assumption; the real question is *what share* of risk is linked with those factors that can be included in a scorecard, and what share is linked with subjective or idiosyncratic factors that do not (or cannot) appear in the scorecard.

For credit-card loans to salaried people in high-income countries with comprehensive credit bureaus, a very large share of risk is linked with quantified factors. For microfinance loans to self-employed people in low-income countries without comprehensive credit bureaus, a much smaller share of risk is linked to characteristics recorded in the data base. No one knows exactly how much risk can only be judged by loan officers, but scoring will certainly never replace loan officers in microfinance. At the same time, experience in Colombia and Bolivia (Schreiner, 2001b and 2000a) proves that scoring for microfinance is powerful enough to be worth doing.

3.10 Statistical scoring assumes the future will be like the past

For example, the simplest statistical scorecard might observe that, historically, 10 percent of loans to farmers went bad and that 7 percent of loans to manufacturers went bad. If a farmer applies for a loan today, the scorecard would produce a risk forecast of 10 percent, because historical risk to farmers was 10 percent.

But if the data base includes only loans from non-drought years, and if a drought hits this year, then the risk of farmers would skyrocket. Scoring would blithely continue to predict 10-percent risk even as arrears among farmers doubled or tripled. The point is that intelligent, careful management must still adjust the use of scoring for changes in context, competition, or even in the lender's own policies. Scoring cannot predict something that has not already happened many times and that has not been recorded in the data base.

At the same time, even though things change, scoring usually continues to predict *relative risk* well, even though it may lose power in terms of *absolute risk*. For example, the risk of arrears of 30 days or more for a Bolivian microlender increased by more than 100 percent from 1996 to 1997 due to the unprecedented entrance of Chilean finance companies in the microfinance market (Rhyne, 2001). A scoring model built on data up to the end of 1996 would have completely missed the shift in 1997. Still, borrowers who had lower predicted risk had lower observed risk (Schreiner, 1999). In other words, although scoring did not correctly predict the level of risk, it still successfully distinguished between low-risk and high-risk borrowers. Experience with U.S. credit-card loans through the business cycle suggests that this result is general (Gross and Souleles, 2000; McCorkell, 1999; Lewis, 1990).

3.11 Statistical scoring works in probabilities, not certainties

The output of scoring is a percentage, the predicted risk that a loan will go "bad" before it is repaid. Although predicted risk is always more than zero and always less than one, realized risk is always either zero (did not go bad) or one (did go bad). Thus, for any single loan, scoring never "works" or "hits the target". Scoring can only work on average for a large group of loans. For example, scoring works if, among 1,000 loans, each with predicted risk of 10 percent, average observed risk is close to 10 percent.

Loan officers and credit managers tend to judge scoring on the basis of its performance for single (usually exceptional) loans. They are wont to say, "Scoring does not work; this person had a predicted risk of 60 percent for her eighth loan, but she repaid all seven previous loans without even one day of arrears". Or, "Scoring does not work; this person had a predicted risk of 2 percent but never made a single payment and is now a year overdue." Scoring always misses some share of individual cases; after all, even if scoring works, half the loans with a predicted risk of 50 percent turn out good. Scoring works on average, and this can be hard for some employees to grasp and accept. Skeptics will point out that with a super-risky threshold of 50 percent, about half of auto-rejects will have turned out good. Of course, this is true also for the current subjective, implicit scoring; some rejected borrowers under the current technology (probably much more than 50 percent) would have been good, had they received a loan. Statistical scoring is more susceptible to this critique only because it makes the trade-off explicit (and thus susceptible to management). Subjective scoring also rejects applicants who would have been good, but the extent of the loss is not measurable.

3.12 Statistical scoring is susceptible to abuse

Scoring provides management with a predicted risk; it does not tell management what to do with that information. The most common abuse of scoring is neglect; managers ignore risk forecast and continue to do what they have done all along. The cure for neglect is education and follow-up.

Another common abuse is excessive *overrides*, when managers make exceptions to scoring policy. For example, managers might override a super-risky threshold of 50 percent by approving an application with a 60-percent predicted risk. Of course, managers sometimes know something that the scorecard does not (in particular, they often know that someone with low-risk quantitative characteristics nevertheless has high-risk qualitative characteristics), so some overrides are to be expected and accepted. Not all super-riskies, however, can be overrides, and, in any case, managers must track overrides and then compare their performance with predicted performance to see whether managers or scoring was closer to the truth. Overrides in the first few months of implementation may be very common, but they should dwindle as employees see that scoring works. In U.S. credit-card lending, the goal is to override less than 10 percent of super-riskies.

Finally, if loan officers know the weights of characteristics in the scorecard, then they can abuse scoring by cooking the data. For example, a loan officer, after a field visit, might believe that a carpenter qualifies for a loan. The loan officer knows, however, that the scorecard counts carpenters as more risky than shoemakers. The loan officer does not want to be contradicted and embarrassed by the risk forecast, so he falsely records that the applicant is a shoemaker. This changes the predicted risk, but of course true risk remains that of a carpenter, so, in the long term, loan officers who cook data pay the price in arrears. At some point, the lender will have to make a policy choice about how much of the scorecard to reveal to loan officers. Better knowledge helps loan officers to understand scoring and thus to accept it and to use it properly, but greater knowledge also allows greater abuse. The ideal is to train loan officers well enough that they can be trusted to know what is going on.

Of course, traditional subjective scoring is also subject to all of these abuses. But the explicitness of scoring—and the process of change from traditional evaluation to traditional evaluation plus scoring—makes the dangers more obvious. Furthermore, skeptics and change-blockers seize on any apparent weakness.

3.13 Statistical scoring may use illegal or immoral predictors

Among the most predictive characteristics for scoring in microfinance are gender, marital status, age, place of residence, and ethnicity (Schreiner, 2000b). Borrowers, however, do not choose their gender, age, or ethnicity, and many may not have had much choice in their marital status or place of residence. Throughout the world, these ascribed characteristics have been (and still are) used by lenders to oppress certain groups without any relation to associations with risk. Yet, these simple, easy-to-observe and hard-to-conceal factors are strongly associated with a plethora of other characteristics—whether chosen by the individual or imposed on the individual by society—that are in turn strongly associated with repayment risk.

Should a microlender be expected to ignore this? In principle, the solution is to measure and record all the characteristics that are correlated with the ascribed characteristics, so that the use of the illegal or immoral predictors as proxies does not add any predictive power. In practice, measuring all these factors is prohibitively costly. In the United States, society has decided that lenders cannot use these characteristics to predict risk and thus that lenders must bear some of the costs of redressing past and current injustice. Even if a microlender includes these characteristics in a scorecard, their use would not necessarily prejudice the lender's judgement; a disciplined lender with a strong sense of its social mission might sensibly use the information to make more careful and explicit trade-offs between social mission and profits.

4. Characteristics of borrowers, loans, and lenders

Predictive power increases with the number of characteristics available. Of course, there are diminishing returns to more data, and, furthermore, the marginal cost of collecting additional characteristics may be extremely high, especially if they are not already collected for some other purpose.

The following lists of characteristics are based on experience with scoring for microfinance, and all are predictive of repayment risk to some degree.

4.1 Characteristics of borrowers

- Past arrears
  - Days in arrears per installment
  - Longest spell of arrears in days
  - Number of spells of arrears
  - Subjective grade
- Experience as a borrower
  - Number of previous loans
  - Months since the first disbursement
- Borrower demographics
  - Gender
  - Age
  - Marital status
  - Number of household members
  - Presence of telephone in the house
  - Education
- Business characteristics
  - Type of business
  - Years in current business
  - Years of experience in current type of activity
  - Presence of a telephone in the business
  - Number of employees

- Business and household financial data
  — Monthly sales
  — Monthly expenses
  — Other business income
  — Other business expenses
  — Household (non-business) income
  — Household (non-business) expenses
  — Monthly free cash flow
  — Cash on hand and banks
  — Inventory
  — Fixed assets
  — Other business assets (accounts receivable)
  — Accounts payable
  — Debts
  — Other liabilities
- Other characteristics
  — Home tenure (owner, renter, other)
  — Length of time at current address
  — Presence of salaried job
  — Length of time at salaried job
  — Whether or not the business is run from the home

4.2 Characteristics of loans

- Type of loan
- Date of application
- Date of disbursement
- Date scheduled to be paid-off
- Disbursement amount requested and granted
- Installment amount requested and granted
- Frequency of installments
- Term to maturity requested and granted
- Presence, type, and value of guarantee

4.3 Characteristics of lenders

- Branch
- Loan officer
  — Number of disbursements of experience
  — Gender
  — Age
  — Marital status

5. Sample plan for a scoring project

A typical scoring project has six stages:

- Introduction
- Scorecard construction, reporting, and writing of a “Technical Guide”
- Integration in the MIS and writing of a “Scoring Policy Manual”
- First-round branch training and roll-out in pilot branches
- Second-round branch training, expansion to all branches, and MIS adjustments
- Long-term monitoring and follow-up.

To repeat a central point, the greatest challenge in scoring is not technical but human. People must understand scoring so that they can trust it and then use it appropriately. This requires repeated training and careful follow-up. The statistics and the MIS work are relatively easy, at least in the sense that success can be confirmed before moving to the next stage. Management of the use of scoring is more difficult.

5.1 Introduction to scoring

The project starts with an on-site presentation to introduce the basic concepts of scoring to upper management. This helps to secure the organizational support required in the next stages and begins the educational process needed throughout the organization to ensure proper use of scoring after roll-out.

Following the presentation, the lender defines the risk it wants to predict by answering the question “What is a ‘bad’ loan?” In this process, the lender asks itself:

- When are arrears so costly that the loan is no longer profitable?
- When are arrears so costly that a request for a repeat loan would be refused?
- When are arrears so costly that the lender would like to have rejected the loan?

By these criteria, a loan turns “bad” long before it is charged off.

The introductory visit will also entail meetings with the local systems analyst to coordinate plans for the integration of scoring in the MIS, to learn the structure and content of the data base, and to obtaian a electronic copy of the data base.

When the introductory stage ends, upper management understands scoring and the next stages in the project, the consultant understands the structure and content of the lender’s data base, and the lender has given the consultant a copy of the data base.

5.2 Scorecard construction, reporting, and the “Technical Guide”

The process of scorecard construction entails creating a single master record for each loan that contains fields for all the characteristics of the borrower, loan, and lender. Many characteristics are transformations of the “basic characteristics” that appear directly in the data set and that are listed in the previous section. The consultant constructs and tests various scorecards and consults with the microlender on questions as they arise during data analysis.

After settling on a final scorecard, the consultant writes a report for upper management that covers the basics of scoring and the lender’s specific results. In particular, the report discusses how each characteristic in the scorecard is linked with risk, and it describes the results of the historical test. This report is management’s reference guide to scoring and to their scorecard.

The consultant also writes a “Technical Guide” for the systems analyst and the local programmer. It specifies how to integrate the scorecard and associated reports in the lender’s MIS. The document also describes the logic of the transformation of characteristics as they appear in the lender’s data base to the characteristics as they appear in the scorecard. (For example, the MIS will compute loan-officer experience in terms of number of previous disbursements; this characteristic *per se* does not appear in the lender’s data base, although the information to derive it does.) The consultant also outlines the design of screens and reports and, in particular, how to make them user-friendly. Employees are more likely to use scoring appropriately if predicted and observed risk appear in formats with which they are already comfortable.

This stage ends with the microlender in possession (on paper) of its newly tested scorecard and with the local programmer in possession of the knowledge required to implement the scorecard in the MIS. Scorecard construction is not simple, but once it is complete, it can be tested and its predictive power established.

5.3 Integration of scoring in MIS and “Scoring Policy Manual”

As the local programmer integrates the scorecard and associated reports in the MIS, management writes—with the assistance of the consultant—a “Scoring Policy Manual”. The manual sets the risk levels that correspond to the various policy ranges (super-safe, regular, risky, and super-risky) and the prescribed responses. Management sets the policy ranges using knowledge of trade-offs derived from the historical test. A written, explicit policy statement is essential because without it, loan officers and branch managers can justify ignoring forecast risk by claiming (correctly) they did not know the appropriate response. Just as lenders have a written credit policy, they should have a written scoring policy. The consultant helps management decide what ranges to set, and the consultant also provides guidelines for document content.

As the local programmer integrates the scorecard, the consultant answers questions by telephone and e-mail. The consultant also compares, characteristic-by-characteristic, the derived characteristics generated by the scorecard integrated in the MIS with the derived characteristics expected by the consultant. Likewise, the consultant also compares the risk computed by the lender for each loan with the risk expected to be computed. This check-and-correct process iterates until the scorecard in the MIS is error-free.

Integration lasts from 3 to 5 months, depending on the programmer and MIS complexity. Again, this programming should be tasked to a full-time programmer with no responsibilities of than the integration of the scorecard.

This stage ends with scoring integrated in the MIS and upper management—now with a stronger and more concrete understanding of scoring based on the historical test—ready to have scoring put to a live test in pilot branches.

5.4 First-round branch training and roll-out in pilots

Just as the first on-site presentation introduced scoring to upper management, this stage introduces scoring to branch managers and to loan officers in all the branches. The presentations—held in the branches—cover the concepts of scoring and the specific results from the lender's newly constructed scorecard and the historical test. They also present examples of the screens and reports that deal with scoring from the newly integrated scoring module in the lender's own MIS. The purpose is to begin the process of understanding how scoring works and how it might benefit those employees who must understand scoring and accept it on the front lines.

The consultant also makes a similar presentation to upper management. By now, managers have had more time to think about scoring, and the concrete example afforded by the newly constructed scorecard will elicit questions and reveal areas of weaker understanding.

The microlender also chooses 2-3 pilot branches where scoring will start to be used immediately. The advantage of pilots is that any snags that might affect the confidence of users are discovered and corrected before expansion to all branches.

The head office receives weekly written reports from the branches listing risk forecasts for new applicants and the disposition of the cases. The goal is first to ensure that the pilot branches do not misuse scoring and second to ensure that they do use it.

During the first few months of implementation, the microlender and the consultant remain in close contact. In particular, the microlender should run the "Scoring Follow-Up Report" weekly in each branch and share a copy with the consultant. Regular monitoring helps to guard against disaster.

This stage ends with all front-line users having been introduced to scoring and with pilots well underway in 2-3 branches.

5.5 Second-round branch training, adjustments, and massification

After 2-4 months of experience in the pilot branches, the consultant returns for a second round of training in all branches, starting in the pilots. Now, the loan officers and branch managers are full of questions, things that they did not know to ask (or did not realize that they did not understand) in the first round. These first training sessions in the pilot branches explicitly serve as feedback sessions where front-line workers can make suggestions about how to improve reports and other elements of work-flow.

Users from the pilot branches then accompany the consultant for training in the other branches, explaining in their own words their experiences with scoring. The training will cover many of the same concepts and results as did the first round, but it will have new meaning and urgency for front-line employees, as expansion of scoring to all branches in imminent. Furthermore, the experience of the users in the pilots will have particular credence among fellow front-line workers.

The consultant also works with upper management to fine-tune thresholds and other aspects of scoring policy. In addition, the consultant works with the local programmer to specify adjustments in the reports and other MIS issues that arise out of the feedback from the front lines.

When this stage ends, front-line workers understand how scoring works and have witnessed repeated demonstrations of predictive power. Management has revised the "Scoring Policy Manuel" to take into account the pilot experience.

5.6 Long-term follow-up

Nothing remains but long-term monitoring and follow-up. Unless the country experiences large macroeconomic shifts, and unless the microlender suddenly changes its policies or target groups, and unless competition enters in force, a microfinance scorecard probably will maintain predictive power near its original level for 3 to 5 years. When the scorecard loses enough power (because the near future is less like the distant past than like the recent past), then it can be renovated with an updated data base. The microlender may also want to look into scorecards for other types of risk (such as the risk of desertion). Furthermore, once scoring is successfully up and running, the microlender may find fresh motivation to look into improving data quality and/or gathering additional characteristics to record in the data base.

References

Gross, David B.; and Nicholas S. Souleles. (2000) "Consumer Response to Changes in Credit Supply: Evidence from Credit-Card Data", manuscript, University of Chicago and University of Pennsylvania, souleles@wharton.upenn.edu.

Lewis, Edward M. (1990) *An Introduction to Credit Scoring*, San Rafael: Athena Press, LIC# 90-92258.

McCorkell, Peter. (1999) "Credit Scoring 101", Presentation at a Federal Trade Commission Public Forum on "The Consumer and Credit Scoring", www.ftc.gov/bcp/creditscoring.

Mester, Loretta J. (1997) "What's the Point of Credit Scoring?", *Business Review*, Sept./Oct., pp. 3-16, Federal Reserve Bank of Philadelphia.

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. (2001a) "Scoring at Prizma: How to Prepare", report to Prizma and CGAP, Mostar, Bosnia-Herzegovina, sean@prizma.ba.

_____. (2001b) *Un Sistema de Scoring del Riesgo de Créditos de FIE en Bolivia*, report to Fomento de Iniciativas Económicas, La Paz, gsantacruz@usaid.gov.

_____. (2000a) *A Scoring Model of the Risk of Costly Arrears for Loans from Affiliates of Women's World Banking in Colombia*, report to Women's World Banking, hdellien@swwb.org.

_____. (2000b) "Credit Scoring for Microfinance: Can It Work?", *Journal of Microfinance*, Vol. 2, No. 2, pp. 105-118.

_____. (1999) "A Scoring Model of the Risk of Arrears at a Microfinance Lender in Bolivia", manuscript, Center for Social Development, Washington University in St. Louis, www.microfinance.com.