OccasionalPaper

NO. 7 JANUARY 2003

SCORING: THE NEXT BREAKTHROUGH IN MICROCREDIT?

BY MARK SCHREINER

CGAP

For comments, contributions, or requests for other notes in this series, please contact CGAP

1818 H Street, NW
Washington DC 20433

Tel: 202.473.9594
Fax: 202.522.3744

E-mail:
cgap@worldbank.org

Web:
www.cgap.org

This paper is aimed at microcredit managers who want an initial technical introduction to how scoring works, what it can and cannot do, and how to prepare for implementation. The first sections are more general; later sections are more technical. Readers who want less detail can browse the main text and boxes.

The paper is not a "how-to" manual. The design and implementation of a scoring project require highly specialized expertise and are in general too complex and institution-specific to be explained in a document of this length.

The discussion here is based on some of the first experiments in scoring for microcredit.[1] In places it also draws on the long experience with scoring in high-income countries.[2] The examples used in the paper reflect the author's experience with scoring in Latin America. In turn this reflects the existence of large, sophisticated Latin microlenders with adequate electronic databases. With sufficiently strong organizations and sufficiently large databases, scoring would be just as relevant in Africa, Asia, and Central Europe. In the long term, scoring will spread around the world, although certainly not to every microlender. This paper should help managers judge the likely usefulness of scoring in their own organizations.

Summary

The challenge of microcredit is to judge the risk of whether the self-employed poor will repay their debts as promised. Is scoring—a new way to judge risk—the next breakthrough in microcredit? Scoring does reduce arrears and so reduces time spent on collections; this greater efficiency improves both outreach and sustainability. Scoring, however, is not for most microlenders. It works best for those with a solid individual lending technology and a large database of historical loans. Even when scoring works, it is only a marked improvement, not a breakthrough. In particular, scoring will not replace loan officers in microcredit because much of the risk of the self-employed poor is unrelated to the information available for use in scoring. This paper discusses how scoring works, what microlenders can expect from it, how to use it, and what data is required. Success comes not from technical wizardry but rather from painstaking training of users: loan officers and branch managers will trust scoring to help them make choices only if they understand how it works and only if they see it work in tests. Most importantly scoring changes how microlenders

Dr. Mark Schreiner is a consultant with Microfinance Risk Management, a private U.S. firm that has provided scoring services to microfinance organizations in Bolivia, Bosnia and Herzegovina, Colombia, the Dominican Republic, and Uganda. He is also a Senior Scholar in the Center for Social Development at Washington University in St. Louis. His work helps the poor build assets through greater access to financial services.

think, fostering a culture of analysis in which managers regularly seek to mine their databases for information that addresses business questions.

I. Introduction

Microcredit grew out of two new ways to judge the repayment risk of the self-employed poor: joint-liability groups and loan officers who make detailed personal and financial evaluations of individual borrowers and their homes, businesses, and collateral. Scoring is another new (to microcredit) way to judge repayment risk. It detects historical links between repayment performance and the quantified characteristics of loan applications, assumes those links will persist over time, and then—based on the characteristics of current applications—forecasts future repayment risk. In high-income countries, scoring (through credit cards) has been the biggest breakthrough ever in terms of providing millions of people of modest means with access to small, short, unsecured, low-transaction-cost loans. Is scoring the next breakthrough in microcredit?

For the few microlenders who already are large, are well run, and possess adequate electronic databases, scoring can improve efficiency, expand outreach to the poor, and improve organizational sustainability. Scoring primarily reduces time spent collecting overdue payments from delinquent borrowers. (A typical loan officer might save a half-day per week.) Loan officers can use this time to search for more good borrowers, improving both depth and breadth of outreach.

For large microlenders, scoring can also be profitable. For example, one test with historical data in Bolivia suggested that rejecting the riskiest 12 percent of loans disbursed in the year 2000 would have reduced the number of loans 30 days overdue by 28 percent.[3] Given conservative assumptions about the cost of the scoring project, the net benefit of rejecting loans that would have gone bad, and the net cost of mistakenly rejecting loans that would have been good, scoring would have paid for itself in about one year. It would also have had a net present value of about $1 million.

Scoring is a marked improvement, but it is not a breakthrough on the scale of joint-liability groups and individual evaluations by loan officers. In fact scoring probably will not work for most group lenders or village banks. Furthermore, most microlenders that make loans to individuals are not ready for scoring, either because they must first perfect more basic processes or because their databases are not yet adequate for scoring. Even for microlenders that are ready, scoring will not replace loan officers and their subjective evaluation of risk factors that are not (or cannot be) quantified in a database. Scoring is not the next breakthrough in microcredit, but it is one of a few new ideas (such as tailoring products to demand, offering deposit and payment services, paying attention to governance and incentives, and improving business organization) that promise smaller—but still important—improvements in microcredit for a long time to come.

The central challenge of scoring is organizational change—after all scoring's predictive power can be tested with historical data before it is put to use. Loan officers and branch managers sensibly distrust magic boxes. Before they trust scoring, they need to understand how scoring works in principle and then see it work in practice with their own clients. Understanding and acceptance requires repeated training, careful follow-up, and constant demonstrations of predictive power with currently outstanding loans. In the long term, a good scoring project should change an organization's culture, shifting it toward explicit analysis by managers (with the help of full-time, in-house analysts) of the untapped knowledge in their databases to inform business questions.

Using examples from actual scoring projects, this paper explains how scoring works in principle and in practice. It describes different types of scorecards and—more importantly—tells how to test scorecards before use, how to use them in the field, and how to track their performance. Along the way, the paper discusses strengths and weaknesses of scoring and dispels several myths, in particular the myths that scoring will replace loan officers and will speed the evaluation of loan applications. To help microlenders prepare and take full advantage of scoring, the last section discusses the nuts-and-bolts requirements for the design of data collection.

II. Subjective Scoring versus Statistical Scoring

Microlenders already use subjective scoring, but not statistical scoring. This section presents the basic concepts of scoring—whether subjective or statistical—and tells why the two approaches are complementary. Any technique that forecasts future risk from current characteristics using knowledge of past links between risk and characteristics is scoring. Two approaches to linking characteristics to risk are subjective scoring and statistical scoring. Figure 1 lays out a general comparison of the two.

Figure 1: Comparison of Subjective Scoring and Statistical Scoring

Dimension	Subjective Scoring	Statistical Scoring
Source of knowledge	Experience of loan officer and organization	Quantified portfolio history in database
Consistency of process	Varies by loan officer and day-to-day	Identical loans scored identically
Explicitness of process	Evaluation guidelines in office; sixth sense/gut feeling by loan officers in field	Mathematical rules or formulae relate quantified characteristics to risk
Process and product	Qualitative classification as loan officer gets to know each client as an individual	Quantitative probability as scorecard relates quantitative characteristics to risk
Ease of acceptance	Already used, known to work well; MIS and evaluation process already in place	Cultural change, not yet known to work well; changes MIS and evaluation process
Process of implementation	Lengthy training and apprenticeships for loan officers	Lengthy training and follow-up for all stakeholders
Vulnerability to abuse	Personal prejudices, daily moods, or simple human mistakes	Cooked data, not used, underused, or overused
Flexibility	Wide application, as adjusted by intelligent managers	Single-application; forecasting new type of risk in new context requires new scorecard
Knowledge of trade-offs and "what would have happened"	Based on experience or assumed	Derived from tests with repaid loans used to construct scorecard

Box 1: Scoring, Group Loans, and Village Banks

Because of data issues and the nature of group lending, statistical scoring probably will not work well for joint-liability groups or village banks. A fundamental data issue is that most group lenders do not accept partial payments: either everyone in the group pays on time, or no one does. This is a sensible policy, but it means that the database does not record whether individuals in the group were willing and able to make their payments on time. There is no data on individual risk. In this case, scoring can predict the risk of the group, but not the risk of an individual in the group. Unfortunately, group risk is much less strongly linked to group characteristics (such as whether the members are the same gender, or their average age) than individual risk is linked to individual characteristics.

Even if a lender does accept individual payments, the essence of joint liability is that the individual risk of group members is largely decoupled from individual characteristics. The group can increase an individual's willingness to pay (through peer pressure and social sanctions), and the group can increase an individual's ability to pay (through peer mentoring and informal insurance). On the other hand, the group—through "domino default"—can destroy an individual's willingness to pay. Thus, regardless of an individual's characteristics, repayment risk depends in large part on interactions among group members, and the outcome of these interactions is not likely to be well proxied by quantified characteristics.

In summary, quantified characteristics are less indicative of risk for groups than for individuals. This is not bad; it is the purpose of the group. It does, however, make scoring more difficult and less powerful for lenders to groups or for village banks.

Subjective Scoring

Microlenders currently judge risk with subjective scoring, forecasting repayment based on their quantified knowledge (measured in numbers and recorded in their electronic database) and their qualitative knowledge (not measured in numbers and/or not recorded in their electronic database) of the characteristics of the client and the loan contract. The loan officer and credit manager—as well as the microlender as an organization—share their experience through written policy, training, and simple word-of-mouth.

While subjective scoring does use quantitative guidelines—for example, disqualifying anyone with less than a year in business—it focuses on the loan officer's sense of the personal character of the client. Based mostly on intuition, subjective scoring produces a qualitative judgment of "not very risky, disburse" versus "too risky, reject."

Subjective scoring works, as the history of microcredit demonstrates. But is there room for improvement? For example, loan officers must spend time absorbing the lessons of the organization's experience and developing a sixth sense for risk. Also the predictive accuracy of subjective scoring can vary by officer and by a loan officer's mood on a given day. Subjective judgment also allows for discrimination or mistakenly focusing on too few (or the wrong) characteristics.

Statistical Scoring

Statistical scoring forecasts risk based on quantified characteristics recorded in a database. Links between risk and characteristics are expressed as sets of rules or mathematical formulae that forecast risk explicitly as a probability. For example, a 25-year-old male carpenter applying for his first loan might have a 20 percent predicted risk of having arrears of 30 days, whereas a 50-year-old female seamstress, who had no late payments in three previous loans, might have a predicted risk of 5 percent. Finance is risk management, and statistical scoring facilitates risk management by making risk evaluation consistent and explicit. The predictive accuracy of statistical scoring can be tested before use.

Scoring's weakness is its newness; only a handful of microlenders have tried it. The use of quantitative knowledge in a database to help judge risk runs counter to the two breakthrough innovations (joint-liability groups and one-on-one relationships with loan officers) that define microcredit, both of which take advantage of people's subjective knowledge of creditworthiness. To adopt something so different as statistical scoring requires a long period of training and adjustment, as well as constant demonstrations of predictive power. Even after microlenders accept scoring, they must guard against depending on it too much. Unfortunately statistical scoring is probably more relevant for individual loans than for group loans or village banks, as Box 1 explains.

Scoring for microcredit also has limited application because it requires an electronic database that records repayment behavior for a large number of past loans as well as characteristics of the client and the loan contract. Furthermore, the data must be reasonably accurate. Some microlenders have accumulated adequate data in the course of their normal portfolio management. Many others, however, do not have electronic databases, do not record enough information on each loan, or do not record accurate data. One aim of this paper is to help managers think about how to redesign their information systems so that in the future their databases will be adequate to support scoring.

Subjective Scoring and Statistical Scoring are Complements

Statistical scoring ignores everything but quantified characteristics, while subjective scoring focuses mostly on qualitative characteristics. Which approach is best? In microcredit both have a place because they complement each other. Subjective scoring can consider what statistical scoring ignores, and statistical scoring can consider relationships too numerous, too complex, or too subtle for subjective scoring. Both approaches to scoring assume that the future will be like the past and that characteristics are linked with risk. These assumptions, of course, are never completely true, but they come close enough to make scoring worthwhile.

Scoring—be it statistical or subjective—presumes that some knowledge of the past is better than none. Subjective scoring—because it relies on experienced people who can spot patterns and combine knowledge from many sources—can respond quickly and flexibly when trends break with the past. Statistical scoring is more consistent and picks up more (and subtler) trends, but it can only forecast what has already happened many times.

Some risk is undoubtedly linked with quantified characteristics, such as indebtedness and previous arrears. Not all characteristics are quantifiable, however, and even quantifiable characteristics are not always quantified. Most relevant for microcredit, some (unknown) share of risk depends on personal character that can be judged only by getting to know the client. What share of risk is linked with quantified characteristics? This paper, buttressed by the tests in Sections III and IV, argues that the share is large enough to make statistical scoring worthwhile. The tests in Sections III and IV also show that the share is too small to discard subjective scoring.

Some risk is linked with quantified characteristics best considered by statistical scoring; some risk is linked with qualitative characteristics best considered by subjective scoring. In microcredit the qualitative share is too large for statistical scoring to replace loan officers and their subjective scoring. Likewise, statistical scoring will not relieve credit managers of the responsibility for credit decisions. For example, it cannot detect whether borrowers know their business or whether they will squander the loan

Figure 2: Four-Leaf Tree, 1992–99 Data (Tree Form)

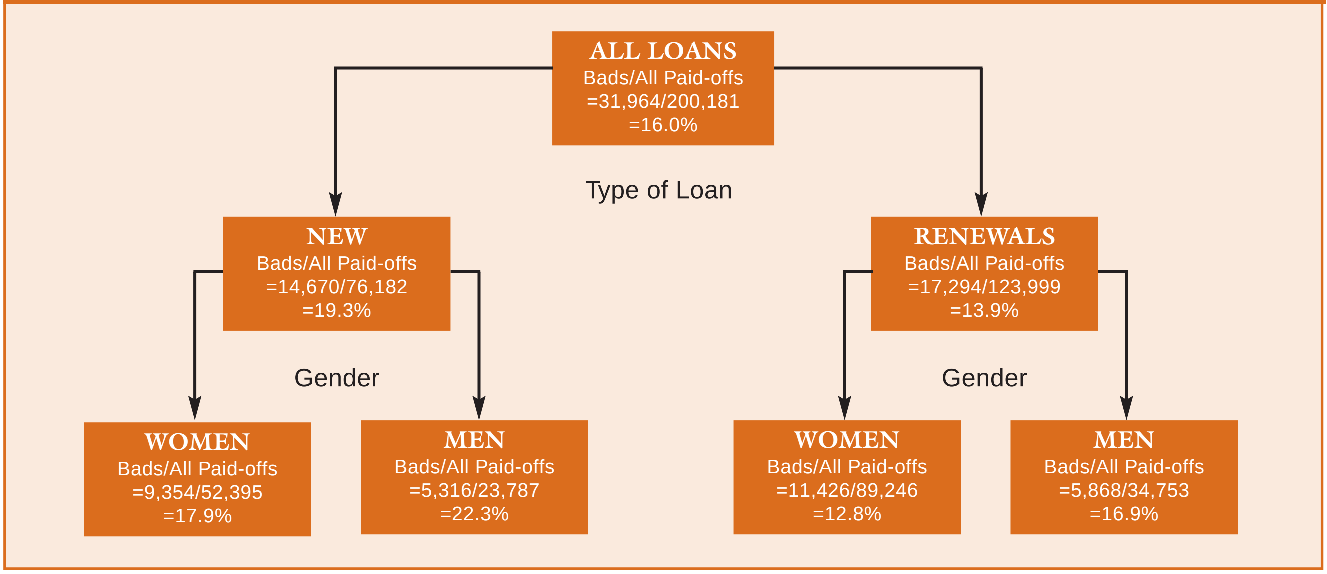


Figure 3: Four-Leaf Tree with Historical Risk, 1992–99 Data (Table Form)

	Branch of Tree		Construction Sample, 1992–99				
Leaf	First	Second	Bads	Goods	Total Cases in Leaf	% Bad	% of All Cases in Leaf
1	New	Woman	9,354	43,041	52,395	17.9	26.2
2	New	Man	5,316	18,471	23,787	22.3	11.9
3	Renewal	Woman	11,426	77,820	89,246	12.8	44.6
4	Renewal	Man	5,868	28,885	34,753	16.9	17.4
All Loans			**31,964**	**168,217**	**200,181**	**16.0**	**100.0**

Source: Latin American microlender

proceeds. Statistical scoring is nothing more than a third voice in the credit committee to remind the credit manager and the loan officer of elements of risk that they might have overlooked.

III. How Scorecards Work and How to Test Them

A scorecard specifies the expected links between future risk and the current characteristics of the borrower, the loan, and the lender. Whereas subjective scorecards combine explicit credit-evaluation guidelines with implicit judgments made by loan officers, statistical scorecards are explicit sets of rules or mathematical formulae. This section presents an example tree, the simplest type of statistical scorecard, and shows how to test scorecards before they are used.

A Four-Leaf Tree

The four-leaf tree scorecard in Figures 2 and 3 was constructed using data on paid-off loans at a large microlender in Latin America. The lender defines as "bad" all loans with at least one spell of arrears of 30 days, or with an average of at least seven days of arrears per installment.[4]

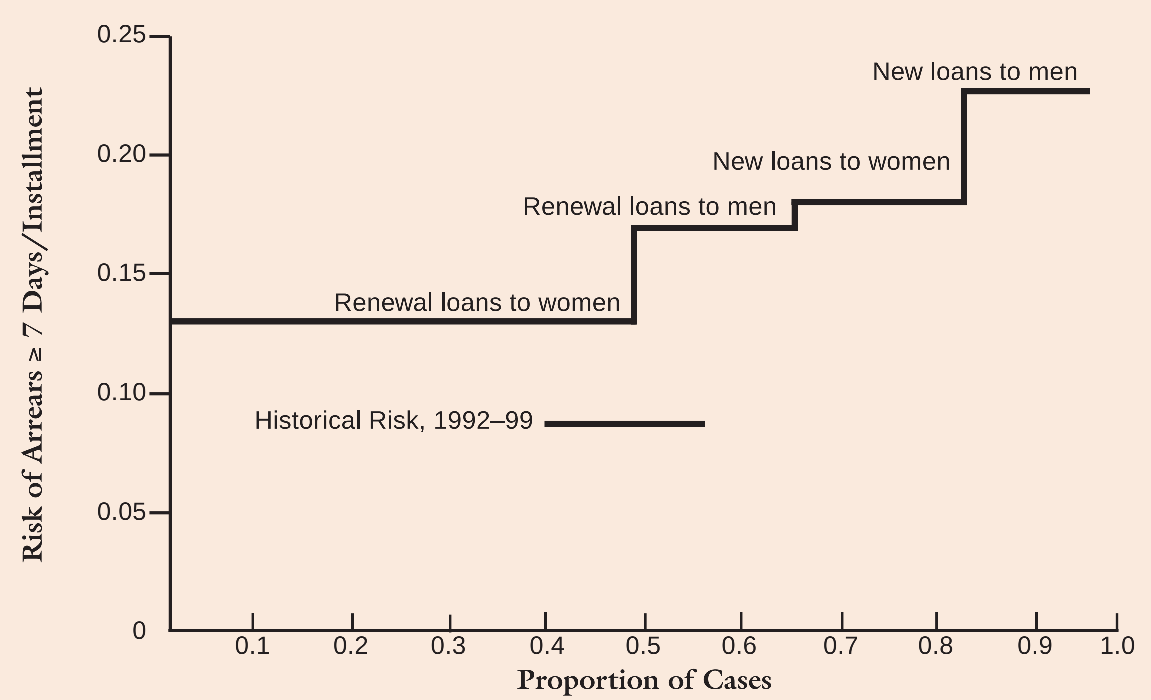
The tree root at the top of Figure 2 shows that 31,964 of 200,181 loans paid off in 1992–99 were "bad." Historical risk was thus 16 percent, the number of bad loans divided by the number of all loans. Tree branches below the root in Figure 2 split "paid-off loans" (which include both paid-off loans and written-off loans) into four leaves according to the type of loan (new or renewal) and then according to the gender of the applicant (woman or man). For new loans to women (lower left leaf), historical risk was 17.9 percent—9,354 bad loans divided by 52,395 total loans. For new loans to men, historical risk was 22.3 percent—5,316 bad loans divided by 23,787 total loans. For renewal loans to women, historical risk was 12.8 percent, and for renewal loans to men, historical risk was 16.9 percent.

Figure 4 depicts the same tree as Figures 2 and 3. The four segments represent the four leaves. The segments are ordered from least risk (left) to most risk (right). Their height depicts their historical risk, and the length of each segment depicts the share of the leaf among all paid-off loans. For example, renewal loans to women account for 89,246 divided by 200,181 = 44.6 percent of paid-off loans (see Figure 3, Leaf 3, right-most column).

This simple four-leaf tree offers several insights for this microlender:

- For a given gender, new loans had more risk than renewals.
- For new loans and renewals, loans to men had more risk than loans to women.
- The least risky segment (repeat loans to women) had about half as much risk as the most risky segment (new loans to men).
- The largest segment (repeat loans to women, with almost half of all loans) had the least risk.
- The smallest segment (new loans to men, with about 12 percent of all loans) had the most risk.

Figure 4: Four-Leaf Tree with Historical Risk, 1992–99 Data (Graph Form)



How might the microlender act on these insights? Because new loans—especially to men—are risky, the lender might want to screen applications in this segment with extra care. The lender might also reduce the analysis required of loan officers, or the requirements for clients, for applicants in the low-risk segments. Scoring only predicts risk; it does not tell the lender how to manage it.

The results from this simple four-leaf tree are not too surprising. Most microlenders probably know that new loans are riskier than repeat loans and that men have higher risk than women. Some might be surprised, however, to discover that new loans to men are almost twice as risky as repeat loans to women. This simple example merely illustrates the concepts of scoring rather than providing deep insights into previously unknown links between characteristics and repayment risk.

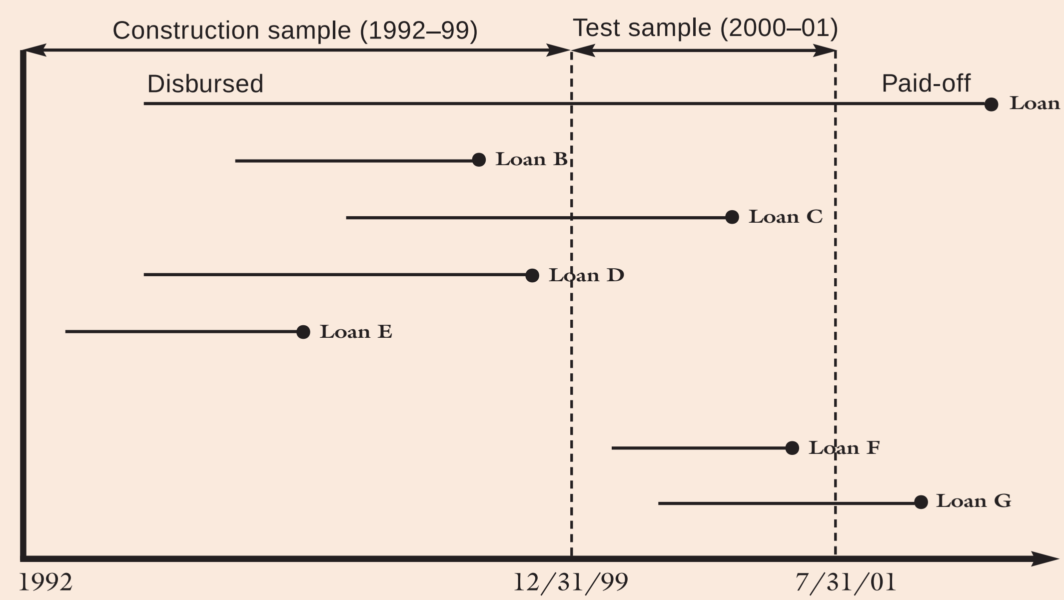
How Does a Tree Forecast Risk?

Scoring assumes that past relationships between risk and characteristics will still hold in the future. Thus, historical risk in a segment becomes predicted risk for the segment. Suppose, for example, that the microlender with the four-leaf tree in Figure 4 receives a renewal application from a woman and, after a traditional credit evaluation process, provisionally approves it. Historical risk for renewal loans to women is 12.8 percent, so the risk forecast derived from the tree scorecard is 12.8 percent. An application for a new loan from a man—if provisionally approved by the lender's traditional norms—would have a risk forecast of 22.3 percent, the historical risk of that segment.

Scoring makes forecasts—whether by means of trees or more complex scorecards—by assuming that the future risk of an application with given characteristics will be the same as the historical risk of applications with the same characteristics. Subjective scoring also does this, but it measures historical relationships qualitatively and implicitly rather than quantitatively and explicitly.

Any scorecard can forecast risk, but not all do it well. Fortunately predictive power can be tested

Figure 5: Construction Sample and Test Sample



- Construction sample: Loans B, D, and E
- Test sample: Loans C and F
- Outstanding loans as of 7/31/01: Loans A and G

Figure 6: Four-Leaf Tree with Realized Risk, 2000–01

	Branch of Tree							
Leaf	First	Second	Bads	Goods	Total Cases in Leaf	Predicted % Bad	Realized % Bad	% of All Cases in Leaf
1	New	Woman	5,740	26,589	32,329	17.9	17.8	23.9
2	New	Man	3,281	11,674	14,955	22.3	21.9	11.1
3	Renewal	Woman	7,752	56,575	63,327	12.8	12.1	47.6
4	Renewal	Man	3,770	19,627	23,397	16.9	16.1	17.3
All Loans			**20,543**	**114,465**	**135,008**	**16.0**	**15.2**	**100.0**

Source: Latin American microlender

before use. Historical tests reveal how well the scorecard would have performed had it been used in the past. The assumption is that scoring will have similar predictive power from now on.

Suppose someone who plays the stock market or the horses concocts a new system to beat the market or the track. Before staking their own cash, they would be foolish not to test the new system with historical data to see how it would have worked in past years. Likewise, microlenders should test their scorecards before use. This prevents disasters and helps convince skeptical personnel that scoring really works.

The historical test uses the scorecard to predict risk for loans already paid off (or written off), based on the characteristics known for those loans at disbursement. The test then compares predicted risk with realized risk, that is, whether the loan (after disbursement) turned out good or bad. Historical tests are a central feature of scoring; no lender should score without first testing predictive power.

Historical Tests

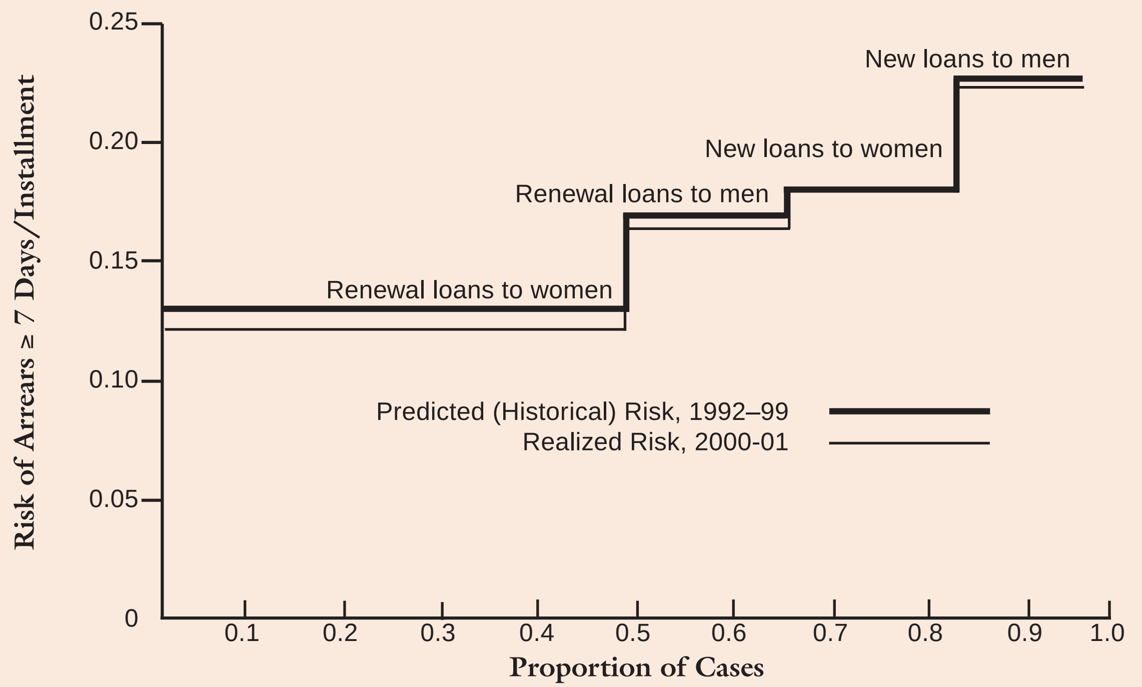
Historical tests have three steps: deriving a scorecard from loans in the construction sample, using the scorecard to forecast risk for loans in the test sample, and comparing predicted (historical) risk with realized risk.

A historical test divides paid-off loans (including written-off loans) into two samples. Loans that were paid off by a specified deadline in the past make up the construction sample used to build the scorecard. In Figure 5, loans B, D, and E were paid off before the deadline and so fall into the construction sample.

Loans paid off after the deadline, but before the last date in the database, make up the test sample used to test the predictive power of the scorecard. In Figure 5, the test sample is loans C and F because they were paid off after the construction deadline but before the database cut-off. Loans outstanding as of the database cut-off—loans A and G in Figure 5—are omitted from both the test sample and the construction sample because their good/bad status is still unknown.

To mimic real-life scoring, the test should follow three principles. First, a given loan may be used in either construction or testing, but not both. Using the same loan in both stages overstates predictive power. The construction stage tailors the scorecard to fit apparent patterns of association between characteristics and risk in the construction sample. Some of these patterns, however, change over time, or are not real patterns at all but the results of chance in a finite sample. These patterns are absent in loans outside the construction sample. Thus the scorecard predicts more accurately for loans in the construction sample than for other loans. In real life, what

Figure 7: Test of Four-Leaf Tree, Comparing Predicted Risk with Realized Risk, 2000-01



matters is prediction for loans not in the construction sample.

Second, test loans must be paid off after construction loans. An actual scorecard forecasts risk for loans paid off after the cut-off date for loans in the construction sample, and the test should mimic this situation.

Third, the test must base forecasts only on characteristics known at disbursement. Any information acquired after disbursement must be ignored because real-life forecasts cannot take advantage of this data.

In the four-leaf tree in Figure 3, the construction sample is the 200,181 loans paid off in 1992–99, and the test sample, Figure 6, is the 135,008 loans paid off between January 1, 2000, and July 31, 2001. Given the type of loan (new or renewal) and the gender of the borrower (woman or man), the scorecard predicts that future risk for test cases will be the same as historical risk for construction cases with the same characteristics.

For example, in Figure 6, predicted risk for renewal loans to women is the historical risk for the segment, 12.8 percent (Leaf 3, "Predicted % Bad" column). It turns out that realized risk in 2000–01 was 12.1 percent (Leaf 3, "Realized % Bad" column). The accuracy of the scorecard is seen in Figure 7 as the distance between the lines for predicted (historical) risk and realized risk.[5]

Predicted risk for new loans to men (the highest-risk segment) is 22.3 percent (Leaf 2, "Predicted % Bad" column). This again comes close to the realized risk of 21.9 percent (Leaf 2, "Realized % Bad" column). In fact the tree's risk forecast was close to the realized risk in all four segments, as the graph in Figure 7 shows.

Scoring forecasts risk by assuming that past links between risk and characteristics will hold in the future. Historical tests of predictive power compare predicted risk with realized risk for loans paid off in the past. Scoring works much like the arrears-based grades that many microlenders already use, but scoring, once it has been developed, is easier and more powerful to use (see Box 2).

IV. How to Use Scorecards

How would loan officers and credit managers use scoring in their daily work? This section uses a 19-leaf tree to illustrate a policy for application decisions based on four risk classes into which the applications fall. The section then shows how to use the historical test of predictive power to set policy thresholds and to estimate trade-offs between risk, disbursements, and profits.

A 19-Leaf Tree

Like the four-leaf tree, the 19-leaf tree in Figure 8 is constructed from data on paid-off loans gathered by a large microlender. The microlender defines "bad" as a loan with a 30-day spell of arrears or an average of seven days of arrears per installment. The 19-leaf tree has more leaves than the four-leaf tree, but the concepts are the same. More leaves allow finer-grained forecasts and greater distinctions between high-risk cases and low-risk cases. The 19 leaves are defined by up to four splits on seven variables that most microlenders record as part of their traditional evaluation process:

- type of loan (new or renewal)
- number of telephone numbers (none, 1, or 2)
- age of applicant (years)
- experience of loan officer (number of disbursements)
- days of arrears per installment in last paid-off loan
- indebtedness (liabilities divided by assets)
- guarantee coverage (resale value of chattel guarantee divided by amount disbursed)

Leaf 11 is the largest segment, 15.0 percent of all loans ("% of All Cases in Leaf" column), and also the least risky, 4.5 percent ("% Bad" column). Segment 11 contains renewals from applicants who averaged less than 1.5 days of arrears per installment in their last paid-off loan, reported zero or one telephone number, and were more than 40 years old.

Box 2: Scoring versus Arrears-Based Grading

Many microlenders grade applicants based on their arrears during the previous loan. Scoring is similar to grading, only scoring is more accurate and, because differences between forecasts have known meanings, it is easier to use. If grading is useful, scoring is more useful for three reasons.

First, scoring quantifies risk as a probability; grading merely ranks risks. For example, grade A might mean "offer special incentives to keep loyal," grade B, "accept and allow increased amount and term-to-maturity," grade C, "accept with no change in terms," and grade D, "reject." The lender, however, has no expectation of what share of those graded A will go bad, nor does the lender know how much more risk "is implied by an A than a C." In contrast, scoring not only ranks risks but also—once adjusted for absolute accuracy (see Section V)—specifies precise differences in risk. For example, among loans with predicted risk of 10 percent, 10 percent are expected to go bad, half as many as among loans with predicted risk of 20 percent.

Second, scoring accounts for links between risk and a wide range of characteristics (including arrears), but grading ignores everything except arrears. While grading is useless for new loans because they do not have an arrears record, scoring works nearly as well for new loans as for repeat loans.

Third, scoring uses the historical database and statistical techniques to optimally link risk to a wide range of characteristics. In contrast, grading links risk to arrears based on the judgment and experience of the managers who concoct the system. Of course, some simple analyses of the database could inform the design of grading systems, but managers rarely do such analysis. Likewise, historical tests of predictive power are standard for scoring, but are virtually non-existent for grading.

Figure 8: 19-Leaf Tree with Historical Risk, 1992–99 Data

Branch of Tree					Construction Sample, 1992–99				
Leaf	First	Second	Third	Fourth	Bads	Goods	Total Cases	% Bad	% of All Cases in Leaf
1	**New**	No telephone	N/A	N/A	186	453	639	29.1	0.8
2		1 telephone	Age ≤ 40	Loan officer exp. ≤ 500	603	2,459	3,062	19.7	4.0
3				Loan officer exp. > 500	613	4,980	5,593	11.0	7.4
4			Age > 40	Loan officer exp. ≤ 150	158	746	904	17.5	1.2
5				Loan officer exp. > 150	446	4,962	5,408	8.2	7.1
6		2 telephones	Age ≤ 40	Loan officer exp. ≤ 700	993	3,032	4,025	24.7	5.3
7				Loan officer exp. > 700	614	3,590	4,204	14.6	5.5
8			Age > 40	Loan officer exp. ≤ 700	490	2,029	2,519	19.5	3.3
9				Loan officer exp. > 700	319	2,395	2,714	11.8	3.6
10	**Renewal**	Days of arrears/ installments ≤ 1.5	0 or 1 telephone	Age ≤ 40	670	9,463	10,133	6.6	13.4
11				Age > 40	513	10,879	11,392	4.5	15.0
12			2 telephones	Age ≤ 40	980	7,895	8,875	11.0	11.7
13				Age > 40	706	7,945	8,651	8.2	11.4
14		1.5 < Days of arrears/ installments ≤ 7	0 or 1 telephone	Loan officer exp. ≤ 2,100	476	1,655	2,131	22.3	2.8
15				Loan officer exp. > 2,100	100	960	1,060	9.4	1.4
16			2 telephones	Guarantee/amt. disb. ≤ 2.7	777	1,698	2,475	31.4	3.3
17				Guarantee/amt. disb. > 2.7	207	1,036	1,243	16.7	1.6
18		Days of arrears/installments > 7	Libs./Assets ≤ 0.03	N/A	108	293	401	26.9	0.5
19			Libs./Assets > 0.03	N/A	195	233	428	45.6	0.6
All Loans					**9,154**	**66,703**	**75,857**	**12.1**	**100.0**

Source: Latin American microlender

In contrast Leaf 19 is one of the smallest segments, 0.6 percent of all loans ("% of All Cases in Leaf" column), and also the most risky, 45.6 percent ("% Bad" column). It contains renewals from applicants who averaged more than seven days of arrears per installment in the previous loan, and had an indebtedness ratio in excess of 0.03.

A quick analysis of the 19-leaf tree in Figure 8 provides several lessons for the microlender. For example, although the portfolio is concentrated in low-risk segments, some segments are very risky. The worst, Leaf 19 with a 45.6 percent risk, is almost ten times as risky as the best segment, Leaf 11 with a 4.5 percent risk. The microlender probably would want to treat applicants from the highest-risk segments differently than applicants from the lowest-risk segments.

Characteristics related with risk are as follows:

- Youth signals more risk than age.
- More arrears in the last paid-off loan signals more risk than less arrears.
- Smaller guarantees signal more risk than larger guarantees.
- More indebtedness signals more risk than less indebtedness.
- Greater loan officer experience signals more risk than less experience.
- The presence of one phone number signals more risk than none or two (perhaps because the services of this microlender in this country fit better the demands of the "average" poor [with one phone] than for the poorest [with no phone] or the not-so-poor [with two phones]).

These patterns fit the lender's experience. This confirms the potential of scoring and also the lender's intuition. Scoring does more, however, than tell the lender what it already knows; it quantifies links with risk. For example, the lender already knows that risk increased with arrears in the last paid-off loan, but it does not know by how much. The tree suggests that risk for renewals with 0 to 1.5 days of arrears per installment in the last paid-off loan is 7.3 percent (computed as the total number of "Bads" in Leaves 10–13 divided by the total number of loans in those segments). This is 15.3 percentage points less than the risk of renewals having 1.5 to 7 days of arrears (segments 14–17), and it is 29.3 percentage points less than renewals with more than 7 days of arrears (segments 18 and 19).

The 19-Leaf Historical Test

The historical test with the 19-leaf tree follows the same process as the four-leaf tree. The construction sample covers the period 1992–99, and the test sample covers 2000–01. As before, historical risk in a segment from 1992–99 is taken as predicted risk for loans in that segment in 2000–01. The test then compares the predicted risk with realized risk.

How well does the 19-leaf tree, constructed with 1992–99 data, predict risk in 2000–01? Figure 8 shows historical risk for the 19 segments in 1992–99, and Figure 9 shows realized risk in 2000–01. Figure 10 compares the predicted risk with realized risk. Predictive power can be looked at in three ways.

First, *absolute accuracy* looks at the distance between predicted risk and realized risk. In Figure 10, some distances are narrow and some are wide. For example, predicted risk for segment 11 (lower left corner) was 4.5 percent, and realized risk was 4.1 percent, an error of about 9 percent ([4.5 – 4.1] divided by 4.5 = 0.09). In segment 13 (two steps up from the lower left corner), however, predicted risk was 8.2 percent and realized risk was 11.5 percent, a 40-percent error ([11.5 – 8.2] divided by 8.2 = 0.40).

Second, *relative accuracy* looks at whether loans with lower predicted risk have lower realized risk than do loans with higher predicted risk. A scorecard with relative accuracy correctly rank-orders loans even though it may lack absolute accuracy. For the 19-leaf tree, relative accuracy was high:

Figure 9: 19-Leaf Tree with Realized Risk, 2000–01

Branch of Tree					Construction Sample, 1992–99					
Leaf	First	Second	Third	Fourth	Bads	Goods	Total Cases	Predicted % Bad	Realized % Bad	% of All Cases in Leaf
1	New	No telephone	N/A	N/A	61	116	177	29.1	34.5	0.4
2		1 telephone	Age ≤ 40	Loan officer exp. ≤ 500	460	1,827	2,287	19.7	20.1	4.6
3				Loan officer exp. > 500	508	3,920	4,428	11.0	11.5	9.0
4			Age > 40	Loan officer exp. ≤ 150	126	436	562	17.5	22.4	1.1
5				Loan officer exp. > 150	387	4,271	4,658	8.2	8.3	9.4
6		2 telephones	Age ≤ 40	Loan officer exp. ≤ 700	573	1,293	1,866	24.7	30.7	3.8
7				Loan officer exp. > 700	483	1,603	2,086	14.6	23.2	4.2
8			Age > 40	Loan officer exp. ≤ 700	311	1,005	1,316	19.5	23.6	2.7
9				Loan officer exp. > 700	227	1,164	1,391	11.8	16.3	2.8
10	Renewal	Days of arrears/ installments ≤ 1.5	0 or 1 telephone	Age ≤ 40	477	6,980	7,457	6.6	6.4	15.1
11				Age > 40	340	8,027	8,367	4.5	4.1	16.9
12			2 telephones	Age ≤ 40	612	3,465	4,077	11.0	15.0	8.3
13				Age > 40	490	3,761	4,251	8.2	11.5	8.6
14		1.5 < Days of arrears/ installments ≤ 7	0 or 1 telephone	Loan officer exp. ≤ 2,100	447	1,526	1,973	22.3	22.7	4.0
15				Loan officer exp. > 2,100	144	1,079	1,223	9.4	11.8	2.5
16			2 telephones	Guarantee/Amt.disb. ≤ 2.7	527	1,015	1,542	31.4	34.2	3.1
17				Guarantee/Amt.disb. > 2.7	243	627	870	16.7	27.9	1.8
18		Days of arrears/ installments > 7	Libs./Assets ≤ 0.03	N/A	68	106	174	26.9	39.1	0.4
19			Libs./Assets > 0.03	N/A	423	257	680	45.6	62.2	1.4
All Loans					**6,907**	**42,478**	**49,385**	**12.1**	**14.0**	**100.0**

Source: Latin American microlender

Box 3: How Do Abrupt Changes Affect Scoring?

When the context changes, scoring loses absolute accuracy,[a] but it usually retains relative accuracy. In microcredit, change is constant: competition sharpens, police start to enforce laws, or the economy weakens. Even without external changes, microlenders grow and constantly adjust internally.

For example, the success of microcredit in Bolivia attracted competition from Chilean consumer-finance companies in 1995–96.[b] The battle for market share tripled arrears and doubled drop-out rates.

Can scoring staunch the flow of drop-outs? A desertion scorecard (see Section VII) was constructed with data from 1988–96 and tested on data from 1997.[c] The construction sample and test sample straddled the abrupt market shift. Absolute accuracy was low, but relative accuracy was still usefully high.

[a] Edward M. Lewis, *An Introduction to Credit Scoring* (San Rafael, Calif.: Athena Press, 1990).
[b] Elisabeth Rhyne, *Mainstreaming Microfinance: How Lending to the Poor Began, Grew, and Came of Age in Bolivia* (Bloomfield, Ind.: Kumarian, 2001); and Jeffrey Poyo and Robin Young, "Commercialization of Microfinance: The Cases of Banco Económico and Fondo Financiero Privado FA$$IL, Bolivia" (Bethesda, Md.: Microenterprise Best Practices, 1999).
[c] Mark Schreiner, "Scoring Drop-out at a Microlender in Bolivia" (manuscript, Center for Social Development, Washington University, St. Louis, Mo., 2001).

except for a few segments, realized risk consistently increased as predicted risk increased (see Figure 10). In general the line of realized risk slopes up from left to right. Relative accuracy matters more than absolute accuracy because, as discussed in Section V, managers can use the "Global Follow-up Report" to convert relatively accurate scores into absolutely accurate scores. Also abrupt changes in the market or macroeconomy affect relative accuracy less than absolute accuracy (see Box 3).

Figure 10: Test of 19-Leaf Tree, Comparing Predicted Risk with Realized Risk, 2000–01

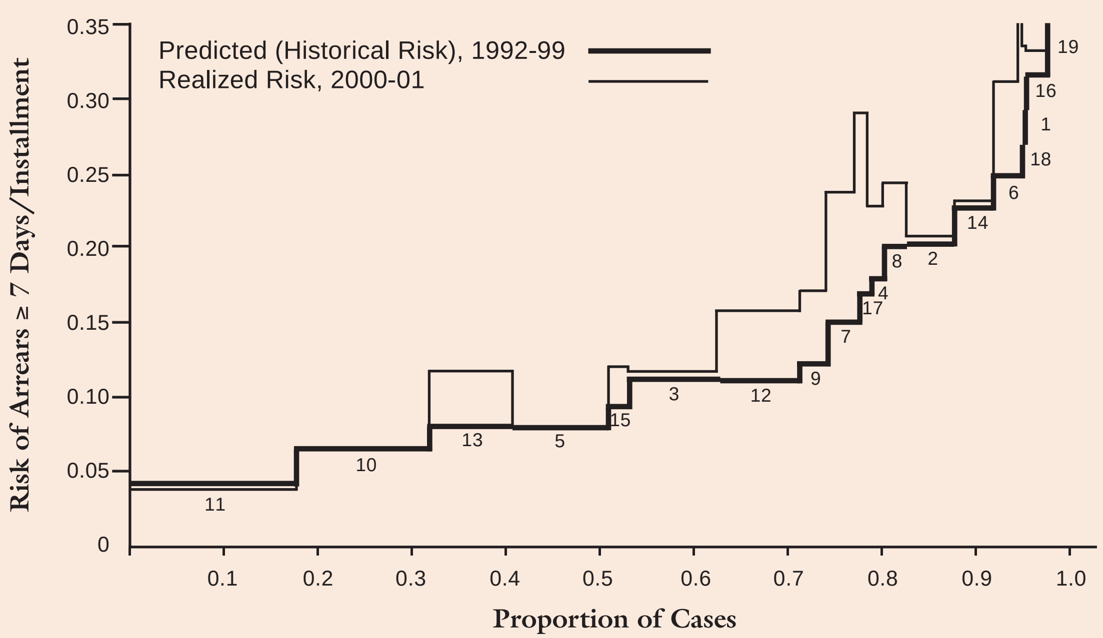
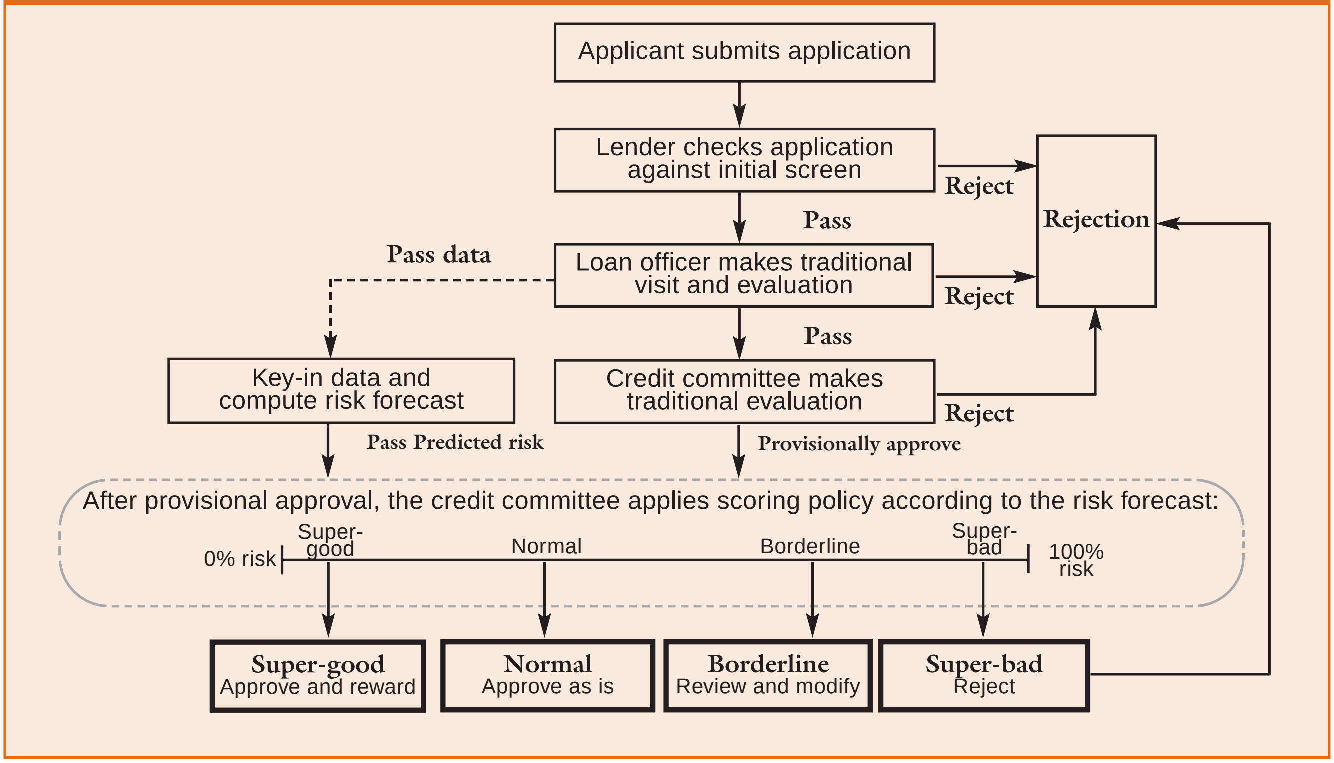


Figure 11: A Four-Class Scoring Policy and the Traditional Evaluation Process



Third, *tail accuracy* looks at relative and absolute accuracy where it matters most, among loans with very low or very high predicted risk. After all, most loans are about average, and scoring policy does not affect average loans. Scoring does, however, affect the lowest-risk applicants (they might receive special rewards) and the highest-risk applicants (their applications might be modified or even rejected). The 19-leaf tree had excellent tail accuracy: cases with the lowest predicted risk also had the lowest realized risk, and cases with the highest predicted risk also had the highest realized risk. For example, the two segments with the lowest predicted risk (11 and 10 in the lower left corner of Figure 10) also had the lowest realized risk and very small prediction errors. The five segments with the highest predicted risk (6, 18, 1, 16, and 19 in the upper right corner of Figure 10) had large prediction errors, but they also had the highest realized risk. (Trees often have systematic and variable prediction errors, especially for small segments.[6])

Using Scoring with Four Risk Classes

Before scoring an application, the microlender must first approve it using the same credit evaluation process that it would use if it did not have scoring. Given the characteristics of a provisionally approved loan, scoring then forecasts risk. The credit committee acts on the predicted risk according to the policies the microlender has established for four risk classes of loan applicants: super-bad, borderline, normal, and super-good. The lender sets the four thresholds to meet its mission, given trade-offs among breadth, depth, and length of outreach.[7]

Because scoring ignores qualitative characteristics and considers only quantified characteristics, it cannot replace any part of the traditional evaluation (see

Box 4). Scoring simply adds a step at the end of the traditional process, just before disbursement.

Figure 11 depicts a typical evaluation process for a microlender using scoring. It starts when a client submits an application. Before the loan officer makes a field visit, the application is screened against basic policy rules, such as having at least one year of experience in the business. If the application clears this hurdle, the loan officer makes the field visit and—perhaps after some analysis in the office—decides whether to present the case to the credit committee. Applications that pass this stage are then keyed into the information system. The system computes a score and prints scoring reports (the Scoring Simulator and Effects of Characteristics Reports are discussed in Section V) to be included

Box 4: Why Score Only Cases Provisionally Approved by the Traditional Process

The share of risk missed by scoring but captured by subjective evaluation is large, and vice versa. In principle, scoring could come before or after subjective evaluation. If scoring is first and predicts low risk, then the lender may be tempted to skimp on the (more costly) subjective evaluation. This could be disastrous because loans that seem low risk based on quantitative factors may be very high risk after accounting for subjective factors. Thus microlenders should score only cases already provisionally approved under the subjective evaluation process.

Overall repayment risk can be broken into three parts according to how it is linked with the quantified characteristics of the borrower, the loan, and the lender:

- Random risk is not linked at all with any characteristics, quantified or not.
- Proxied risk is linked with quantified characteristics.
- Qualitative risk is linked with non-quantified characteristics.

Random risks (like lightning bolts) are unpredictable. Scoring measures proxied risk and only proxied risk. Scoring reveals correlations, not causes; it does not reveal why an attribute of a characteristic is associated with risk, only that it is. Finally, traditional evaluation in microcredit looks at both proxied risk and qualitative risk. Compared with scoring, traditional evaluation does better with qualitative risk (scoring ignores qualitative risk) and worse with proxied risk.

A microlender that uses scoring to skip (or skimp on) traditional evaluation gambles that the qualitative risk of through-the-door applicants is about the same as the qualitative risk of applicants who have been provisionally approved by traditional evaluation. This supposes—in stark contrast to most current microlending technologies—that qualitative risk is unimportant or unmeasurable.

Just how important is qualitative risk? Performance is known only for disbursed loans, so no historical test can reveal how loans rejected for qualitative reasons under the traditional process would have performed, had they been booked.

Microlenders who substitute scoring for subjective screening do so at their own peril. Unless qualitative risk does not matter at all, forecasts will be too low. The only way to know exactly how low is to book some loans without subjective screening and then see how they turn out.

With time credit bureaus will become better, more widespread, and more complete, and microlenders will quantify more characteristics. With more and better data, perhaps scoring can preclude the need for subjective risk evaluation, but no one knows yet. One finance company that entered Bolivia and judged the risk of microcredit borrowers only with scoring went bankrupt.[a] For now scoring complements—but does not replace—loan officers and traditional evaluation.

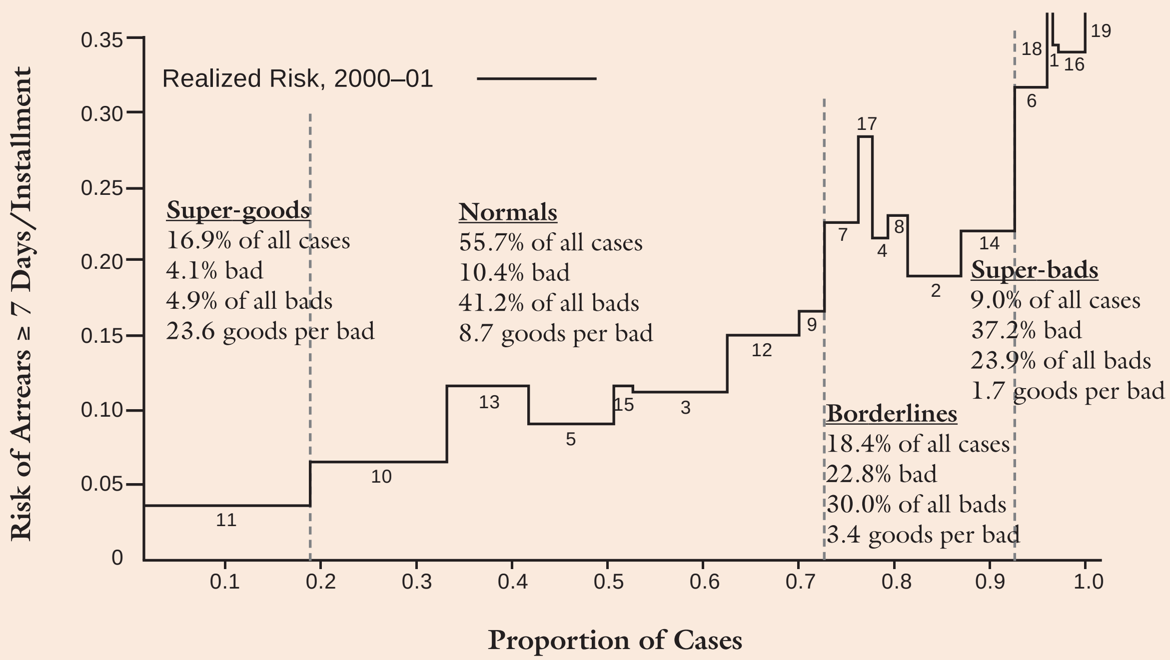
[a] Elisabeth Rhyne, *Mainstreaming Microfinance: How Lending to the Poor Began, Grew, and Came of Age in Bolivia* (Bloomfield, Ind.: Kumarian, 2001).

Figure 12: Results of a Four–Class Scoring Policy Used in 2000–01 with a 19-Leaf Tree Constructed from 1992–99 Data

Branch of Tree					Test Sample, 2000–01							
Leaf	First	Second	Third	Fourth	Bads	Goods	Total Cases	Predicted % Bad	Realized % Bad	% of All Cases in Leaf	% of All Bads in Leaf	Goods per Bad
All Loans					**6,907**	**42,478**	**49,385**	**12.1**	**14.0**	**100.0**	**100.0**	**6.1**
	<u>Super-goods</u>											
11	Renewal	Days of arrears /installments ≤ 1.5	0 or 1 telephone	Age > 40	340	8,027	8,367	4.5	4.1	16.9	4.9	23.6
	TOTAL for Super-goods				**340**	**8,027**	**8,367**	**4.5**	**4.1**	**16.9**	**4.9**	**23.6**
	<u>Normal:</u>											
10	Renewal	Days of arrears /installments ≤ 1.5	0 or 1 telephone	Age ≤ 40	477	6,980	7,457	6.6	6.4	15.1	6.9	14.6
13	Renewal	Days of arrears /installments ≤ 1.5	2 telephones	Age > 40	490	3,761	4,251	8.2	11.5	8.6	7.1	7.7
5	New	1 telephone	Age > 40	Loan officer exp. > 150	387	4,271	4,658	8.2	8.3	9.4	5.6	11.0
15	Renewal	1.5 < Days of arrears /installments ≤ 7	0 or 1 telephone	Loan officer exp. > 2,100	144	1,079	1,223	9.4	11.8	2.5	2.1	7.5
3	New	1 telephone	Age ≤ 40	Loan officer exp. > 500	508	3,920	4,428	11.0	11.5	9.0	7.4	7.7
12	Renewal	Days of arrears /installments ≤ 1.5	2 telephones	Age ≤ 40	612	3,465	4,077	11.0	15.0	8.3	8.9	5.7
9	New	2 telephones	Age > 40	Loan officer exp. >700	227	1,164	1,391	11.8	16.3	2.8	3.3	5.1
	TOTAL for Normals				**2,845**	**24,640**	**27,485**	**9.0**	**10.4**	**55.7**	**41.2**	**8.7**
	<u>Borderlines</u>											
7	New	2 telephones	Age ≤ 40	Loan officer exp. > 700	483	1,603	2,086	14.6	23.2	4.2	7.0	3.3
17	Renewal	1.5 < Days of arrears /installments ≤ 7	2 telephones	Guarantee/Amt. disb. > 2.7	243	627	870	16.7	27.9	1.8	3.5	2.6
4	New	1 telephone	Age > 40	Loan officer exp. ≤ 150	126	436	562	17.5	22.4	1.1	1.8	3.5
8	New	2 telephones	Age > 40	Loan officer exp. ≤700	311	1,005	1,316	19.5	23.6	2.7	4.5	3.2
2	New	1 telephone	Age ≤ 40	Loan officer exp. ≤500	460	1,827	2,287	19.7	20.1	4.6	6.7	4.0
14	Renewal	1.5 < Days of arrears/ installments ≤ 7	0 or 1 telephone	Loan officer exp. ≤ 2,100	447	1,526	1,973	22.3	22.7	4.0	6.5	3.4
	TOTAL for Borderlines				**2,070**	**7,024**	**9,094**	**18.1**	**22.8**	**18.4**	**30.0**	**3.4**
	<u>Super-bads</u>											
6	New	2 telephones	Age ≤ 40	Loan officer exp. ≤ 700	573	1,293	1,866	24.7	30.7	3.8	8.3	2.3
18	Renewal	Days of arrears /installments > 7	Libs./Assets ≤ 0.03	N/A	68	106	174	26.9	39.1	0.4	1.0	1.6
1	New	No telephone	N/A	N/A	61	116	177	29.1	34.5	0.4	0.9	1.9
16	Renewal	1.5 < Days of arrears /installments ≤ 7	2 telephones	Guarantee/Amt. ≤ 2.7	527	1,015	1,542	31.4	34.2	3.1	7.6	1.9
19	Renewal	Days of arrears /installments > 7	Libs./Assets > 0.03	N/A	423	257	680	45.6	62.2	1.4	6.1	0.6
	TOTAL for Super-bads				**1,652**	**2,787**	**4,439**	**28.4**	**37.2**	**9.0**	**23.9**	**1.7**

Source: Latin American microlender

Figure 13: Graph of Results of Four-Class Scoring Policy Used in 2001–01 with a 19–Leaf Tree Constructed from 1992–99 Data



with other reports normally prepared for the credit committee.

To this point, scoring has changed nothing in the traditional evaluation process; the use of scoring still awaits provisional approval of the application. When is that? If the credit committee rubber-stamps almost all applications that reach it, then provisional approval takes place when the loan officer decides to present an application to the committee. In this case, the committee uses the score to determine which applications to review in depth and which to pass unchanged. If, however, provisional approval takes place in the committee itself, then the score must be ignored until the traditional screening is done. (If the committee peeks at the score early, it may be tempted to approve loans without screening them for qualitative risk.) Score in hand, the committee applies a four-class scoring policy (see bottom row in Figure 11).

"Super-good" Risk Class

Applicants with predicted risk below the lowest threshold are classified as super-good. To keep these low-risk clients loyal, the lender might adopt a policy to enhance the loan value for them by offering lines of credit, reduced fees, rebates for perfect repayment, or lower guarantee requirements. Scoring only identifies super-goods, which is not the best way to keep the applicant class loyal. It merely forecasts risk, leaving managers to decide what to do next. Should they want to use risk-based pricing, then they must decide how to adjust interest rates, given predicted risk.

For the sake of discussion, suppose that the super-good threshold is 5 percent for the 19-leaf tree in Figure 12—that is, all cases with a risk forecast of 5 percent or less qualify as super-good. All of the super-goods are in Leaf 11 with a predicted risk, based on 1992–99, of 4.5 percent. Super-goods represent 16.9 percent of all cases.

Box 5: Does Scoring Policy Apply to Renewals?

Renewal applicants have a repayment record, so scoring works even better for them than for new applicants. Some microlenders, however, are loath to consider modifying borderline renewals—let alone rejecting super-bad renewals—partly because they doubt the power of scoring and partly because they want to maintain a reputation for rewarding faithful repayment with access to additional loans.

What to do? The scorecard should consider the type of loan (new or renewal) and the repayment record. If repeat borrowers with low arrears in previous loans actually have less risk, then an accurate scorecard will reflect that. Nevertheless scoring may finger as bad risks a few cases with spotless records. If the historical test did not break down for renewals, then these applications probably do in fact have high risk.

Still lenders cannot reject these applicants, both because it would send the wrong signal to current borrowers and because the credit committee would sooner reject scoring than reject renewals with perfect records. In these cases, the policy for managing super-bads should specify careful review of the evaluation, modifications to the loan contract, and preventive "courtesy visits" after disbursement.

How well would this 5 percent, super-good threshold have worked? In 2000–01, scoring would have qualified 16.9 percent of loans approved under the traditional evaluation process as super-good (see Figures 12 and 13). Among these, 4.1 percent went bad, accounting for 4.9 percent of all bad loans. Seen another way, among the super-goods, there were 23.6 good loans for each bad loan.

Scoring identifies both low-risk cases and high-risk cases, helping proactive lenders manage risk at both extremes. Lenders that do not want to reward low risks can choose to set the super-good threshold to zero, as predicted risk is never that low.

"Normal" Risk Class

Applicants with predicted risk in excess of the super-good threshold but below the borderline threshold are normal. Scoring confirms the provisional approval of these cases, and they immediately leave the credit committee and are disbursed as is. Most provisionally approved applications qualify as normal, so in most cases scoring does not affect the evaluation nor impose additional costs on the credit committee.

Again, suppose that the normal threshold for the 19-leaf tree in Figures 12 and 13 is 12 percent (Leaves 10, 13, 5, 15, 3, 12, and 9). In 2000–01 more than half (55.7 percent) of all cases were normal, with a risk forecast greater than the 5 percent super-good threshold, but less than the 12 percent normal threshold. Of these normals, 10.4 percent went bad, which was 41.2 percent of all bads. Among normals there were 8.7 good loans per bad loan.

"Borderline" Risk Class

Applicants with predicted risk in excess of the borderline threshold but below the super-bad threshold are borderline. The credit committee reviews these cases with extra care and, if warranted, modifies the amount disbursed, the term-to-maturity, guarantee requirements, or interest rates or fees (risk-based pricing). The committee may also decide to reject some borderline cases.

Scoring increases the time that the credit committee spends evaluating borderlines. This increases costs, although most forewarned lenders welcome the chance to manage borderline cases before booking them.

In Figures 12 and 13, suppose that the borderline threshold for the 19-leaf tree was 23 percent (Leaves 7, 17, 4, 8, 2, and 14). In 2000–01, 18.4 percent of all cases were borderline (risk forecast greater than the normal threshold of 12 percent but less than the borderline threshold of 23 percent). Of these borderlines, 22.8 percent went bad, accounting for 30 percent of all bads, and there were 3.4 good loans per each bad loan.

"Super-bad" Risk Class

Applicants with predicted risk in excess of the highest threshold are super-bad. Except for rare cases (see Box 5), super-bads are summarily rejected. The committee may review super-bads to see what they missed or to check if there are any overwhelming, positive qualitative factors to justify overriding scoring policy.

Returning to Figures 12 and 13, suppose that cases with risk greater than 24 percent are super-bad (Leaves 6, 18, 1, 16, and 19). In 2000–01, 9 percent of all cases had risk in excess of 24 percent and so qualified as super-bad. Of these super-bads, 37.2 percent went bad, which was 23.9 percent of all bads. Among super-bads there were 1.7 good loans for each bad.

Those lenders that skip historical tests would be mortified to reject high-risk cases that, without scoring, would have been approved (see Box 6). They can effectively eliminate the super-bad threshold by setting it at 100 percent, as risk never gets that high.

A four-class scoring policy rewards low-risk cases and reviews, modifies, or rejects high-risk cases. Most cases have about average risk and for them scoring has no effect. Scoring can only confirm the provisional approval conferred by the loan officer or credit committee, so loans rejected by traditional standards are still rejected by scoring.

Setting Thresholds for Scoring Policy

The choice of thresholds depends on the predictive power of scoring for a specific microlender and how the microlender values trade-offs between different aspects of its mission:[8] breadth of outreach (number of loans), depth of outreach (poverty of borrowers), and length of outreach (organizational permanence through profits).

A microlender must make these value judgments for itself. After that, the historical test can guide the lender in setting scoring policy to optimize its goals. It does this by showing how different hypothetical thresholds would affect the numbers of loans approved, good loans missed, and bad loans avoided. (The assumption is that the historical test indicates future results in actual use.)

For example, Figure 14 shows the results for the 19-leaf tree with a range of super-bad thresholds. If the lender had set a super-bad threshold of 24 percent in

Box 6: Scoring Throws Out the Goods with the Bads

Some applicants rejected as super-bad would have been good, just as some borderlines would have been fine without modification. For some people, knowing this makes it almost impossible to accept statistical scoring. Of course, traditional subjective evaluation also modifies some loans unnecessarily and mistakenly rejects some applicants. That is, subjective scoring also throws out the goods with the bads. With statistical scoring, however, the historical test quantifies prediction error and thus improves the choice between a strict or lax policy. With subjective scoring, prediction error is unknown, so choices are less apt to be optimal.

In Latin America for example, some microlenders who make individual loans are as strict as Scrooge. One renowned microlender in Colombia rejects half of all applicants and two-thirds of all new applicants. An even more well-known Bolivian lender almost never grants the requested amount or term-to-maturity. Given such strictness, it is possible that if lenders better understood the true risk/outreach trade-offs, they might better meet demand and maintain—or even decrease—risk.

Figure 14: Ratio of Good Loans to Bad Loans Avoided for a Range of Super-bad Thresholds for the 19-Leaf Tree

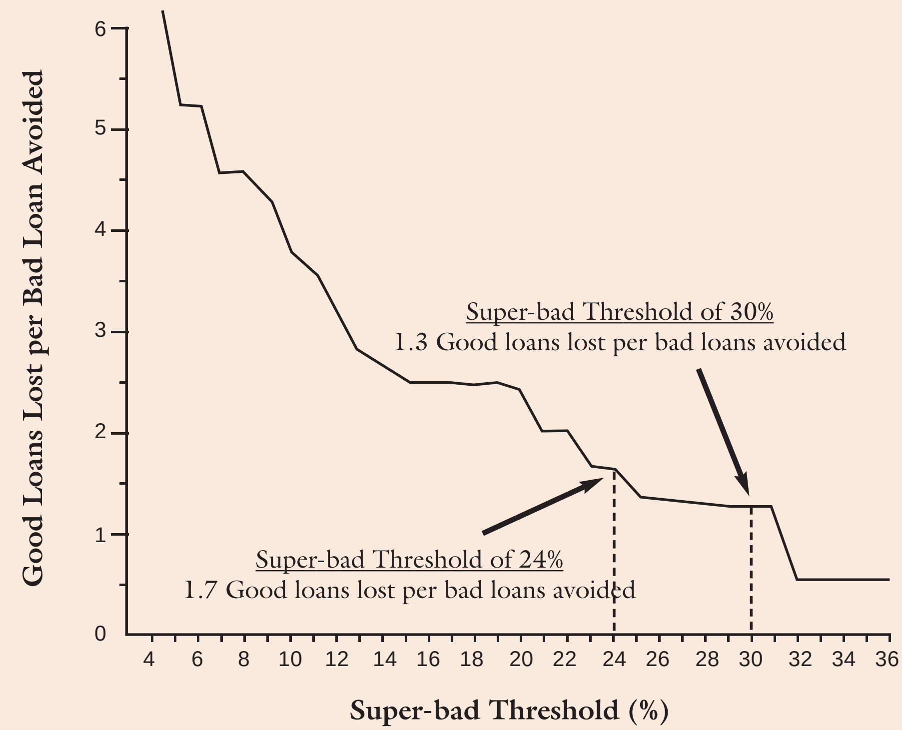
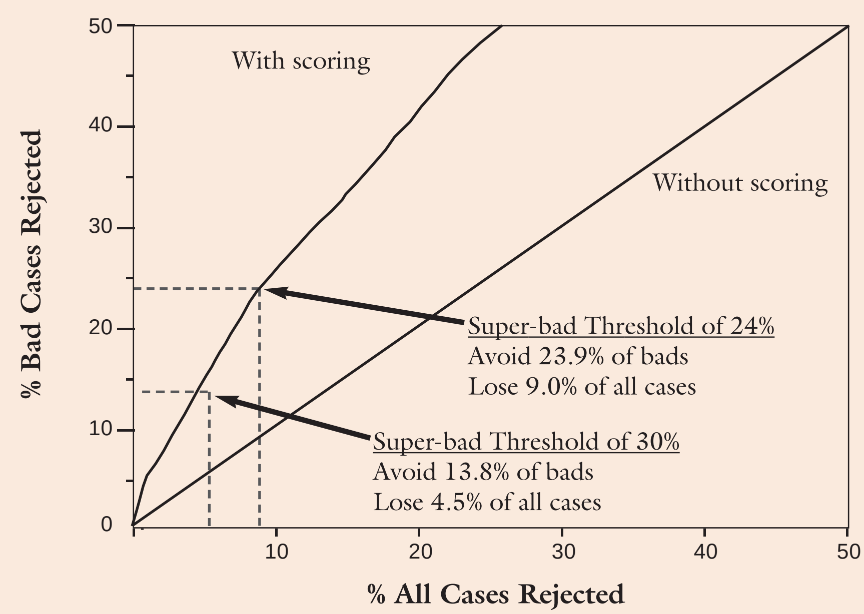


Figure 15: Share of Cases Rejected versus Share of Bad Loans Avoided for a Range of Super-bad Thresholds for the 19-Leaf Tree



2000–01, then 1.7 good loans would have been lost for each bad one avoided. About 9 percent of all cases would be policy rejects, with 23.9 percent of all bads avoided (see Figure 15).

How would things change if the super-bad threshold were moved, for example, to 30 percent? Figure 14 shows that 1.3 goods are lost for each bad avoided, and Figure 15 shows that 4.5 percent of all cases are policy rejects and that 13.8 percent of all bads are avoided. Given the likely outcomes of different possible thresholds, the historical test allows the microlender to choose the threshold that best fits its mission.

Scoring also shows how risk is linked with characteristics that mark depth of outreach (such as gender, income, or age). This indicates the trade-offs between depth of outreach and risk. For example, scoring may indicate that subsistence farmers—all else constant—are 2 percentage points more likely to have a spell of arrears of 30 days. This knowledge allows the microlender to explicitly trade off depth (lending to subsistence farmers) against both breadth (reaching more borrowers by avoiding the worst risks) and length (making more profit by avoiding the worst risks). Of course, knowing that a case is risky does not obligate a lender to reject it.

Nothing forces poverty-oriented microlenders to reject high-risk cases, but they do not want to ignore risk forecasts either. No one lends with utter disregard for risk, and even the most dedicated microlender limits the cost it will accept to reach a given depth of outreach. Scoring merely sheds light on trade-offs; the lender must still decide what to do. In addition rejection need not always hurt applicants. Microcredit is sometimes more harmful than helpful, especially for the poorest.[9] Some high-risk cases, even if they do not go bad, will struggle so much to pay their loan installments on time that they would have been better off being rejected in the first place.

Scoring—given estimates of the net financial cost of a good loan missed and of the net financial benefit of a bad loan avoided—can help to estimate the direct, first-round trade-offs between breadth of outreach and length of outreach (profits). The impact can be surprisingly large. Given reasonable assumptions, a 24 percent super-bad threshold for the 19-leaf tree in 2000–01 would have saved the lender more than $200,000 (see Box 7).

In practice (such as with the example 19-leaf tree in Figures 12 and 13), most microlenders will probably aim for thresholds that result in about 10 percent of cases being super-good, 60 percent being normal, 20 percent borderline, and 10 percent super-bad. This broad pattern has four advantages. One, it keeps the share of super-goods low, enabling the lender to offer special incentives to their best clients, yet control the cost of incentives. Two, most cases are normal, so scoring will not change the standard loan evaluation process for most of them. This can be reassuring to front-line personnel and encourage them to accept scoring. Three, most risky cases are borderline. Loan officers and credit managers are reluctant to reject applicants solely on the basis of scoring. With most risky borrowers classified as borderline, the credit committee is encouraged not to reject but to review risky cases and consider adjusting the terms of the loan contract. Four, the share of super-bads is low. The few super-bads included are extremely risky. Because a very large share would have turned out bad, loan officers are apt to notice the difference in repayment performance (and in their bonuses). Over time this builds confidence in scoring.

With thresholds that produce a distribution of cases in these broad ranges, scoring may simultaneously increase breadth, depth, and length of outreach. Breadth of outreach may increase because rejecting a few extremely risky cases can save enough time in collections that loan officers can increase disbursement to more than compensate for the rejected cases. Length of outreach (permanence via profits) may increase because the revenue from the increased lending volume will likely exceed the costs of scoring. Depth of outreach may increase because some share of the additional loan volume will likely accrue to poorer borrowers. In sum, scoring is an innovation that boosts efficiency and thus has the potential to skirt the normal trade-offs between aspects of outreach.[10] If scoring helps the microlender to do more with less, then it can make everything better without making anything worse.

Costs of Scoring

Scoring has fives types of costs: data accumulation, set-up, operational, policy-induced, and process costs. First, collecting and entering the data to construct a scorecard incurs data accumulation costs. For the least sophisticated microlenders, this involves not only inputting application data as it is received but also beefing up the information system to handle the additional data. For these lenders, scoring should not be a priority; improvements to their information systems are worthwhile quite apart from their usefulness to scoring.

Box 7: Estimating the Effects of Scoring on Profit

A lender can estimate the effects of scoring on profit, even before scoring is implemented. Such profitability estimates can help convince stakeholders that scoring is worthwhile.[a]

Given a super-bad threshold, the historical test shows the number of good loans lost for each bad loan avoided. Suppose then that the lender knows the average net financial benefit of booking a good loan as well as the average net financial cost of booking a bad loan. (This cost is mostly the opportunity cost of the time that loan officers spend in collections rather than in marketing, evaluation, and disbursement.)

In fact few microlenders have measured these benefits and costs, even though they drive profitability and thus (if only implicitly) drive lending policy, with or without scoring. Lenders do know, however, that the cost of a bad loan far exceeds the benefit of a good one. For example, credit-card lenders in rich countries commonly assume that it takes more than ten good loans to make up for one bad loan.

If a lender implements a scorecard, the number of bads decreases (decreasing costs), and the number of goods—at least as a first-round effect—also decreases (decreasing benefits). The net effect of scoring on profits may be computed as:

(Costs per bad x Bads avoided) – (Benefit per good x Goods lost).

For the 19-leaf tree, the assumed cost of a bad loan is $300, and the assumed benefit of a good loan is $100. With a super-bad threshold of 24 percent, the historical test (Figure 12, bottom row, "Total Cases" column) shows that 4,439 cases would have qualified as super-bad. Of these, 1,652 turned out bad ("Bads" column), and 2,787 turned out good ("Goods" column). Among super-bads there were 1.7 goods for each bad. If all super-bads had been rejected as a matter of policy in 2000-01, the estimated change in profit would have been:

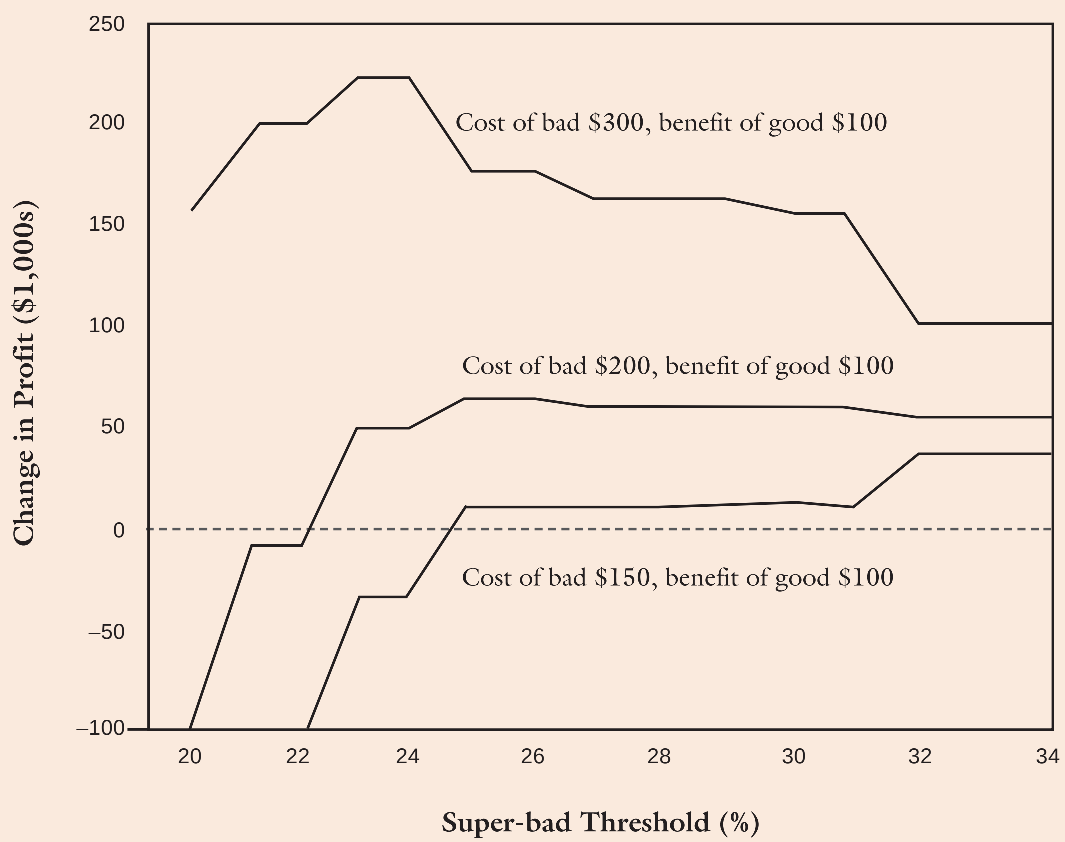
($300 x 1,652) – ($100 x 2,787) = $216,900.

Even rejecting only the 1.4 percent of applicants in leaf 19 (the riskiest segment, see Figure 12, "% of All Cases in Leaf" column) would have boosted profits by $101,200: ($300 x 423) – ($100 x 257) = $101,200.

Figure 16 shows changes in profits for the 19-leaf tree for three possible sets of assumptions about the cost of a bad loan and the benefit of a good one. Two lessons are noted here. First, a carelessly set super-bad threshold, blindly followed, could quickly bankrupt a lender. Second, the greater the ratio of the cost of a bad loan to the benefit of a good loan, the greater the potential profitability of scoring.

[a] John Coffman, "An Introduction to Scoring for Micro and Small Business Lending" (paper presented at the World Bank conference on "Profiting from Small Business Lending," Washington, D.C., April 2–3, 2001).

Figure 16: Estimated Change in Profit by Using a 19-Leaf Tree Scorecard in 2000–01



For more sophisticated microlenders, most data accumulation costs are already sunk; all applications are standardly input as they are received. For these lenders, scoring is possible as soon as the database has enough cases to support scorecard construction. There is a third group of lenders that have adequate information systems, but do not yet key in applications. Rather than hire an army of data-entry personnel to input archived paper applications, they can start to capture data in electronic form immediately.

Second, the scoring project itself—scorecard construction, integration with the information system, training, and follow-up—produces one-time set-up costs. In particular adjusting the information system to automatically compute and report risk forecasts can be an unexpectedly long and difficult process that consumes a large share of the project budget. In fact many scoring projects fail at this stage.

Third, daily scoring takes time from data-entry personnel, loan officers, and credit managers, incurring operational costs. These costs are low. For example, loan officers already collect most of the characteristics used in the scorecard. The information system computes the score, so the main operational costs are the extra time the credit committee spends reviewing borderline cases and the costs of on-going training of personnel.

Fourth, rewarding super-goods or rejecting super-bads induces policy costs. Rewards are not always effective; and some super-bads, had they been approved, would have been good.

Fifth and most importantly, the advent of scoring puts the organization in flux, inducing process costs. Some power shifts from the credit department to the information department. Some employees may openly oppose the changes produced by scoring; others may subtly skirt scoring by cooking data or ignoring policy rules. Still others may inadvertently sabotage scoring by skimping on the traditional evaluation. Training and follow-up (see Section V) are the best ways to manage these process costs.

Benefits of Scoring

The benefits of scoring include reduced loan losses, greater client loyalty, and ability to adjust interest rates and fees according to risk (risk-based pricing). Most importantly scoring can reduce time in collections and introduce the microlender to explicit, quantitative analysis as an aid to decision making by managers.

Reduced loan losses are probably the smallest benefit of scoring, if only because most microlenders who could use scoring suffer very few defaults. Greater loyalty from super-goods is probably a greater benefit than reduced loan losses.

Given a score, the microlender can manage risk by rejecting the loan application or modifying the loan contract. One such modification attempts to compensate for risk by increasing the interest rate or fees. In practice, however, knowing how much to adjust prices can be complicated, especially without accurate estimates of the various components of costs and revenues.

The greatest benefit of scoring results from loan officers' spending less time in collections and more time generating new business. Bad loans are costly mostly because collections eat up a lot of time. Scoring affects profit (see Box 7) because rejecting super-bads and modifying borderlines means that loan officers must chase down fewer bads. They are then free to spend the time saved on marketing, evaluation, and disbursement. Many microlenders expect scoring to save them more time in evaluation than in collections. Most loan officers, however, spend as much time in collections as in evaluation, and it must be reiterated that scoring cannot substitute for qualitative evaluation (see Box 4).

Hypothetically loan officers may spend two to three days per week on collections. Suppose—given the 19-leaf tree with a 24 percent super-bad threshold—that scoring reduces disbursements by about 10 percent and reduces bads by about 25 percent (see Figure 17). Also suppose (conservatively) that before scoring, loan officers spent two days a week on collections. Scoring then saves them half a day (25 percent of two days) per week.

Suppose further that loan officers used to spend two days a week on marketing, evaluation, and disbursement. If they use the extra half-day to drum up new clients as productively as they did before, then disbursements will increase by 25 percent. After netting off the 10 percent of super-bads rejected by scoring, scoring ends up decreasing bads by 25 percent and increasing disbursements by about 12.5 percent. Box 7 discusses a possible bottom-line impact.

Scoring, even though it may cause some loans to be rejected that otherwise would have been approved, can improve breadth and length of outreach. What about depth? In high-income countries, scoring has increased depth.[11] Most households have access to the most flexible microcredit product ever—the credit card—because scoring can inexpensively evaluate the risk of massive numbers of small, short-term, unsecured loans.

In microcredit scoring should also increase depth. First, the extra half-day per week to search for new clients will likely allow loan officers to increase the number of poor borrowers in their portfolios. (Even if most new borrowers are relatively well-off, at least some will be poorer.) Second, scoring will protect some poor borrowers from their own worst judgment.

Rejections or modifications of high-risk cases not only reduce lender costs but also help borrowers who otherwise would worry, endure collections visits, and sell off assets as they struggle to pay their debts. Scoring can help microcredit to do less harm. Third and most fundamentally, microcredit started from the premise that the poor are creditworthy but that lenders lacked the right tools to judge their risk. Scoring improves the risk-evaluation tool kit and thus helps to purge prejudice and mistakes from the evaluation process (see Box 8). If the poor really are creditworthy, then scoring will help reveal that better than ever, deepening outreach.

Perhaps the most important benefit of scoring in the long term is to whet management's appetite for decision making aided by explicit, quantitative knowledge of trade-offs derived from analysis of the database. For example, once managers establish a scoring policy with the knowledge of the trade-offs, such as those in Figure 12, they will only reluctantly go back to vague seat-of-the-pants judgments of the consequences of alternative credit policies.

Finance is all about information, and the information in the databases of many microlenders is an untapped gold mine. The experience of scoring may prompt microlenders to dedicate an employee or two to informing business decisions through data mining—the use of historical information to predict future behavior. Forecasting repayment risk (credit scoring) is one example, but data mining can also predict drop-out risk[12] or the types of potential clients most likely to respond to a marketing campaign.[13] In-house data mining need not be extremely sophisticated. For example, simple cross-tabs (such as the example trees here) can be inexpensive yet informative. Simple, useful analyses with quick turn-around encourage managers to stop thinking only within the bounds of what the information system currently produces and to start thinking about what type of information would help them to make better decisions.

Figure 17: Benefit of Scoring—Less Time Spent by Loan Officers in Collections

Activity	Before Scoring		After Scoring	
	% Time	Days per week	% Time	Days per week
Meetings and administration	20	1	20	1
Marketing, evaluation, disbursement	40	2	50	2.5
Collections	40	2	30	1.5

Changes:	Increase in applications due to increase in loan officer time	+25%
	Decrease in approvals due to use of scoring:	–10%
Result:	Net increase in approved applications:	+12.5%

Source: Hypothetical example

V. Training Staff, Establishing Scoring Policies, and Monitoring Performance

In purely technical terms, scoring for microcredit works as the previous section demonstrates. In human terms, however, scoring is not so straightforward. Using scoring to improve choices requires not only cerebral knowledge of how scoring can work but also gut faith that scoring does work, plus the heart to try to change. Belief comes from understanding, and willingness to change comes from seeing benefits. In the end, success in scoring hinges less on technical finesse than on training and follow-up.

Training is central to scoring because stakeholders—funders, upper-level managers, credit managers, and loan officers—may have a healthy skepticism. To absorb and accept the paradigm shift implicit in scoring requires repeated training, spread over the course of months.

In the first place, a consultant—likely with a strange accent if he or she can speak the language at all—parachutes in and, without having met the microlender's employees or clients, claims to have a secret computer formula that can help front-line personnel in their most difficult job: figuring out who to trust with money.

Second, scoring breaks from traditional microcredit evaluation via joint-liability groups or personal visits by loan officers. The new approach relies not on personal knowledge of character, rather it relies on quantified knowledge of characteristics.

Box 8: Is Statistical Scoring Discrimination?

Statistical scoring does discriminate: it assumes that each applicant is another instance of the same old thing, not a unique individual who might differ from other apparently similar cases in the database. Subjective scoring, however, discriminates just as much if not more. Loan officers evaluate risk based on what they and their mentors learned from other borrowers, not on some magical intelligence that developed apart from experience and prejudice. In truly unique cases (or if the microlender or loan officer is just starting out), there is no experience, so decisions can only proceed from random guesses or prejudices.

It is unfair to evaluate one person according to the experience with others thought to be similar, but the alternative is not to evaluate at all. The only non-discriminating lenders are those who approve all applicants. Thus the question is not whether to discriminate but rather how to discriminate as fairly as possible.

Fair discrimination compares like with like. For example, statistical scoring matches applicants with previous borrowers at the same lender with similar quantified characteristics. If women have a better repayment history than men, then the scorecard says so. In contrast, subjective scoring draws on the experience of microcredit in general, the experience of the organization, and the experience of the particular loan officer and credit manager. Inevitably part of this experience comes from outside the microlender's own history, if only because it takes time to build a history.

Fair discrimination consciously chooses what characteristics it uses. The characteristics used in subjective scoring (and their links with risk) are explicit; in statistical scoring, they are at least partly implicit. Awareness of the discrimination inherent in all evaluation helps ensure that the evaluation process does not perpetuate the very oppression that microcredit seeks to abolish (see Box 10).

Fair discrimination uses only characteristics that are truly linked with risk. Furthermore fair discrimination seeks to discover new characteristics linked with risk, to measure experience more accurately, and to better convert experience into risk evaluation. Historical tests are key to fair discrimination because they show whether supposed links are real. Compared with subjective scoring, statistical scoring is much easier to test.

Overall, scoring promotes fair discrimination because it increases the microlender's knowledge of its own experience. This can only decrease prejudice and correct mistaken inferences.

Third, loan officers and credit managers judge risk for a living. Not surprisingly they are loath to trust their livelihood to a magic box. Trust requires more than just seeing the scorecard; people need to understand the source of scoring forecasts, see the forecasts hit the mark, and have follow-up as they use the forecasts.

Like all projects, scoring needs management buy-in and an in-house champion. Like all projects, someone has to demonstrate how scoring works and what problems scoring resolves. This is nothing new, just work.

Introducing Scoring

Most upper-level managers and funders have heard of scoring, but some of them see it as a cure-all, others a gimmick, and all believe some common myths. Much like this paper, an introductory presentation has the practical purpose of correcting misperceptions and setting realistic expectations from the start. Although the engine of scoring is mathematical, scoring is much more than just a chalkboard exercise; it is a seismic change in organizational culture. Setting up a scoring project can be larger, longer, and more difficult than most managers imagine. As managers grasp scoring, some get excited, others huddle defensively as they sense a threat to their turf, and all stay skeptical. To nurture acceptance of change, the scoring project must constantly ask managers questions, get input, and invite feedback along such lines as:

- What is your mission?
- How would scoring promote your mission?
- What characteristics do you find matter most for risk?
- How important are qualitative characteristics?
- What is a “bad” loan for you?
- What risk do you want to forecast?
- How many good loans would you sacrifice to avoid a bad loan?
- How far back can you go until the past is unlike the future?
    - When did lending policy change?
    - When did the credit evaluation process change?
    - When did target niches shift?
    - When did competition start?
    - When did recent macroeconomic booms and busts occur?
- What parts of the database would you distrust?
- How well can the information system (and the personnel of the information department) adjust to accommodate scoring?
- What roadblocks do you expect to affect a scoring project?

Constructing and Testing the Scorecard

The next step is to construct the scorecard and run the historical test. Results in hand, the scoring project manager meets again with upper management to review basic concepts and to present concrete, lender-specific results, including the outcome of the historical test and the links detected between risk and characteristics. The scoring project then tours the branches to introduce scoring to all loan officers and credit managers. This introduction focuses less on abstract concepts and more on concrete examples from the historical test and from the constructed scorecard.

These meetings are costly, but skipping them would be a mistake. Even when loan officers and credit managers see that scoring works in the historical test, it may still be hard for them to accept it. Front-line personnel must pass through denial and disbelief. It is better to give them time to do this before the scorecard is installed.

Here too the key is to ask questions and invite responses:

- Do the links between risk and characteristics square with your experience?
- What real-world causes do you think explain the links?

- What do you look for when you make a field visit?
- What data do you gather in the field that is untrustworthy in the database?
- What characteristics would you recommend recording for use in future scorecards?
- When do you provisionally approve an application?
- How can you modify terms and conditions of the loan contract to manage risk?
- How much time per week do you spend in collections?
- How much time per week do you spend in marketing, evaluation, and disbursement?

Producing Scores and Reports

The next step is to automate the production of scores and scoring reports. Managers generally prefer to avoid changing their information systems, but to use scoring in the branches, there is no alternative to automation. There are two broad approaches. In the first, the microlender buys a ready-made scoring system—software and possibly hardware—from the consultant. This is quick and easy but expensive. It may also require entering data twice, once for the regular information system and once for the scoring system. In addition the parallel scoring system will not run on its own as soon as a case is entered; someone must start it manually. Finally, predicted risk in the parallel system cannot easily be integrated into the periodic reports that the lender already uses. If users must work to use scoring, then they are likely to ignore it.

In the second approach to automation, the microlender integrates the scorecard and associated reports directly into its existing information system. This is no small task. The microlender (or its software provider) must be able to modify the system and ded-

Figure 18: Example Scoring Simulator Report of Risk Forecasts after Modifying Loan Terms

Client: Jane Doe | **Branch:** Central | **App. No.:** 12345
Loan Officer: John Smith | **Committee:** 1/03/01 | **App. Date:** 1/1/01

	Amount Disbursed	Term-to-Maturity	Guarantee (% amt.)	Predicted Risk (%)
Requested Terms:	1,000	10	100	40
Amount Disbursed:	900	10	100	38
	800			33
	700			29
Term-to-Maturity:	1,000	9	100	37
		8		32
		7		27
Guarantee (% amt.):	1,000	10	125	39
			150	37
			200	36

Source: Author's example

icate a programmer full-time to scoring. Depending on the system, integration requires three to six person-months; the lender's systems manager cannot do this work on evenings and weekends. The technical challenges of integration will vary by lender, so all issues cannot be anticipated in advance. Integration, however, has important advantages: data is entered only once, scores are produced automatically, and risk forecasts can be easily integrated into the standard reports used by the lender. Weighing both pros and cons, integration is the preferred approach.

Once the scorecard is automated, the project enters a no-obligation, get-to-know-scoring phase. For several months, the system produces a score for all cases, but the project explicitly instructs loan officers and credit managers to do nothing in response to predicted risk and to look at the score only after the credit committee makes a decision. This allows loan personnel to acclimate to scoring slowly, and encourages them to think about how to use risk forecasts without pressure to change immediately.

This phase must create explicit time and space for feedback. People may seize on any apparent weakness or mistake to discount scoring, so concerns must be heard and addressed. This may require a second branch tour to review concepts, showing new tests of predictive power for disbursements made since the scorecard was installed, and asking more questions:

- Did the forecasts make sense to you?
- How often did the forecasts turn out to be accurate?
- Were there cases with high predicted risk that you knew from your experience to be in fact low-risk?
- What could you do to manage high-risk cases?
- How could you reward low-risk cases?
- What reports would help you to take advantage of scoring?
- What changes to the scoring process would you suggest?
- How would your performance bonus have changed if you had acted on scoring predictions for super-bad and borderline applicants?

Two Useful Reports

Credit committees commonly request to see how modifying borderline cases would affect the risk forecast. The Scoring Simulator Report responds to this. For example, Figure 18 shows how predicted risk might change as elements of the loan contract are varied one-by-one. These risk forecasts are the result of running the application through the scorecard again after modifying one of the terms of the loan contract. The Scoring Simulator Report comes in two forms. The first is an option within the information system for the credit committee to test modifications on the fly. The second is a paper report included in the bundle produced each day for the credit committee.

A second report, the Effects of Characteristics, responds to the request to know the reasons behind a risk forecast. For the given application, it shows the characteristics whose deviations from average historical values most increase risk and the characteristics that most decrease risk. (Figure 19 is an example.) A paper print-out of this report would also be included in the credit committee's daily reports.

Instituting Scoring Policies

Once loan personnel have had several months to familiarize themselves with scoring, the microlender institutes a scoring policy, distributes a written scoring policy manual, and begins using scoring in actual cases. Why a written policy? Without a written policy and explicit rules, it can be difficult to prevent staff from reverting to traditional credit evaluations. An explicit policy also helps minimize incorrect and inconsistent use of scoring. Just as with traditional credit evaluation, scoring needs a written policy.

Figure 19: Example Effects of Characteristics Report

Client: Jane Doe | **Case:** A 12345 | **Risk:** 30 Days of Arrears in a Row
Loan Officer: John Smith | **App. Date:** 6/2/02 | **History:** 1/1/95 to 5/1/02

Characteristic	Actual Value	Historical Average	Effect (% pts.)
1. Days of Arrears/Installments (in last paid-off loan)	8.7	1.7	+5.8
2. Number of Late Installments (in last paid-off loan)	6	4	+4.2
3. Experience of Loan Officer (# of loans disbursed)	77	535	+3.4
4. Type of Business Activity	Carpentry	N/A	+1.5
5. Telephone in Residence	No	Yes	+1.1
6. Term-to-Maturity (in last paid-off loan; # of months)	8	10.5	+0.6
7. Rotation of Capital (%)	Missing	326	+0.3
8. Repayment Burden (%)	20	18	+0.1
. . . . . .	. . .	. . .	
36. Guarantee Coverage (%)	350	300	–0.4
37. Client Gender	Woman	Woman	–0.7
38. Number of Employees	0	0.25	–1.9
39. Experience of Client (# of months)	36	14	–2.3
40. Client Age	55	43	–4.4
Risk Forecast:	23.2	9.3	+13.9

Source: Author's example

A written scoring policy should specify risk thresholds as well as actions for each threshold. For example, the policy establishes the risk level below which cases qualify as super-good and the risk level above which cases qualify as super-bad. It also establishes the risk levels that correspond to normal and borderline. Furthermore the written scoring policy tells how to reward super-goods. For borderlines it specifies how the credit committee should prioritize attempts to mitigate risk—whether by decreasing loan size, decreasing term-to-maturity, and/or increasing guarantee coverage. It also guides them in using the Scoring Simulator Report (see Figure 18) to see the likely effects of these possible modifications to the loan contract. Finally, the written scoring policy emphasizes that super-bad cases must be rejected.

Box 9: Why Was Scoring Wrong for This Borrower?

Like good weather forecasts, good scoring forecasts work on average, not for each day or for each individual loan. In fact the risk forecast never hits the mark for any single case; predicted risk is always greater than 0 percent and less than 100 percent, but realized risk is always 0 percent (did not go bad) or 100 percent (did go bad). For a given loan, it does not make sense to say scoring was right or wrong.

Forecasts from scoring are probabilities, not certainties. Accuracy is measured by comparing average predicted risk for a group with average bad loan rates (realized risk). If scoring works as it should, then some cases with high predicted risk will stay good and some cases with low predicted risk will go bad. For example, if scoring works, then half of all borrowers with 50 percent risk will stay good, and 1 in 20 of all borrowers with 5 percent risk will go bad.

Of course, scoring policy (unlike scoring forecasts) can turn out right or wrong for individual cases. Just as the choice to carry an umbrella because the weather forecast calls for a 60 percent chance of rain can be right (if it rains) or wrong (if it does not rain), the choice to approve or reject with the help of scoring can turn out to be right or wrong (although the correctness of reject decisions will never be known).

Override Policy

Scoring is most valuable as a way to identify high-risk cases that the credit committee thinks are safe bets. Loan officers and credit managers, however, are human, and when scoring contradicts their judgment, they may scoff and search for any small quirk to discredit scoring (such as one low-risk loan that went bad or one high-risk loan that stayed good—see Box 9). In the same vein, they may demand to know why risk is so high.

Choices that go against scoring policy are overrides. (In microcredit overrides are approved super-bads and unreviewed borderlines.) Override policy deals with this in three ways. First, it constantly tests predictive power via the Global Follow-up Report. Second, the override policy shows how risk is linked with characteristics via the Effects of Characteristics Report. Third, override policy does more than just urge users not to ignore scoring; it specifies consequences.

For example, microlenders can sanction excessive overrides through the performance bonus.[14] If overrides exceed *x* percent of super-bads, then the bonus is cut. In the long term, explicit sanctions are less necessary as loan officers realize that abuse of overrides leads to greater arrears and a smaller performance bonus.

Careful overrides do have their place. The credit committee may know that a particular case is exceptional, and only human judgment can evaluate the qualitative characteristics ignored by the scorecard. The point is moderation. Just as not all people can be above average, neither can all high-risk loans be overridden. In high-income countries, lenders try to limit overrides to 10 percent of super-bads. In microcredit a good goal might be 25 percent.

The microlender must track overrides to provide feedback to loan officers. In general overrides end up with less risk than was forecast (both because the credit committee does know something that the scorecard does not and because loan officers work extra hard to make their prophecies come true), but more risk than other loans (because the scorecard knows something that the credit committee does not).

Underride Policy

Override policy seeks to prevent too little dependence on scoring; underride policy seeks to prevent too much dependence on scoring. In particular written policy must stress (as does this paper) that scoring works only for applications already provisionally approved by the traditional evaluation process. Constant reminders are needed to help once-skeptical people find balance. After people see scoring work, they may neglect traditional evaluation. Scoring will understate risk—perhaps drastically—if it is applied to loans that have not already been provisionally approved under the standards of the lender's traditional subjective evaluation. To repeat a central point of this paper, a microlender cannot replace its subjective evaluation with scoring. It should add scoring after the subjective evaluation is completed, otherwise arrears may skyrocket.

The Global Follow-up Report

This report tracks the on-going performance of scoring. Like a historical test, it compares predicted risk with realized risk, but unlike a historical test, it applies to outstanding loans. The Global Follow-up Report is the central scoring report, even more useful than the historical test. It checks whether scoring works with live loans. Like other scoring reports, it is produced automatically by the system. In the first months of scoring, the lender consults it weekly to check the predictive power of the scoring and guide adjustments to policy. After that monitoring takes place monthly.

The first Global Follow-up Report covers outstanding loans that were not scored before disbursement, and like a historical test, it shows hypothetical predictive power. After a few months, the report reveals its predictive power for loans that were indeed subject to scoring before disbursement.

Figure 20 is a Global Follow-up Report based on a regression scorecard (discussed in Section VI) of a Latin American microlender. "Bad" is defined as an average of four days of arrears per installment due or a spell of arrears of 30 days.

The left-most column in Figure 20 ("Forecast Risk %") defines the range of predicted risk for each row. The lender defines the number of ranges as well as their boundaries. The second column from the left ("Number of Loans Outstanding") is the share of outstanding loans whose predicted risk falls within a row's range. It shows the distribution of predicted risk in the outstanding portfolio. For example, 0.5 percent of loans outstanding as of July 31, 2001, had predicted risk in the range of 0–2 percent. Likewise 9.5 percent of loans had predicted risk in excess of 40 percent (adding down the columns), and 19.5 percent had predicted risk in excess of 30 percent. (Numbers in the loans-outstanding column add to 100.)

The four center columns ("Realized Risk % by Days since Disbursement") show realized risk for outstanding loans given predicted risk and age. Comparing realized risk with predicted risk row-by-row reveals the scorecard's power. The closer predicted risk is to realized risk, the greater the predictive power. (The numbers in these columns do not add to 100.)

For example, realized risk was 5.3 percent for loans with predicted risk of 8–10 percent and aged 0–90 days (see Figure 20). That is, of the 1,394 outstanding loans that met the two criteria, 74 (5.3 percent) were bad as of the date of the report. In another example, loans with predicted risk above 70 percent and aged 271+ days had realized risk of 77.9 percent.

Figure 20 illustrates a general point: realized risk increases with age after disbursement. Two factors explain this. First, some recent loans have not had an installment come due yet, so they have not had a chance to go bad. Second, arrears increase toward the end of the loan.[15] Thus the best test of predictive power looks at recently paid-off loans and/or well-aged outstanding loans.

The right-most column of the example Global Follow-up Report shows realized risk for recently paid-off loans. (The lender determines how many months back the report will cover; the example uses 12.) This is the key column, both because it covers loans of all terms-to-maturity and because recently paid-off loans have had time to go bad.

Checking Predictive Power

The Global Follow-up Report checks whether a scorecard works. Absolute accuracy means that realized risk is close to predicted risk. In Figure 20, recently paid-off loans with predicted risk of 0–2 percent had realized risk of 3.2 percent (first row, right column). This is outside the predicted range, but it is close. Realized risk is within the predicted range for 2–4 percent, 4–6 percent, and 6–8 percent, and realized risk is higher than the top boundary in all other ranges. Absolute accuracy is good but not perfect because predicted risk is somewhat lower than realized risk for cases with high predicted risk.

Relative accuracy means that realized risk is lower for loans with lower predicted risk than for loans with higher predicted risk. The scorecard in Figure 20 has very good relative accuracy. Except for the lowest two ranges, realized risk increases with each range from the top of the figure to the bottom.

Tail accuracy means that absolute and relative accuracy are good in the extremes (tails) of the risk distribution. Tail accuracy matters because scoring policy does not affect cases with about average risk (normals). Scoring affects only the very low risks (super-goods) and the very high risks (borderlines and super-bads).

The scorecard in Figure 20 has excellent tail accuracy. For example, realized risk for recently paid-off loans with predicted risk of 0–2 percent was 3.2 percent. Realized risk for the ranges of 2–4, 4–6, and 6–8 percent were within the predicted range. On the high end, 75.4 percent of recently paid-off loans with predicted risk in excess of 70 percent went bad (bottom right corner). Among paid-off loans with predicted risk in excess of 40 percent, more than half went bad.

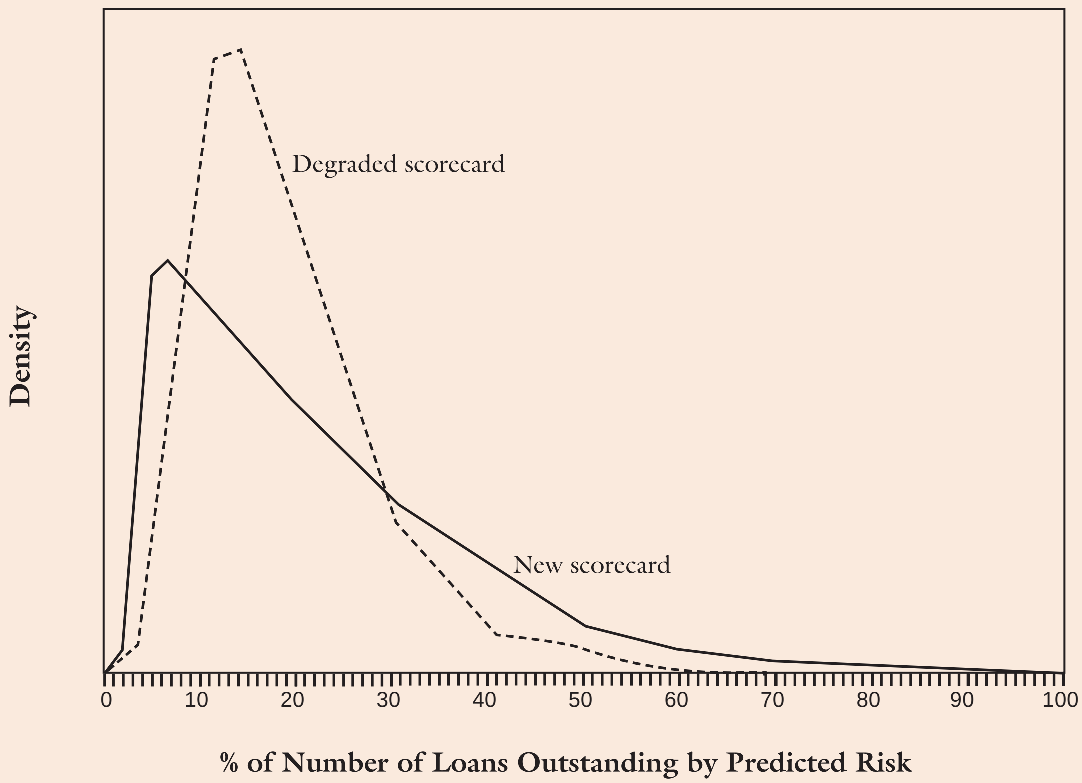
Figure 20: Example Global Follow-up Report

Risk: 4 Days/Installment or 30 in a Row **Quantity at Risk:** Number of Loans **Branch:** All
Date Tested: 6/2/02 **Date Scorecard Constructed:** 7/31/01

Forecast Risk (%)	Number of Loans Outstanding (%)	Realized Risk (%) by Days since Disbursement 0–90	91–180	181–270	271+	Realized Risk (%) for Loans Paid Off in Last 12 Months
0–2	0.5	1.4	2.0	0.0	4.0	3.2
2–4	5.1	2.8	2.8	2.1	3.5	3.1
4–6	7.8	3.0	4.0	4.0	5.1	4.7
6–8	8.1	3.9	4.8	5.5	8.1	7.8
8–10	7.7	5.3	6.7	6.4	11.5	10.6
10–15	17.0	5.5	8.1	11.6	18.1	16.3
15–20	14.5	6.8	12.1	17.9	27.6	24.7
20–25	11.4	9.0	16.9	23.8	33.1	27.2
25–30	8.4	11.4	19.4	30.4	37.8	36.3
30–40	10.0	14.6	25.0	37.3	45.8	43.1
40–50	5.1	18.4	30.4	50.9	53.6	52.6
50–60	2.7	23.0	42.3	57.2	60.4	60.1
60–70	1.2	32.4	42.6	65.2	70.5	70.3
70–100	0.5	34.3	62.9	65.5	77.9	75.4

Source: Scorecard applied to portfolio of a Latin American microlender

Figure 21: Example of Change in Distribution of Predicted Risk in New and Degraded Scorecards



Tracking Overrides

Loans disbursed with predicted risk greater than the super-bad threshold are by definition overrides. Overrides can be abused, so managers must track their outcomes. They do this by examining changes through time in realized risk among disbursed super-bads. The baseline for comparison is realized risk before scoring began. If, as loans disbursed under scoring age, realized risk among super-bads is far less than predicted risk, then overrides have been successfully limited, on average, to cases where predicted risk was greatly overestimated. If the reduction in realized risk is so great that the lender would want to approve loans known to have that level of risk, then the current limits on overrides should be maintained. Otherwise the limits should be tightened until realized risk among overrides is acceptably low. For example, suppose that the super-bad threshold is 70 percent, and suppose that the Global Follow-up Report run on the first day after scoring is launched shows 78 percent realized risk among past loans that would have qualified as super-bad. After a year of scoring, suppose that the Global Follow-up Report reveals that realized risk among overrides (loans disbursed with predicted risk in excess of 70 percent) was 35 percent. This suggests that the credit committee limited, on average, overrides to cases with overestimated risk. This 35 percent may be more risk than the lender wants to bear, and if so the lender would tighten override limits. If the lender is willing to make loans this risky, then the current override policy would be maintained.

Fixing Absolute Inaccuracies

Scorecards with absolute accuracy are easier to use. Relative accuracy merely orders loans by expected risk. For example, loans with 10 percent predicted risk have less realized risk than loans with 20 percent predicted risk, but realized risk for the two groups might turn out to be 7 percent and 25 percent. With absolute accuracy, loans with 10 percent predicted risk not only have 10 percent realized risk but also have exactly half the risk of loans with 20 percent predicted risk.

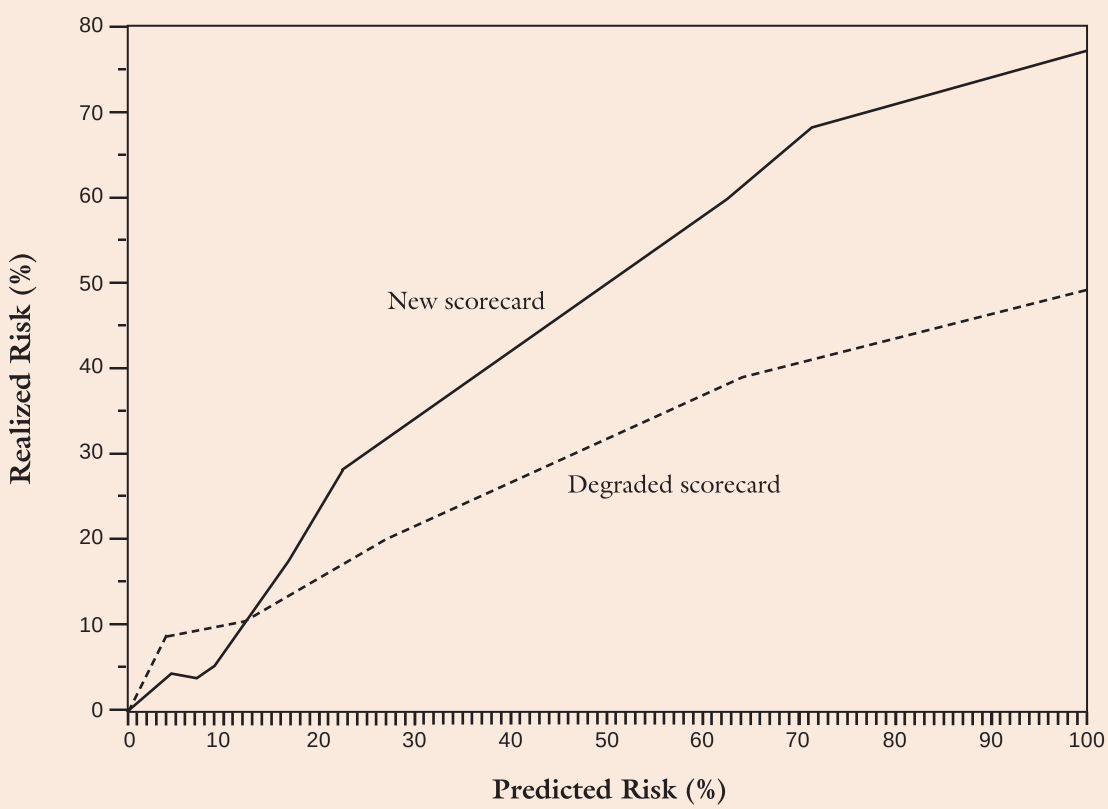
Unfortunately no scorecard has perfect absolute accuracy. The Global Follow-up Report, however, shows the levels of realized risk that correspond to given levels of predicted risk. With this information, the user can adjust the levels of predicted risk so that the adjusted predictions are absolutely accurate.

Suppose that the Global Follow-up Report shows that predicted risk is always 5 percentage points too high. The lender then simply acts as if loans with, for example, 25 percent predicted risk had 20 percent predicted risk. In real life, the patterns of inaccuracies are more complex, but the conversion principle still works, and the information system can make the conversion automatically.

Setting or Adjusting Policy Thresholds

The Global Follow-up Report shows the share of loans in each risk range and the level of realized risk that corresponds to a given level of predicted risk. Thus the microlender can use the Global Follow-up Report to set or adjust policy thresholds.

Figure 22: Example of Change in Relationship between Predicted Risk and Realized Risk in New and Degraded Scorecards



For the scorecard in Figure 20, a super-good threshold of 2 percent would have affected 0.5 percent of outstanding loans (second column from the left, first row), whereas a super-good threshold of 4 percent would have affected 5.6 percent of outstanding loans. A super-bad threshold of 70 percent would have rejected 0.5 percent of loans now outstanding. Furthermore such a super-bad policy would have avoided three bad loans for each good loan lost (because realized risk in this range is about 75 percent—see bottom right corner). If the super-bad threshold were reduced to 30 percent, then 19.5 percent of loans would have been rejected, and about half would have been bad.

Detecting Scorecard Degradation

Because the future resembles the recent past more than it resembles the distant past, the predictive power of a scorecard degrades with time. The Global Follow-up Report shows this in two ways. The first is a more peaked (less spread out) distribution of predicted risk. Degradation moves the typical prediction closer to the average prediction. Figure 21 is a hypothetical example in which the distribution of predicted risk for the new scorecard is based on the first two columns of the Global Follow-up Report in Figure 20.

The second indicator of degraded predictive power is a less steeply sloped (flatter) relationship between predicted risk and realized risk. With degradation realized risk exceeds predicted risk at low levels of predicted risk. Furthermore degradation means that realized risk is less than predicted risk at high levels of predicted risk. Figure 22 is a hypothetical example in which the relationship of predicted risk to realized risk for the new scorecard is based on the second-to-last column of the Global Follow-up Report in Figure 20.

To detect the extent of degradation, managers compare the distribution of predicted risk (and/or the relationship between predicted risk and realized risk) in the Global Follow-up Report when a given scorecard was new against the most recent report. Graphs such as Figures 21 and 22 make the changes in the data in the Global Follow-up Report stand out.

The speed of degradation depends on the rate of change in lending policy, target niches, competition, portfolio growth, the macroeconomy, and other variables that both affect risk and change over time. Before degradation advances too far (probably after two to four years), the microlender should renovate the scorecard. Renovation is simpler and quicker than the initial scoring project. A new scorecard is constructed, including data accumulated since the first scorecard, and plugged into the existing system.

The Loan Officer Follow-up Report

The Global Follow-up Report is central to scoring, but for loan officers and credit managers, it may be too abstract (because it compares predicted and realized risks for groups of loans) and too broad (because it covers all outstanding loans and all recently paid-off loans). Technically the Global Follow-up Report is the best test of the predictive power of scoring, but front-line personnel seem to prefer simpler reports that allow them to compare predicted risk with repayment performance for the individual borrowers they know personally.

One such report, the Loan Officer Follow-up Report, adds measures of predicted risk and repayment performance (realized risk) to the portfolio reports that loan officers and credit managers receive daily or weekly. Figures 23 and 24 are simple reports from a regression scorecard (see Section VI) of a Latin American microlender who defines “bad” as at least one spell of arrears of 30 days during the lifetime of the loan. These Loan Officer Follow-up Reports differ from historical tests by covering outstanding loans, and they differ from the Global Follow-up Report by including the names of individual borrowers.

Figure 23: Example Loan Officer Follow-up Report with 30 Highest-Risk Cases Disbursed More Than 270 Days

Report Date: 7/31/01 **Branch:** All **Risk:** 1 Spell≥30 Days **Loans:** Outstanding>270 Days **List:** 30 Most Risky

						Realized Risk					
Loan Code	Client Name	Days Out	Amount Out ($)	Monthly Payment	Next Due	Current Arrears	No. of Spells	Days of Arrears/ Installments	Longest Spell	Bad?	Predicted Risk (%)
79922	Javela, María	308	2,106	83	03 Aug	23	2	42.5	77	Bad	90
50973	Posada, María	334	1,860	71	29 Aug	0	3	21.1	36	Bad	81
71596	Arboleda, Nivelly	336	1,323	132	29 Aug	2	3	14.8	25	Good	80
80816	Beltrán, Dioselina	304	1,032	48	29 Aug	0	3	14.8	42	Bad	80
62037	Nuñez, Dolly	337	5,683	316	02 Aug	0	1	22.7	28	Good	72
45638	Cruz, Leonor	304	377	22	29 Aug	0	3	45.5	101	Bad	71
64823	Rivera, Antonia	304	603	39	29 Aug	23	2	22.2	39	Bad	68
61653	Marín, Graciela	337	5,763	283	02 Aug	0	4	14.5	25	Good	62
78800	Muñoz, Marco	304	2,003	111	29 Aug	0	3	25.7	67	Bad	60
24893	Silva, Oswaldo	304	388	29	29 Aug	86	2	36.0	86	Bad	59
65323	Ruíz, Asia	308	56	12	03 Aug	58	4	24.7	58	Bad	59
59506	Cardona, Graciela	334	188	51	29 Aug	0	2	11.9	18	Good	59
54093	Tejada, María	285	14,638	790	11 Aug	0	1	0.3	2	Good	58
71243	Castillo, Rosa	293	630	70	18 Aug	0	2	6.1	15	Good	58
22692	Tavárez, María	348	143	39	13 Aug	0	1	0.4	2	Good	58
99155	Marroquín, Libia	334	77	41	29 Aug	0	1	11.1	22	Good	58
18634	Rivera, Melida	334	470	50	29 Aug	191	2	82.7	191	Bad	57
74810	Marulanda, Pablo	304	331	27	29 Aug	23	3	25.8	54	Bad	56
20410	Valencia, Claudia	356	323	53	21 Aug	0	4	5.5	14	Good	55
60737	Suárez, Yolanda	335	275	40	03 Aug	0	1	0.5	2	Good	55
85854	Marín, Jorge	308	1,275	106	03 Aug	0	4	7.7	20	Good	55
42074	Lozano, Nevalia	292	251	19	18 Aug	86	2	52.0	93	Bad	54
30986	Berrios, Fanny	318	2,449	136	13 Aug	0	2	4.4	15	Good	54
31208	Gomez, Diafanor	306	6,049	291	01 Aug	0	3	4.5	12	Good	54
89020	Calderón, Editha	319	259	38	14 Aug	0	1	7.0	14	Good	54
8408	Marulanda, María	306	332	42	01 Aug	0	2	61.6	131	Bad	53
36244	Castillo, Brunilda	279	383	46	05 Aug	0	1	0.9	3	Good	52
5699	Ortíz, Nubia	334	570	46	29 Aug	0	2	15.5	39	Bad	52
7719	Montoya, Javier	281	100	17	07 Aug	36	3	12.6	36	Bad	52
40373	Moreno, Peregrino	304	381	50	29 Aug	177	4	68.9	177	Bad	51
								Average Risk:		50	61

Source: Regression scorecard and database of Latin American microlender

Figure 24: Example of Loan Officer Follow-up Report with 30 Lowest-Risk Cases Disbursed More Than 270 Days

Report Date: 1/12/01 **Branch:** All **Risk:** 1 Spell≥30 Days **Loans:** Outstanding>270 Days **List:** 30 Least Risky

						Realized Risk					
Loan Code	Client Name	Days Out	Amount Out ($)	Monthly Payment	Next Due	Current Arrears	No. of Spells	Days of Arrears/ Installments	Longest Spell	Bad?	Predicted Risk (%)
62225	Valencia, Lucero	292	59	60	18 Aug	0	0	0.0	0	Good	0.5
38388	Betancourt, José	305	73	26	01 Aug	0	1	0.1	1	Good	0.5
88687	Valencia, Juan	279	35	36	05 Aug	0	0	0.0	0	Good	0.5
94799	Fernández, Zorrilla	281	289	38	07 Aug	0	0	0.0	0	Good	0.5
8154	Sánchez, Hernán	290	102	36	16 Aug	0	0	0.0	0	Good	0.5
38563	Escobar, Patricia	316	117	32	11 Aug	0	1	7.0	13	Good	0.5
27819	Echandia, Henry	322	102	36	17 Aug	0	0	0.0	0	Good	0.6
21502	Jaramillo, Ema	285	289	103	11 Aug	0	1	0.1	1	Good	0.6
71907	Guervara, César	295	87	31	20 Aug	0	0	0.0	0	Good	0.6
49562	Paz, María	336	768	167	01 Aug	0	1	0.8	5	Good	0.6
93142	Escobar, Mónica	284	35	36	10 Aug	0	0	0.0	0	Good	0.6
11221	Palomino, Fe	287	73	26	13 Aug	0	0	0.0	0	Good	0.7
88301	Garcia, Alberto	308	289	38	03 Aug	0	0	0.0	0	Good	0.7
77258	Arce, Eduardo	305	116	41	02 Aug	0	1	1.0	5	Good	0.7
1582	Contreras, Elena	318	147	77	13 Aug	0	1	0.1	1	Good	0.7
79476	Sánchez, Gonzalo	323	293	65	18 Aug	0	1	1.4	5	Good	0.7
985	Lopez, Flor	295	35	36	20 Aug	0	0	0.0	0	Good	0.7
85657	Torres, Mariá	280	347	46	06 Aug	0	0	0.0	0	Good	0.7
16697	Chacón, Emilsa	293	73	26	18 Aug	0	1	4.0	20	Good	0.7
53165	Gutierrez, Lucila	356	153	55	21 Aug	0	0	0.0	0	Good	0.7
80399	López, Alejandro	291	460	86	17 Aug	0	1	0.1	1	Good	0.7
32949	Castaño, Alvaro	323	68	36	18 Aug	0	0	0.0	0	Good	0.7
94131	Duque, Lucia	287	219	78	13 Aug	0	0	0.0	0	Good	0.7
28050	Polanco, Gerardo	294	76	79	19 Aug	0	1	0.1	1	Good	0.7
30709	Fajardo, Carmen	349	101	103	14 Aug	0	0	0.0	0	Good	0.7
54730	Aristiza, Morena	287	73	26	13 Aug	0	0	0.0	0	Good	0.7
18377	Ceballos, Luís	314	168	45	09 Aug	0	0	0.0	0	Good	0.7
28881	Escobar, José	323	78	41	18 Aug	0	0	0.0	0	Good	0.8
34129	Muñoz, Edisón	283	461	86	09 Aug	0	0	0.0	0	Good	0.8
74078	Tabarez, Jesús	341	50	51	06 Aug	0	1	0.2	1	Good	0.8
								Average Risk:		0.0	0.6

Source: Regression scorecard and data base of Latin American microlender

For super-bad loans, Figure 23 shows the 30 highest-risk outstanding loans that were disbursed at least 270 days before the date of the report. In this group of outstanding loans, average predicted risk is 61 percent (bottom right corner), and average realized risk is 50 percent. Even the 15 "good" loans are not that good; all 15 had some arrears, and all but four had a spell longer than ten days. When loan officers see their own borrowers in such a list, and when they recall the troubles they had collecting from these borrowers, they may start to see the value of scoring.

On the super-good side, Figure 24 shows the 30 lowest-risk loans. Average predicted risk is less than 1 percent (bottom right corner), and not a single case turned out bad. In fact 19 of the 30 cases had no arrears at all. Of the 11 cases with arrears, six had only one day, and only two had more than ten days.[16]

For loan officers and branch managers, seeing their own borrowers in reports, such as Figures 23 and 24, goes a long way toward dispelling doubts that scoring can identify high-risk and low-risk cases among those already approved by the credit committee. Microlenders who score should add Loan Officer Follow-up Reports to the standard daily and weekly reports distributed to loan officers and credit managers.

If employees give scoring a chance, they will see that it works, but they must understand it and believe that success is likely. This then is the task of training and tests. Once scoring is accepted, proper use depends on a written policy, strict control of overrides, and constant monitoring. Follow-up reports that compare predicted risk with realized risk for outstanding loans—both for the global portfolio and for each loan officer—provide the necessary constant reinforcement.

VI. Regression Scorecards and Expert Systems

This section presents regression, a type of scorecard that is more complex—and more powerful—than trees. It also presents expert systems (a third type of scorecard) and then compares and contrasts regression scorecards, trees, and expert systems.

Regression Scorecards

A regression scorecard is a mathematical formula that produces forecasts (probabilities) by adding up the weighted values of the characteristics of the borrower, loan, and lender. The characteristics selected for the formula and their weights are derived from complex statistical techniques not discussed here. Using regression forecasts, however, is like using tree forecasts, and the information system handles all the calculations. Compared with trees and expert systems, regression predicts best and also shows most clearly the links between risk and characteristics.

Suppose statistical work finds that risk decreases with the age of the borrower at a rate of 0.1 percentage points per year. Statistical work further finds that "base risk" is 10 percentage points. The regression formula that forecasts the probability of a loan being bad is thus:

$$\text{Risk} = 10 - 0.1 \times \text{Age}.$$

Given this equation, predicted risk for a 30-year-old borrower is $10 - 0.1 \times 30 = 7$ percentage points. For a 55-year-old, predicted risk is $10 - 0.1 \times 55 = 4.5$ percentage points. (These weights are examples. Real weights are lender-specific.)

In a second example, suppose statistical work finds that risk increases with the term-to-maturity at a rate of 0.25 percentage points per month. Given a base risk of 10 percentage points, the regression forecast is then:

$$\text{Risk} = 10 + 0.25 \times \text{Term-to-Maturity}.$$

Thus predicted risk for a three-month loan is $10 + 0.25 \times 3 = 10.75$ percentage points. For a 12-month loan, predicted risk is $10 + 0.25 \times 12 = 13$ percentage points.

Figure 25: Relationship in Regression Scorecard between Risk and Number of Months since First Disbursement

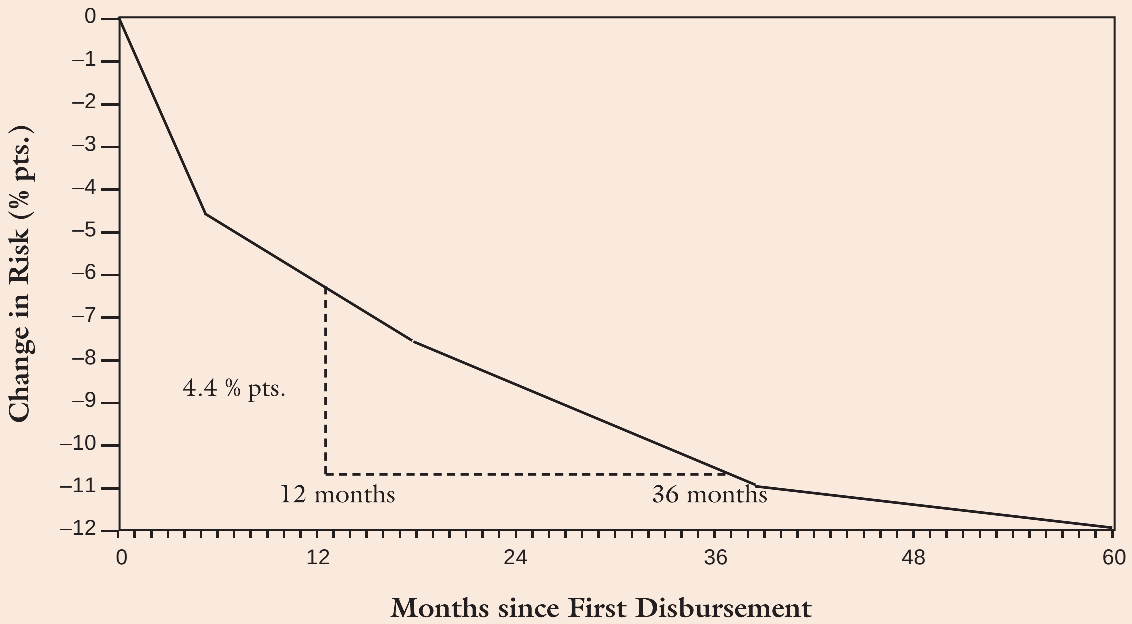
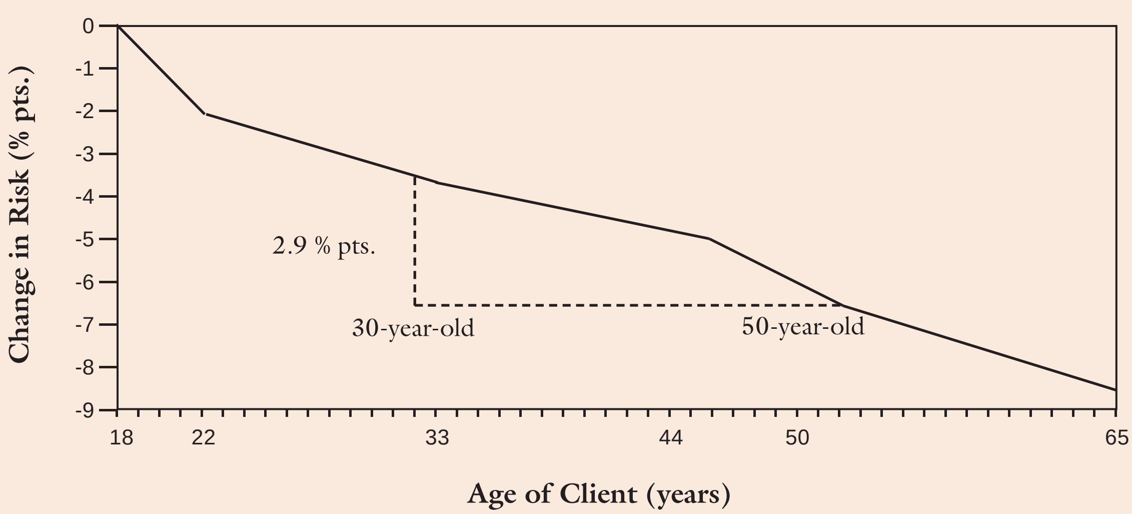


Figure 26: Relationship in Regression Scorecard between Risk and Age of Client



In practice, regression scorecards include a wide range of characteristics. For example, a scorecard combining the two one-characteristic formulae above more finely distinguishes between high and low risks:

Risk = 10 – 0.1 x Age + 0.25 x Term-to-Maturity.

For example, a 30-year-old borrower with a 36-month loan has a predicted risk of 10 – 0.1 x 30 + 0.25 x 36 = 16 percentage points. In contrast a 55-year-old with a three-month loan has a predicted risk of 10 – 0.1 x 55 + 0.25 x 3 = 5.25

percentage points. In practice a regression scorecard might include 30 to 50 characteristics and would derive all weights from a particular microlender's database. After the information system computes the forecast, the lender uses it as described in previous sections.

Links between Risk and Characteristics from Regression Scorecards

Although regression has the best predictive power of all types of scorecards, perhaps its greatest advantage is that it clearly shows the relationship between risk and characteristics. The weight assigned to a

Figure 27: Relationship in Regression Scorecard between Risk and Applicant Indebtedness

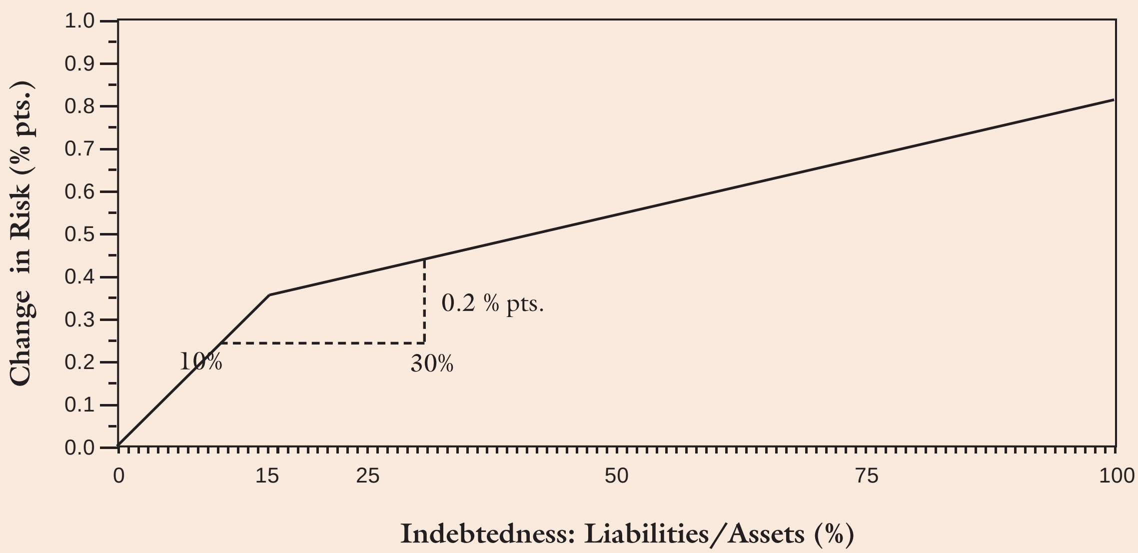


Figure 28: Relationship in Regression Scorecard between Risk and Arrears in Previous Three Loans

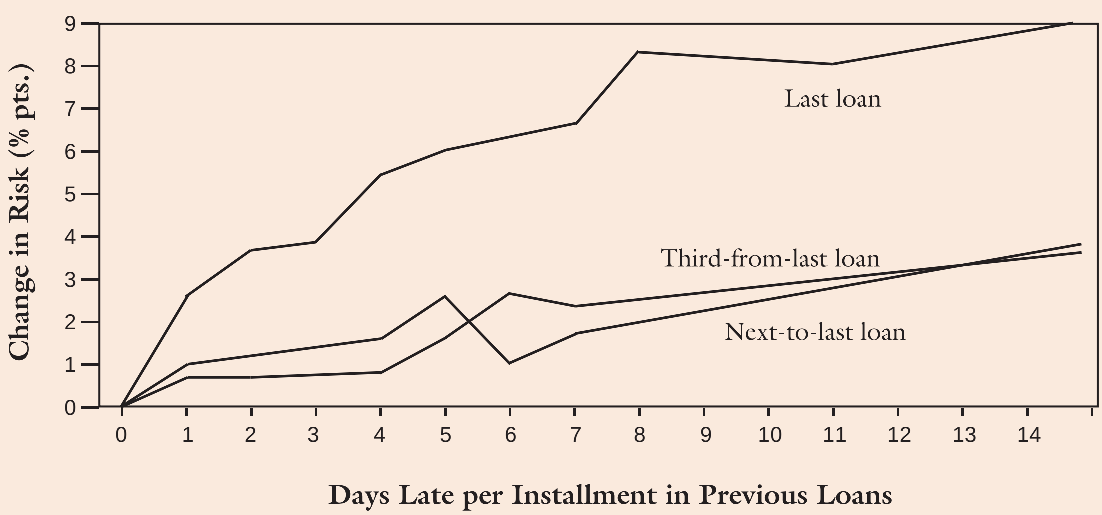


Figure 29: Relationship in Regression Scorecard between Risk and Type of Business

Type of Business	Effect on Risk (%)	Portfolio Share (%)
Trucking and taxi driving	–3.6	0.4
Fruits and vegetables sales	–3.5	2.3
Corner grocery store	–2.6	4.3
Small household items store	–2.1	6.4
Ambulatory sales	–2.0	4.4
Beauty salon	–2.0	2.7
Bakery	–1.9	2.3
Cosmetics sales	–1.9	1.6
Grocery store	–1.7	2.3
Seamstress and clothes-making	–1.3	11.1
Prepared food sales	–1.0	1.0
Schools	–1.0	0.6
Food processing	–1.0	0.6
Auto-parts store	–0.6	0.7
Street fast food	–0.6	0.5
Meat market	–0.5	1.4
Home appliance sales	–0.5	1.0
Clothing store	–0.2	1.6
Other or unknown	0.0	39.5
Shoe stores	+0.1	2.5
Pharmacies	+0.3	1.9
Sit-down restaurants	+0.7	1.7
Hardware stores	+0.8	1.1
General stores	+0.9	4.1
Professional services	+1.0	0.6
Artwork	+1.2	0.8
Locksmith and metalworking	+1.6	0.7
Auto mechanics	+1.7	0.5
Shoemaking	+2.1	1.0
Carpentry	+2.6	0.5

Source: Latin America microlender

characteristic shows not only whether the characteristic increases or decreases risk—other characteristics in the scorecard kept constant—but also by how much. These links hold only after an application is provisionally approved by traditional evaluation. The examples shown here are from a real-life regression scorecard of a Latin American microlender.

Relationships between Risk and Characteristics

The regression scorecard in Figure 25 shows that risk decreases strongly as the number of months since disbursement grows. For example, a borrower at 36 months past disbursement has—all else constant—4.4 percentage points less risk than someone 12 months past. Risk also decreases significantly with age. In Figure 26 for example, a 50-year-old has—all else constant—about 2.9 percentage points less risk than a 30-year-old.

Risk increases with the indebtedness ratio of liabilities to assets in the household/enterprise, as depicted in Figure 27. Someone with 10 percent indebtedness would—all else constant—have 0.2 percentage points less risk than someone with 30 percent indebtedness. Risk also increases with the average days of arrears per installment in each of the three previous loans (see Figure 28). For example, ten days of arrears in the last loan increases current risk by 8 percentage points, and seven days in the next-to-last loan increases current risk by 2 percentage points. The effect on current risk of arrears in the third-to-last loan is very similar to the effect of arrears in the second-to-last loan.

Thus compared to a borrower with a perfect record, someone who averaged 10, 7, and 7 days of arrears in the last three loans would have about 8 + 2 + 2 = 12 percentage points more risk in the current loan (assuming it had already been provisionally approved according to traditional evaluation standards).

Figure 28 offers four broad lessons about the relationship between future arrears and past arrears for a given borrower. One, more realized risk in the past means higher predicted risk in the future. Two, arrears in the distant past are weaker signals than are arrears in the recent past. Three, compared with a perfect record, even short spells of arrears in the past signal much higher risk in the future. For example, a one-day average in the previous loan increases current risk by more than 2 percentage points. Given that the overall bad rate for this microlender is less than 15 percent, a two-percentage-point change is large. Four, risk increases with past arrears but at a diminishing rate (although this relationship holds only for provisional approvals).

Other Links

The type of business is strongly related to risk. For the microlender in Figure 29 ("Effect on Risk" column, in descending order), the lowest risks were:

- taxi and truck drivers
- stores whose inventory rotates quickly (fruits and vegetables, groceries, small household items)
- street food venders (fast foods, bakeries)
- beauty salons and cosmetic stands
- seamstresses

The business types with the highest risks for this lender were:

- manufacturers (carpenters, shoemakers, auto mechanics, and locksmiths)
- professionals and artists
- stores whose inventory rotates slowly (hardware, pharmaceuticals, shoes, clothes, home appliances, and auto parts)
- sit-down restaurants

Figure 30: Relationship in Regression Scorecard between Risk and Individual Loan Officer

Loan Officer	Effect on Risk (%)
Carmen Ochoa	–10.1
Catalina González	–9.0
David Soto de los Santos	–5.7
Rosario Sosa Almanecer	–3.9
Mariangelí Cintrón Ruíz	–2.0
Rosa Justiniano Orñes	–0.2
Others	0.0
Ma. Eugenia Mariscal	+1.1
Marcos Orta	+2.3
Eldo Parra Barriga	+3.0
Oscar Navajas	+3.3
Teresa Guzmán	+4.9
Enrique Flores Santos	+7.0
María Padilla Ruíz	+13.6

Figure 31: Example Expert System Tree

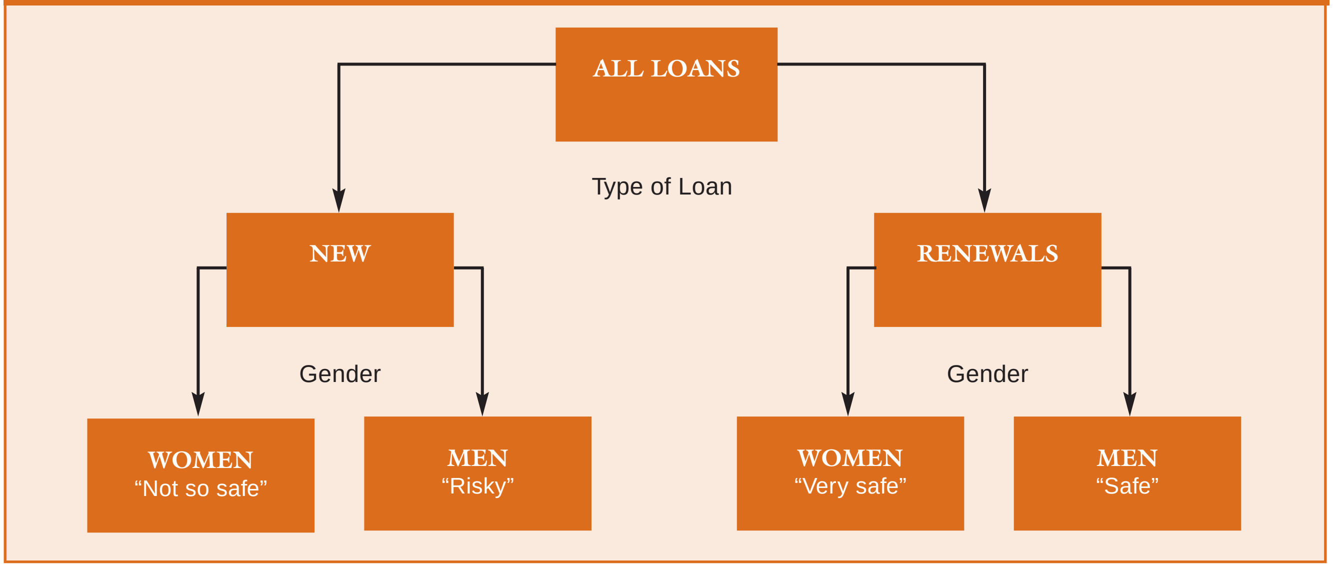


Figure 29 shows the share of the historical portfolio for each type of business. This lender clearly concentrated on low-risk businesses.

Loan Officer

Regression can also reveal the link between risk and a particular loan officer. In Figure 30, the links are strong, with wide ranges between loan officers. In this example, almost 24 percentage points separate the top and bottom loan officers.

The loan officer in charge of a loan affects risk a great deal, but only regressions—not trees or expert systems—use this knowledge to boost the accuracy of predicted risk. Regression can help the lender target training, encouragement, and bonuses.

A caveat applies to interpreting Figure 30. Loan officers manage risk by screening applicants before disbursement and by monitoring loans after disbursement. Regression reveals the effectiveness of monitoring but not the effectiveness of screening. This is because regression measures the effect of the loan officer with all other characteristics in the regression constant, as if all loan officers managed portfolios with the same quantified characteristics. In fact loan officers manage different portfolios, whose composition (both quantified and qualitative) depends on how well the loan officer screens applicants. Some loan officers achieve a given level of portfolio risk by screening for applicants who do not need much monitoring. Others achieve the same level of portfolio risk with less screening and more monitoring. Further, some loan officers are assigned to tough neighborhoods where a given level of skill and effort is less effective than it would be elsewhere. Thus lenders should not immediately fire loan officers who rank low on the regression scorecard but should investigate the reasons for the low ranks and work to address them.

Expert Systems

Scorecards derived from the experience and judgment of managers, and not from statistical analysis of data, are called expert systems. Expert systems differ from traditional subjective scoring in that subjective scoring uses implicit judgments, while expert systems use explicit rules or mathematical formulae. The strength of expert systems is that they do not require a database and that—because they are constructed by a microlender’s managers and loan officers—they are less difficult to sell within the organization. The weakness of expert systems is that they have less predictive power than trees or regressions. Also, because expert systems assume links between risk and characteristics, they cannot reveal links. Most microlenders who claim to use scoring today are running what amount to expert systems.

Expert system trees are like statistical trees, except their splits come not from a statistical analysis of the database by a consultant but from the experience, judgment, and guesswork of the lender's managers and loan officers. The result is a tree whose leaves show qualitative ranks, not quantitative probabilities. For example, the statistical tree in Figure 2 forecasts a risk of 12.8 percent for renewal loans to women, but the expert-system tree in Figure 31 ranks these same renewal loans to women as "very safe." The most common expert-system tree in microcredit today is the arrears-based grade (see Box 2).

Expert-system regressions are mathematical formulae (like statistical regressions), but managers choose the characteristics and their weights rather than derive them from data. Expert-system regressions produce a number which is a rank and not a probability, so scores may exceed 100 or be negative. Thus expert-system regressions lack absolute accuracy, although they may achieve some level of relative accuracy.

All expert systems—be they trees or regressions—can be improved by using tests of predictive power to translate ranks into probabilities. Historical tests and follow-up reports apply to expert systems as they do to statistical scorecards. Rather than compare predicted risk as a probability with realized risk, however, tests of expert systems compare predicted ranks with realized risk. A lender can use the tests to convert

Figure 32: Example Policies for Five Types of Risk

Type of Risk to Be Forecast	Examples of Policy Actions
1. Pre-disbursement: If disbursed, will this loan reach some level of arrears in its lifetime?	Super-bad: Reject Borderline: Modify terms Normal: Disburse as is Super-good: Offer rewards and enhancements
2. Post-disbursement: Will this borrower be late on the next installment?	Presumed guilty: Pay "courtesy visit," make phone call, or write letter Presumed innocent: Wait and see
3. Collections: Will this loan, currently x days in arrears, reach $x + y$ days?	High risk *and* high value-at-risk: Visit now and skip gentle tactics High risk *or* high value-at-risk: Visit now but use gentle tactics Low risk *and* low value-at-risk: Visit later and then dun gently
4. Desertion: Will this borrower apply for another loan once the current one is paid off?	Kick-outs: Cannot repeat due to poor repayment performance Unsafe waverers: Wait and see, no incentives Safe waverers: Offer incentives to repeat Loyalists: Wait and see, no incentives
5. Visit: Will the lender reject the application after the field visit by the loan officer?	Unpromising: Reject without a field visit Promising: Proceed with visit

non-probabilistic ranks into probabilities and then work only with probabilities.

More importantly historical tests and follow-up reports show the extent of predictive power. If managers do choose sub-optimal splits and sub-optimal weights, expert systems may nonetheless be usefully predictive.[17] Further, expert systems may compensate for their low predictive power with their low data requirements and ease of adoption.

Microlenders should feel free to experiment with simple home-grown scorecards,[18] but they should test them before and during use. Incredibly most microlenders that use expert systems have not tested them. Their mistake is not that they use expert systems rather than statistical scorecards but that they neglect to test predictive power. Those who score should walk by sight, not faith.

Regressions have the greatest predictive power and they also reveal links between risk and characteristics, better than trees or expert systems. Regression, however, is complex, and it makes the greatest demands on a database. Only the largest and most sophisticated microlenders are ready for regression scorecards.

Trees—even do-it-yourself trees—forecast surprisingly well, and they require less data than regression. Like expert systems, trees are simple to explain and to sell to personnel, but they do not always clearly reveal links between risk and characteristics.

Expert systems are easy to construct because they do not require data. While this makes them the most relevant type of scorecard for microlenders today, their downside is that they do not predict as well as trees or regressions. Microlenders who lack the data required for statistical scoring might start with an expert system, but they should also begin to collect the data needed to support a better scorecard.

VII. Preparing to Score: What Type of Risk to Forecast?

The first scoring project should simply be to construct a single scorecard. The lender must choose from pre-disbursement scoring, post-disbursement scoring, collections scoring, desertion scoring, or visit scoring (see Figure 32). Most will choose pre-disbursement scoring (the type discussed so far in this paper) both because the four-class policy is simple and useful and because a pre-disbursement risk forecast can stand in for post-disbursement and collections scores.

Pre-Disbursement Scoring

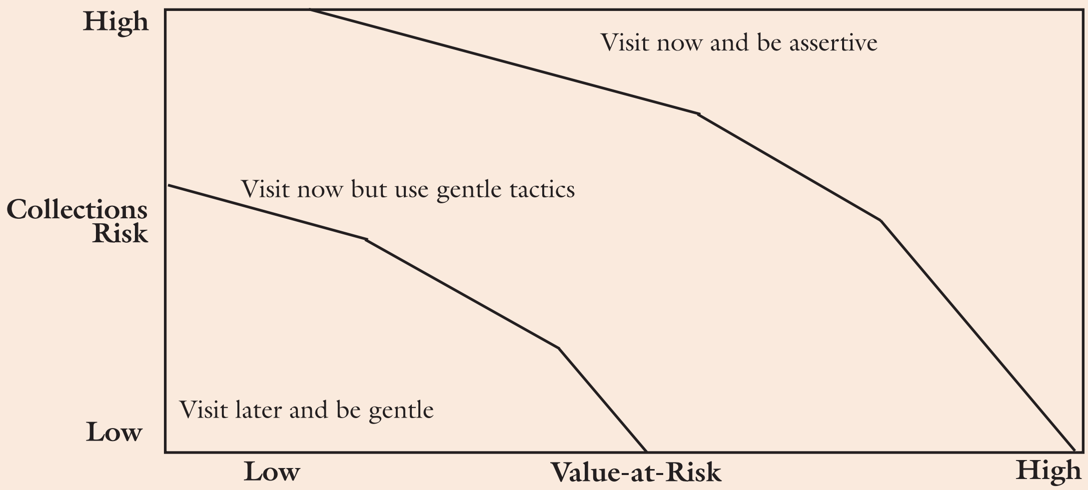
Pre-disbursement scoring predicts the probability that a provisionally approved loan, if disbursed, will go bad sometime in its life. The lender must choose how to define "bad," usually as some combination of:

- a spell of arrears in excess of *x* days
- more than *y* spells of arrears, regardless of length
- more than *z* average days of arrears per installment

Defining "bad" for scoring can be a healthy exercise. It forces the microlender to think carefully about arrears and costs—such as, whether the number of spells or their length matters more and whether numerous short spells can be tolerated. Lenders should also ask themselves what criteria they currently use to determine whether to give another loan to a client with some arrears in the previous loan.

For pre-disbursement scoring, the definition of bad should not be "default." On a technical level, most microlenders have too few historical defaults to reveal relationships between risk and characteristics. More importantly most microlenders consider a loan to be bad long before it goes into default. Loan officers do not ask themselves "if I approve this

Figure 33: A Three-Class Collection Policy



loan, will I eventually collect it?" but rather "if I approve this loan, will I need to work a lot to collect it?" As evidence of this, most microlenders have policies to refuse repeat loans to borrowers who, even though they did not default on the previous loan, had a lot of arrears at some point.

Post-Disbursement Scoring

Post-disbursement scoring predicts the probability that the next installment on an outstanding loan will be late. Risk after disbursement is highly correlated with risk before disbursement. Both types of scorecards forecast from the same set of characteristics, except that the post-disbursement scorecard also includes the repayment record in the current loan, the number of installments already paid in the current loan, and the balance outstanding. A pre-disbursement score is an effective surrogate for a post -disbursement score; loans with high risks before disbursement also have high risks after disbursement. A pre-disbursement score is a poor substitute for a post-disbursement score only in cases where post-disbursement risk is already obvious to the lender—such as outstanding loans with severe arrears since disbursement.

Regardless of the scorecard used to forecast post-disbursement risk, there is a simple two-class policy choice (see Figure 32). The loans with the highest risks (or perhaps the highest value-at-risk) are presumed guilty, a class that might cover 5 percent of all loans. Even before trouble starts, they receive a preventive courtesy visit, phone call, or letter. All others are presumed innocent, and the microlender does nothing special until they actually fall into arrears.

The Loan Officer Follow-up Report (see Figures 23 and 24) helps loan officers decide who to visit. For example, candidates from the list in Figure 23 would include three high-risk, high-value loans that have yet to go bad:

- $6,049 outstanding with predicted risk of 54 percent
- $14,638 outstanding with predicted risk of 58 percent
- $5,683 outstanding with predicted risk of 72 percent

Figure 34: A Four-Class Desertion Scoring Policy

<table>
<tr><th colspan="2" rowspan="3"></th><th colspan="3">Traditional Credit Evaluation Norms</th></tr>
<tr><th rowspan="2">Disqualified</th><th colspan="2">Qualified</th></tr>
<tr><th>High Pre-Disbursement Risk</th><th>Low Pre-Disbursement Risk</th></tr>
<tr><th rowspan="2">Desertion Risk</th><th>High</th><td rowspan="2">Kick-outs:
No incentives</td><td>Unsafe waverers:
No incentives</td><td>Safe waverers:
Incentives offered</td></tr>
<tr><th>Low</th><td colspan="2">Loyalists: No incentives</td></tr>
</table>

In the courtesy visit, loan officers simply call on the borrower—unrelated to any current collection issue—and discuss any non-threatening topic. The loan officer should never let on to clients that they scored as high risks, lest it become a self-fulfilling prophecy. Borrowers in good standing are likely to take offense if they feel suspected. The mere presence of the loan officer is enough to reinforce the importance of timely repayment in the mind of the borrower. Loan officers can take advantage of the visit to get feedback, asking the clients how the disbursement went, what they like or dislike about the lender's service, and whether they have any questions about the loan contract.

Courtesy visits are especially valuable right after a lender starts to use scoring. At this point, many super-bads are already on the books, and although the lender cannot call these loans back, it can do something to manage their risk.

Collections Scoring

Collections scoring predicts the probability that a loan currently x days late will reach $x + y$ days. Most commonly, it predicts the risk that a loan that fell into arrears yesterday and is now one day late will eventually become 30 days late. In practice, the collections score would be added to the daily report on delinquent loans. Then based on collections risk and value-at-risk, loan officers would follow a three-class policy to decide who to visit first and how gently to dun them (see Figures 32 and 33). Cases with high risk and high value-at-risk receive immediate, assertive visits. Cases with high risk or low value-at-risk, but not both, also receive immediate visits delivered with a gentler message. Finally, cases with low risk and low value-at-risk are left alone for a few days, and then the first contact is gentle. Low-risk clients may chafe at contact the day after they miss a payment. They may very well cure themselves, and if not, a friendly nudge may be enough to get them back on track.

Like post-disbursement scorecards, collections scorecards use almost the same characteristics as pre-disbursement scorecards, so a pre-disbursement score can stand in for a collections score. Thus the pre-disbursement scorecard provides one score that has three uses.

Desertion Scoring

Desertion scoring predicts the probability that a borrower will apply for another loan once the current one is paid off.[19] Microlenders want to prevent desertion because profitability usually increases with each repeat loan.[20] If the lender knows which clients are at risk of dropping out, it can offer inducements to repeat, such as reduced interest rates or forgiveness of the disbursement fee—contingent, of course, on satisfactory repayment of the current loan. These incentives are costly to the lender; desertion scoring helps control costs by targeting the incentives to the most likely drop-outs.

In the month before the last scheduled installment of an outstanding loan, the lender computes a desertion score and a pre-disbursement score, assuming that the repeat loan contract would have terms similar to the current one. The lender then applies a four-class policy (see Figures 32 and 34):

Kick-outs. Kick-outs are forced drop-outs. Under traditional evaluation standards, their arrears in the current loan disqualifies them from receiving additional loans.

Unsafe waverers. Even though they have not gone bad in the current loan, unsafe waverers are at risk both of dropping out and of going bad. They can apply to repeat, but the lender does not offer them special incentives.

Safe waverers. Safe waverers are at risk of dropping out but not at risk of repayment problems. These good clients might desert, so the lender offers them incentives to repeat.

Loyalists. Loyalists are neither at risk of dropping out or of going bad. The lender does not offer them special incentives because they probably will repeat anyway.

Visit Scoring

Before the field visit, visit scoring predicts the probability of rejection after the field visit. Such rejected cases cost loan officers a lot of time without producing revenue. Visit scoring cuts down on the number of fruitless visits by forecasting rejection risk based on characteristics in the written application. The two-class policy in Figure 34 rejects unpromising clients (perhaps the worst 5 percent of visit scores) without a visit but does visit promising clients per the traditional evaluation process.

Visit scoring can be used only to reject without a visit, not to accept without a visit. As discussed in Box 4, low-proxied risk does not imply low qualitative risk, but high-proxied risk might make the level of qualitative risk moot.

Rather than forecasting rejection after the visit, a visit scorecard could forecast going bad after disbursement. This is pre-disbursement scoring without the characteristics collected in the visit. Even though repayment performance for rejected applicants is unknown, quantified characteristics linked with high repayment risk for approved applicants are probably also linked to high rejection risk for all applicants, given that expected repayment problems lead to after-visit rejections. Thus visit scoring for repayment can be a surrogate for visit scoring for rejection and vice versa.

Only a live test can reveal the power of visit scoring for repayment. In contrast visit scoring for rejections can be tested beforehand using historical data. Unlike the construction of a visit scorecard for repayment, however, the construction of a visit scorecard for rejection requires characteristics from rejected applications, and few microlenders have already entered this data into their information systems.

Most microlenders will start with pre-disbursement scoring, perhaps also using it as a surrogate for post-disbursement scoring and for collections scoring. Once they have mastered the use of pre-disbursement scoring, they could then add desertion scoring and, for those lenders with adequate data, visit scoring.

VIII. Preparing to Score: What Data to Collect?

Scoring can predict only what has already happened many times, and then only if it is recorded in a database. Cutting-edge risk evaluation is hostage to mundane data collection.

Most microlenders do not have enough quality data to construct a scorecard, so once they settle on a risk to score, the next step is to accumulate more and better data.[21] This task has three parts. The first is simply accumulating more bad cases. This task takes time and—for lenders with small portfolios—growth. The second is collecting additional characteristics on the borrower, loan, and lender; and the third is improving the quality of their data.

Required Number of Bad Loans

There is no way to know exactly how many bad loans are needed to construct a score card. Statistical theory supplies exact sample sizes only for the simplest statistics, such as averages. Even then required sample sizes depend on parameters unknown until after the sample is drawn. There are no sample-size formulae for regressions or trees.[22]

The accepted wisdom in high-income countries is that scorecards require at least 500 bad loans.[23] This assumes, however, that clients have both salaried jobs and credit records in a credit bureau.[24] In this special case, a scorecard with ten to fifteen characteristics (most of them from the credit report) can suffice to construct a powerful scorecard. Currently in microcredit, however, most borrowers are self-employed, and if a credit bureau exists, most borrowers are not yet recorded in its records.

Thus the typical characteristic in a microcredit scorecard is much less predictive than the typical characteristic in a scorecard in a high-income country.[25] Acquiring adequate predictive power in microcredit requires more characteristics, and deriving the links between risk and more characteristics requires a larger number of bad loans.

Constructing a useful microcredit scorecard probably requires at least 1,000 bad loans. This is a guess, more likely too low than too high. While more is better, the exact trade-offs are unknown for scoring in general (and for scoring in microcredit). Trade-offs are also dependent on the lender and the context. Such uncertainty is the price of innovation.

Can microlenders pool their data, as small business lenders do in the United States?[26] Unfortunately in microcredit, one size does not fit all. A pooled-data scorecard might be better than nothing only if the microlenders worked in the same country, had the same target market, and used the same traditional evaluation process. Transferring scorecards across borders would be foolish.

Collecting Appropriate Characteristics

What characteristics should a microlender begin to collect now to be able to construct a scorecard (or a more powerful scorecard) in a few years? In the lists below, the core set of required characteristics is marked with asterisks. Most microlenders who make individual loans already collect most of this core data as part of a traditional evaluation. Additional characteristics that would increase predictive power are also listed below, although powerful scorecards (such as those whose results appear in Figures 20, 23, and 24) can be constructed without them. Most of these additional characteristics could be supplied by the applicant in the initial application.

At a minimum, microlenders who plan to score should quantify loan officers' subjective judgments, enter credit bureau data into their information systems, and record household assets and demographics. Lenders should not go back and collect this data for past loans but should start to record it now.

Characteristics of the Borrower

Demographics. Applicant demographics are among the most predictive characteristics:

- gender*
- year of birth*
- marital status* (married/cohabiting, never married/never-cohabited, divorced/separated, widowed)
  - year of marriage/cohabitation
  - year of divorce/separation/widowhood
- last school grade completed*

Dates of marriage or separation are convenient proxies for family stability. Some microlenders may choose to ignore risk linked with demographic characteristics that applicants do not choose for themselves (see Box 10).

Contact information. The presence of phone numbers and contact information in the database is predictive of risk:

- phone number to contact at home* (may be a neighbor's phone)
- phone number to contact at business* (may be a neighbor's phone)
- distance from home (and from the business) to nearest branch
  - minutes spent in travel
  - money spent for public transport (if used)

The distance to the nearest branch bank (and the presence of a telephone) proxies for transaction costs.[27] Greater transaction costs increase arrears by borrowers and make monitoring more difficult for loan officers.

Household demographics. Household composition affects cash flows and risk:

- number of people age 18 or older (including applicant)
- number of people age 17 or younger

Household assets. The presence of household assets (and changes in assets over time) signal risk:

- Home tenure (owner, renter, other)
  - year moved to current residence
  - year moved to previous residence
  - number of rooms (excluding bathrooms and kitchen) in current residence
- Land ownership
  - homestead land with title (present or absent)
  - homestead land without title (present or absent)
  - other land with title (number of hectares)
  - other land without title (number of hectares)
- Housing construction
  - tin roof (present or absent)
  - concrete floor (present or absent)
  - connection to water lines (present or absent)
  - connection to sewage lines (present or absent)

    connection to electricity (present or absent)
- Vehicles that run
  - automobile, tractor, truck, or bus (present or absent)
  - motorcycle (present or absent)
  - bicycle (present or absent)
- Appliances
  - refrigerator (present or absent)
  - gas or electric stove (present or absent)
  - working color television (present or absent)
  - electrical generator (present or absent)
- Formal savings account (present or absent)

Relevant household assets depend on local context. Assuming that assets would not change in the absence of loans, these data indicate "impact." Also many of these assets appear in poverty assessment tools, so the lender may want to collect them for a range of reasons beyond their usefulness in scoring.

Scoring may show that poorer clients (for example, those with fewer assets) have greater risk. The microlender's policy on poverty targeting may lead it to exclude some poverty-linked characteristics from the scorecard or to accept greater risks for poorer clients. Scoring does not change the risk of borrowers; it only improves knowledge of the risk that already exists.

Business demographics. The basic features of the business are predictive of repayment:

- sector* (manufacturing, services, trade, agriculture)
- specific type of business*

- year started in this line of business
- year started in this specific enterprise*
- official registration (present or absent)
- written records of cash flows (present or absent)
- type of locale (storefront, mobile, lock-box, home-based, other)
- tenure of locale (owned, rented, other)
- year moved to current locale
- number of person-months of full-time-equivalent workers employed per year
  - paid family members
  - unpaid family members
  - paid non-family members

Many microlenders already record "number of employees," but this is often useless data for scoring. Such data mixes seasonal with permanent, part-time with full-time, family with non-family, and paid with unpaid employees. Employees should be measured in terms of person-months per year for each type of worker.

Financial flows of the household/enterprise. The strength of monthly cash flows are strongly predictive of credit risk:

- business revenue*
- household income from salaries*
- household income from other sources*
- business expenses for goods purchased*
- business salary expense*
- other business expenses*
- rent payment
- other household expenses*
- monthly installments due on other debts (including home mortgage)*

Because cash flows fluctuate, the microlender should also ask about the share of sales in cash versus credit. Financial data must be collected by a loan officer during the field visit. Most microlenders currently record sales, other income, business expenses, and household expenses. Greater disaggregation is useful for scoring because risk depends partly on whether cash flows are regular versus irregular or obligatory versus voluntary.

Stocks of the enterprise. Most microlenders already record the value of assets and liabilities:

- Total assets*
  - fixed assets*
  - inventory*
  - cash and bank accounts*
- Total liabilities*
  - informal debts*
  - formal debts*

Repayment record. The best predictor of future performance is past performance. For each installment due on each loan, lenders should record the date due and the date paid. This allows the derivation of the following measures of aspects of arrears:

- longest spell*
- days of arrears per installment*
- number of installments paid late*

After each loan is paid off, the lender should also ask the loan officer to grade overall repayment performance subjectively on a scale of 1 (best) to 5 (worst).

Credit bureau. Credit bureau data are powerfully predictive.[28] If lenders receive credit bureau reports for some borrowers, they should enter the following data into their information systems:

- identity of current and past creditors
- dates disbursed (and dates paid off) for current and past loans
- amounts disbursed for current and past loans
- monthly installments for current and past loans
- maximum line of credit with current and past creditors
- arrears in current and past loans
- amount owed to current creditors
- number of inquiries

Proxies for personal character. Microlenders serious about scoring should seek to record characteristics that proxy for personal character traits that may

Box 10: Should Scoring Use Protected Characteristics?

No one chooses their gender, ethnicity, native language, or age, and many people—especially women and ethnic minorities—have limited choice regarding marital status or place of residence. Yet all these characteristics are visible at a glance and thus can be—have been, and are—used to oppress one group for the benefit of another. Traditional lenders have disproportionately excluded people with these markers ("protected characteristics"), both because lenders participated in their oppression and because their oppression made these applicants worse risks. A central purpose of microcredit is to help change this.

In some high-income countries, laws prohibit using protected characteristics in scorecards. The laws aim to purge explicit oppression from non-statistical scoring and prevent statistical scoring from using the knowledge that oppression elsewhere in society has linked risk to protected characteristics. In most low-income countries, however, no such laws exist. Protected characteristics are predictive of repayment risk; should microcredit scorecards use them?

There is no easy answer. One approach is to collect more and better data. Genes, after all, do not cause risk directly. Protected characteristics are associated indirectly with risk because they are associated with socially produced characteristics that are, in turn, directly linked with risk. For example, the absence of a Y chromosome does not affect a woman's repayment risk, but the fact that society allows women to be seamstresses—but not blacksmiths—does. With more and better data on characteristics directly linked with risk, the importance of protected characteristics as indirect proxies would decrease.

Of course, this does not resolve the issue. Even if women are more (or less) risky—not because they are women but because society limits women—they did not choose their characteristics. To some extent, non-protected characteristics can be involuntary. For example, poor people did not choose to be poor. Even apparently chosen characteristics result from some unobserved clash between choice and constraint. Yet some people believe that there are no choices, only the inexorable clockwork of physical laws.

In the end, there is risk, much of it linked with unchosen characteristics. Microlenders must decide how to evaluate risk, given that any evaluation must necessarily be based on experience and prejudice. There will always be a trade-off between better prediction and reinforcing unfair discrimination. Ultimately, the microlender must make a value judgment about what data to collect and how to use it. Scoring can improve this judgment by quantifying the trade-offs between the use of certain characteristics and predictive accuracy.

correlate highly with repayment discipline. In Latin America for example, someone who has a personal policy not to drink alcohol may be more likely to take debt repayment seriously. Likewise weekly (or daily) attendance at religious services may mark someone as likely to follow a repayment regimen faithfully. Religion or vices may be sensitive (or irrelevant or illegal) in some places, so lenders should adapt these guidelines to the local context:

- number of alcoholic drinks in the past year
- number of cigarettes smoked in the past year
- number of lottery tickets bought in the past year
- number of times attended religious services in the past year
- current membership in neighborhood committee or church group
- date of last salaried employment
- participation in Rotating Savings and Credit Associations (ROSCAs)
    - date of most recent participation
    - amount of periodic contribution
    - frequency of contribution

Participation in a ROSCA signals experience as a saver and a debtor. A ROSCA may also serve as a fallback source of funds to finance installments paid to the microlender.

Quantified subjective judgments. The only way to screen for qualitative risk is to send loan officers to the field to get to know applicants as people (see Box 4). Yet a loan officer's subjective judgment can be quantified. This would allow scoring to reveal, for example, how the probability of going bad is linked with the subjective judgment of "average" versus "above-average."

Microlenders who want to score in the future should start to quantify subjective judgments regarding the following criteria on a three-point scale ("below-average," "average," and "above-average"):

- overall credit risk
- honesty and transparency of responses
- quality of references
- entrepreneurial ambition and creativity
- business prospects
- variability of cash flows
- extent of recent investment in home or business
- grasp of the rules in the loan contract
- family relationships and informal support
- organization and cleanliness of home and business

For obvious reasons, this does not work if all accepted applicants are rated above-average.

Characteristics of the Loan

Microlenders already record most of the predictive characteristics of the loan:

- date application submitted*
- date loan disbursed*
- date paid-in-full*
- amount requested*
- amount disbursed*
- amount of installments*
- number of installments*
- frequency of installments*
- interest rate*
- fees and commissions*
- grace period*
- rescheduled status*
- purpose of loan*
- type of guarantee*
- appraised value of guarantee*
- identity of cosigner

The date of application is used to measure days between application and disbursement. Knowing the cosigner allows scoring to incorporate their credit record (if they have one) in the applicant's score. If the cosigner later applies for their own loan, then the repayment record of the loans that they guaranteed can also be used as a predictor.

Characteristics of the Lender

The characteristics of the lender, that is, the specific branch and the assigned loan officer, strongly influence risk. The lender should also record a few simple characteristics of the loan officer. Scoring will then not only reveal the profile of the ideal loan officer but also better predict the risk of loans from loan officers hired after scorecard construction:

- gender
- year of birth
- marital status (married or not married)
- number of people in household
- subject studied in college
- last school grade completed

The Value of Additional Data

Given enough bad loans, a useful and powerful scorecard can be constructed from the core characteristics marked with asterisks above, most of which microlenders already collect. A scorecard compiled from all the characteristics listed above would probably predict 20-40 percent better than a scorecard with just the core characteristics.

Box 11: Does Scoring Work with "Noisy" or "Dirty" Data?

Microcredit data—like all data—always have some "dirt" (errors) and "noise" (random variation around the true value). For example, the value of fixed assets is noisy because it is difficult to appraise. It can also be dirty because the loan officer may manipulate the appraisal so that an application that the loan officer deems worthy satisfies the financial ratios required by the lender's evaluation policy.

The statistical work in scorecard construction filters whatever signal (the link between risk and a characteristic) it can from the dirt and noise. If there is no signal (or if a characteristic is simply not linked with risk), then the statistical process reveals this and drops the characteristic from the scorecard. In many cases, data known to be dirty and noisy still contain useful signals.

Accumulating additional data will provide greater predictive power but will also incur greater costs. These costs are mainly from redesigning application forms, modifying the information system to accept the added data, and entering the new data. Although loan officers would have to do a bit more work, a literate client can easily supply most of the additional items on the initial application.

Guidelines for Warehousing Better-Quality Data

After human resources, a microlender's greatest asset is information. Often, however, formal information systems are weak, having been used for little besides tracking loans. The advent of scoring and the more intense use of an electronic database rewards greater attention to data quality.

Most microlenders have collected the core set of characteristics for years but never used the data. As a result, loan officers and data-entry personnel have learned that paying attention to data quality costs them time but offers no reward. With scoring, data quality matters. To make the requisite effort, front-line personnel need to know that old habits are no longer acceptable, why such habits are no longer acceptable, and how they will personally benefit from the change.

Establish Consistent Definitions for Type of Business

The type of business is one of the three most predictive characteristics, along with past arrears and the identity of the loan officer. Unfortunately the quality of data on business type is often poor because a given code may encompass too wide a range of businesses and thus does not distinguish sharply between high and low risks. Nevertheless "dirty" and "noisy" data are better than no data (see Box 11).

The business type is often coded poorly for three reasons. First, loan officers do not share common definitions. One officer's bar is another's restaurant. Second, loan officers look at products rather than activities, for example lumping shoemakers, shoe repairers, and shoe sellers under "shoes," even though these activities are in manufacturing, service, and commerce, respectively, and have very different risks. Third, data-entry personnel tend to lump information under general headings, such as "food sales" or "stores," rather than search for a match through a long list of codes.

Making front-line personnel aware of this issue is the first step toward improvement. The second step is to make a list of 50 or more of the most common business types, define each one carefully, and teach loan officers and data-entry personnel to stick to them. About 90 percent of businesses will fall under one of these 50 codes, and the remaining 10 percent or so can be coded as "other." The third step is to define types of activities (sectors) precisely:

- trade (sale of untransformed items)
- manufacturing (sale of transformed items; like traders, manufacturers buy and sell, but what they buy differs from what they sell)
- services (sale of specialized labor or of the use of physical items)
- agriculture (manufacture of plants, animals, or minerals directly from the land)

The fourth step is to establish a formal, written policy to code each enterprise as one of the 50 business types and as one of the four activities. The fifth step is to include a checklist of all the sectors (with definitions) and all the business types on the form that the loan officer fills out. The sixth and final step is to monitor the use of the new system.

This is a lot of work; however, the type of business, if recorded properly, is highly predictive. Without salaried borrowers and without credit bureau data, microcredit scorecards cannot afford to lose one of three top characteristics.

Do Not Throw Away Data

Compared with waiting years to construct a scorecard because old data was discarded, electronic storage is inexpensive. Long unused data is the lifeblood of scoring today and the key to future market research and client monitoring.[29] The rule is: once keyed in, keep it in.

Collecting Data from Rejected Applications

Many microlenders would like to use visit scoring to shorten (or skip) some field visits. This means forecasting either repayment problems or post-visit rejections. Forecasting repayment problems before the visit might work, but only a live test can confirm predictive power. (The Global Follow-up Report cannot help.) Because a visit scorecard is constructed only from approved borrowers who pass a qualitative screen, forecasts for unscreened borrowers have unknown accuracy (see Box 4). Loan officers will still have to visit applicants who pass the visit score because without a qualitative screen scoring cannot approve, only reject.

Forecasting rejection after the field visit is a better alternative. To do this, microlenders must first enter data from several thousand rejected applications into their information systems. Once they have data on both post-visit rejects and post-visit approvals, they can construct scorecards to forecast rejections based on characteristics known before the visit. (Even with data from rejected applications, a visit scorecard for repayment risk still cannot approve applicants without a visit because the repayment behavior of unscreened borrowers is still unknown.)

Recording Characteristics of Both Screening and Monitoring Officers

One of the three most predictive characteristics is the identity of the loan officer. The officer in charge of a loan sometimes changes due to internal reorganizations, workload reallocations, or job changes. When this happens, most systems delete the original screening officer's characteristics and record only the current monitoring officer's. This reduces the predictive power of scoring in two ways. First, the risk ascribed by the scorecard to the monitoring officer erroneously depends in part on the original screening officer. Second, the risk ascribed to the screening officer ignores the results of loans that were transferred to others.

The solution is to add a field to the database that records the screening officer. The original "loan officer" field continues to record the current monitoring officer. If one officer stays with a loan from start to finish, the screening officer is the same as the monitoring officer.

Recording both loan officers may seem trivial because most loans have but one loan officer. In practice loan officers fingered by scoring as high risks often point out that they inherited many bad loans or that they had to give away all their good loans. The identity of the loan officer has a strong effect on predicted risk. To convince loan officers and credit managers to accept this requirement means accounting for transferred loans during scorecard construction. In turn this requires tracking both the screening officer and the monitoring officer in the database.

Missing Values Are Unknown, Not Zero

Sometimes an applicant leaves a blank space on an application, or a loan officer forgets to write down an item from the field visit, or a data-entry operator accidentally skips a field. The result is a missing (unknown) value. For example, if an applicant leaves "year of birth" blank, his age is not zero but unknown.

The presence of missing values is often very predictive. For example, loan files missing data on business revenue may be more risky than loan files with revenue recorded. Often missing data and repayment risk have a common cause (such as a skipped field visit or an applicant with something to hide). Unfortunately most microcredit information systems do not record missing values properly. They either change blanks to zeroes or force each field to have an entry, leading data-entry operators to change blanks to zeros, make up data, or invent inconsistent codes for missing values. (One large microlender, for example, evidently lends to hundreds of nonagenarians, all born November 11, 1911.)

Failure to record missing values properly harms scoring in two ways. First, it precludes using the presence of missing values as a predictive characteristic. Second, it confuses the risk associated with missing values with the risk associated with true zero values. For example, the number of children recorded can often be non-zero, zero, or missing. The risk of people who do not report the number of children probably differs from the risk of people who report zero children. Replacing unknown values with zero, however, forces scoring to assign the same risk to both groups.

The solution is to establish an explicit code for missing values and then to train loan officers and data-entry operators to use it. Some database languages already reserve a code for missing values. For other languages, the microlender can use "–99."

Regardless of the type of risk to be forecast, statistical scoring requires a great deal of good-quality data. Even the few microlenders who have adequate databases should start to enter loan officer judgments, credit bureau reports, and rejected applications into their information systems. As for the remaining majority of microlenders, they must revamp their systems now if they want use scoring in a few years. Improving the quality of the database is hard work, but not quite as hard as forever judging risk without the help of scoring.

IX. Conclusion

Scoring quantifies the risk that the self-employed poor will not pay as promised. Scoring also makes explicit the links between repayment and the characteristics of borrowers, loans, and lenders. Most importantly scoring provides a taste of decision making based on quantified risks and explicit trade-offs. This may prompt a shift in organizational culture as managers seek greater knowledge and precision about alternatives to their decisions and their consequences. Although simple data analyses can inform decisions, most microlenders have yet to invest in—let alone take advantage of—an accurate, comprehensive database as an asset.

Scoring in microcredit on average comes close to the target. About 20 percent of loans with a predicted risk of 20 percent, for example, do indeed turn out bad. The number and range of mistakes around the average, however, are much greater than for scoring in high-income countries. Unfortunately much of the risk of the self-employed poor is unrelated to quantifiable characteristics. Thus scoring complements—but does not replace—loan officers and their subjective evaluations. Scoring is a "third voice" in the credit committee, a supplement to the judgment of the loan officer and the credit manager.

The purpose of scoring is to forecast risk. But for a microlender who wants to implement scoring, its predictive power is of secondary concern because scoring can be tested beforehand with historical data. The microlender's primary concern is to convince board members, managers, and loan officers to accept scoring. In the end, statistical weaknesses do not kill scoring projects, people do.[30] Scoring—even if it works like a dream—represents a change that some people will resist. Acceptance requires repeated training for stakeholders at all levels and persistent follow-up with constant demonstrations of predictive power for currently outstanding loans.

Scoring may not be the next breakthrough in microcredit, but for a few microlenders, scoring can reduce time in collections and thus boost efficiency, outreach, and sustainability. As more organizations learn about scoring and set up processes to accumulate adequate data, scoring will likely become a part of microcredit best practice.

Some might argue that scoring is a new-fangled gadget that microcredit can do without. "If it ain't broke, don't fix it," is a common response. Lenders in high-income countries said the same thing for decades, and scoring has now all but replaced manual evaluation, especially for small, short-term, uncollateralized loans that closely resemble microcredit.[31] Microcredit is good, but it can still improve, and growth and competitive pressures increasingly mean that the best microlenders must seek change proactively. Credit scoring is one way to keep ahead.

Notes

1 See Mark Schreiner, "Scoring Drop-out at a Microlender in Bolivia," (manuscript, Center for Social Development, Washington University, St. Louis, Mo., 2001); "Credit Scoring for Microfinance: Can It Work?" *Journal of Microfinance* 2, no. 2 (2000): 105-18; and "A Scoring Model of the Risk of Arrears at a Microfinance Lender in Bolivia" (manuscript, Center for Social Development, Washington University, St. Louis, Mo., 1999). See also Ulrike Vogelgesang, "Microfinance in Times of Crisis: The Effects of Competition, Rising Indebtedness, and Economic Crisis on Repayment Behavior," *Gk Working Paper Series,* No. 2001-06 (University of Mannheim, 2001); and Laura Viganò, "A Credit Scoring Model for Development Banks: An African Case Study," *Savings and Development* 17, no. 4 (1993): 441-82.

2 See Elizabeth Mays, *Handbook of Credit Scoring* (Chicago: Glenlake, 2000), and *Credit Risk Modeling: Design and Application* (New York: Glenlake, 1998); Lyn C. Thomas, "A Survey of Credit and Behavioral Scoring: Forecasting the Financial Risk of Lending to Consumers," *International Journal of Forecasting* 16 (2000): 149-72; Peter McCorkell, "Credit Scoring 101" (paper presented at the Federal Trade Commission public forum, "The Consumer and Credit Scoring," 1999); David J. Hand and W. E. Henley, "Statistical Classification Methods in Consumer Credit Scoring: A Review," *Journal of the Royal Statistical Association, Series A* 160, part 3 (1997): 523-41; Loretta J. Mester, "What's the Point of Credit Scoring?" *Business Review* (September/October 1997): 3-16; Lyn C. Thomas, J. N. Crook, and D. B. Edelman, *Credit Scoring and Credit Control* (Oxford: Clarendon Press, 1992); and Edward M. Lewis, *An Introduction to Credit Scoring* (San Rafael, Ca.: Athena Press, 1990).

3 Mark Schreiner, *Un Sistema de Scoring del Riesgo de Créditos de FIE en Bolivia* (report to Fomento de Iniciativas Económicas, La Paz, Bolivia, 2001).

4 See page 48 for a discussion of why "bad" is defined as serious delinquency rather than default.

5 Changes over time in the proportion of paid-off loans in each of the four segments in Figure 7 (and also in Figure 10) are ignored to avoid clutter.

6 David J. Hand, Heikki Mannila, and Padhraic Smyth, *Principles of Data Mining* (Cambridge, Mass.: MIT Press, 2001).

7 Mark Schreiner, "Aspects of Outreach: A Framework for the Discussion of the Social Benefits of Microfinance," *Journal of International Development* 14 (2002): 591–603.

8 Schreiner, "Aspects of Outreach."

9 Paul Mosley, "Microfinance and Poverty in Bolivia," *Journal of Development Studies* 37, no. 4 (2001): 101-32.

10 Elisabeth Rhyne, "The Yin and Yang of Microfinance: Reaching the Poor and Sustainability," *MicroBanking Bulletin* 2 (1998): 6-8; and Claudio Gonzalez-Vega, "Microfinance: Broader Achievements and

New Challenges," *Economics and Sociology Occasional Paper,* No. 2518 (Columbus, Ohio: The Ohio State University, 1998).

11 W. Scott Frame, Michael Padhi, and Lynn Woosley, "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 (2001); Peter McCorkell, "Credit Scoring 101"; and Edward M. Lewis, *An Introduction to Credit Scoring.*

12 Mark Schreiner, "Scoring Drop-out at a Microlender in Bolivia."

13 Rhonda Delmater and Monte Hancock, *Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence* (Boston, Mass.: Digital Press, 2001); and Michael J. A. Berry and Gordon Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management* (New York: John Wiley and Sons, 2000).

14 Martin Holtmann, "Designing Financial Incentives to Increase Loan-Officer Productivity: Handle with Care," *MicroBanking Bulletin* 6 (2001): 5-10.

15 Ulrike Vogelsang, "Microfinance in Times of Crisis."

16 Did scoring just get lucky? Given the lender's historical bad rate of 9.6 percent, the chance of picking 30 loans at random and none being bad—as in the low-risk Loan Officer Follow-up Report—is less than 1 in 20. The chance of 15 of 30 being bad—as in the high-risk Loan Officer Follow-up Report—is less than one in a billion.

17 A. D. Lovie and P. Lovie, "The Flat Maximum Effect and Linear Scoring Models for Prediction," *Journal of Forecasting* 5: 159-68; Peter Kolesar and Janet L. Showers, "A Robust Credit Screening Model Using Categorical Data," *Management Science* 31, no. 2 (1985): 123-33; William G. Stillwell, F. Hutton Barron, and Ward Edwards, "Evaluating Credit Applications: A Validation of Multiattribute Utility Weight Elicitation Techniques," *Organizational Behavior and Human Performance* 32 (1983): 87-108; and Howard Wainer, "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin* 83 (1976): 213-17.

18 Mark Schreiner, "Do It Yourself Scoring Trees for Microfinance," (presentation at Tercer Seminario sobre Banca y Microfinanzas en Latinoamérica y el Caribe, Santo Domingo, Dominican Republic, November 11–12, 2001).

19 Mark Schreiner, "Scoring Drop-Out at a Microlender in Bolivia."

20 Craig F. Churchill and Sahra S. Halpern, "Building Customer Loyalty: A Practical Guide for Microfinance Institutions," *Microfinance Network Technical Guide* No. 2 (2001); Richard Rosenberg, "Measuring Client Retention," *MicroBanking Bulletin* 6 (2001): 25–6.

21 Parts of this section draw on Mark Schreiner, "Scoring at Prizma: How to Prepare" (report to Prizma and CGAP, Mostar, Bosnia and Herzegovina, sean@prizma.ba).

22 William G. Cochran, *Sampling Techniques,* 3rd ed. (New York: John Wiley and Sons, 1977).

23 Edward M. Lewis, *An Introduction to Credit Scoring*.

24 Credit bureaus are databases that work with multiple lenders to collect, store, and distribute information on the repayment behavior of individual borrowers.

25 Mark Schreiner, "Credit Scoring for Microfinance."

26 Latimer Asch, "Improving Efficiencies in SME Lending with Scoring," (manuscript, San Rafael, Calif.: Fair, Isaac, 2000).

27 Mariano Rojas and Luis Alejandro Rojas, "Transaction Costs in Mexico's Preferential Credit," *Development Policy Review* 15 (1997): 23-46; and Carlos E. Cuevas, "Transaction Costs of Financial Intermediation in Developing Countries," Economics and Sociology Occasional Paper, No. 1469 (Columbus, Ohio: The Ohio State University, 1998), ccuevas@worldbank.org.

28 Michael E. Staten, "The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience" (paper presented at the World Bank conference on "Profiting from Small Business Lending," Washington, D.C., April 2–3, 2001); and Elinor Haidor, "Credit Bureaus: Leveraging Information for the Benefit of Microenterprises," *Microenterprise Development Review* 2, no. 2 (2000): 1, 5-8.

29 On market research, see Inez Murray, "What Do MFI Customers Value? A Comparative Analysis from Three Continents" (manuscript, Women's World Banking, Columbia, 2001). On client monitoring, see Gary Woller, "Summary of Important Issues to Consider in Designing and Implementing an Impact Monitoring System" (manuscript, Brigham Young University, Provo, Utah, 2001).

30 Mona A. Mayr, "Successful Implementations: To Be, or Not to Be," in *Handbook of Credit Scoring,* ed. Elizabeth Mays (Chicago: Glenlake, 2000): 337-52; Kevin J. Leonard, "Credit Scoring and Quality Management," in *Statistics in Finance,* ed. David J. Hand and Saul D. Jacka (London: John Wiley and Sons, 1998): 105-26; Leonard J. McCahill, "Organizational Issues in Building and Maintaining Credit Risk Models," in *Credit Risk Modeling,* ed. Elizabeth Mays (Chicago: Glenlake, 1998): 13-22; and D. B. Edelman, "The Introduction of Credit Scoring into Branch Banking," in *Credit Scoring and Credit Control,* ed. L. C. Thomas, J. N. Crook, and D. B. Edelman (Oxford: Clarendon Press, 1992): 161-77.

31 Edward M. Lewis, *An Introduction to Credit Scoring.*

Bibliography

Asch, Latimer. "Improving Efficiencies in SME Lending with Scoring." Manuscript. San Rafael, Calif.: Fair, Isaac, 2000.

Berry, Michael J.A., and Gordon Linoff. *Mastering Data Mining: The Art and Science of Customer Relationship Management.* New York: John Wiley and Sons, 2000.

Churchill, Craig F., and Sahra S. Halpern. "Building Customer Loyalty: A Practical Guide for Microfinance Institutions." *Microfinance Network Technical Guide* No. 2. Washington, D.C.: MicroFinance Network, 2001.

Cochran, William G. *Sampling Techniques*. 3d ed. New York: John Wiley and Sons, 1977.

Coffman, John. "An Introduction to Scoring for Micro and Small Business Lending." Paper presented at the World Bank conference on "Profiting from Small Business Lending," Washington, D.C., April 2–3, 2001.

Cuevas, Carlos E. "Transaction Costs of Financial Intermediation in Developing Countries." *Economics and Sociology Occasional Paper,* No. 1469. Columbus, Ohio: The Ohio State University, 1988.

Delmater, Rhonda, and Monte Hancock. *Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence.* Boston, Mass.: Digital Press, 2001.

Edelman, D. B. "The Introduction of Credit Scoring into Branch Banking." In *Credit Scoring and Credit Control.* Edited by L. C. Thomas, J. N. Crook, and D. B. Edelman. Oxford: Clarendon Press, 1992.

Frame, W. Scott, Michael Padhi, and Lynn Woosley. "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. 2001.

Gonzalez-Vega, Claudio. "Microfinance: Broader Achievements and New Challenges." *Economics and Sociology Occasional Paper,* No. 2518. Columbus, Ohio: The Ohio State University, 1998.

Haidor, Elinor. "Credit Bureaus: Leveraging Information for the Benefit of Microenterprises." *Microenterprise Development Review* 2, no. 2 (2000): 1, 5-8.

Hand, David J., Heikki Mannila, and Padhraic Smyth. *Principles of Data Mining*. Cambridge: MIT Press, 2001.

_____, and W. E. Henley. "Statistical Classification Methods in Consumer Credit Scoring: A Review." *Journal of the Royal Statistical Association, Series A* 160, part 3 (1997): 523-41.

Holtmann, Martin. "Designing Financial Incentives to Increase Loan-Officer Productivity: Handle with Care." *MicroBanking Bulletin* 6 (2001): 5-10.

Kolesar, Peter, and Janet L. Showers. "A Robust Credit Screening Model Using Categorical Data." *Management Science* 31, no. 2 (1985): 123-33.

Leonard, Kevin J. "Credit Scoring and Quality Management." In *Statistics in Finance*. Edited by David J. Hand and Saul D. Jacka. London: John Wiley and Sons, 1998.

Lewis, Edward M. *An Introduction to Credit Scoring*. San Rafael, Calif.: Athena Press, 1990.

Lovie, A. D., and P. Lovie. "The Flat Maximum Effect and Linear Scoring Models for Prediction." *Journal of Forecasting* 5 (1986): 159-68.

Mayr, Mona A. "Successful Implementations: To Be, or Not to Be." In *Handbook of Credit Scoring.* Edited by Elizabeth Mays. Chicago: Glenlake, 2000.

Mays, Elizabeth. *Handbook of Credit Scoring.* Chicago: Glenlake, 2000.

_____. *Credit Risk Modeling: Design and Application.* Chicago: Glenlake, 1998.

McCahill, Leonard J. "Organizational Issues in Building and Maintaining Credit Risk Models." In *Credit Risk Modeling.* Edited by Elizabeth Mays, Chicago: Glenlake, 1998.

McCorkell, Peter. "Credit Scoring 101." Presentation at the Federal Trade Commission public forum on "The Consumer and Credit Scoring," Washington, D.C., 1999.

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

Mosley, Paul. "Microfinance and Poverty in Bolivia." *Journal of Development Studies* 37, no. 4 (2001): 101-32.

Murray, Inez. "What Do MFI Customers Value? A Comparative Analysis from Three Continents." Manuscript. Women's World Banking, 2001.

Poyo, Jeffrey, and Robin Young. "Commercialization of Microfinance: The Cases of Banco Económico and Fondo Financiero Privado FA$$IL, Bolivia." Bethesda, Md.: Microenterprise Best Practices, 1999.

Rhyne, Elisabeth. *Mainstreaming Microfinance: How Lending to the Poor Began, Grew, and Came of Age in Bolivia.* Bloomfield, Ind.: Kumarian, 2001.

_____. "The Yin and Yang of Microfinance: Reaching the Poor and Sustainability." *MicroBanking Bulletin* 2 (1998): 6-8.

Rojas, Mariano, and Luis Alejandro Rojas. "Transaction Costs in Mexico's Preferential Credit." *Development Policy Review* 15 (1997): 23-46.

Rosenberg, Richard. "Measuring Client Retention." *MicroBanking Bulletin* 6 (2001): 25-6.

Schreiner, Mark. "Aspects of Outreach: A Framework for the Discussion of the Social Benefits of Microfinance." *Journal of International Development* 14 (2002).

_____. *Un Sistema de Scoring del Riesgo de Créditos de FIE en Bolivia*. Report to Fomento de Iniciativas Económicas, La Paz, Bolivia, 2001.

_____. "Scoring Drop-out at a Microlender in Bolivia." Manuscript. Center for Social Development, Washington University, St. Louis, Mo., 2001.

_____. "Do-It-Yourself Scoring Trees for Microfinance." Paper presented at Tercer Seminario sobre Banca y Microfinanzas en Latinoamérica y el Caribe, Santo Domingo, Dominican Republic, November 11–12, 2001.

_____. "Scoring at Prizma: How to Prepare." Report to Prizma and CGAP, Mostar, Bosnia and Herzegovina, 2001.

_____. "Credit Scoring for Microfinance: Can It Work?" *Journal of Microfinance* 2, no. 2 (2000): 105-18.

_____. "A Scoring Model of the Risk of Arrears at a Microfinance Lender in Bolivia." Manuscript. Center for Social Development, Washington University, St. Louis, Mo., 1999.

Staten, Michael E. "The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience." Paper presented at the World Bank conference on "Profiting from Small Business Lending," Washington, D.C., April 2–3, 2001.

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

Thomas, Lyn C. "A Survey of Credit and Behavioral Scoring: Forecasting the Financial Risk of Lending to Consumers." *International Journal of Forecasting* 16 (2000): 149-72.

_____, J. N. Crook, and D. B. Edelman. *Credit Scoring and Credit Control*. Oxford: Clarendon Press, 1992.

Viganò, Laura. "A Credit Scoring Model for Development Banks: An African Case Study." *Savings and Development* 17, no. 4 (1993): 441-82.

Vogelgesang, Ulrike. "Microfinance in Times of Crisis: The Effects of Competition, Rising Indebtedness, and Economic Crisis on Repayment Behavior." *Gk Working Paper Series,* No. 2001-06. University of Mannheim, 2001.

Wainer, Howard. "Estimating Coefficients in Linear Models: It Don't Make No Nevermind." *Psychological Bulletin* 83 (1976): 213-17.

Woller, Gary. "Summary of Important Issues to Consider in Designing and Implementing an Impact Monitoring System." Manuscript. Brigham Young University, Provo, Utah, 2001.

Occasional Paper
No. 7

Many people have helped the author learn about scoring for microcredit. The author is grateful to CGAP for its support and to Rich Rosenberg in particular for his insightful suggestions. Elizabeth Littlefield and Brigit Helms also made constructive comments. Additional thanks go to Hólmer Aguirre, Clara de Akerman, Lilián Alomía Ortíz, Oydén Cantillo Rincón, Margarita Correa, Jorge Alberto Díaz, Angela Escobar, María Mercedes Gómez, Diego Guzmán, Jorge Enrique Londoño, Leonor Melo de Velasco, María Isabel Muños Reyes, Nohora Niño, Manuel Enrique Olago Villamizar, Nestor Raúl Plata, Teresa Eugenia Prada, Javier Salazar Osorio, and Oscar Iván Tafur of Women's World Banking in Colombia; María del Carmen Pereira, Elizabeth Salinas Nava, and Enrique Soruco Vidal of FFP-FIE in Bolivia; Maja Gizdic and Sean Kline of Prizma in Bosnia and Herzegovina; and Levi Cenác, Pedro Jiménez, Manual Ovalle, and Guillermo Rondón of BancoADEMI in the Dominican Republic. The author is especially grateful to Hans Dellien of Women's World Banking for being the first to take a risk on scoring for microcredit and for contributing to the development of many of the concepts in this Occasional Paper.