Scoring: The Next Breakthrough in Microcredit?

Paper prepared for the Consultative Group to Assist the Poorest

Mark Schreiner[1]

June 11, 2002

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

Abstract

The challenge of microcredit is to judge the risk of whether the self-employed poor will repay their debts as promised. Scoring is a new way to judge risk; is it the next breakthrough in microcredit? Scoring does reduce arrears and so reduces time spent on collections, and 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 with a large data base 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 fundamentally, scoring changes how microlenders think, fostering a culture of analysis in which managers regularly seek to mine their data bases to for information that addresses business questions.

---

[1] Mark Schreiner is a consultant with Microfinance Risk Management and a Senior Scholar in the Center for Social Development at Washington University in St. Louis. His work helps the poor build assets through greater access to financial services.

Acknowledgments

Many people have helped me to learn about scoring for microcredit. I am grateful to CGAP for funding this paper and to Rich Rosenberg managing it and providing very helpful comments. Elizabeth Littlefield and Brigit Helms also made useful comments. I also thank 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 Gizdić and Sean Kline of Prizma in Bosnia-Herzegovina; and Levi Cenác, Pedro Jiménez, Manual Ovalle, and Guillermo Rondón of BancoADEMI in the Dominican Republic. I am 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 paper.

Table of Contents

List of Boxes

List of Figures

Scoring: The Next Breakthrough in Microcredit?

1. 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 of 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 through 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?

1.1 Scoring improves both outreach and sustainability

For the few microlenders who are already large, who are already well run, and who already possess adequate electronic data bases, scoring can indeed expand the efficiency frontier and so improve both poverty outreach and organizational sustainability. Scoring does this mostly by reducing time spent collecting overdue payments from delinquent borrowers; a typical loan officer might save about half a day per week. Loan officers can then use some of their new-found time to search for more good borrowers, expanding 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 2000 would have reduced the number of loans that reached 30 days overdue by 28 percent (Schreiner, 2001a). Given conservative assumptions about the cost of the scoring project, the net benefit of successfully rejecting loans that would have gone bad, and the net cost of mistakenly rejecting loans that would not have gone bad, scoring would have paid for itself in about one year and would have had a net present value of about $1 million.

1.2 Scoring is a marked improvement, but not a breakthrough

Scoring is not, however, 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 nor for village banks. Furthermore, most microlenders that make loans to individuals are also not ready for scoring, either because they must perfect more basic processes first or because their data bases are not yet adequate for scoring. Even for microlenders who are ready now, scoring will not replace loan officers and their subjective evaluation of risk factors that are not (or cannot be) quantified in a data base. 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 plain-old business organization) that promise smaller—but still important—improvements in microcredit for a long time to come.

1.3 Scoring is about organizational change, not statistics

The central challenge of scoring is organizational change; after all, predictive power can be tested with historical data before scoring is put to use. But loan officers and branch managers sensibly distrust magic boxes. To trust scoring, field personnel must understand how scoring works in principle and then see it work for their own clients in practice. 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 in which managers regularly seek (with the help of full-time, in-house analysts) to mine the untapped knowledge in their data bases 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 myth that scoring will replace loan officers and the myth that scoring will speed the evaluation of loan applications. To help microlenders prepare to take full advantage of scoring, the last section discusses the nuts-and-bolts requirements for the design of data collection.

1.4 Audience and intent of this paper

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, and later sections are more technical. Readers who want less detail can sample from the main text and boxes.

This 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 (Schreiner, 2001b, 2000, 1999; Vogelgesang, 2001; Viganò, 1993). In places, it also draws on the long experience with scoring in high-income countries (Mays, 2000 and 1998; Thomas, 2000; McCorkell, 1999; Hand and Henley, 1997; Mester, 1997; Thomas, Crook, and Edelman, 1992; Lewis, 1990).

The examples here reflect the author's experience with scoring in Latin America. In turn, this reflects the existence of large, sophisticated Latin microlenders with adequate electronic data bases. Provided a sufficiently strong organization and a sufficiently large data base, 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.

2. Subjective scoring versus statistical scoring

All microlenders already use scoring, although it is subjective scoring, not statistical scoring. This section presents the basic concepts of scoring—whether subjective or statistical—and tells why the two approaches are complementary.

2.1 What is scoring?

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).

2.1.1 Subjective scoring

Microlenders currently judge risk with subjective scoring, forecasting repayment based on their quantified knowledge (measured in numbers and recorded in their electronic data base) and their qualitative knowledge (not measured in numbers and/or not recorded in their electronic data base) of the characteristics of the client and of 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 client's personal character. Based mostly on intuition, subjective scoring produces a qualitative judgement 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 a lot of time to absorb the lessons of the organization's experience and to develop a sixth sense for risk. Also, the predictive accuracy of subjective scoring varies by officer and also varies with a given loan officer's mood. Subjective judgement also allows for discrimination or for mistakenly focusing on too few characteristics or on the wrong characteristics.

2.1.2 Statistical scoring

Statistical scoring forecasts risk based on quantified characteristics recorded in a data base. These 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 woman seamstress who was not once late 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.

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 is on time, or no one is. This is a sensible policy, but it means that the data base does not record whether individuals 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 all members are of the same gender, or the average age of members) 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). Of course, the group—through "domino default"—can also 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 groups lenders or village banks.

Scoring's weakness is its newness; only a handful of microlenders have tried it. The use of quantitative knowledge in a data base 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 (Box 1).

Scoring for microcredit also has limited application because it requires an electronic data base that records, for a very large number of past loans, repayment behavior as well as the characteristics of the client and of the loan contract. Furthermore, the data must be reasonably accurate. Some microlenders happen to have accumulated adequate data in the course of their normal portfolio management over time; many others do not have electronic data bases, 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, at some point in the future, their data bases will be adequate to support scoring.

2.2 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. 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. Of course, these assumptions are never completely true, but they usually come close enough to make scoring worthwhile.

2.2.1 The future will be like the past

Nothing is constant but change, and the future is certainly more like the recent past than the distant past. Still, scoring—be it statistical or subjective—presumes that some knowledge of the past is better than none.

Subjective scoring—because it relies on intelligent 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.

2.2.2 Risk is linked with quantified characteristics

Some share of risk is undoubtedly linked with quantified characteristics such as indebtedness and previous arrears. But not all characteristics are quantifiable, 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 after getting to know the client as a person.

What share of risk is linked with quantified characteristics? This paper (buttressed by the tests in Sections 3 and 4) argues that the share is large enough to make statistical scoring worthwhile. The tests in Sections 3 and 4 also show that the share is too small to discard subjective scoring.

2.2.3 Statistical scoring will not replace loan officers

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 a borrower knows her business or whether she plans to drink the loan proceeds. Statistical scoring is nothing more than a third voice in the credit committee that reminds the credit manager and the loan officer of elements of risk that they might have overlooked.

3. How scorecards work and how to test them: A simple four-leaf tree

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 judgements 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. It also shows how to test scorecards before they are used.

3.1 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 7 days of arrears per installment.[2]

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 16.0 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 should be understood to 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 bads ÷ 52,395 loans). For new loans to men, historical risk was 22.3 percent, (5,316 bads ÷ 23,787). 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 stand for 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 ÷ 200,181 = 44.6 percent of paid-off loans (leaf 3, right-most column, Figure 3.)

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

---

[2] Page 46 discusses why "bad" is defined as serious delinquency rather than default.

- 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

How might the microlender act on these insights? For example, because new loans—especially new loans to men—are risky, the lender might want to screen applications in this segment with extra care. Or the lender might reduce the analysis required of loan officers (or the requirements made of clients) for applicants in the low-risk segments. As usual, scoring only predicts risk, without directly telling the lender how to manage risk.

The results from this simple 4-leaf tree are not too surprising; most microlenders probably knew that new loans were riskier than repeat loans and that men were riskier than women. Some might be surprised, however, to discover that new loans to men are almost twice as risky as repeat loans to women. In any case, this simple example is meant to illustrate the concepts of scoring rather than to provide deep insights into previously unknown links between characteristics and repayment risk.

3.2 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.

For example, suppose that the microlender with the four-leaf tree receives a renewal application from a woman and, after its 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 with trees or with 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. Of course, subjective scoring also does this, but it measures historical relationships qualitatively and implicitly rather than quantitatively and explicitly.

3.3 How to test predictive power

Any scorecard can forecast risk, but not all can do it well. Fortunately, predictive power can be tested before use. Historical tests reveal how well the scorecard would have performed, had it been used in the past. The assumption is then that scoring will have similar predictive power from now on.

3.3.1 Why test?

Suppose a stock-picker or horse-bettor concocts a new system to beat the market or the track. Before putting her own cash at stake, she would be foolish not to test the 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 also helps to 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 was observed (after disbursement) to turn out good or bad. Historical tests are a central feature of scoring; no lender should score without first testing predictive power.

3.3.2 Historical tests

Historical tests have three steps:

- Derive a scorecard from loans in the construction sample
- Use the scorecard to forecast risk for loans in the test sample
- Compare predicted risk with realized risk

A historical test divides paid-off loans (along with written-off loans) into two samples. Loans that were paid off up to a deadline in the past make up the *construction sample* used to build the scorecard. In the example 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 data base 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 as they were paid off after the construction deadline but before the data base cut-off. Loans outstanding as of the data base cut-off—such as 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 construction or in 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 through time or are not real patterns at all but rather the results of chance in a finite sample. These patterns are absent from loans outside the construction sample. Thus, the scorecard predicts more accurately for loans in the construction sample than for other loans. In real life, of course, what matters is prediction for loans not in the construction sample.

Second, test loans must be paid-off after construction loans. In real life, the 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 its 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.

3.3.3 A test with the four-leaf tree

For the example four-leaf tree, the construction sample is the 200,181 loans paid off in 1992-99, and the test sample 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, predicted risk for renewal loans to women is the historical risk for the segment (12.8 percent, leaf 3, column "Predicted bads/cases (%)" in Figure 6). It turns out that realized risk in 2000-01 was 12.1 percent (leaf 3, column "Realized bads/cases (%)" in Figure 6). The accuracy of the scorecard is seen in Figure 7 as the distance between the lines for predicted (historical) risk and realized risk.[3]

Predicted risk for new loans to men (the highest-risk segment) is 22.3 percent (leaf 2, column "Predicted bads/cases (%)" in Figure 6). This again comes close to realized risk (21.9 percent in leaf 2, column "Realized bads/cases (%)" in Figure 6). In fact, the tree's risk forecast was close to realized risk in all four segments (Figure 7).

3.4 Summary

This section showed how scoring works, how to test it, and that even a simple scorecard can be useful. 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 a lot like the arrears-based grades that many microlenders already use, but, once scoring has been developed, it is easier to use and more powerful (Box 2).

---

[3] Changes through time in the proportion of paid-off loans in each of the four segments in Figure 7—and also later in Figure 10—are ignored to avoid clutter.

Box 2: Scoring versus arrears-based grading

Many microlenders grade applicants based on their arrears in the previous loan. Scoring is like grading, only more accurate and, because differences between forecasts have known meanings, 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, does not know what share of A’s to expect to go bad, nor does the lender know how much worse an A is than a C. In contrast, scoring not only ranks risk but also—once adjusted for absolute accuracy (Section 5)—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 data base and statistical techniques to optimally link risk to a wide range of characteristics. In contrast, grading links risk to arrears based on the judgement and experience of the managers who concoct the system. Of course, some simple analyses of the data base could inform the design of grading systems, but such analysis is rare. Likewise, historical tests of predictive power are standard for scoring but virtually non-existent for grading.

4. How to use scorecards: A 19-leaf tree

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

4.1 A 19-leaf tree

Like the four-leaf tree above, a 19-leaf tree (Figure 8) was constructed from paid-off loans at a large microlender who defines "bad" as a loan with a spell of arrears of 30 days *or* an average of 7 days of arrears per installment.

The 19-leaf tree has more leaves than the 4-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 already record as part of their traditional evaluation process (Figure 8):

- 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 ÷ assets)
- Guarantee coverage (resale value of chattel guarantee ÷ amount disbursed)

For example, segment 11 is the largest segment (15.0 percent of all loans, column "Cases in leaf/all cases (%)") and also the least risky (4.5 percent, column "bads/goods (%)"). 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 numbers, and
- Were more than 40 years old

In contrast, segment 19 was one of the smallest segments (0.6 percent of all loans, column "Cases in leaf/all cases (%)") and also the most risky (45.6 percent, column "bads/goods (%)"). It contains renewals from applicants who:

- Averaged more than 7 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, with the worst (segment 19, risk 45.6 percent) about 10 times as risky as the best (segment 11, risk 4.5 percent). The microlender probably would want to treat applicants from the highest-risk segments differently than applicants from the lowest-risk segments.

Characteristics were related with risk 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 zero 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 phones) or the not-so-poor (with two phones)

These patterns fit the lender's experience. This confirms the potential of scoring, and it also confirms the lender's intuition. But scoring does more than tell the lender what it already knew; it quantifies links with risk. For example, the lender already knew that risk increased with arrears in the last paid-off loan, but it did 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 was 7.3 percent (computed as the total number of bads in segments 10-13 divided by the total number of cases in those segments). This is 15.3 percentage points less than the risk of renewals with 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).

4.2 Historical test

The test for the 19-leaf tree follows the same process as for the four-leaf tree. Once again, the construction sample covers 1992-99 and the test sample covers 2000-01.

Box 3: How do abrupt changes affect scoring?

When the context changes, scoring loses absolute accuracy, but it usually retains relative accuracy (Lewis, 1990). 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 (Rhyne, 2001; Poyo and Young, 1999). The battle for market share tripled arrears and doubled drop-out.

Can scoring staunch the flow of drop-outs? A desertion scorecard (Section 7) was constructed with data from 1988-96 and tested on data from 1997 (Schreiner, 2001b). The construction and test samples thus straddled the abrupt market shift. Absolute accuracy was low, but relative accuracy was still usefully high.

As before, historical risk in a segment in 1992-99 is taken as predicted risk for loans in that segment in 2000-01. The test then compares predicted risk with realized risk.

How well would the 19-leaf tree, constructed with data from 1992-99, have predicted 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 depicts the comparison of 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 ($[4.5 - 4.1] \div 4.5 = 0.09$). 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 ($[11.5 - 8.2] \div 8.2 = 0.40$).

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; except for a few segments, realized risk consistently increased as predicted risk increased (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 5, 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 (Box 3).

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 in that cases with the lowest predicted risk also had the lowest realized risk and in that 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—see Hand, Mannila, and Smyth, 2001.)

4.3 Acting on scoring according to 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: "super-bads", "borderlines", "normals", and "super-goods." The lender sets the four thresholds to meet its mission, given trade-offs among breadth, depth, and length of outreach (Schreiner, forthcoming).

4.3.1 Where scoring fits in the credit-evaluation process

Scoring ignores qualitative characteristics and considers only quantified characteristics. Thus, scoring cannot replace any part of the traditional evaluation (Box 4); it 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, then the loan officer makes the field visit and—perhaps after some analysis back at 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 the "Effects of Characteristics Report", see Section 5) along with the 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 rubber-stamp. 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 permits itself to peek at the score early,

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, however, and if scoring predicts low risk, then the lender may be tempted to skimp on the (more costly) subjective evaluation. This could be disastrous, as 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 both at 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 abandon (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 under 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 (Rhyne, 2001). For now, scoring complements—but does not replace—loan officers and traditional evaluation.

then it will be tempted to approve loans without screening them for qualitative risk.) Score in hand, the committee applies a four-class scoring policy (bottom of Figure 11).

4.3.2 "Super-goods"

Applicants with predicted risk below the lowest threshold are *super-goods*. To keep these very low risks loyal, the lender might adopt a policy to enhance value for them, for example by offering lines of credit, reduced fees, rebates for perfect repayment, or lower guarantee requirements. Of course, scoring only identifies super-goods without identifying the best way to keep them loyal. Scoring merely forecasts risk; managers must decide what to do next. For example, if they want to use risk-based pricing, then they must decide how to adjust interest rates, given predicted risk.

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

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-goods (Figures 12 and 13). Among these, 4.1 percent went bad, accounting for 4.9 percent of all bad cases. Seen another way, among super-goods, there were 23.6 goods for each bad.

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

4.3.3 "Normals"

Applicants with predicted risk in excess of the super-good threshold but below the borderline threshold are *normals*. 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 normals, so, in most cases, scoring does not affect the evaluation nor impose additional costs on the credit committee.

For the sake of discussion for the 19-leaf tree, suppose that the normal threshold was 12 percent (segments 10, 13, 5, 15, 3, 12, and 9 in Figures 12 and 13). In 2000-01, more than half (55.7 percent) of all cases were normals (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, accounting for 41.2 percent of all bads. Among normals, there were 8.7 goods per bad.

4.3.4 "Borderlines"

Applicants with predicted risk in excess of the borderline threshold but below the super-bad threshold are *borderlines*. The credit committee reviews these cases with

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 scoring's power 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 previous arrears really do 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 signals 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.

extra care and, if warranted, modifies the amount disbursed, term to maturity, guarantee requirements, and interest rates or fees (risk-based pricing). Of course, the committee may decide to reject some borderline cases.

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

For the sake of discussion for the 19-leaf tree, suppose that the borderline threshold was 23 percent (segments 7, 17, 4, 8, 2, and 14 in Figures 12 and 13). 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 30.0 percent of all bads, and there were 3.4 goods per bad.

4.3.5 "Super-bads"

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

For the sake of discussion for the 19-leaf tree, suppose that cases with risk greater than 24 percent are super-bads (segments 6, 18, 1, 16, and 19 in Figures 12 and 13). In 2000-01, 9.0 percent of all cases had risk in excess of 24 percent and so qualified

> **Box 6: Scoring throws out the goods with the bads**
>
> Some applicants rejected as super-bads would have been good, and 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 traditional subjective evaluation also 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 strict/lax policy. With subjective scoring, prediction error is unknown, so choices are less apt to be optimal.
>
> In Latin America, some microlenders who make individual loans are as strict as Scrooge. For example, one renowned Colombian microlender 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 this strictness, it is possible that if lenders knew the true risk/outreach trade-offs better, then they might meet demand better and maintain—or even decrease—risk.

as super-bad. Of these super-bads, 37.2 percent went bad, accounting for 23.9 percent of all bads. Among super-bads, there were 1.7 goods for each bad.

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

4.3.6 Summary of four-class scoring policy

A four-class scoring policy rewards low-risk cases and reviews and modifies (or rejects) high-risk cases. Most cases have about average risk; 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.

4.4 How to set thresholds for scoring policy

This section describes, in broad terms, how to set policy thresholds. The choice of thresholds depends on the predictive power of scoring for a specific microlender and on how the microlender values the trade-offs between aspects of its mission (Schreiner, forthcoming):

- Breadth of outreach (number of loans)
- Depth of outreach (poverty of borrowers)
- Length of outreach (organizational permanence through profits)

Box 7: Must lenders reject risky clients?

Nothing forces a poverty-oriented lender to reject high-risk cases. But a poverty lender would not want to ignore risk forecasts either. No one lends with utter disregard for risk, and even the most hard-core poverty lender limits the cost it will accept to reach a given depth. As always, scoring merely sheds light on trade-offs; the lender still must decide what to do. In addition, rejection need not always hurt applicants. Microcredit sometimes harms more than it helps, especially for the poorest (Mosley, 2001). Some high-risk cases, even if they do not go bad, will struggle so much to pay their debts on time that they would have been better off being rejected in the first place.

A microlender must make these value judgements 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, goods lost, and bads avoided. (As always, the assumption is that the historical test indicates future results in actual use.)

For example, Figure 14 shows 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 2000-01, then 1.7 goods would have been lost for each bad avoided. About 9.0 percent of all cases would be policy rejects, with 23.9 percent of all bads avoided (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 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) versus both breadth (reaching more borrowers by avoiding the worse 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 (Box 7).

Finally, scoring—given estimates of the net financial cost of a good lost and of the net financial benefit of a bad 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 (Box 8).

Box 8: 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 (Coffman, 2001).

The historical test shows—given a super-bad threshold—the number of goods lost for each bad 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 far exceeds the benefit of a good. For example, credit-card lenders in rich countries commonly assume that it takes more than 10 goods to make up for one bad.

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:

(Cost per bad x Bads avoided) – (Benefitpergood x Goods lost).

For the 19-leaf tree, the assumed cost of a bad is \$300, and the assumed benefit of a good is \$100. With a super-bad threshold of 24 percent, the historical test (bottom row of Figure 12, column "Cases") shows that 4,439 cases would have qualified as super-bad. Of these, 1,652 turned out bad (column "bads"), and 2,787 turned out good (column "goods"). 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, then 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 segment 19 (the riskiest segment, Figure 12, column "Cases in leaf/all cases (%)")) would boost profit by (\$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 and the benefit of a good. 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 to the benefit of a good, the greater the potential profitability of scoring.

In practice—as in the example for the 19-leaf tree (Figures 12 and 13)—most microlenders will probably aim for thresholds that result in about 10 percent of cases being super-goods, 60 percent being normals, 20 percent borderlines, and 10 percent super-bads. This broad pattern has four advantages. First, it keeps the share of "super-goods" low. Thus, the lender can offer special incentives to keep the best clients loyal and still control the cost of incentives. Second, most cases are "normal". This means that scoring will not change the standard loan-evaluation process for most cases. This also helps front-line personnel to accept scoring. Third, most risky cases are "borderlines". Loan officers and credit managers in the branches are reluctant to reject applicants purely on the basis of scoring. With most risky borrowers classified as "borderlines", the credit committee is encouraged not to reject but rather to review risky cases and then perhaps adjust the terms of the loan contract. Fourth, the share of "super-bads" is low. The few super-bads 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 bonus). Through 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 so much time in collections that loan officers can increase disbursement enough to more-than-compensate for the rejected cases. Length of outreach (permanence via profits) may increase because the revenue from increased lending volume likely exceeds 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 that thus has the potential to skirt the normal trade-offs between aspects of outreach (Rhyne, 1998; Gonzalez-Vega, 1988). That is, if scoring helps the microlender to do more with less, then it can make everything better without making anything worse.

4.5 Costs of scoring

Scoring has fives types of costs: data-accumulation, set-up, operational, policy-induced, and process.

First, collecting and entering the data required to construct a scorecard incurs *data-accumulation costs*. For the least sophisticated microlenders, this involves not only keying-in 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, and improvements to the information system are probably worthwhile quite apart from their usefulness in scoring. For some other, more sophisticated microlenders, most data-accumulation costs are already sunk; they already key in all applications as they are received anyway. For these lenders, scoring is possible as soon (possibly right away) as the data base has enough cases to support scorecard construction. Finally, a third group of lenders has adequate information systems but does not yet key in applications. Rather than hire an army of data-entry personnel to key in archived paper applications, they can start to capture data in electronic form from now on.

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

Fourth, rewarding super-goods or rejecting super-bads induces *policy costs*. After all, 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 and so induces *process costs*. Some power shifts from the credit department to the information department. Some employees openly oppose scoring's changes, and others try to subtly skirt scoring by cooking data or by ignoring policy rules. Still others inadvertently sabotage scoring by skimping on the traditional evaluation. Training and follow-up (Section 5) are the best ways to manage these process costs.

4.6 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 help a microlender develop a taste for explicit, quantitative analysis as an aid to decision-making by managers.

4.6.1 Reduced loan losses and greater client loyalty

Reduced loan losses is 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.

4.6.2 Risk-based pricing

Given a score, the microlender can manage risk by rejecting the loan application or by 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.

4.6.3 Less time in collections means more time in disbursement

The greatest benefit of scoring results from loan officer's 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 (Box 8) because rejecting super-bads and modifying borderlines means that loan officers must chase down fewer bads (and fewer "almost bads"). They then can spend the time saved on marketing, evaluation, and disbursement.

Many microlenders expect scoring to save them more time in evaluation than in collections. But most loan officers spend at least as much time in collections as in evaluation, and scoring cannot substitute for qualitative evaluation (Box 4).

For example, loan officers at the example microlenders in this paper spend 2 to 3 days per week on collections. Suppose—as in 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 (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.

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 8 discusses a possible bottom-line impact.

Scoring, even though it leads to some loans being 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 (Frame, Padhi, and Woosley, 2001; McCorkell, 1999; Lewis, 1990). After all, 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, 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 judgement. 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

Box 9: Is statistical scoring discrimination?

Statistical scoring does discriminate; it assumes each applicant is another instance of the same old thing, not a unique individual who might differ from other apparently similar cases in the data base. Of course, subjective scoring discriminates just as much, if not more. After all, 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 past borrowers at the same lender with similar quantified characteristics. If women had better repayment than men, then the scorecard says so. In contrast, subjective scoring draws on the experience of microcredit in general, th 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 statistical 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 (Box 11).

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 evaluations. Historical tests are a key to fair discrimination because they show whether supposed links are real. Compared with subjective scoring, statistical scoring is much easier to put to the 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.

judge their risk. Scoring improves the risk-evaluation tool kit and thus helps to purge prejudice and mistakes from the evaluation process (Box 9). If the poor really are creditworthy, then scoring will help reveal that better than ever, deepening outreach.

4.6.4 A taste for explicit, quantitative analysis

The final benefit of scoring—and perhaps the most important benefit 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 data base. For example, once managers set scoring policy with the knowledge of the trade-offs embodied in something like Figure 12, they will only reluctantly go back to vague seat-of-the-pants judgements of the consequences of alternative credit policies.

Finance is all about information, and the information in the data bases 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 (Schreiner, 2001b) or predict what types of potential clients are most likely to respond to a marketing campaign (Delmater and Hancock, 2001; Berry and Linoff, 2000). 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 make better decisions.

5. Training and follow-up: How to get people to use scoring and to use it correctly

In purely technical terms, scoring for microcredit works, as the previous section demonstrates. In human terms, however, scoring is not so straightforward. For people to use scoring to improve choices requires not only cerebral knowledge of how scoring *can* work but also gut faith that scoring *does* work, as well as 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. This section discusses how to do it.

5.1 Why training matters

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

In the first place, a consultant—likely with a strange accent if he or she can even 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 but rather on quantified knowledge of characteristics.

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's forecasts, to see the forecasts hit the mark, and to have follow-up as they use the forecasts.

5.2 Project management for healthy organizational change

Like all projects, scoring needs management buy-in and an in-house champion. Like all projects, this means showing how scoring works and what problems scoring resolves. This is nothing new, just work.

5.2.1 Introductory presentation

Most upper 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, the introductory presentation aims to get them on the right page. This has a very practical purpose. Although the engine of scoring is mathematical, scoring is much more than just a chalkboard exercise; it is a seismic change in organizational culture. This makes a scoring project larger, longer, and more difficult than most managers had imagined. The introductory presentation aims to bring expectations down to earth—better to nip the project in the bud than to do it and be disappointed.

As managers grasp scoring, some get excited, others huddle in defense 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:

- What is your mission?
- How would scoring promote your mission?
- What characteristics do you find to 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 goods would you sacrifice to avoid a bad?
- 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 were the recent macroeconomic booms and busts?
- What parts of the data base 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?

5.2.2 Scorecard construction and testing

The next step is to construct the scorecard and to run the historical test. Results in hand, the scoring project 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 after loan officers and credit managers see that scoring works in the historical test, they cannot bring their hearts to believe. Before they can accept scoring, front-line personnel must pass through denial and disbelief. It is better to give them time to do this before the scorecard is installed.

In this step, the key again 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 data base?
- 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?

5.2.3 Production of scores and reports

The next step is to automate the production of scores and scoring reports. Of course, managers prefer to avoid changing anything in the information system, but, to use of 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 also 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 in its existing information system. This is no small task. First, the microlender (or its software provider) must be able to modify the system. Second, the lender must dedicate a programmer full-time to scoring. Depending on the system, integration requires 3-6 person-months; the lender's systems manager cannot do it on evenings and weekends. Third, the technical challenges of integration vary by lender, so all issues cannot be anticipated in advance. Still, integration has important advantages: data is entered only once, scores are produced automatically, and risk forecasts can be straightforwardly integrated in all the reports that the lender already uses. Weighing both pros and cons, integration is the preferred approach.

5.2.4 Getting to know scoring

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 is over.

This lets people acclimate to scoring slowly, and it lets potential users think about how to use risk forecasts without imposing any pressure to change just yet. Of course, some branches may even not look at the score, and other branches may start to act on it even though they were told not to.

This phase must make explicit time and space for feedback. People seize on any apparent weakness or mistake to discount scoring, so concerns must be heard and addressed. This means a second branch tour to review concepts, show new tests of predictive power for disbursements since the scorecard was installed, and ask 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's predictions for super-bads and for borderlines?

5.2.5 Useful additional reports

At users' request, the project should make adjustments or add reports. Often these changes are merely cosmetic, but two useful ones are described below.

5.2.5.1 The "Scoring Simulator"

Credit committees commonly request to see how modifying borderline cases would affect the risk forecast. The "Scoring Simulator" 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" 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 for each day's credit committee.

5.2.5.2 The "Effects of Characteristics Report"

This report 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 is included in the bundle for each day's credit committee.

5.2.6 Set scoring policy

Once people have had several months to get to know scoring, the microlender sets policy, distributes a "Scoring Policy Manual", and starts to use scoring in earnest.

Why a written policy? Without one, a scorecard is useless at best and harmful at worst. Without explicit rules, most users will take the path of least resistence, falling back on traditional credit evaluation without scoring. Even if some of them would use scoring without an explicit policy, they would probably use it incorrectly (or at least inefficiently), and certainly no two people would use scoring in the same way. Just like traditional credit evaluation, scoring needs a written policy. What should it cover?

5.2.6.1 Policy thresholds and actions

The 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-goods and the risk level above which cases qualify as super-bads. It also establishes the risk levels that correspond to normals and borderlines.

Furthermore, the written scoring policy tells how to reward super-goods. For borderlines, it tells how the credit committee should prioritize attempts to mitigate risk—whether by decreasing loan size, decreasing term to maturity, and/or increasing guarantee coverage—and how to use the "Scoring Simulator" (Figure 18) to see the likely effects of these possible modifications to the loan contract. Finally, the written scoring policy should emphasize that super-bads really are to be rejected.

5.2.6.2 Override policy

Choices that go against scoring policy are *overrides*. In microcredit, overrides are super-bads approved and borderlines left unreviewed.

Scoring is most valuable as a way to identify high-risk cases that the credit committee thought were safe bets. Loan officers and credit managers, however, are human, and when scoring contradicts their judgement, they may scoff and go on, 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, Box 10), or demand to know why risk is high.

Override policy deals with this in three ways. First, it constantly tests predictive power via the "Global Follow-up Report" (see below). 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 (Holtmann, 2001): if overrides exceed x percent of super-bads, then

Box 10: Why was scoring wrong for this borrower?

Like good weather forecasts, good scoring forecasts work on average, not for each day nor 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 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 stay good, and 1 in 20 of all borrowers with 5-percent risk 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 turn out 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 right or wrong (although the correctness of reject decisions will never be known).

cut the bonus. In the long term, explicit sanctions are less necessary, as loan officers come to realize that abuse of overrides leads only to greater arrears and a smaller performance bonus.

Of course, careful overrides have their place: the credit committee sometimes does know that a particular case is exceptional, and of course only human judgement can evaluate the qualitative characteristics ignored by the scorecard. The point is moderation; just as not all people can be above-average, not all high-risk loans can 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.

In any case, 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 to make their prophecies come true) but more risk than other loans (because the scorecard does know something that the credit committee does not).

5.2.6.3 Underride policy

Override policy seeks to prevent depending on scoring too little; underride policy seeks to prevent depending on scoring too much. In particular, written policy must stress (as does this paper) that scoring works only for applications already provisionally approved in the traditional evaluation process (Box 4). Constant reminders are needed because after once-skeptical people see scoring work, they may go too far to the other side and neglect traditional evaluation. But 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 can only add scoring after the subjective evaluation is completed. Otherwise, arrears may skyrocket.

5.3 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 report of scoring, more useful even than the historical test. It checks whether scoring works with live loans.

Like other scoring reports, the "Global Follow-up Report" is produced automatically by the system. In the first months of scoring, the lender consults it weekly to check predictive power and to guide adjustments to policy. After that, monitoring takes place monthly.

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

5.3.1 An example "Global Follow-up Report"

Figure 20 is a "Global Follow-up Report" based on a regression scorecard (Section 6) of a Latin American microlender. "Bad" is defined as an average of 4 days of arrears per installment due so far *or* a spell of arrears of 30 days.

The left-most column ("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 ("# Loans out. (%)") is the share of loans outstanding whose predicted risk falls in 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 had predicted risk in excess of 40 percent (adding down columns), and 19.5 percent had predicted risk in excess of 30 percent. Numbers in this 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 (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. As 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 (Vogelgesang, 2001). Thus, the best test of predictive power looks at recently paid-off loans and/or at 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 to go; 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 a full chance to go bad.

5.3.2 Uses of the Global Follow-up Report

5.3.2.1 Check 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, as 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 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 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 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.

5.3.2.2 Track 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 started. If, as loans disbursed under scoring age, realized risk among super-bads turns out to be far less than predicted risk, then it means that 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 the first day after starting scoring shows 78-percent realized risk among past loans that would have qualified as super-bads. 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. Still, 35 percent might be more risk than the lender wants to bear; if so, it would tighten override limits. Or perhaps the lender is willing to make loans this risky, in which case the current override policy would be maintained.

5.3.2.3 Fix absolute inaccuracies

Scorecards with absolute accuracy are easier to use. Relative accuracy merely orders loans by expected risk; 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 predicted risk. Given this information, the user can adjust the levels of predicted risk so that the adjusted predictions are absolutely accurate.

As a simple example, 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, say, 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.

5.3.2.4 Set or adjust policy thresholds

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

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 bads for each good lost (because realized risk in this range is about 75 percent, 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.

5.3.2.5 Detect 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 in 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 with 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 “Global Follow-up 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 everything else that both affects risk and changes through time. Before degradation advances too far (probably after 2-4 years), the microlender should renovate the scorecard. Renovation is simpler and quicker than the initial scoring project; just construct a new scorecard—including data accumulated since the first scorecard—and plug it into the existing system.

5.4 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). In a technical sense, 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 individual borrowers whom 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 already receive daily or weekly. Figures 23 and 24 are simple reports from a regression scorecard (Section 6) 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.

On the super-bad side, 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 10 days. When loan officers see their own borrowers in such a list, and when they recall the troubles they have had in collections with these borrowers, then 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 10 days.[4]

For loan officers and branch managers, seeing their own borrowers in reports such as those in 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 reports that they already distribute daily or weekly to loan officers and credit managers.

[4] 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"—are 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 1 in a billion.

5.5 Wrap-up

If employees give scoring a chance, then they will see that it works. But to give scoring a chance, they must understand it and believe that success is likely; this is the role of training and tests. Once people accept scoring, 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.

6. 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, trees, and expert systems.

6.1 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 just 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.

6.1.1 Examples of simple one-characteristic regression scorecards

6.1.1.1 Age

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

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

For example, predicted risk for a 30-year-old is 10 – 0.1 x 30 = 7 percentage points. For a 55-year-old, predicted risk is 10 – 0.1 x 55 = 4.5 percentage points. (These weights are examples. Real weights are lender-specific.)

6.1.1.2 Term to maturity

As a second example, suppose that the 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:

$$\text{Risk} = 10 + 0.25 \text{ x Term-to-maturity.}$$

For example, predicted risk for a 3-month loan is 10 + 0.25 x 3 = 10.75 percentage points. For a 12-month loan, predicted risk is 10 + 0.25 x 12 = 13 percentage points.

6.1.2 Age and term to maturity combined

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

$$\text{Risk} = 10 - 0.1 \text{ x Age} + 0.25 \text{ x Term-to-maturity.}$$

For example, a 30-year-old 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 3-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 30-50 characteristics and would derive all weights from the particular microlender's data base. After the information system computes the forecast, the lender uses it as described in previous sections.

6.2 Predictive power of regression

Of all types of scorecards, regression usually has the best predictive power. For example, the results in the "Global Follow-up Report" in Figure 20 in the previous section come from a real-life regression scorecard. Likewise, the "Loan Officer Follow-up Reports" that look at the 30 highest-risk cases and the 30 lowest-risk cases (Figures 23 and 24 in the previous section) also come from a real-life regression scorecard.

6.3 Links between risk and characteristics from regression

Although regression is powerfully predictive, perhaps its greatest advantage is that it clearly shows the relationships between risk and characteristics. The weight assigned to a characteristic shows not only whether the characteristic increases or decreases risk—other characteristics in the scorecard kept constant—but also by how much. As always, these links hold only after an application is provisionally approved under traditional evaluation standards. The examples below are from a real-life regression scorecard of a Latin American microlender.

6.3.1 Experience of the applicant as a borrower

The regression scorecard shows that risk decreases strongly with the number of months since disbursement (Figure 25). For example, someone with 36 months has—all else constant—4.4 percentage points less risk than someone with 12 months.

6.3.2 Age of the applicant

Risk decreases strongly with age (Figure 26). For example, a 50-year-old has—all else constant—about 2.9 percentage points less risk than a 30-year-old.

6.3.3 Indebtedness

Risk increases with the ratio of liabilities to assets in the household/enterprise (Figure 27). For example, someone with 10-percent indebtedness would—all else constant—have 0.2 percentage points less risk than with 30-percent indebtedness.

6.3.4 Arrears in previous loans

Risk increases with the average days of arrears per installment in each of the three previous loans (Figure 28). For example:

- 10 days of arrears in the last loan increases current risk by 8 percentage points
- 7 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 with someone 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 is already provisionally approved under traditional evaluation standards).

Figure 28 offers four broad lessons about the relationship between future arrears and past arrears for a given borrower. First, more realized risk in the past means more predicted risk in the future. Second, arrears in the distant past are weaker signals than are arrears in the recent past. Third, compared with a perfect record, even very small amounts 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 at this microlender is less than 15 percent, a 2-percentage-point change is large. Fourth, risk increases with past arrears, but at a diminishing rate. (Of course, this relationship holds only for provisionally approvals.)

6.3.5 Type of business

The type of business is strongly related with risk (Figure 29). For this microlender, the lowest risks were (top of column "Effect on risk (%)" downward):

- Taxis and truck drivers
- Stores whose inventory rotates quickly (fruits and vegetables, groceries, small household items)
- Venders of street food (fast foods, bakeries)
- Beauty salons and cosmetic stands
- Seamstresses

The highest risks for this lender were (bottom of column "Effect on risk (%)" upward in Figure 29):

- 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

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

6.3.6 Loan officer

Just as regression can reveal the link between risk and a particular type of business, it can also reveal the link between risk and a particular loan officer (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 lot, but only regressions—not trees or expert systems—can use this knowledge to boost the accuracy of predicted risk. The lender can use this knowledge to target training, encouragement, and bonuses.

A caveat applies to interpreting Figure 30. Loan officers manage risk by screening before disbursement and by monitoring 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*, that is, as if all loan officers managed portfolios with the same quantified characteristics. In fact, loan officers manage different portfolios, and their composition (both quantified and qualitative) depends on how well the loan officer screens applicants. On the one hand, some loan officers achieve a given level of portfolio risk by screening for applicants who do not need much monitoring. On the other hand, some loan officers achieve the same level of portfolio risk with less screening and more monitoring. Furthermore, 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 the loan officers who rank worst in the regression scorecard but rather investigate the reasons for the low ranks and then work to address them.

6.4 Expert systems

Scorecards derived not from the statistical analysis of data but rather from the experience and judgement of managers are called *expert systems.* Expert systems differ from traditional subjective scoring in that subjective scoring uses implicit judgements but expert systems use only explicit rules or mathematical formulae. The strength of expert systems is that they do not require a data base and that—because they are constructed by the 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.

6.4.1 Examples

6.4.1.1 Expert-system trees

Expert-system trees are like statistical trees, except their splits come not from a statistical analysis of the data base by a consultant but rather from the experience, judgement, 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 (Box 2).

6.4.1.2 Regression trees

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, but it is a rank, not a probability, so scores may exceed 100 or even be negative. Thus, expert-system regressions lack absolute accuracy, although, they may achieve some level of relative accuracy.

6.4.2 Improving expert systems

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 just 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 could use the tests to convert 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. Even though managers do choose sub-optimal splits and sub-optimal weights, expert systems may nonetheless be usefully predictive (Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Wainer, 1976). Furthermore, expert systems may compensate with their low predictive power with their low data requirements and their ease of adoption.

Microlenders should feel free to experiment with simple home-grown scorecards (Schreiner, 2001c), but they should test them, both before and during use. Incredibly, most microlenders who use expert systems today have not tested them. Their mistake is not that they use expert systems rather than statistical scorecards but rather that they neglect to test predictive power. Those who score should walk by sight, not faith.

6.5 Comparison of regressions, trees, and expert systems

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

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

Expert systems are easy to construct. They do not require data, so they are probably the most relevant type of scorecard for most microlenders today. Expert systems, however, 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 start to plan to collect the data needed to support a better scorecard.

7. How to prepare to score: What risk to forecast and what data to collect

Scoring can predict only what has already happened many times, and even then only if it is recorded in a data base. Thus, cutting-edge risk evaluation is hostage to mundane data collection. Right now, the data bases of most microlenders are inadequate for statistical scoring. Because data requirements depend on the exact risk to be forecast, this section first describes five types of risk. It then discusses what data microlenders should collect now so that they are prepared to score in a few years.

7.1 What risk to forecast?

To stay simple, the first scoring project should construct only one scorecard, so the lender must choose from pre-disbursement scoring, post-disbursement scoring, collections scoring, desertion scoring, or visit scoring (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.

7.1.1 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":

- 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
- Some combination of the above

Defining "bad" for scoring can be a healthy exercise because it forces the microlender to think carefully about arrears and costs. What matters more, the number of spells or their length? Can numerous short spells be tolerated? Lenders should also ask themselves what criteria they currently use to determine whether to allow a client with arrears in the previous loan to repeat.

For pre-disbursement scoring, the definition of "bad" should not be "default." On a merely 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; after all, loan officers do not ask themselves "If I approve this 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.

7.1.2 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 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—such as outstanding loans with severe arrears since disbursement—in which post-disbursement risk is already obvious to the lender.

Regardless of the scorecard used to forecast post-disbursement risk, there is a simple two-class policy choice (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" (such as in 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

In the "courtesy visit", loan officers simply pay a visit—unrelated to any current collection issue—and discuss any non-threatening topic. By no means should the loan officer should let on to clients that scoring fingered them as high risks, lest it become a self-fulfilling prophecy. Anyway, 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, so loan officers need not discuss repayment at all. Instead, they 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, or 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.

7.1.3 Collections scoring

Collections scoring predicts the probability that a loan currently x days late will reach $x + y$ days. Most commonly, it would predict the risk that a loan that fell into arrears yesterday and is now 1 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 on value-at-risk, loan officers would follow a three-class policy to decide who to visit first and how gently to dun them (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, but these are delivered with a gentler tone. 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 might very well cure themselves, and if not, a friendly nudge might 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 three for the price of one.

7.1.4 Desertion scoring

Desertion scoring predicts the probability that a borrower will apply for another loan once the current one is paid off (Schreiner, 2001b). Microlenders seek to prevent desertion because profitability usually increases with each repeat loan (Churchill and Halpern, 2001; Rosenberg, 2001). If the lender knew which clients were at-risk to drop out, then it could encourage them to repeat, perhaps offering them a reduced interest rate or forgiveness of the disbursement fee, contingent, of course, on satisfactory repayment of the current loan. These incentives, however, are costly to the lender; desertion scoring controls costs by targeting the incentives to 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 (Figures 32 and 34):

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

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

- *Loyalists* are not at-risk to drop out nor to go bad. The lender does not offer them special incentives because they probably will repeat anyway

7.1.5 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 any 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 (Figure 32) rejects *unpromising* clients (perhaps the worst 5 percent of visit scores) without a visit but does visit *promising* clients per the traditional evaluation process.

Of course, 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 very 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, *sans* 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 with 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 a 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 rejection can be tested before use on 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.

7.1.6 Summary

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, then they could add desertion scoring and, for those lenders with adequate data, visit scoring.

7.2 Data requirements[5]

Most microlenders do not yet 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. This has three parts. The first is the simple accumulation of more bad cases. This takes time and—for lenders with small portfolios—growth. The second is the collection of additional characteristics on the borrower, loan, and lender. The third part is the improvement of data quality.

7.2.1 Required number of bads

There is no way to know exactly how many bads are needed to construct a scorecard. 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 (Cochran, 1977).

The received wisdom in high-income countries is that scorecards require at least 500 bads (Lewis, 1990). This assumes, however, that clients have both salaried jobs and credit records in a credit bureau.[6] In this special case, a scorecard with 10-15 characteristics (most of them from the credit report) can suffice to construct a powerful scorecard. At the moment in microcredit, however, most borrowers are self-employed, and if there is a credit bureau, most borrowers are not in it yet.

Thus, the typical characteristic in a microcredit scorecard is much less predictive than the typical characteristic in a scorecard in a high-income country (Schreiner, 2000). To get adequate predictive power in microcredit requires more characteristics, and to derive the links between risk and more characteristics requires a larger number of bads.

To construct a useful microcredit scorecard probably requires at least 1,000 bads. 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) and, they also depend on the lender and the context. Such uncertainty is the price of innovation.

Might microlenders pool their data? After all, this is what small-business lenders do in the United States (Asch, 2000). Unfortunately, in microcredit one size does not fit all. A pooled-data scorecard might be better than nothing only if the microlenders involved worked in the same country, with the same target market, and with the same traditional evaluation process. Transferring scorecards across borders would be foolish.

---

[5] Parts of this section draw on Schreiner (2001d).

[6] Credit bureaus are data bases that work with multiple lenders to collect, store, and distribute information on the repayment behavior of individual borrowers.

7.2.2 Collection of appropriate characteristics

What characteristics should a microlender collect now so as to construct a scorecard (or a more powerful scorecard) in a few years? In the list below, the core set of required characteristics is marked with asterisks. Most microlenders who make individual loans already collect most of this core as part of 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 form.

At a minimum, microlenders who plan to score should quantify loan officer's subjective judgements, enter credit-bureau data into their information system, and record household assets and demographics. Lenders should not go back and collect this data for past loans, but they should start to record it from now on.

7.2.2.1 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 grade completed in school*

Dates of marriage or separation are convenient proxies for family stability. Of course, some microlenders may choose to ignore risk linked with demographic characteristics that the applicant did not choose for herself (Box 11).

Contact information

The presence of phone numbers in the data base 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

Box 11: 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 as to their marital status or place of residence. Yet all these characteristics are visible at a glance and thus can be—and 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 the lenders participated in their oppression and because their oppression made them 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 aim is to purge explicit oppression from non-statistical scoring and to prevent statistical scoring from using the knowledge that oppression elsewhere in society has linked risk with 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. After all, genes 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.

But of course this does not really resolve the issue. Even if women are more (or less) risky not because they are women but because society limits women, they still did not choose to their characteristics. To some extent, even non-protected characteristics are involuntary; for example, surely many poor people did not choose to be poor. Even apparently chosen characteristics result from some unobserved clash between choice and constraint. In the limit, 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 judgement about what data to collect and how to use it. Scoring can improve this judgement by quantifying the trade-offs between the use of certain characteristics and predictive accuracy.

The distance to the nearest branch (and the presence of a telephone) proxies for transaction costs (Rojas and Rojas, 1997; Cuevas, 1988). Greater transaction costs increase arrears by borrowers and make monitoring more difficult for loan officers.

Household demographics

Household composition affects cash flows and risks.

- 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 through 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 (presence or absence)
  — Homestead land without title (presence or absence)
  — 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)

Of course, the relevant household assets depend on the local context. Assuming that assets would not change in the absence of loans, then 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. Of course, 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 (presence or absence)
- Written records of cash flows (presence or absence)
- 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 usually useless for scoring because it mixes seasonal with permanent, part-time with full-time, family with non-family, and paid with unpaid. 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 expense 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:

- Share of sales in cash (versus credit)

Financial data must be collected by the loan officer in 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.

<u>Stocks of the enterprise</u>

Most microlenders already record the value assets and liabilities:

- Total assets*
  — Fixed assets*
  — Inventory*
  — Cash and bank accounts*
- Total liabilities*
  — Informal debts*
  — Formal debts*

<u>Repayment record</u>

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 will allow 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 (Staten, 2001; Haidor, 2000); if lenders receive credit-bureau reports for some borrowers, then they should start to enter them 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 be highly correlated 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 probably marks someone as likely to follow a repayment regimen faithfully. Of course, 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 (yes or no)
- Date of last salaried employment
- Participation in Rotating Savings and Credit Associations
  — Date of most recent participation
  — Amount of periodic contribution
  — Frequency of contribution

Participation in a Rotating Savings and Credit Association signals experience as a saver and a debtor. A Rotating Savings and Credit Association may also serve as a fall-back source of funds to finance installments paid to the microlender.

Quantified subjective judgements

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

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

- Overall credit risk
- Honesty and transparency of responses
- Quality of references
- Entrepreneurial oomph
- Business prospects
- Variability of cash flows
- Extent of recent investment in the home or business
- Grasp of the rules in the loan contract
- Family relationships and informal support
- Organization and cleanliness of the home and business

Of course, this does not work if all accepted applicants are rated above-average.

7.2.2.2 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 installment*
- 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 the cosigner

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

7.2.2.3 Characteristics of the lender

The branch and the loan officer strongly influence risk:

- Branch*
- Loan officer*

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 predict better 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 grade completed

7.2.2.4 Are more data worth it?

Given enough bads, a usefully powerful scorecard can be constructed from the core characteristics marked with asterisks above. Microlenders already collect most of these core characteristics. Should they collect more, getting greater predictive power but also incurring greater costs for data collection and data entry?

A scorecard with all the characteristics listed above would probably predict 20-40 percent better than a scorecard with just the core characteristics marked with asterisks. The main costs of collecting additional data are to redesign paper forms, to modify the information system, and to enter the additional data. Although loan officers would have to do a little extra work, the client (if literate) can be trusted to supply most of the additional items on an initial written application quickly and truthfully.

7.3 Guidelines for warehousing better-quality data

After human resources, information is a microlender's greatest asset. 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 the electronic data base rewards greater attention to data quality.

Box 12: 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 is also sometimes dirty because the loan officer may manipulate the appraisal so that an application that he or she deems worthy satisfies the financial ratios required in 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 usefully strong signals.

7.3.1 Discuss data quality with front-line personnel

Most microlenders have collected the core set of characteristics marked above for years, but they never used this data for anything. Loan officers and data-entry personnel have therefore learned that care with data quality costs them time but offers no reward in return. With scoring, however, data quality matters. To make the requisite effort, front-line personnel need to know that old habits are no longer acceptable, why, and how they will benefit.

7.3.2 Establish consistent definitions for the type of business

The type of business tends to be one of the three most-predictive characteristics (past arrears and the identity of the loan officer are the other two). Unfortunately, data quality for type of business is often very poor in that a given code encompasses a wide range of businesses and thus does not distinguish sharply between high and low risks. Still, even dirty and noisy data is better than no data (Box 12).

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

The first step toward improvement is to make front-line personnel aware of the issue. The second step is to make a list of the 50 or so most-common business types, define each one carefully, and teach them to loan officers and data-entry personnel.

About 90 percent of businesses will fall under one of these 50 codes, with the rest 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 (for example, "shoes") and as one of the four activities (for example, "trade"). The fifth step is to include a checklist of all the sectors (with definitions) and of all the business types on the paper 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 the three top characteristics.

7.3.3 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 today the lifeblood of scoring and perhaps tomorrow of market research (Murray, 2001) and client monitoring (Woller, 2001). The rule is: once keyed in, keep it in.

7.3.4 Enter rejected applications into the information system

Many microlenders would like to use visit scoring to shorten (or skip) some field visits. This means forecasting either repayment problems or post-visit rejection.

Forecasting repayment problems before the visit might work, but only a live test confirms predictive power (the "Global Follow-up Report" cannot help). Because such a visit scorecard is constructed only from approved borrowers who passed a qualitative screen, forecasts for unscreened borrowers have unknown accuracy (Box 4). In any case, loan officers would 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 several thousand written applications of rejected clients into their information system. Once they have data on both post-visit rejects and post-visit approvals, they can construct scorecards to forecast rejection based on characteristics known before the visit. (Even after lenders enter rejected applications, a visit scorecard for repayment risk still cannot approve without a visit because the repayment behavior of unscreened borrowers is still unknown.)

7.3.5 Record both the screening loan officer and the monitoring loan officer

One of the three most-predictive characteristics is the identity of the loan officer. The officer in charge of a loan, however, sometimes changes, perhaps because the original (screening) officer no longer works for the microlender, because the lender draws up new work zones or opens new branches, or because a prolific loan officer transfers some outstanding loans to new or less-productive colleagues. When this happens, most systems delete the screening officer and record only the current (monitoring) officer. 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 (unidentified) 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 data base that records the screening officer. The original "loan officer" field continues to record the current (monitoring) officer. Of course, 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; after all, most loans have but one loan officer. In practice, loan officers fingered by scoring as high risks seemingly always 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 requires accounting for transferred loans during scorecard construction. In turn, this requires tracking both the screening officer and the monitoring officer in the data base.

7.3.6 Record missing values as missing, not as zero

Sometimes an applicant leaves a blank space on a written application, 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 rather unknown.

The presence of missing values is often very predictive. For example, loan files missing data on business revenue may be more risky than loans 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, to make up data, or to invent (inconsistent) codes for missing values. (For example, one large microlender 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 is often non-zero, often zero, and often 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 data base languages already reserve a code for missing values. For other languages, the microlender can use "–99".

7.4 Wrap-up of data requirements

Regardless of the type of risk to be forecast, statistical scoring requires a lot of good-quality data. Even the few microlenders who are already have adequate data bases should start to enter loan-officer judgements, credit-bureau reports, and rejected applications into their information systems. The rest of microlenders, if they are to use scoring a few years down the road, must revamp their systems now. Improving the quality of the data base is hard work, but not quite as hard as forever judging risk without the help of scoring.

8. Summary

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 start to seek greater knowledge and precision about alternatives and consequences in all their decisions. Although even simple data analyses can inform decisions, most microlenders have yet to invest in—let alone take advantage of—the asset that is an accurate, comprehensive data base.

On average, scoring in microcredit comes close to the target; for example, about 20 percent of loans with a predicted risk of 20 percent 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, no more (and no less) than a support for the judgement of the loan officer and the credit manager.

The purpose of scoring is to forecast risk, but, for a microlender who wants to start to score, predictive power is a secondary concern, as it can be tested with historical data before being implemented. The primary concern is acceptance by board members, managers, and loan officers. In the end, statistical weaknesses don’t kill scoring projects; people do (Mayr, 2000; Leonard, 1998; McCahill, 1998; Edelman, 1992). After all, 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 is not the next breakthrough in microcredit. For a few microlenders, however, scoring can reduce time in collections and so boost efficiency, outreach, and sustainability. As more organizations learn about scoring and set up processes to accumulate adequate data, scoring will likely become part of best-practice microcredit.

Some might argue that scoring is a new-fangled gadget that microcredit can do without. “If it ain’t broke, don’t fix it,” they say. Of course, lenders in high-income countries said the same thing for decades, and scoring has now all but driven out manual evaluation, especially for the small, short, uncollateralized loans that most closely resemble microcredit (Lewis, 1990). 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.

References

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

Berry, Michael J.A.; and Gordon Linoff. (2000) *Mastering Data Mining: The Art and Science of Customer Relationship Management*, New York: John Wiley and Sons, ISBN 0-471-33123-6.

Churchill, Craig F.; and Sahra S. Halpern. (2001) *Building Customer Loyalty: A Practical Guide for Microfinance Institutions*, Microfinance Network Technical Guide No. 2, www.bellanet.org/partners/mfn/publications.html.

Cochran, William G. (1977) *Sampling Techniques, Third Edition*, New York: Wiley, ISBN 0-471-16240-X.

Coffman, John. (2001) "An Introduction to Scoring for Micro and Small Business Lending", paper presented at the World Bank conference on "Profiting from Small Business Lending", April 2-3, Washington, D.C., www.worldbank.org/wbi/banking/creditscoring/.

Cuevas, Carlos E. (1988) "Transaction Costs of Financial Intermediation in Developing Countries", Economics and Sociology Occasional Paper No. 1469, The Ohio State University, ccuevas@worldbank.org.

Delmater, Rhonda; and Monte Hancock. (2001) *Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence*, Boston, MA: Digital Press, ISBN 1-55558-231-1.

Edelman, D.B. (1992) "The Introduction of Credit Scoring into Branch Banking", pp. 161-177 in L.C. Thomas, J.N. Crook, and D.B. Edelman (eds) *Credit Scoring and Credit Control*, Oxford: Clarendon Press, ISBN 0-19-853-651-8.

Frame, W. Scott; Padhi, Michael; and Lynn Woosley. (2001) "The Effect of Credit Scoring on Small Business Lending in Low- and Moderate-Income Areas", Federal Reserve Bank of Atlanta Working Paper No. 2001-6, www.frbatlanta.org/publica/work_papers/index.html.

Gonzalez-Vega, Claudio. (1998) "Microfinance: Broader Achievements and New Challenges", Economics and Sociology Occasional Paper No. 2518, The Ohio State University, `aede.osu.edu/Programs/RuralFinance/publications.htm`.

Haidor, Elinor. (2000) "Credit Bureaus: Leveraging Information for the Benefit of Microenterprises", *Microenterprise Development Review*, Vol. 2, No. 2, pp. 1, 5-8, `www.iadb.org/sds/doc/2156ENG.pdf`.

Hand, David J.; Mannila, Heikki; and Padhraic Smyth. (2001) *Principles of Data Mining*, Cambridge, MA: MIT Press, ISBN 0-262-08290-X.

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

Holtmann, Martin. (2001) "Designing Financial Incentives to Increase Loan-Officer Productivity: Handle with Care", *MicroBanking Bulletin*, Vol. 6, pp. 5-10, `www.microbanking-mbb.org`.

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

Leonard, Kevin J. (1998) "Credit Scoring and Quality Management", pp. 105-126 in David J. Hand and Saul D. Jacka (eds) *Statistics in Finance*, London: John Wiley and Sons, ISBN 0-471-24634-4.

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

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

Mayr, Mona A. (2000) "Successful Implementations: To Be, or Not to Be", pp. 337-352 in E. Mays, *Handbook of Credit Scoring*.

Mays, Elizabeth. (2000) *Handbook of Credit Scoring*, Chicago: Glenlake, ISBN 0-8144-0619-X.

_____. (1998) *Credit Risk Modeling: Design and Application*, New York: Glenlake, ISBN 0-8144-0475-8.

McCahill, Leonard J. (1998) “Organizational Issues in Building and Maintaining Credit Risk Models”, pp. 13-22 in E. Mays, *Credit Risk Modeling*.

McCorkell, Peter. (1999) “Credit Scoring 101”, presentation at a Federal Trade Commission public forum on “The Consumer and Credit Scoring”, www.ftc.gov/bcp/creditscoring.

Mester, Loretta J. (1997) “What’s the Point of Credit Scoring?”, *Business Review*, Sept./Oct., pp. 3-16, Federal Reserve Bank of Philadelphia, www.phil.frb.org/files/br/brso97lm.pdf.

Mosley, Paul. (2001) “Microfinance and Poverty in Bolivia”, *Journal of Development Studies*, Vol. 37, No. 4, pp. 101-132.

Murray, Inez. (2001) “What do MFI Customers Value? A Comparative Analysis from Three Continents”, manuscript, Women’s World Banking, imurray@swwb.org.

Poyo, Jeffrey; and Robin Young. (1999) “Commercialization of Microfinance: The Cases of Banco Económico and Fondo Financiero Privado FA$$IL, Bolivia”, Microenterprise Best Practices, www.dai.com.

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

_____. (1998) “The Yin and Yang of Microfinance: Reaching the Poor and Sustainability”, *Microbanking Bulletin*, Issue 2, pp. 6-8, www.microbanking-mbb.org.

Rojas, Mariano; and Luis Alejandro Rojas. (1997) “Transaction Costs in Mexico’s Preferential Credit”, *Development Policy Review*, Vol. 15, pp. 23-46.

Rosenberg, Richard. (2001) “Measuring Client Retention”, *MicroBanking Bulletin*, No. 6, pp. 25-26, www.microbanking-mbb.org.

Schreiner, Mark. (forthcoming) “Aspects of Outreach: A Framework for the Discussion of the Social Benefits of Microfinance”, *Journal of International Development*, www.microfinance.com.

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

_____. (2001b) "Scoring Drop-out at a Microlender in Bolivia", manuscript, Center for Social Development, Washington University in St. Louis, `www.microfinance.com`.

_____. (2001c) "Do-It-Yourself Scoring Trees for Microfinance", Tercer Seminario sobre Banca y Microfinanzas en Latinoamérica y el Caribe, Santo Domingo, Nov. 11-12, `www.microfinance.com`.

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

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

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

Staten, Michael E. (2001) "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", April 2-3, Washington, D.C., `www.worldbank.org/wbi/banking/creditscoring/`.

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

Thomas, Lyn C. (2000) "A survey of credit and behavioral scoring: forecasting the financial risk of lending to consumers", *International Journal of Forecasting*, Vol. 16, pp. 149-172.

_____; Crook, J.N.; and D.B. Edelman. (1992) *Credit Scoring and Credit Control*, Oxford: Clarendon Press, ISBN 0-19-853651-8.

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

Vogelgesang, Ulrike. (2001) "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, www.vwl.uni-mannheim.de/gk/_ulrike.

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

Woller, Gary. (2001) "Summary of Important Issues to Consider in Designing and Implementing an Impact Monitoring System", manuscript, Brigham Young University, gmw@ucs-exch.byu.edu.

Figure 1: Comparison of subjective scoring and statistical scoring

Dimension	Subjective scoring	Statistical scoring
1. Source of knowledge	Experience of loan officer and organization	Quantified portfolio history in data base
2. Consistency of process	Varies by loan officer and day to day	Identical loans scored identically
3. 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
4. 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
5. 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
6. Process of implementation	Lengthy training and apprenticeships for loan officers	Lengthy training and follow-up for all stakeholders
7. Vulnerability to abuse	Personal prejudices, daily moods, or simple human mistakes	Cooked data, not used, underused, or overused
8. Flexibility	Wide application as adjusted by intelligent managers	Single-application; forecasting new type of risk in new context requires new scorecard
9. Knowledge of trade-offs, "what would have happened"	Based on experience or assumed	Derived from tests with loans repaid after the loans used to construct scorecard

Figure 2: Four-leaf tree with data from 1992-99 (tree form)

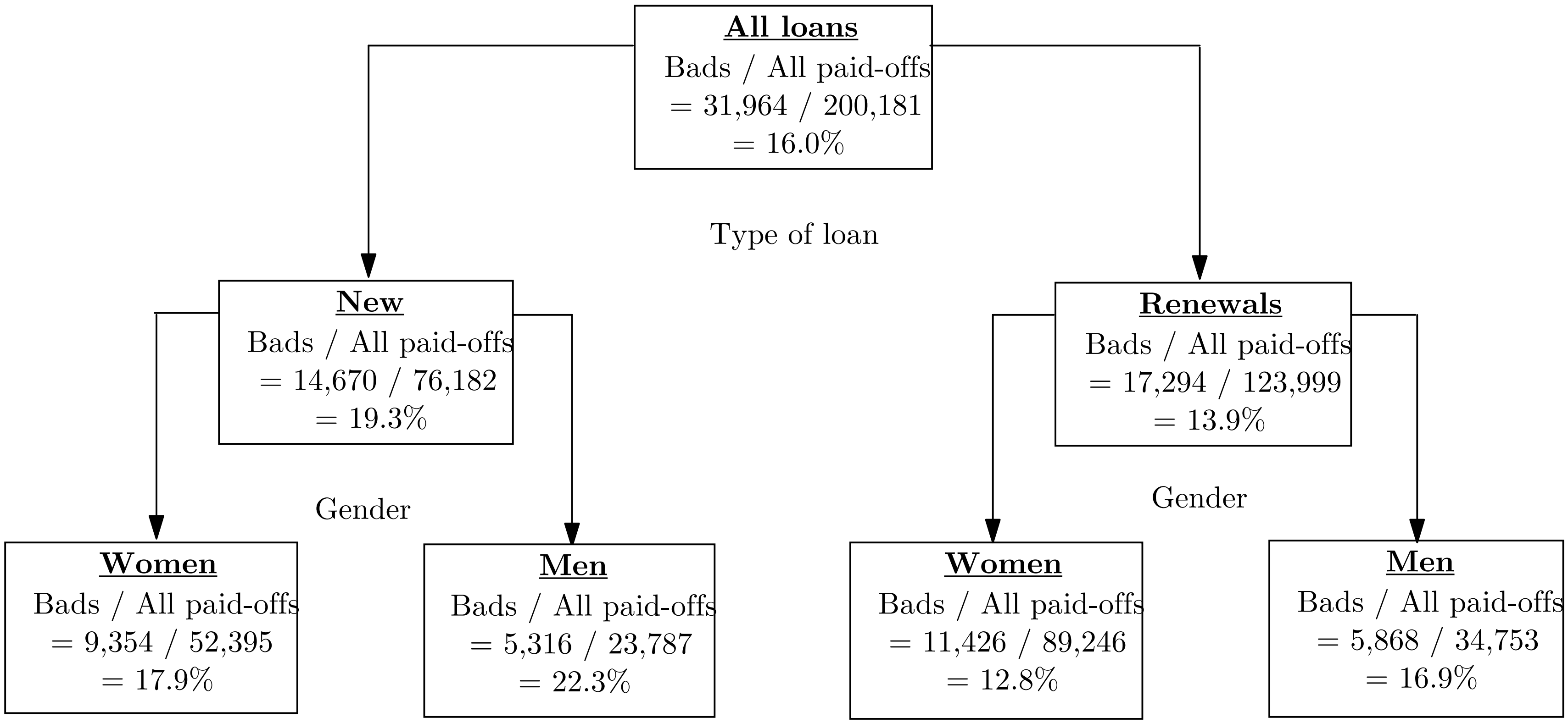


Figure 3: Four-leaf tree, historical risk in 1992-99 (table form)

	Branch of tree		Construction, 1992-1999				
Leaf	**First**	**Second**	**Bads**	**Goods**	**Cases**	**Bads / cases (%)**	**Cases in leaf / all cases (%)**
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.

Figure 4: Four-leaf tree, historical risk in 1992-99 (graph form)

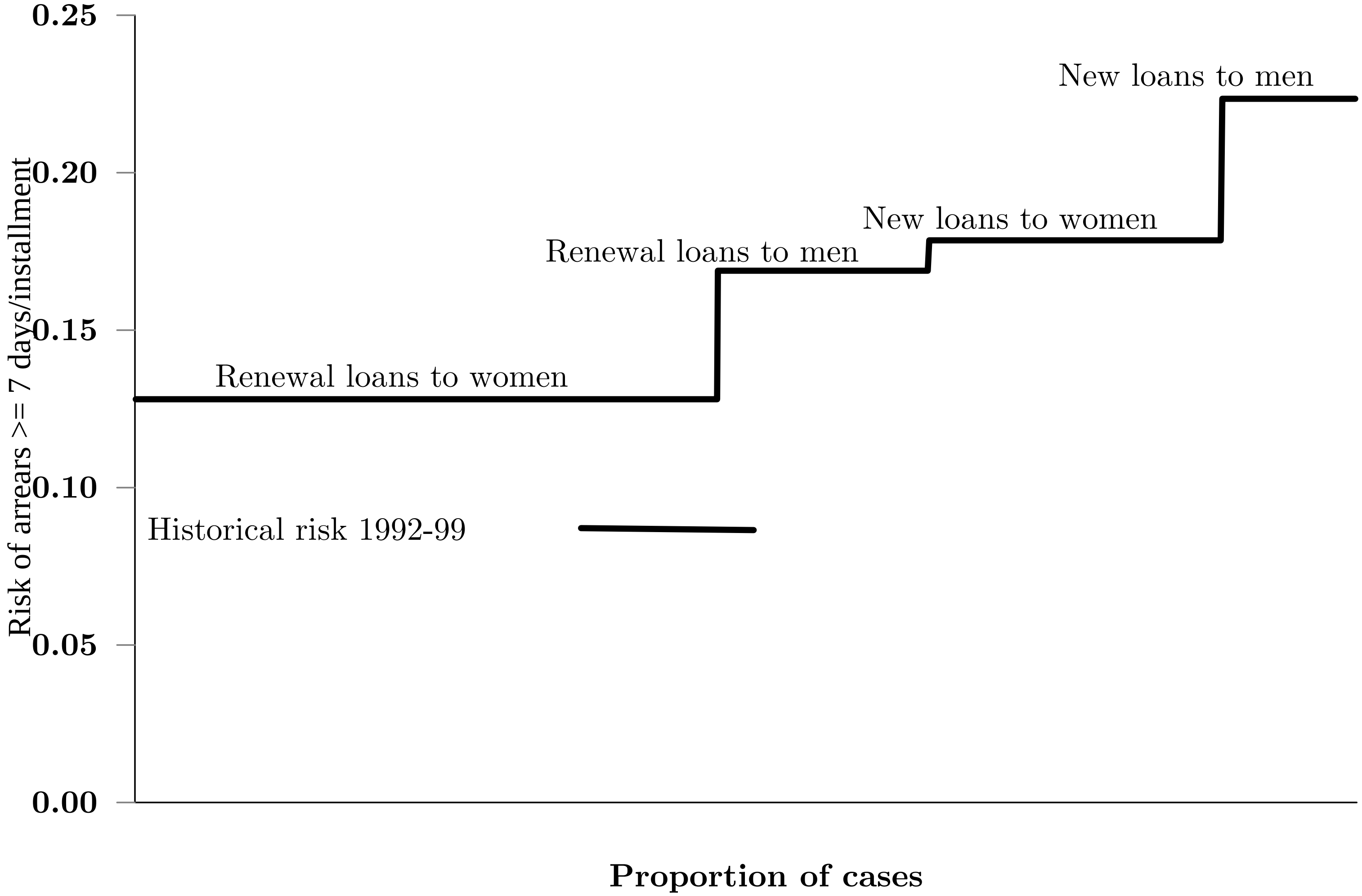
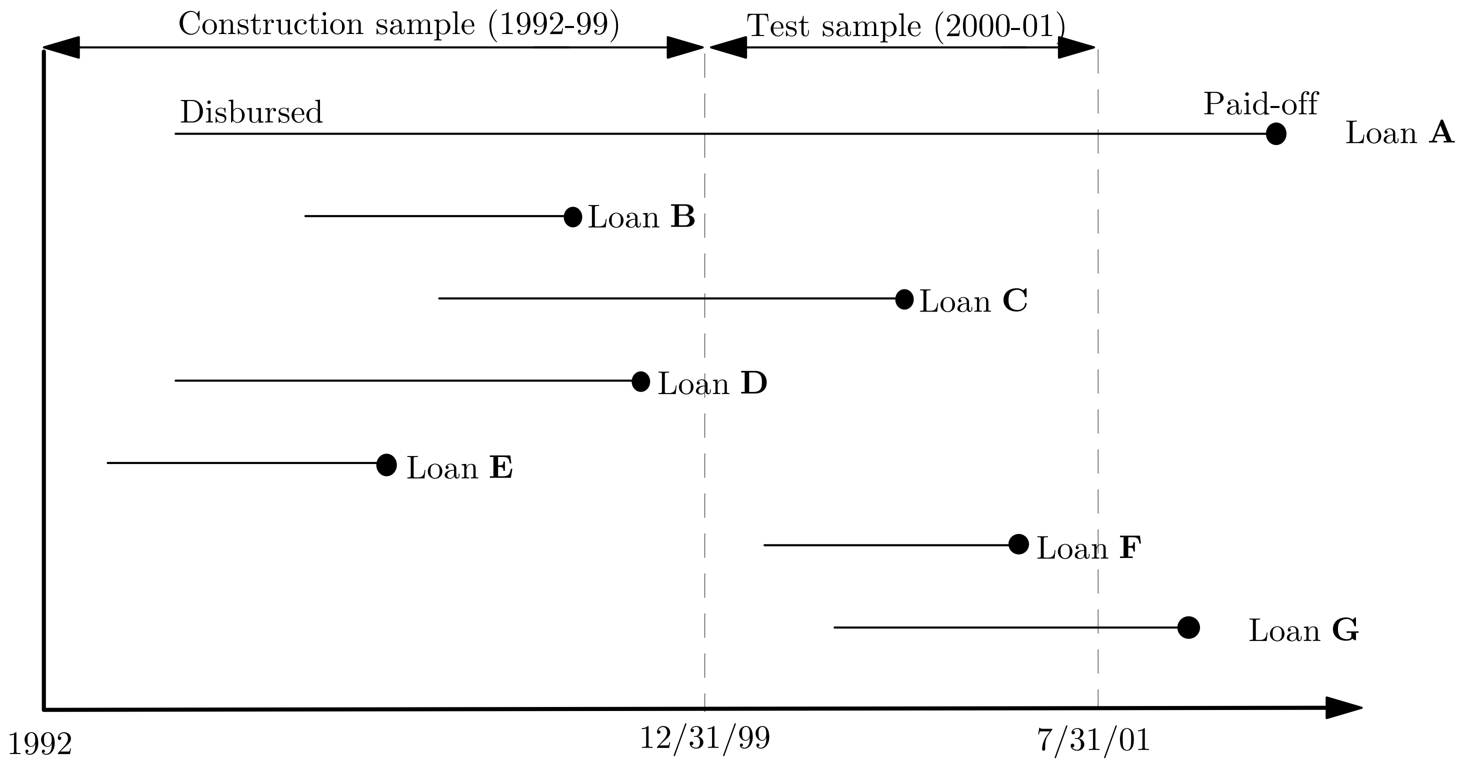


Figure 5: Cases in construction sample and in test sample



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

Figure 6: Four-leaf tree, realized risk in 2000-01

Leaf	Branch of tree: First	Branch of tree: Second	Bads	Goods	Cases	Predicted bads / cases (%)	Realized bads / cases (%)	Cases in leaf / all cases (%)
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	64,327	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.

Figure 7: Test of four-leaf tree, historical risk from 1992-99 (predicted risk for 2000-01) compared with realized risk in 2000-01

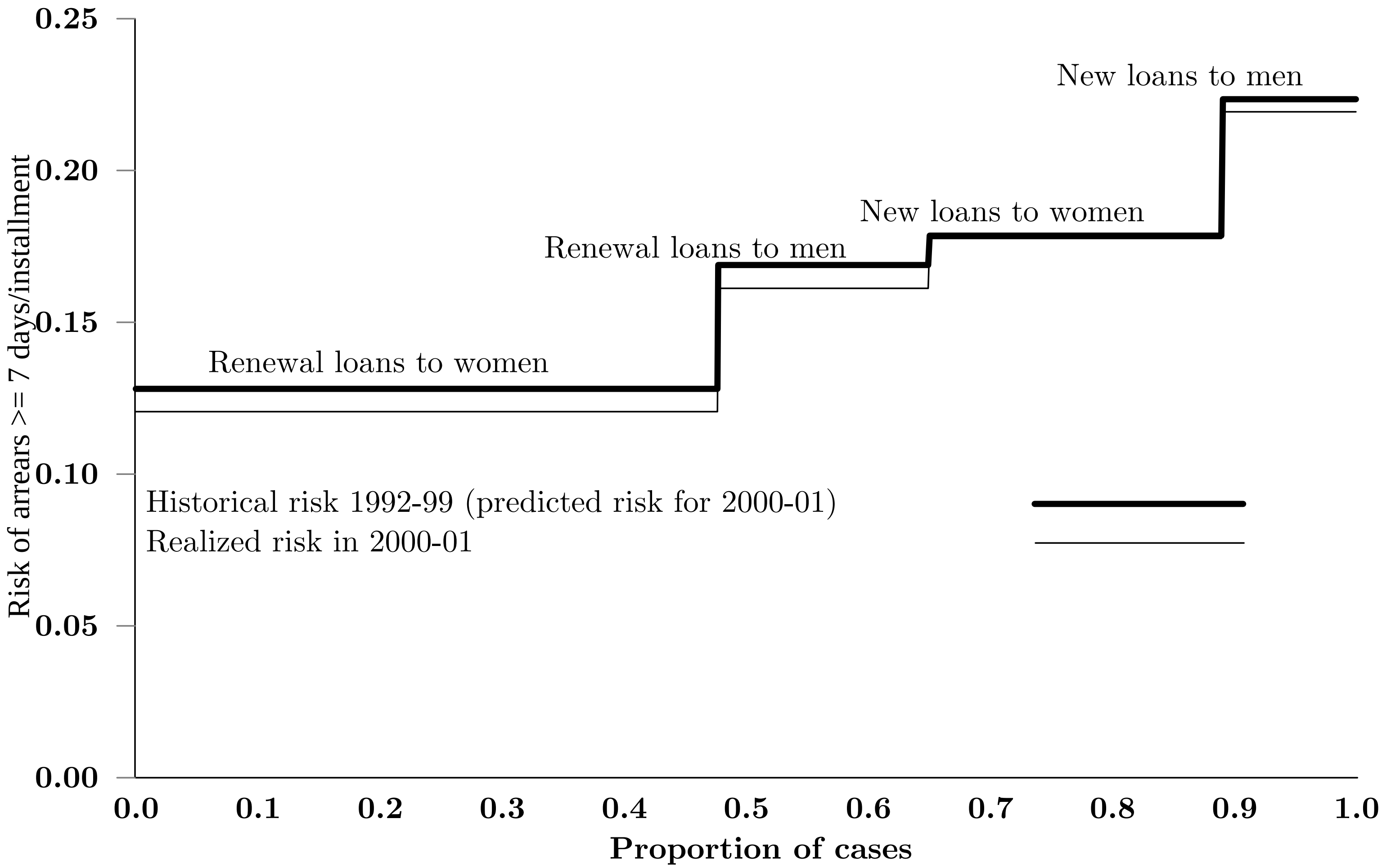


Figure 8: 19-leaf tree, historical risk in 1992-99

	Branch of tree				Construction sample, 1992-1999				
Leaf	First	Second	Third	Fourth	Bads	Goods	Cases	Bads / cases (%)	Cases in leaf / all cases (%)
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.	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: Data base of a Latin American microlender.

Figure 9: 19-leaf tree, realized risk in 2000-01

	Branch of tree							Test sample, 2000-2001		
Leaf	First	Second	Third	Fourth	Bads	Goods	Cases	Predicted bads / cases (%)	Realized bads / cases (%)	Cases in leaf / all cases (%)
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.	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: Data base of a Latin American microlender.

Figure 10: Test of 19-leaf tree, historical risk from 1992-99 (predicted risk for 2000-01) compared with realized risk in 2000-01

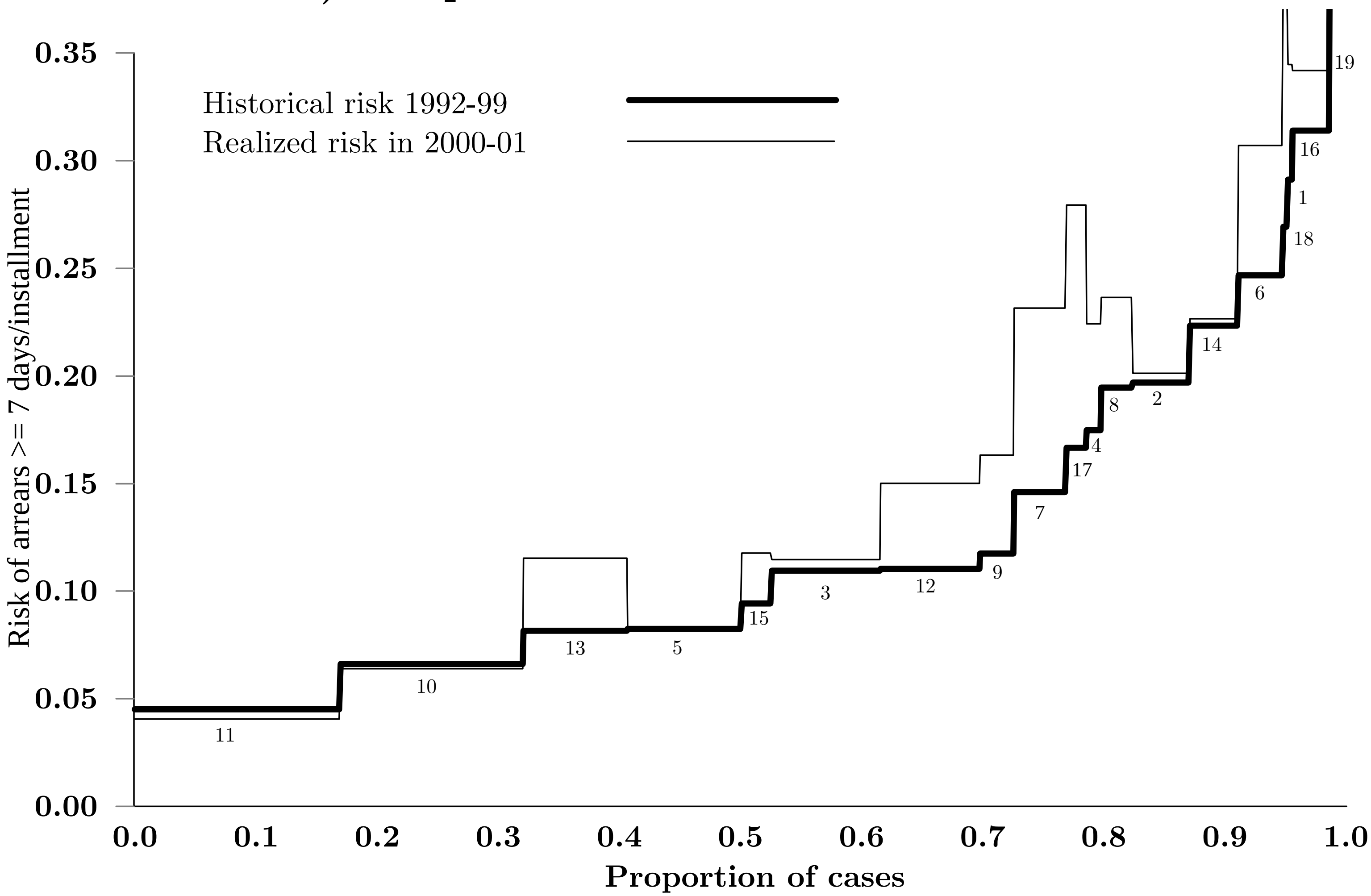


Figure 11: Where four-class scoring policy fits in the traditional evaluation process

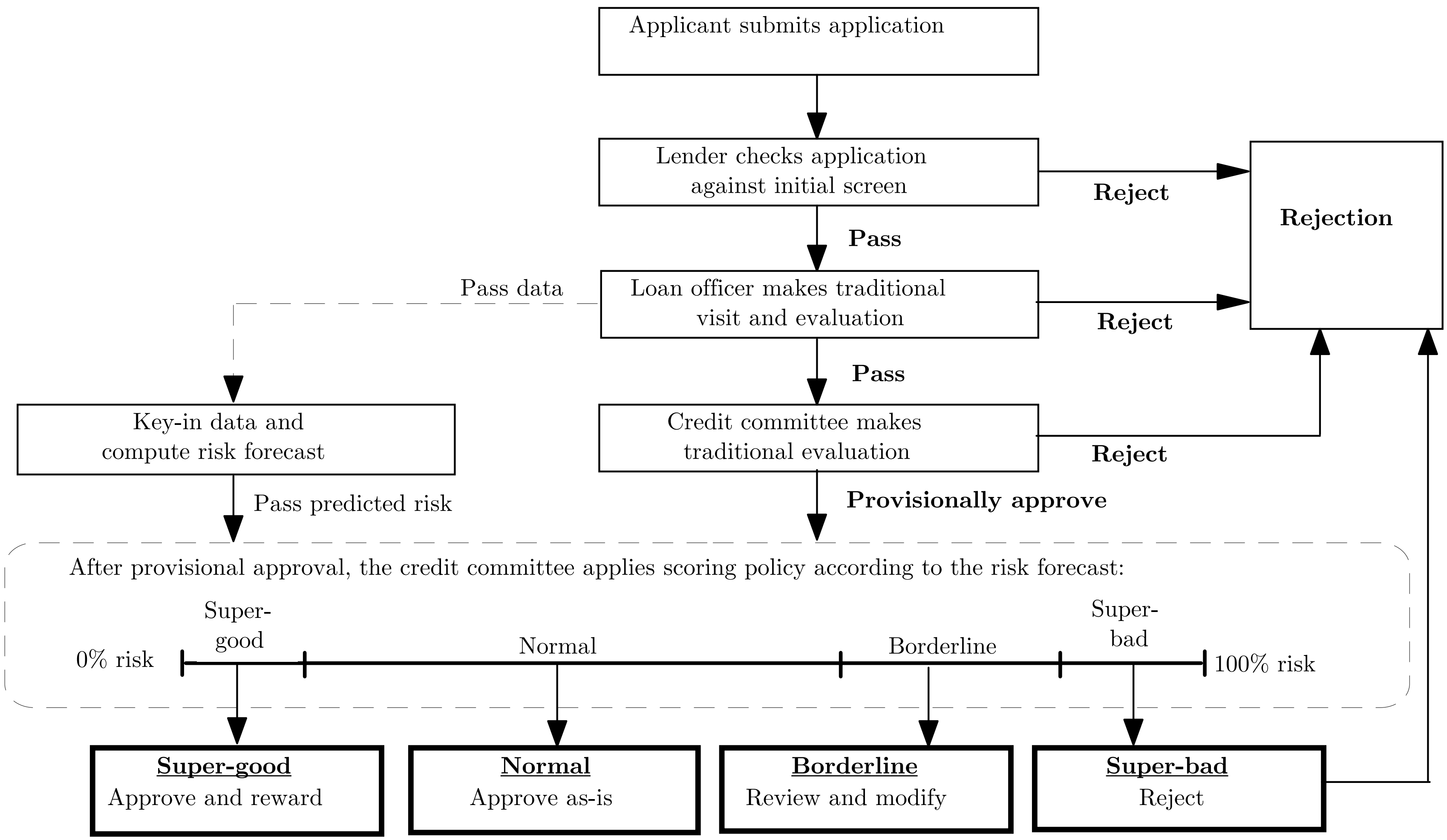


Figure 12: Table of results of four-class scoring policy used in 2000-01 with 19-leaf tree constructed with data from 1992-1999

Leaf	Branch of tree: First	Second	Third	Fourth	Test sample, 2000-2001: Bads	Goods	Cases	Predicted bads / cases (%)	Realized bads / cases (%)	Cases in leaf / all cases (%)	Bads in leaf / all bads (%)	Goods/ bad
All loans					6,907	42,478	49,385	12.1	14.0	100.0	100.0	6.1
	Super-goods											
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
	Normals											
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
	Borderlines											
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
	Super-bads											
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. disb. <= 2.	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

Figure 13: Graph of results of four-class scoring policy used in 2000-01 with 19-leaf tree constructed with data from 1992-1999

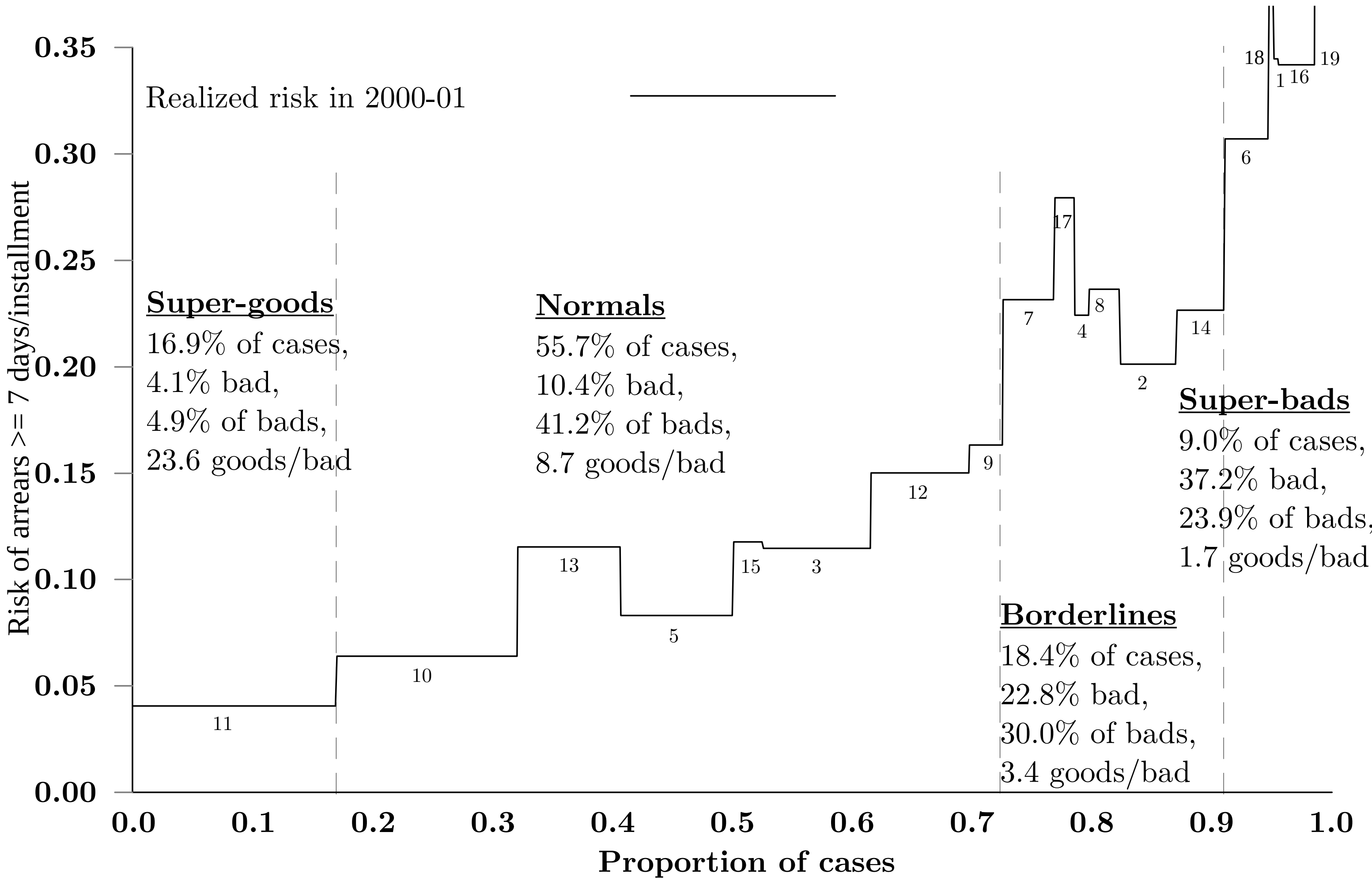


Figure 14: Ratio of goods lost to bads avoided for a range of super-bad thresholds for the 19-leaf tree

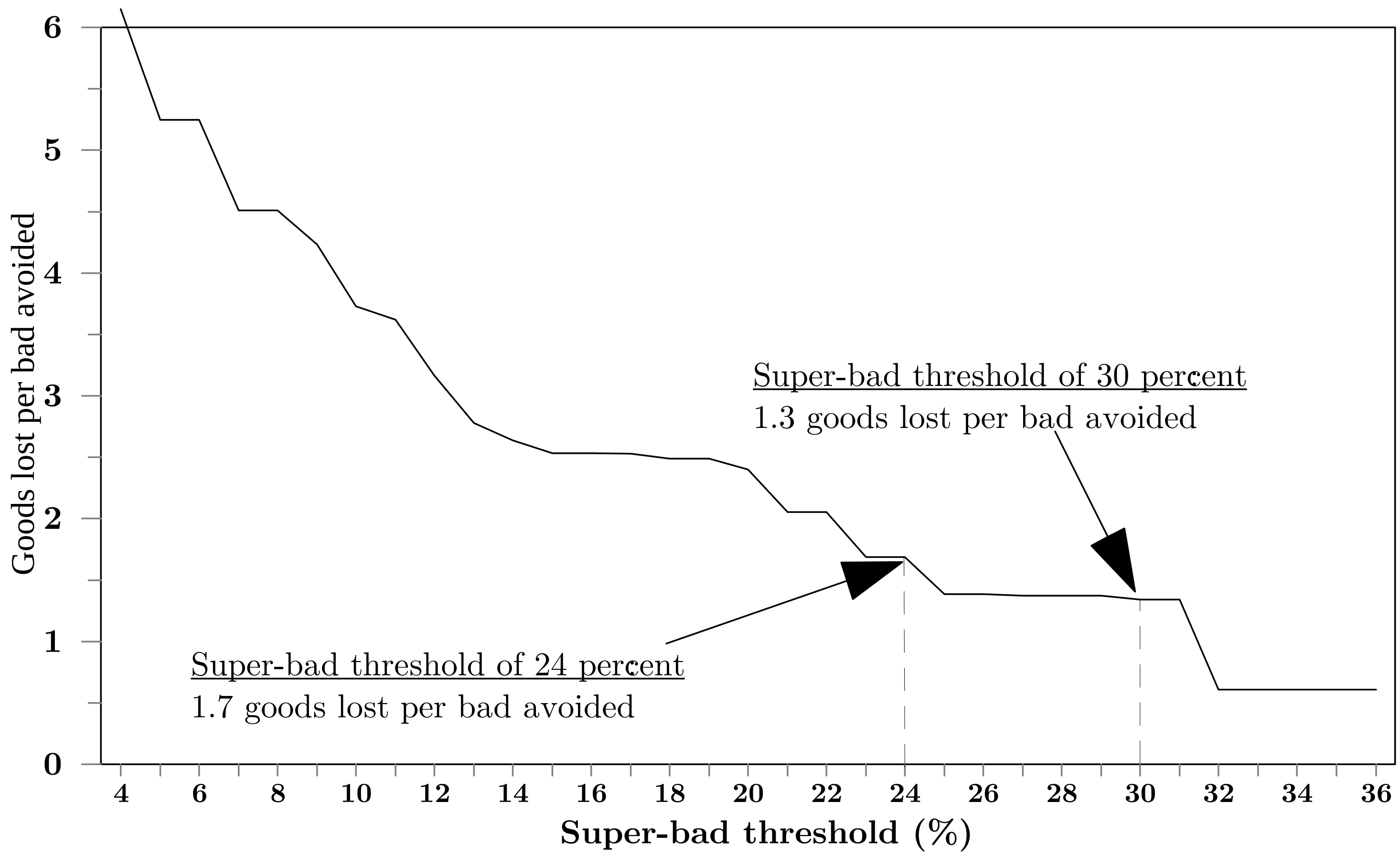


Figure 15: Share of cases rejected versus share of bads avoided for a range of super-bad thresholds for the 19-leaf tree

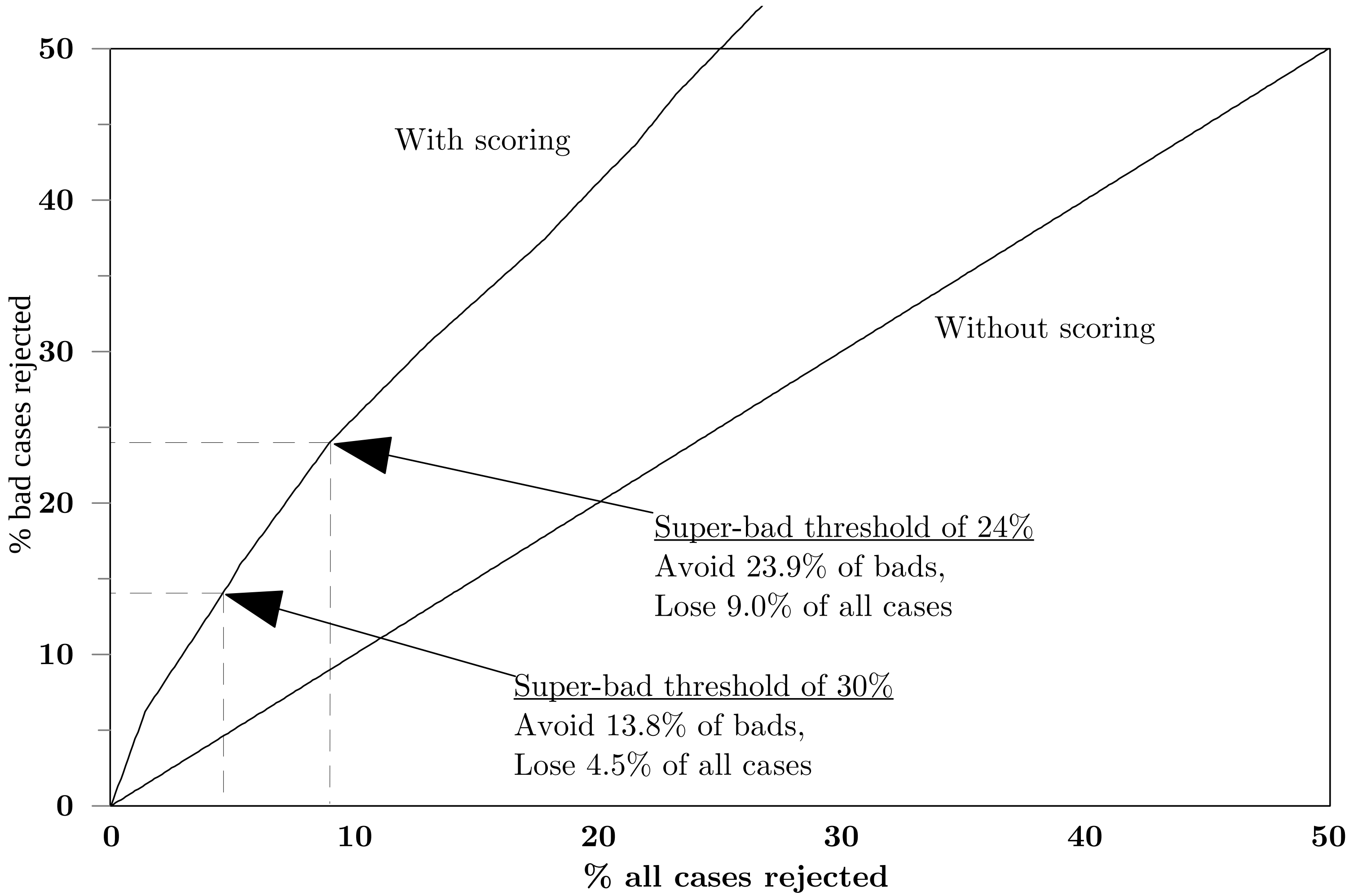


Figure 16: Estimated change in profit due to use of 19-leaf tree scorecard in 2000-01

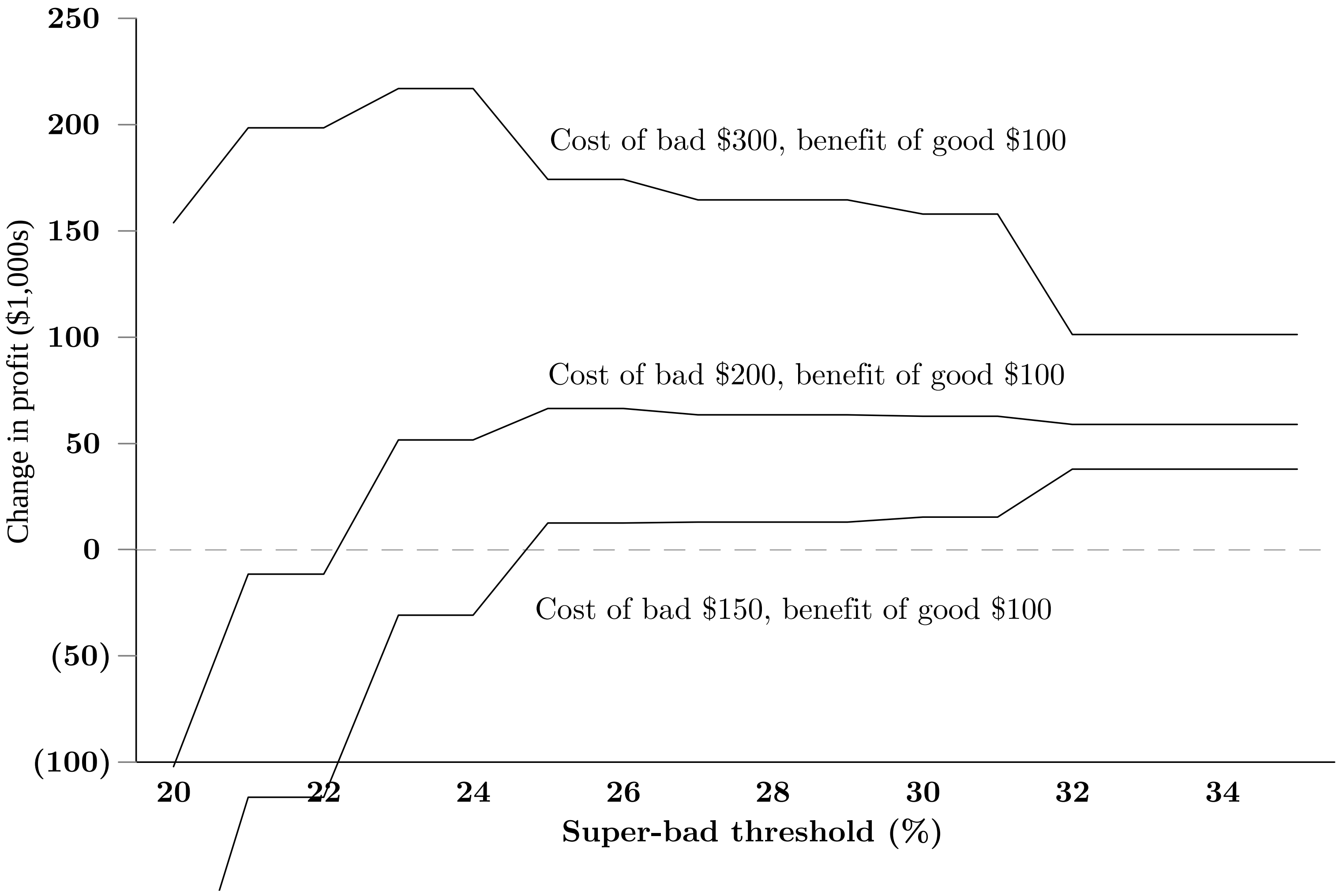


Figure 17: Benefits of scoring from decreased time spent by loan officers in collections

Activity	Before scoring		After scoring	
	% time	**Days**	**% time**	**Days**
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.

Figure 18: Example "Scoring Simulator" of risk forecasts after modifications to loan terms

Client: Jane Doe		Branch: Central		App. no.: 12345
Loan officer: John Smith		Committee: 01/03/01		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: Example of the author.

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, last paid-off loan	8.7	1.7	+5.8
2. # installments late, last paid-off loan	6	4	+4.2
3. Exp. loan officer (# disbursed)	77	535	+3.4
4. Type of business activity	Carpentry	N/A	+1.5
5. Telephone in the residence	No	Yes	+1.1
6. Term to maturity, last paid-off loan (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. Total # of employees	0	0.25	–1.9
39. Exp. client (# months)	36	14	–2.3
40. Client age	55	43	–4.4
Risk forecast:	23.2	9.3	+13.9

Source: Example of author.

Figure 20: Example "Global Follow-up Report"

Risk: 4 days/installment or 30/row Date tested: 6/2/02		Quantity at-risk: Number of loans Date scorecard constructed: 31/07/01			Branch: All	Realized risk (%) for loans paid off in last 12 months
Forecast risk (%)	# Loans out. (%)	Realized risk (%) by days since disbursement				
		0-90	91-180	181-270	271+	
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 change in distribution of predicted risk for a new versus degraded scorecard

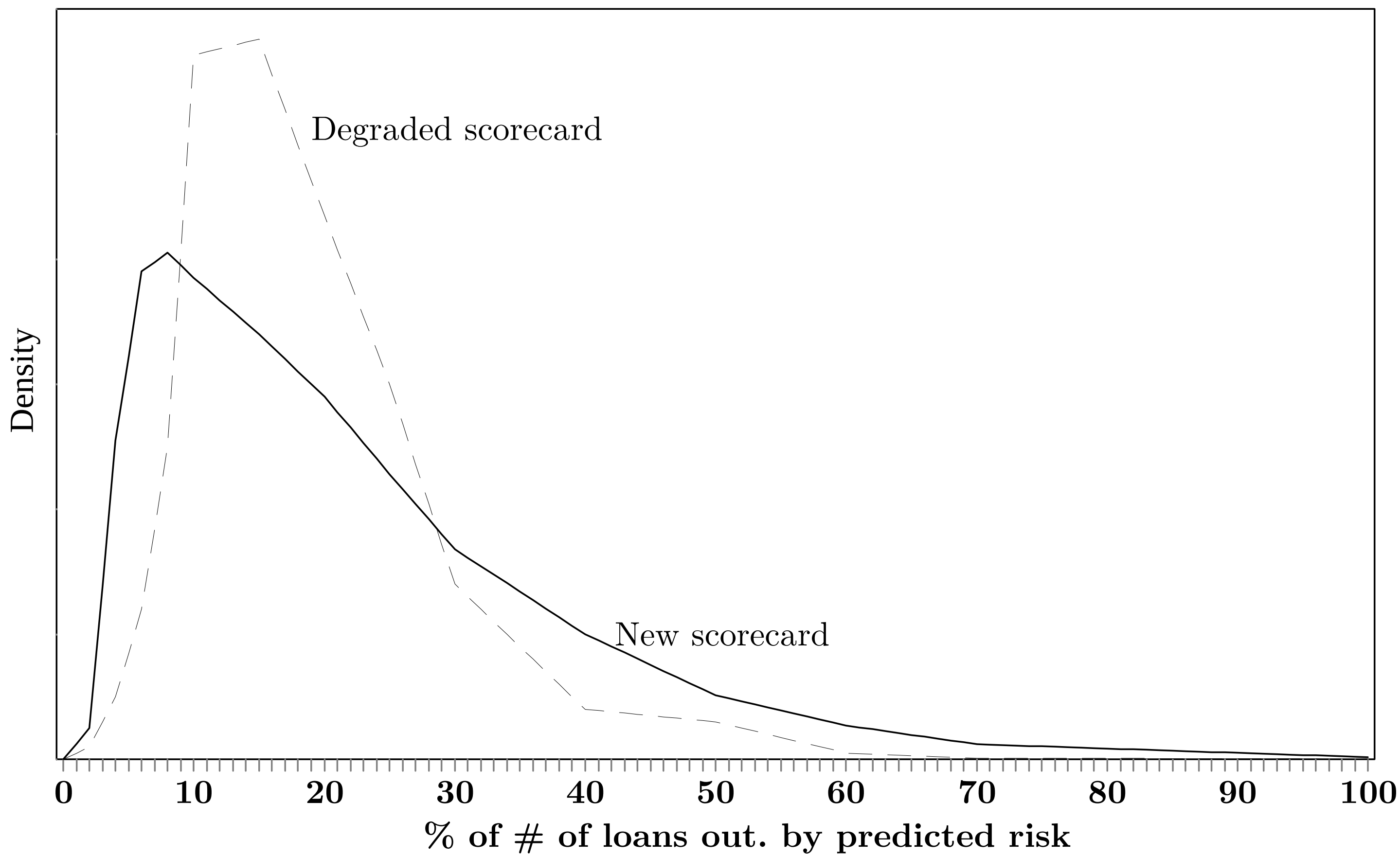


Figure 22: Example change in relationship between predicted risk versus realized risk for a new versus degraded scorecard

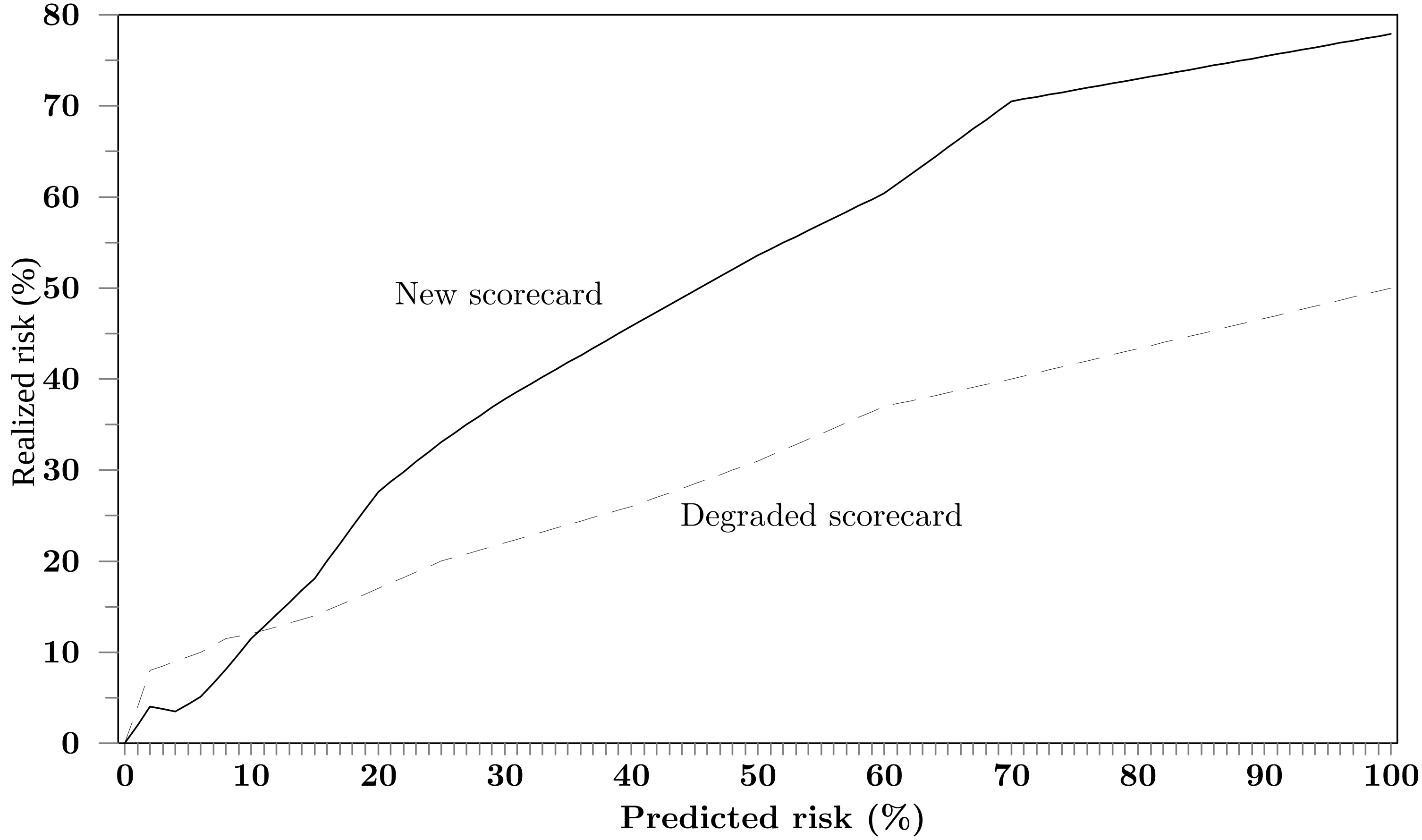


Figure 23: Example "Loan Officer Follow-up Report", 30 highest-risk cases disbursed more than 270 days ago

Report date : 31/07/2001 Branch : All Risk : 1 spell >=30 days Loans : out. >270 days List : 30 most risky

Loan Code	Client name	Days out.	$ out.	Monthly payment	Next due	Current arrears	# spells	Realized risk: 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, Diofanor	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
								Ave. risk:		50	61

Source: Regression scorecard and data base of Latin American microlender.

Figure 24: Example "Loan Officer Follow-up Report", 30 lowest-risk cases disbursed more than 270 days ago

Report date : 01/12/01		Branch : All		Risk : 1 spell >=30 days				Loans : out. >270 days	List : 30 least risky		
								Realized risk			
Loan Code	Client name	Days out.	$ out.	Monthly payment	Next due	Current arrears	# 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	Guevara, 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, María	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
								Ave. risk:		0.0	0.6

Source: Regression scorecard and data base of Latin American microlender.

Figure 25: Relationship in regression scorecard between risk and number of months since the first disbursement

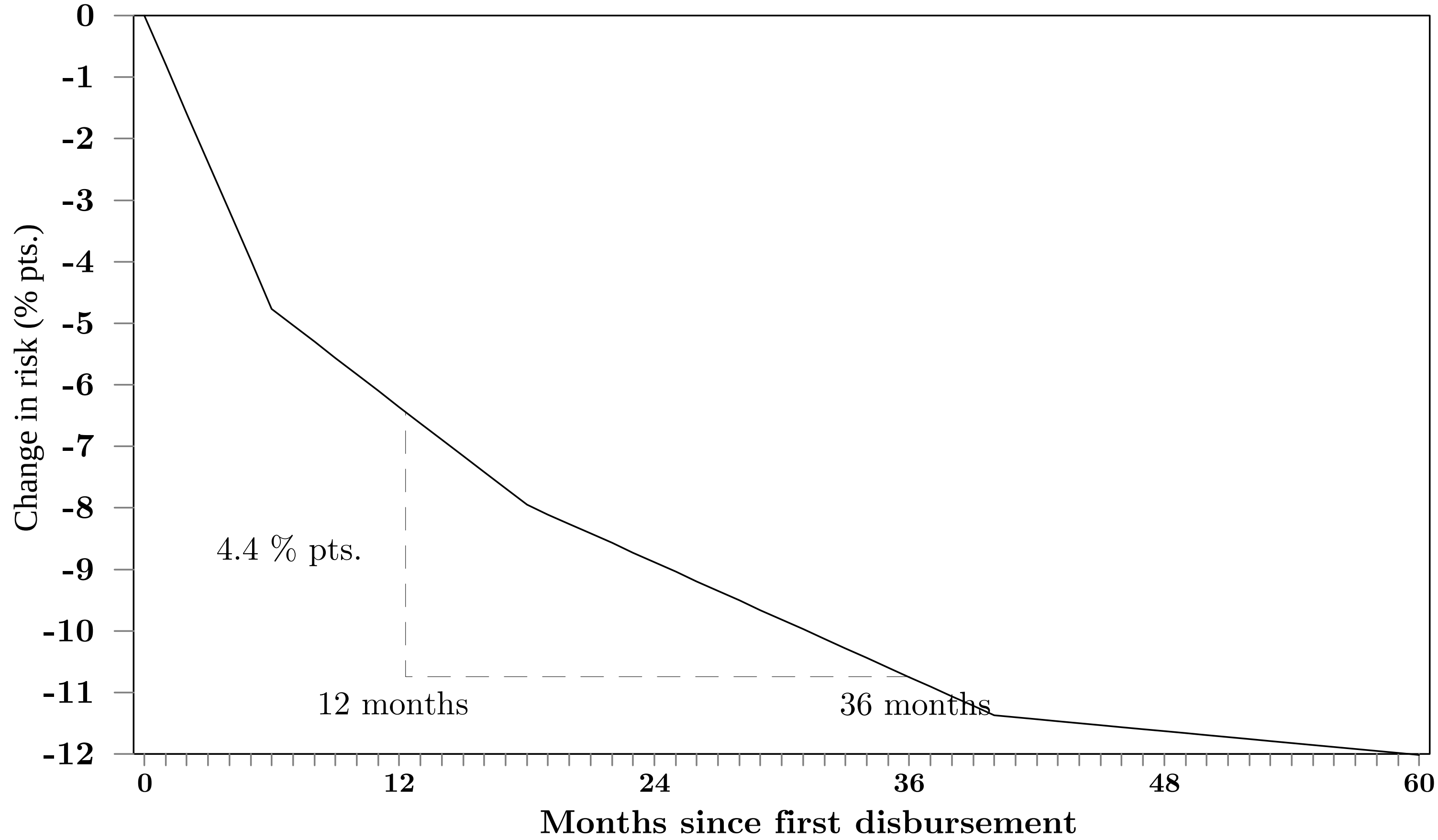


Figure 26: Relationship in regression scorecard between risk and age of client in years

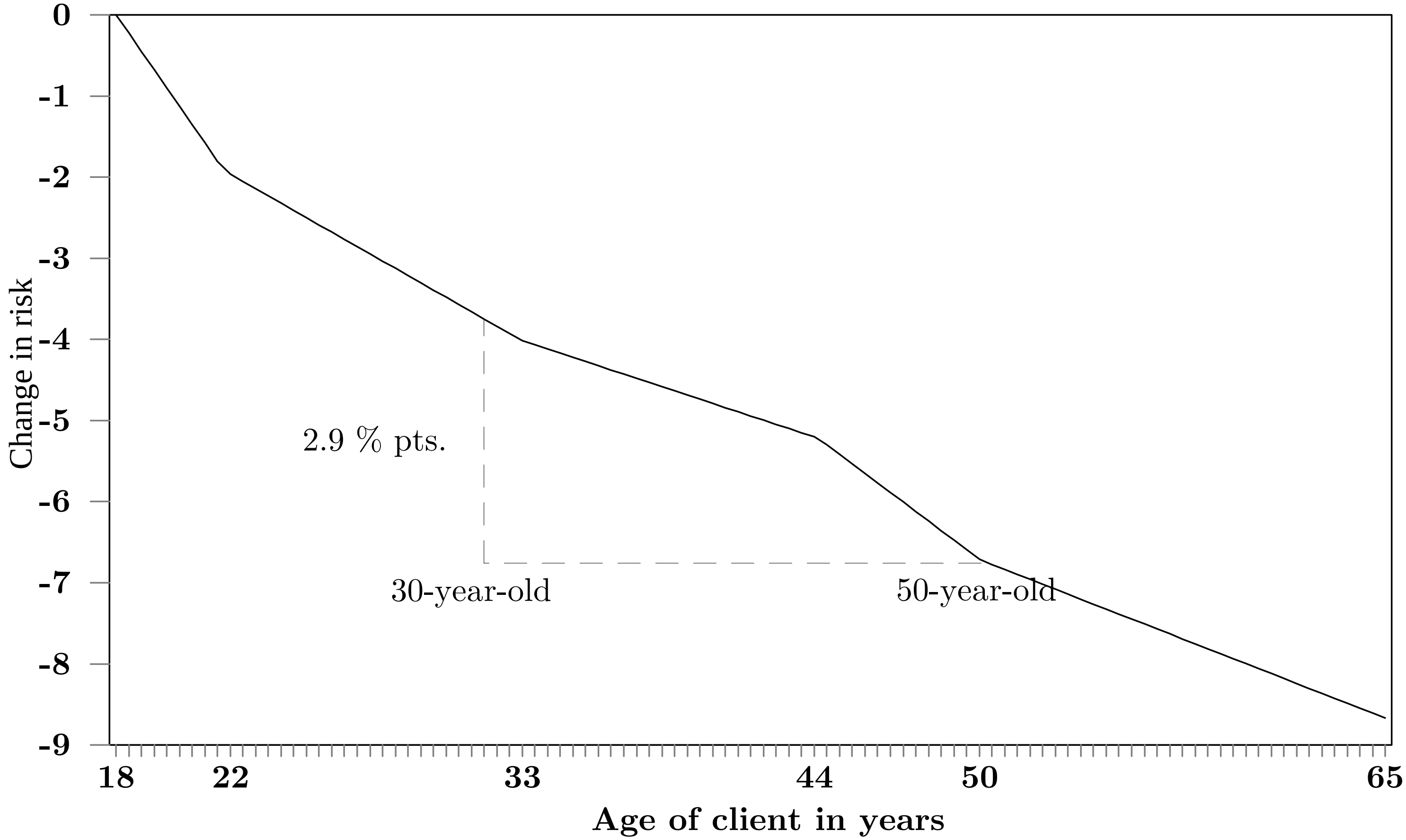


Figure 27: Relationship in regression scorecard between risk and applicant indebtedness

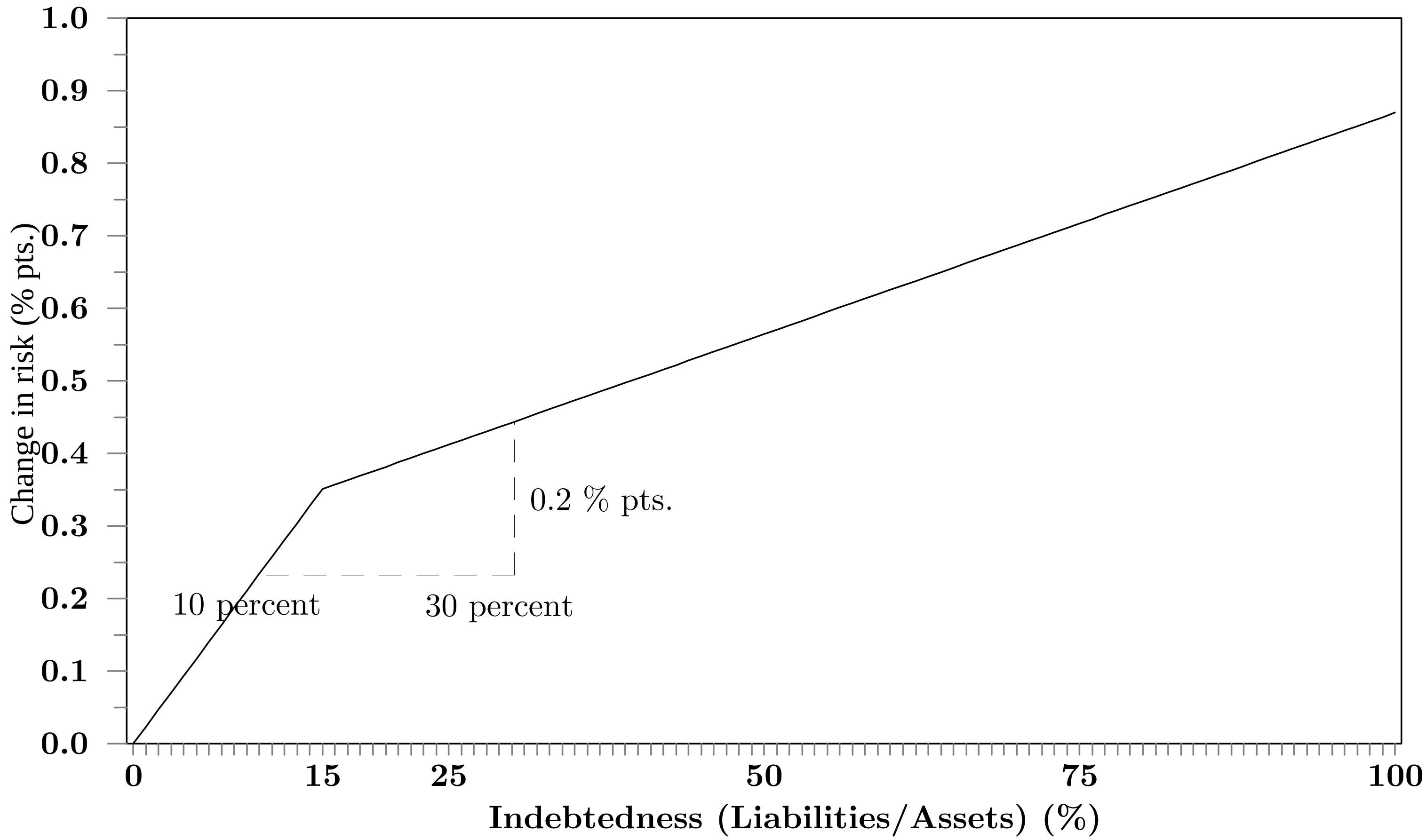


Figure 28: Relationship in regression scorecard between risk and arrears in the previous three loans

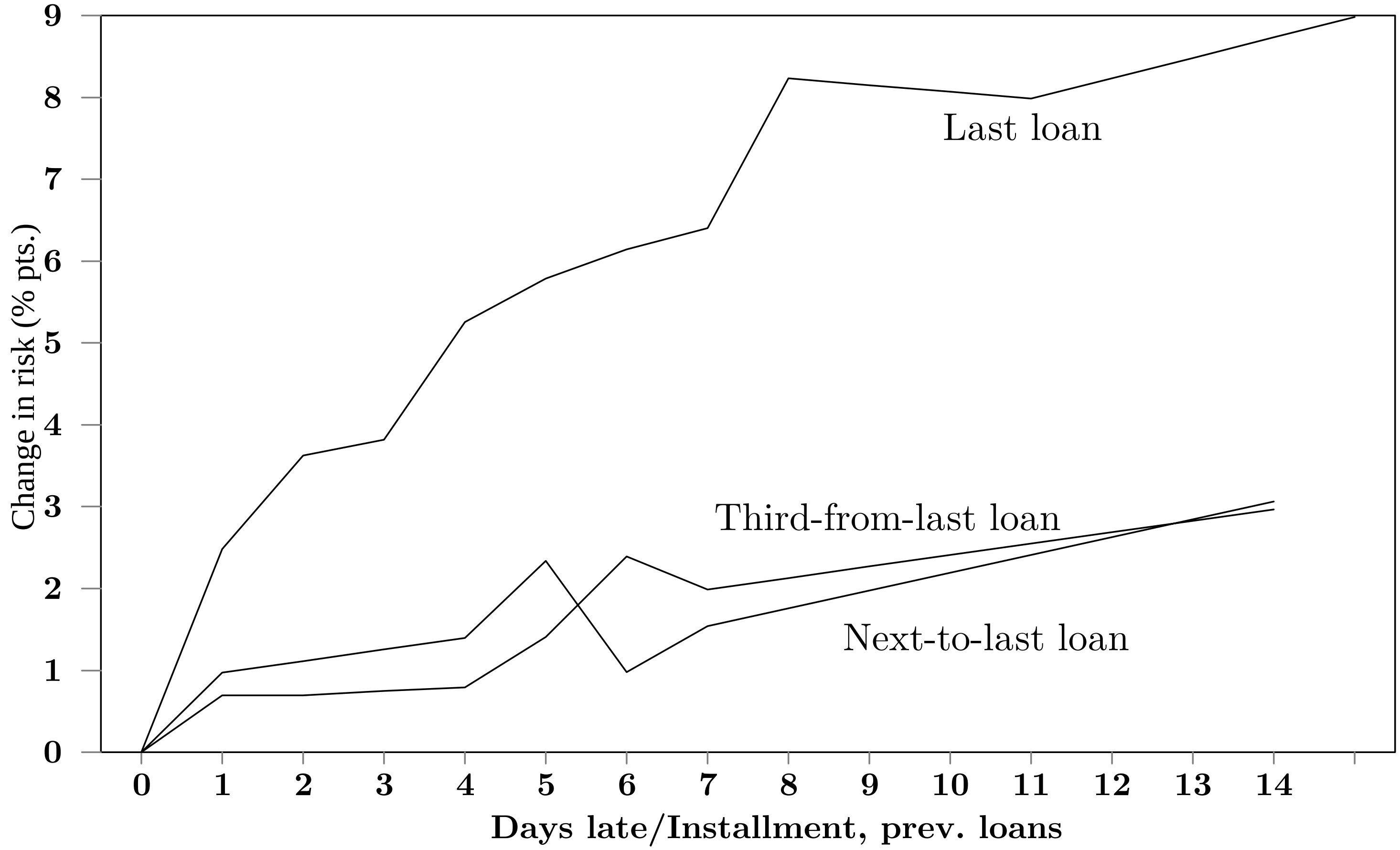


Figure 29: Relationship in regression scorecard between risk and type of business

Type of business	**Effect on risk (%)**	**Portfolio share (%)**
Trucking and taxi driving	-3.6	0.4
Sales of fruits and vegetables	-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
Sales of cosmetics	-1.9	1.6
Grocery store	-1.7	2.3
Seamstress and clothes-making	-1.3	11.1
Sale of prepared food	-1.0	1.0
Schools	-1.0	0.6
Food processing	-1.0	0.6
Auto-parts store	-0.6	0.7
Fried fast food in the street	-0.6	0.5
Meat market	-0.5	1.4
Sale of home appliances	-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

Figure 30: Relationship in regression scorecard between risk and the 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

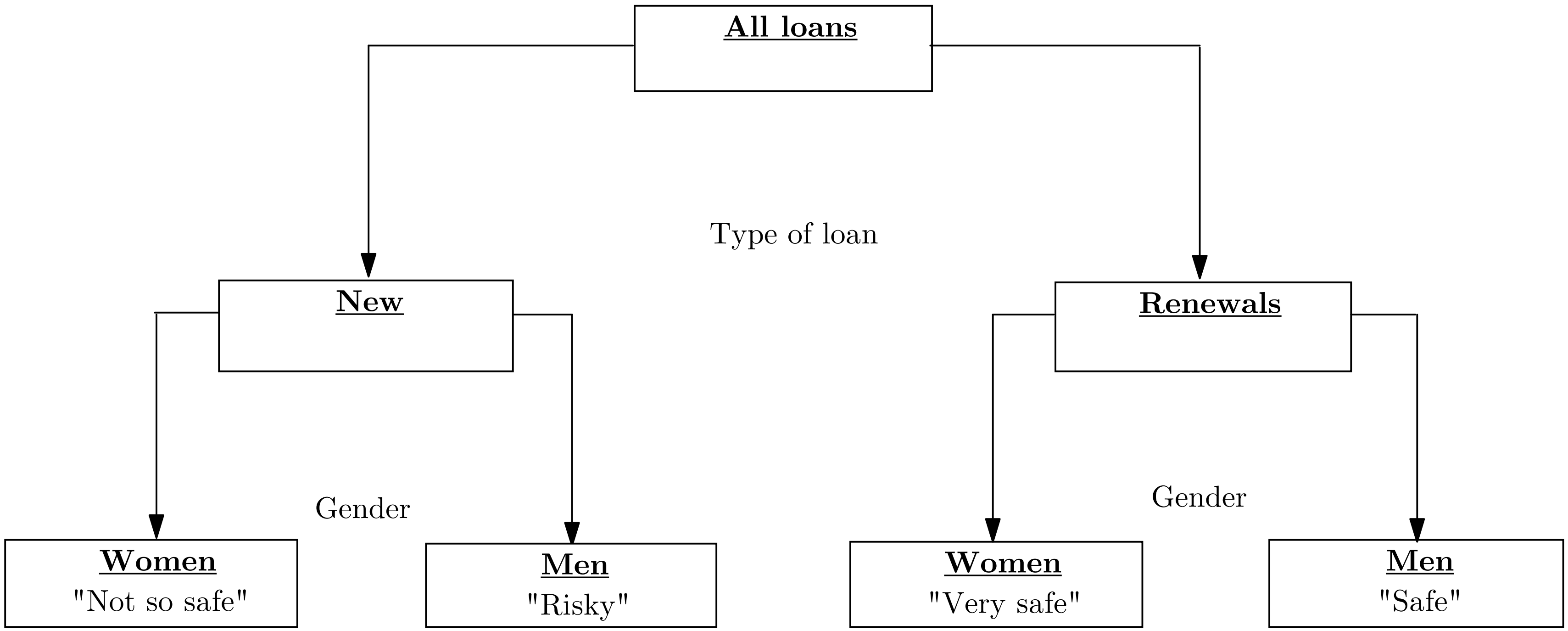


Figure 32: Example policies for five types of risk

Type of risk to be forecast	Example 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", 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 Safe waverers: Offer incentives to repeat Loyalists: Wait and see
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

Figure 33: A three-class collections policy

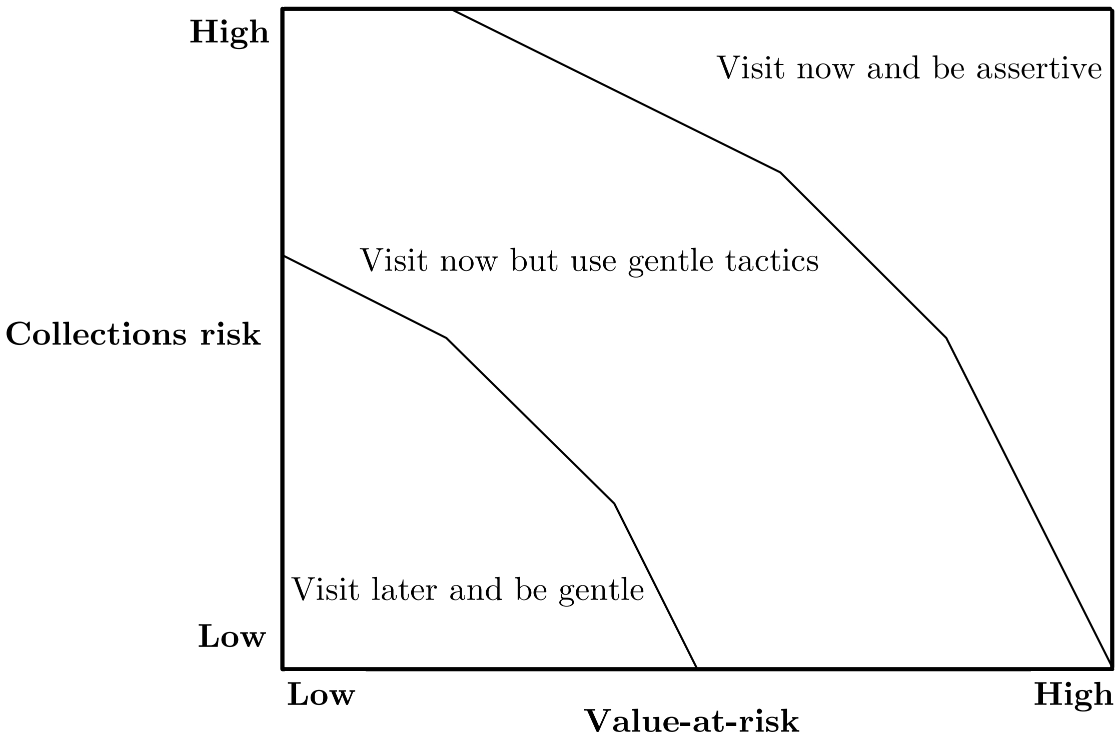


Figure 34: A four-class desertion scoring policy

<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><td rowspan="2">Desertion risk</td><td>High</td><td rowspan="2">Kick-outs: No incentives</td><td>Unsafe waverers:
No incentives</td><td>Safe waverers:
Incentives offered</td></tr>
<tr><td>Low</td><td colspan="2">Loyalists: No incentives</td></tr>
</table>