Consider for a moment a simple example. An applicant applies for a 36-month loan with a fixed payment each month. If the customer makes all their repayments, we should be able to calculate our administrative and funding costs and subtract them from the interest income, and the resultant figure is profit.

As it stands, this is a simple and simplistic scenario. Let us consider some of the sources of complexity. The profit will clearly be affected by how we fund the loan. Indeed, if we fund it on a rolling basis rather than buying three-year money when we grant the loan, then at the beginning of the loan, we do not know our costs exactly. On the other hand, if we fund the loan on a rolling basis but have a variable interest rate applied to the loan, we may be able to assume a fixed margin.

In assessing profit and loss, we also need to consider our marketing acquisition costs. It is possible to debate this issue at length, but there are at least two different points of view that can be held. The first is that we should assume that acquisition costs are fixed overheads and should not be included in any consideration of profit or loss of individual cases. This might be valid in cases where we have a fixed marketing cost and we can apply a portion of this to a new loan retrospectively only when we can tell how many loans were generated by the marketing expenditure. The second is that we should factor an allocation of the acquisition cost of each loan into the calculation.

Now, if the customer repays the loan early, we may lose some of our interest income. On the other hand, we have the money returned to us and can lend it to someone else. We may need to consider timing of early payments.

Thus far, this is still relatively simple. However, especially if this is a new customer to our organization, we may also have to consider the cross-selling opportunities. While the customer may take out a small loan that barely breaks even, if at all, even if they make all their payments on time, we could include in our assessment of profit and loss the opportunities to sell them other products—savings account, insurance, etc. Clearly, once we have an existing customer and a relationship with them, there may be a greater propensity for them to buy other products. In any case, there is greater income to be set against the acquisition cost.

The debate about acquisition costs can carry over into a debate about other overhead-type costs. How much of our branch costs should we include in our calculation of profit and loss? For example, if the customer had not taken out the loan, the branch costs would still be there. On the other hand, if we used that argument across the whole spectrum of products and services, we would never have any money set aside to pay for the branch network. Moving to telephone-based processes or to Internet-based processes merely changes the size of the overhead. It does not change the argument, although in the case of the Internet, one could argue that there is minimal manual cost. In other words, once we have set up and developed a site and advertised its existence and address, handling 1000 inquiries may cost the same as handling 1 million inquiries. This leads to the next issue, as we also need to consider how we account for and recover our system development costs.

Another finance issue is that we need to convert everything into the value of money today, typically using an NPV calculation, as outlined in Chapter 3. In calculating an NPV, we need to know not only the probability of a loan not being repaid but also the timing of the likely default. We also need to know the timing of an early settlement and the timing and size of any other income streams arising from cross-selling.

A further finance issue is the fact that we may be able to measure some costs and factors very accurately, while some will be measured very roughly. In some cases, there will be no measurement, just an allocation of costs. These and other related issues are discussed by Hopper and Lewis (1991).

Thus we have seen that a consideration of the elements in our profit and loss calculation is not simple. However, we also need to decide on a profit threshold, and this introduces the issue of what profit measure to use. If we use a performance definition of good and bad—e.g., bad = ever three payments in arrears—then we might set the cutoff at the point where our marginal cases break even (in some sense). We could do the same when we consider profit once we decide what constitutes breaking even. Once we have definitions of profit and breaking even, we can then adopt a number of acceptance strategies, including the following:

- Accept all loans where the expected income is greater than the expected cost. (If this ignores overheads, the accountants might say that these cases are making a positive contribution.)
- Accept all loans where the expected income is greater than the expected cost by a fixed amount. An example of this would be to accept loans only if we expect to make at least £100 on them. This margin could be introduced to cover the uncertainty in the calculations.
- Accept all loans where the expected income is greater than the expected cost by a fixed percentage. An example of this would be to accept loans only where we expect income to exceed costs by at least 10% (of the costs). This is getting us a little bit closer to calculating a return.
- Accept all loans where the expected profit is at least a fixed percentage of the amount of the loan. This is a form of return. In other words, we are risking, say, £10,000 and will do so only if we expect to make a profit of at least, say, 20%. We could also have a loan proposal of £30,000 with an expected profit of £5,000 but would reject it as three loans of £10,000, each with an expected profit of £2000, would produce more profit for the same amount exposed.

Now, once again, we need to turn to our finance colleagues to get a calculation of return. Often it is the return on equity or return on capital employed or risk-adjusted return on capital (known in some circles as RAROC). These take into account the fact that when we lend money in aggregate, we do not have all that money in the organization. We have to borrow some of it from investors—shareholders—who require a considerable return for their money to compensate them for their risk.

Once we have a calculation of our return, we still need to resolve how we introduce our overheads, what minimum return we should set, and whether we are looking at cases that reach this required return marginally, i.e., in their own right, or in aggregate. Specifically, we could take on cases using a cutoff that gives the maximum return. On the other hand, we should probably set our cutoff to select those cases that maximize our expected profit, subject to the minimum return being achieved, either in aggregate or in each case.

In considering the overheads, we need to allocate them but may run into some difficulties until we have set the cutoffs. For example, with a cutoff of, say, 240, we may have 7,000 loans accepted per month, so we can spread our fixed costs across these 7,000 loans or the 84,000 loans per annum. In this case, our fixed costs might include the cost of our building and the salaries of senior management and our systems staff, as these are the same whether we have 7,000 loans per month or 8,000 per month. However, if we have a cutoff of, say, 225, we may end up taking on 8,000 loans per month and can spread the same fixed costs over more loans, although the quality of the loans taken on will decrease, in terms of the average profit per loan. One way to get out of this potential trap is to take a pragmatic business or accounting decision. Another way is to try to iterate to an optimal position, i.e., set a cutoff, calculate the marginal profit, identify declined cases (i.e., below cutoff) that would be profitable, and change the cutoff to include them, and so on.

Once we go down the route of scoring to assess profit, we then are faced with the issue of what to do with applicants whose performance is predicted to be good in terms of having low arrears but whose transactions may not be sufficient to generate a profit. We could consider several actions, including

- declining them, because they are, in some sense, too good;
- try to convert them to another product, better suited to their needs, that will generate a profit;
- accept them and accept the loss as either a cost of being in business or as the cost of a marketing error;
- rather than risk price them, profit price them; in other words, offer them the product at a higher price that will generate a profit.

None of these is completely satisfactory. Often the approach taken is a mixture of the second and the third options, although some companies, especially those trying to break into a market, may be better placed to adopt the first option, although even for them the third option looks likely. Whatever stance an organization takes, improved prescreening and targeting may help to reduce the scale of the problem.

Of course, some lending organizations do not adhere as strictly to the profit ethos as others. Cooperative societies, and the Coop Bank in the U.K., may be prepared to take a longer-term view of profitability. Similarly, credit unions may be more willing to lend to applicants who represent marginal profit, or even likely loss, as they may wish to achieve a high acceptance rate and a high penetration among their members. With these organizations, there may be some social or political capital to be made in lieu of financial profit.

In summary, from the viewpoint of the scoring methodology, using a profit-loss definition of good and bad rather than a repayment performance definition should present no technical difficulties. The main difficulties arise in the calculation of profit and loss, in handling applicants who are creditworthy but unlikely to generate sufficient profit, and in the (probably greater) number of assumptions that are required for the profitability assessment. In Chapter 14, we begin to look at modeling approaches that might allow us to improve our profit scoring.

11.4 Tax inspection

For many years in the U.S. and over the past few years in the U.K., there has been in place what we might refer to as an honor system of tax assessments for consumers. In essence, each consumer or each taxpayer is responsible for completing a tax return and calculating the total amount of tax they owe for the tax year in question. (In the U.K., if one submits by September 30, Inland Revenue will calculate the tax due.) If one is due to pay more tax than has been withheld through salary payments, etc., then one either encloses a check with the return or makes the payment by a specified date. If more tax has been withheld than one owed, one is due a refund, and one may be able to opt to receive this either as a lump sum from the tax authorities or as a reduction in the amount withheld in the current and subsequent tax years.

Now with such a system, the key to being able to manage the volumes is that not all tax returns are fully audited by the tax authorities. In effect, they are likely to check more fully the returns of those people who have had gross inaccuracies in past returns, especially if these inaccuracies were in the favor of the individual. For the others, they check less fully and may indeed sample among them to decide which ones to audit. Some authorities use a form of scoring to decide which ones to audit.

They can use past experience and past cases to identify those characteristics of the returns that are associated with cases where the tax reported has been a gross underestimate. In scoring terminology, they have a large set of previous cases on which to build a model. The allocation of good and bad can be fairly arbitrary and can generally be changed around, but let us consider a bad case to be one where an audit reveals that the tax due is greater than that on the return. A good case is one where the tax due is either correct or is overstated. (This book is not the appropriate place, but one could enter into a debate concerning the responsibility of the tax authorities, as a public sector body, to identify cases where the individual pays too much tax.)

This could be considered to be akin to application credit scoring, to behavioral scoring, and to fraud scoring. In some cases, the difference between the tax assessment on the return and the tax properly due is caused by errors, misunderstandings of the taxation rules, and allowances, i.e., honest mistakes. In some other cases, there is planned deception to evade large portions of tax.

As we discussed above, the definition of good and bad could depend on whether there is an underpayment of tax. On the other hand, we could adopt a different definition that defines a bad to be a case where the expected underpayment is at least the expected cost of further investigation. If we compare this to scoring in the lending environment, we find that the scoring methodology would be similar if not identical and the monitoring and tracking reports could also be utilized in the same way.

In fact, almost any decision on whether to carry out an audit and the depth of that audit could be supported by scoring methodology.

11.5 Payment of fines and maintenance payments

In the U.K., one of the tasks of the Child Services Agency is to supervise the maintenance of payments by nonresident parents (NRPs) to the parent charged with care for the child or children, often after the breakdown of a marriage. There has been some investigation into whether scoring models will help. The investigations have looked at whether it is possible to build robust models predicting NRP noncompliance with court orders, etc., and, based on these models, whether it is possible to build predictive application scoring–type models.

The first type of model would be a behavioral type of model and uses characteristics about the NRP such as gender, age, employment, and whether they were in debt when the court order was granted. It also includes whether the money comes direct from the employer, whether they have a mortgage, their income, and the number of nights each week they spend with the child. Other factors are included, some of which relate to the parent who has custody of the child, including whether they are receiving state benefits.

The application-type scorecard development is intended to be able to predict, at the outset, from the characteristics known to the agency, how much resource will be required to ensure that the NRP is compliant with their maintenance schedule. If we contrast this with a lending situation, here a bad case is one that will require considerable amounts of time to make the NRP compliant. It is akin to assuming that all moneys lent will be recovered in full but that some will take more collections effort than others to achieve full recovery.

Thus far in this section, we have considered maintenance payments. However, scoring can be used in a very similar way when we consider the payment of fines. We need to get payment of a fine from an individual, and there are several methods of doing so. At one extreme, we hold the individual in a court building until the fine is paid in cash. At the other extreme, we agree to their making payments of the fine over several years. It is possible to build a scorecard that can be used to determine the appropriate level of risk, bearing in mind the objectives of maximizing receipts of fines, minimizing the effort in chasing fines, and dealing fairly with members of the public. (With regard to the last point, it might not be considered fair to hold an elderly person in the court or an adjacent building for several hours until a friend or relative arrives to pay a small fine.)

11.6 Parole

While use of a scoring approach for parole has not become widespread, scoring has been used in reaching the decision on whether to allow a prisoner to be released on parole. Here we can consider two variants of parole. First, the prisoner is let out of prison for a short period, perhaps 48 or 72 hours, as part of the Christmas period, or as a step in rehabilitation, or on compassionate grounds such as a family funeral. Second, the choice is whether to release the prisoner from their sentence early, perhaps for good behavior, either freeing them completely or in terms of a conditional release. Common conditions are continuing good behavior, no reoffense, or regular reporting to a police station.

Once again, the methodology is more or less the same. There is a large set of historical cases, each with the data that were available at the time the decision was made. With the first type of parole, a good case is a prisoner who is let out on parole and returns as agreed. A bad case is one who does not return as agreed. With the second type of parole, a good case is one where the prisoner is released and does not reoffend, either ever or for a specific period. Bad cases are where there is some reoffense. In both cases, rejects are those who are not granted parole.

Similar to many lending environments, one of the main objectives is to minimize the percentage of bad cases, as a released convict may be bad for public safety as well as for public relations.

It is possible to build a scorecard using as predictive data such pieces of information as the prisoner's sentence and crime, how long they have to go until the end, and how long they have already served. Other pieces of information include the proximity of their home from the prison, the frequency of visits from their family, their previous parole history, their age, their recent behavior, and an assessment of their current mental health.

As was stated earlier, standard approaches to monitoring and tracking apply to determine how well the scorecard is working. Perhaps, the only major difference here might be that if there is any doubt about a prisoner, the prisoner be refused and, also, if there is any doubt about the parole process, parole can be withdrawn. However, the prison system continues to operate. This contrasts with the lending environment, where a lender cannot simply decide not to lend any more money because it is too risky. To do so for more than a short period would mean that the lender is withdrawing from this line of business.

11.7 Miscellany

If we can use scoring to assess the probability of recovery of a debt, we can also use it to assess the probability of recovery from a medical operation. This assessment might be used to decide whether to treat a patient or how to treat a patient.

Government security agencies may use a scoring methodology to assess the risk represented by a potential recruit. This could be extended to a general employment position.

To use scoring to assess the risk associated with recruiting an individual as an employee, we clearly need to define the risks with which we are dealing. A similar approach could also be used in selection for university acceptance.

Chapter 12

New Ways to Build Scorecards

12.1 Introduction

Part of the excitement of being involved in credit scoring is the constant search for new ways to improve scorecard development and for new areas where scorecard building can be applied. The former is motivated by the realization that with the volume of consumer borrowing, even a 0.25% drop in the bad debt rate can save millions. The latter is motivated by the success of credit and behavioral scoring and the recognition that using data on customers and their transactional behavior to identify target groups is generic to many areas of retail business. It is the raison d'être for much of the current investment in data warehouses and data-mining tools.

In Chapters 4, 5, and 6, we discussed not only the industry standard methods for developing scorecards, such as logistic regression, classification trees, and linear programming, but also those that have been piloted in the last decade, such as neural networks and nearest-neighbor methods. In this chapter, we discuss approaches that are still in the research stage. They involve ways to combine existing credit-scoring methodologies and to use approaches and models that have proved successful in other areas of statistical and operational research modeling.

In section 12.2, some of the work on generic scorecards and building scorecards on small samples is discussed. If the sample available to a lender is not sufficient to build a robust scorecard, what can one do? Can one put together samples of different populations to construct a scorecard that is robust on any one of the populations? If not, are there other ways to deal with small samples? Sections 12.3 and 12.4 look at the problems of combining classifiers. Section 12.3 investigates simulation work on what happens if there are two different scorecards that a customer has to pass to get a loan. This describes the situation in mortgage scoring, when consumers have to pass a loan-to-value condition and be credit scored. Section 12.4 describes methods that have been suggested to combine all types of credit scoring classifiers, not just scorecards.

An alternative approach to combining credit-scoring classifiers is to use the statistical procedures to estimate intermediate objectives like outstanding balance and amount in excess of overdraft limit and then to use these intermediates to estimate likelihood of defaulting. Section 12.5 discusses this approach. This idea of estimating a number of interdependent quantities rather than just default risk leads one to consider the use of Bayesian networks and statistical graphical techniques in section 12.6. These calculate the strength of the relationships between different variables. Some of these can be the independent variables, like the application-form questions, which are in the usual credit scorecards. Others are variables that one is seeking to estimate, both the final outcomes, like default probability or profitability, and the intermediate outcomes, like credit balance. Finally, there are exogenous variables that describe the current economic conditions, like interest rate and unemployment rate. Bayesian networks give a picture of how these variables are connected, which in turn suggests the variables (and the way they should be interrelated) that are needed to produce a successful credit prediction system.

These approaches not only increase the number of outcome variables but could also allow them to be continuous rather than the standard binary outcome of whether the customer defaults in the next x months. One obviously useful extension is to estimate the distribution of when the customer will default—so instead of estimating if the customer will default, one estimates when they will default. Section 12.7 discusses how survival analysis techniques used in reliability modeling can be translated to deal with this extension.

Not all these ideas will prove to be useful in practical credit scoring, but they give an indication of the way the subject is moving. Analysts are always seeking ways to extend the scope and improve the accuracy of credit scoring.

12.2 Generic scorecards and small sample modeling

It is always assumed that for scoring to be successful, one needs a homogeneous application population. If the population is not homogeneous, then in practice one often segments the population into more homogeneous subpopulations and builds a separate scorecard on each subpopulation. However, at times one might deliberately put together different populations to build a generic scorecard. This might be because the numbers in the individual populations are too small to build a sensible scorecard or when for legal or economic reasons the same scorecard is needed for all the populations. In this section, we describe one such experiment.

U.S. credit unions lend only to their members, who usually have common employment or live in a common locality. They tend to have memberships in the hundreds or low thousands—often not enough on which to build a scorecard. Overstreet, Bradley, and Kemp (1992) looked at what would happen if a generic scorecard was built using the data from 10 such credit unions from the southeastern U.S. They took the scorecards that had been built for the five largest unions and averaged the scores for each attribute in the five scorecards to define a generic scorecard. Their results showed that although the generic scorecard was not as good a discriminator on each population as the scorecard built on that population, in all but one case the results were not too disappointing. At the same acceptance rate, while the individual scorecard identified 70% of the bads, the generic scorecard identified 62% of the bads.

Another way to build a generic scorecard is to pool the populations and use the pooled sample to build the generic scorecard. In a subsequent analysis, Overstreet and Bradley (1996) employed this approach with similar results. At the level where 10% of the goods were misclassified, the generic scorecard correctly identified 53% of the bads, while the individual scorecards correctly classified an average of 58% of the bads.

If the sample population is small and there are no other similar populations to augment it with, other methods may have to be used. The cross-validation, jackknifing, and bootstrapping approaches of section 7.4 allow one to use the whole sample for development rather than leaving some of it out for validation. In some cases, the sample population is still too small for one to have confidence in the results. In that case, one has to resort to a series of ad hoc approaches to combine all the knowledge one has about the situation and to compensate for putting too much emphasis on statistical results derived from very little data.

One example of this was described by Lucas and Powell (1997). They had a sample of only 120 customers with which to build a scorecard for credit cards for the self-employed. These were part of a larger sample who had been given a behavior score and so they took the definition of good to be a behavior score above the median of the larger sample and bad to be a behavior score below the median. There were 84 characteristics available on each applicant in the sample, but using expert opinion, the discrimination tests of section 8.7, and correlations, these were cut to 13 characteristics. For each attribute of each of these characteristics, the sample good rate $\frac{g_i}{g_i+b_i}$, where g_i and b_i, the numbers of goods and bads with attribute i, were adjusted in two ways—one to limit the importance of the statistical results obtained from such a small sample and one to include the subjective opinion of experts on credit lending to the self-employed. First, the estimates were shrunk to allow for the fact that in small data sets, the variation is too high and one wants to shrink the estimate towards the overall mean. The James–Stein estimate (Hoffmann 2000) is one way to do this and is essentially a combination of the original estimate and the overall mean. This estimate of the good rate is

$$\tilde{r} = \lambda\left(\frac{g_i}{g_i + b_i}\right) + (1-\lambda)\left(\frac{g}{n}\right), \tag{12.1}$$

where g_i, b_i, and n_i are the numbers of goods and bads and the total in the sample with attribute i and g and n are the total of goods in the sample and the total in the sample. The choice of λ is discussed in statistical books (Hoffmann 2000).

Second, an expert on credit lending to the self-employed was used to give a prior estimate of the good rate for each attribute i. His belief was given in the form of a beta distribution with parameters (r, m) (recall section 6.7). This was then adjusted to include $\tilde{r}$, the estimate of the good rate obtained from the sample, so that the posterior belief for the good rate is a beta distribution with parameters $(r + n_i\tilde{r}, m + n_i)$. The score for the attribute is then taken as its weight of evidence with this adjusted good rate, i.e.,

$$\log\left(\frac{r + n_i\tilde{r}}{m + n_i} \cdot \frac{n}{g}\right), \tag{12.2}$$

and the scorecard is taken to be the sum of these scores. Thus this approach used subjective estimates in two different ways and used a shrinkage of the estimate to compensate for the small sample sets. Each of these ideas makes sense in its own right, but there is no rationale for the overall methodology. It reflects the ad-hoc nature of scorecard building with very small samples.

12.3 Combining two scorecards: Sufficiency and screening

There are several examples in consumer lending where more than one scoring system is available to the lender. For example, many lenders will build a scoring system based only on the application-form variables and then another one using credit bureau data. The reason is that the lender has to pay a fee to acquire the latter information and so may hope that the results of the application-only score are sufficiently clear that this expense is not necessary. In mortgage lending, lenders often use a loan-to-income ratio as an initial rule of thumb and then use a mortgage score later in their credit decision process. A number of investigations looked at whether these scoring systems should be combined and how.

Zhu, Beling, and Overstreet (1999a) looked at the problem of an application score S_1 and a credit score S_2 for a data set of 600,000 automobile loan applications. They calculated a combined score $s_c = a_0 + a_1 s_1 + a_2 s_2$ using logistic regression. They wanted to show that whether the two individual scorecards S_1 and S_2 were better than each other and if the combined scorecard was better than both. In Chapter 7, we discussed ways to measure the performance of scorecards, but here we want to introduce a new way to look at the performance of these three scorecards, which originated in measuring probabilistic forecasts (Clemen, Murphy, and Winkler 1995, Zhu, Beling, and Overstreet 2001)

Assume that the credit-scoring system gives a score s and that if necessary a nonlinear transformation has been applied to the score so that the probability of being good is the score s, $P(G|s) = s$. Suppose we have a number of such scorecards $A, B, C, \ldots$ based on the same sample; then we can think of the scores s_A, s_B, s_C as random variables and define $f_A(s_A|G)$, $f_A(s_A|B)$, $f_A(s_A)$ to be, respectively, the conditional probability of a good (bad) having a score s_A under A and the distribution of scores under A.

Definition 12.1. *If A and B are two scorecards, then A is sufficient for B if there exists a stochastic transformation h so that*

$$0 \le h(s_A|s_B) \quad \text{for all } s_A, s_B, \qquad \sum_{s_B} h(s_B|s_A) = 1 \quad \text{for all } s_A, \quad \text{and}$$
$$\sum_{s_A} h(s_B|s_A) f_A(s_A|G) = f_B(s_B|G) \quad \text{for all } s_B, \tag{12.3}$$

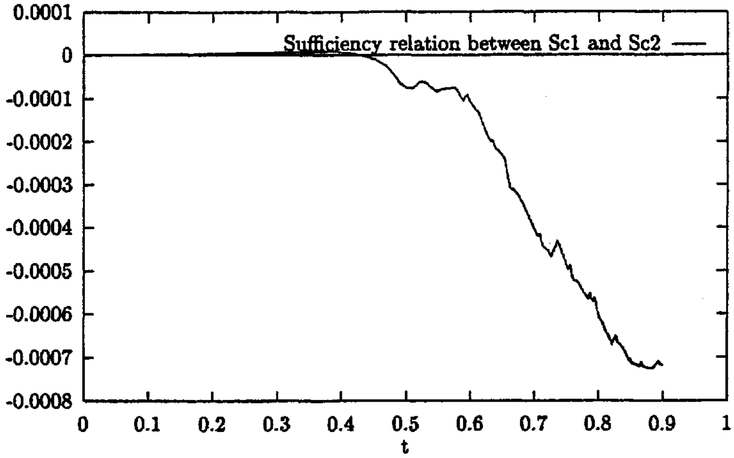
This means that B can be thought of as A with extra uncertainty added and so is dominated by A in that A will be superior under any of the normal business measures. Sufficiency is useful because, as Zhu, Beling, and Overstreet (1999b) pointed out, there is a relationship between sufficiency and the actual loss rates defined by (7.5). Scorecard S_1 is sufficient for scorecard S_2 if and only if for every expected default cost D, every expected profit L, and every score distribution $f(x)$ in (7.5), the actual loss rate l_{s_1}(Actual) is less than or equal to the actual loss rate l_{s_2}(Actual). However, in many pairs of scorecards neither is sufficient for the other.

The definition of sufficiency in (12.3), although making things clear, is very difficult to check. A much easier condition was given by De Groot and Eriksson (1985), namely, that A is sufficient for B if and only if, when $F_A(s) = \sum_{t \le s} f_A(t)$, $F_B(s) = \sum_{t \le s} f_B(t)$,

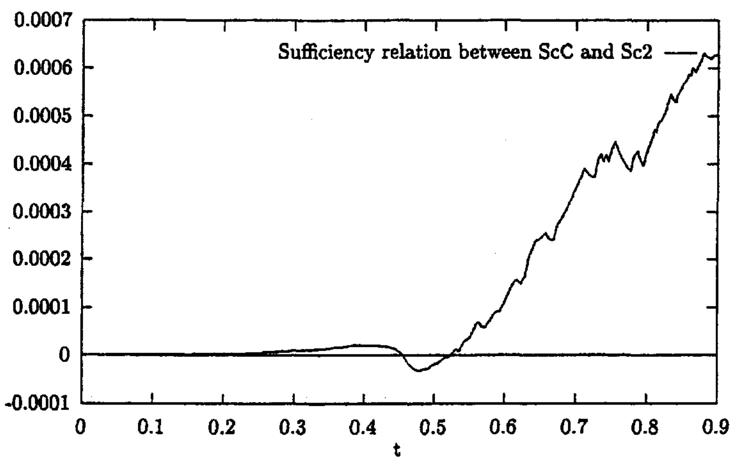
$$\int_0^p F_A(s)ds - \int_0^p F_B(s)ds \ge 0 \quad \text{for all } 0 \le p \le 1.$$

This implies that the distribution f_A has a second-order stochastic dominance over f_B.

When these measurements were applied to the scorecards developed by Zhu, Beling, and Overstreet (2001) (recall that $s_c = a_0 + a_1 s_1 + a_2 s_2$), the results were as in Figure 12.1. What these show is that scorecard 1—the application score—essentially dominates scorecard 2—the pure bureau score, but the combined scorecard is again much better than the application score. These results can be confirmed using the other performance measurements of Chapter 7, and hence we have the not-unexpected result that the combined scorecard is superior to the individual ones. One can show (Zhu, Beling, and Overstreet 1999) that scorecard S_1 is sufficient for scorecard S_2 if and only if the ROC curve of S_1 lies above that of S_2 everywhere. The results in Figure 12.1 do give a warning about sufficiency being used to decide whether to combine scorecards. S_2 is essentially dominated by S_1 and so one might feel that the second scorecard cannot enhance the first one, and yet the combined scorecard is superior to S_1. These are not startling results, but the results give a warning about the use of domination as a way to dismiss scores. Just because one scorecard dominates another does not mean that a combined scorecard may not be better still.



(a)



(b)

Figure 12.1. (a) *Graph of* $\int_0^p F_1(s)ds - \int_0^p F_2(s)ds$. (b) *Graph of* $\int_0^p F_c(s)ds - \int_0^p F_2(s)ds$.

Another approach to investigating whether to combine two scorecards is to use the analogies with screening methods suggested by Tsai and Yeh (1999). Screening models are used in quality control, psychology, education, and medicine to select items whose performance variables are within specifications by observing correlated screening variables rather than the performance variable. In credit scoring, the performance variable would be whether the loan defaults, while the screening variables are the application characteristics and the resultant scores. The screening models assume that all the variables form a multidimensional normal distribution, and the operational characteristics of using the screening—the scorecards in this case—can be obtained partly analytically and partly by simulation.

Tsai, Thomas, and Yeh (1999) look at the problem where there are two scorecards with S_1 and S_2 being the resultant score variables and Y the propensity to be good. It is assumed that (Y, S_1, S_2) have a standard trivariate normal distribution with covariance matrix:

$$\begin{pmatrix} 1 & \rho_{01} & \rho_{02} \\ \rho_{01} & 1 & \rho_{12} \\ \rho_{02} & \rho_{12} & 1 \end{pmatrix}. \tag{12.4}$$

Two different approaches for using the two scorecards in credit decision making are considered. The first is to have two separate scorecards and accept an applicant only if $s_1 \geq k_1$ and $s_2 \geq k_2$; i.e., they pass the cutoff on both scorecards. Thus the acceptance rate is $a_1 = \text{Prob}(S_1 \geq k_1, S_2 \geq k_2)$, and if we assume that bad applicants are those whose propensity to be good is below k_0, then the overall default rate is $d_1 = \text{Prob}\{Y \leq k_0 | S_1 \geq k_1, S_2 \geq k_2\}$. One can then choose k_1 and k_2 to minimize d_1 subject to a_1 equal to some agreed constant.

The alternative is to have a combined scorecard of $w_1 S_1 + w_2 S_2$ and assume that $Y = w_1 S_1 + w_2 S_2 + e$. Using standard regression analysis, one can show that

$$\begin{aligned} w_1 &= \frac{\rho_{01} - \rho_{02}\rho_{12}}{1 - \rho_{12}^2}, \qquad w_2 = \frac{\rho_{02} - \rho_{01}\rho_{12}}{1 - \rho_{12}^2}, \qquad \text{and} \\ \sigma^2 &= \frac{1 + 2\rho_{01}\rho_{02}\rho_{12} - \rho_{01}^2 - \rho_{02}^2 - \rho_{12}^2}{1 - \rho_{12}^2}, \end{aligned} \tag{12.5}$$

where σ^2 is the variance of e. Assuming that the cutoff score is k, so the acceptance rate is $a_2 = \text{Prob}\{w_1 S_1 + w_2 S_2 \geq k\}$, one can show that $k = \Phi^{-1}(z_{1-a_2})\sqrt{1 - \sigma^2}$, where Φ^{-1} is the standard inverse normal distribution. Similarly, one can get an expression for the default rate by

$$d_2 = \text{Prob}\{Y \leq k_0 | w_1 S_1 + w_2 S_2 \geq k\} = \frac{BN(-k(1 - \sigma^2)^{-\frac{1}{2}}, k_0, -(1 - \sigma^2)^{\frac{1}{2}})}{1 - \Phi(k(1 - \sigma^2)^{-\frac{1}{2}})},$$

where BN is the standard bivariate normal distribution function, where values can be compiled using subroutines in the International Mathematical and Statistical Library. This allows one to use simulation to compare the effect of the different ways of using two scorecards in credit decisions.

Using an example where $p_{01} = 0.6$, $p_{02} = 0.4$, $p_{12} = 0.1$, and the overall default rate is 20%, so $\text{Prob}\{Y \leq k_0\} = 0.2$, Tsai, Thomas, and Yeh (1999) compared four different systems. The first used only the scorecard S_1, the second used only the scorecard S_2, the third was where customers were accepted only if they were above the cutoff score on both scorecards, and the last was a new scorecard whose score was a linear combination of S_1 and S_2. Using the analysis above, they found the default rates for different levels of acceptance described in Table 12.1. Again the results showed that combining the scorecards into a single scorecard seemed best.

In this section, we looked at two theoretical approaches that underpin the idea that combining scorecards is in general a sensible thing to do. They are useful tools in trying to understand when such combining of scorecards makes sense and when it does not, but more tools may be necessary. In the next section, we look at other ways to combine the outcome of two scoring systems apart from taking a linear combination of the underlying scores.

Table 12.1. *Default rates for four different scorecards.*

Acceptance rate	S_1	S_2	S_1 and S_2	Linear combination of S_1 and S_2
0.1	.011	.048	.007	.003
0.2	.022	.067	.017	.010
0.3	.035	.082	.029	.018
0.4	.049	.096	.040	.030
0.5	.064	.110	.055	.044
0.6	.082	.124	.076	.061
0.7	.102	.139	.097	.083
0.8	.126	.155	.123	.111
0.9	.156	.173	.155	.147

12.4 Combining classifiers

Suppose that one has n different credit scoring classifiers built on a population where the prior probability of being good (bad) is p_G (p_B). Can we combine these classifiers to produce something more powerful than the best of the individual classifiers? Zhu, Beling, and Overstreet (1999b) showed that this is possible and that one can construct a combined classification that is sufficient for the individual classifiers. As mentioned, this implies that the actual loss rate using the combined classifier will be less than that for using individual classifiers. Let y_i be the random variable describing the outcome of classifier i—either the score or the node of the classification tree—and if a particular consumer comes up with classifier values $y_1, y_2, \ldots, y_n$, let $q(G|y_1, y_2, \ldots, y_n) = q(G|\mathbf{y^n})$ be the probability such a consumer is good. Let $p(\mathbf{y^n}|G)$ and $p(\mathbf{y^n}|B)$ be the conditional probabilities that a good or bad consumer has classifier outcomes $y_1, y_2, \ldots, y_n$. By Bayes's theorem,

$$q(G|\mathbf{y^n}) = \frac{p_G p(\mathbf{y^n}|G)}{p_G p(\mathbf{y^n}|G) + (1 - p_G)p(\mathbf{y^n}|B)}. \tag{12.6}$$

Thus we need to estimate $p(\mathbf{y^n}|G)$. If one believed that the classifiers were stochastically independent, then one could assume that $p(\mathbf{y^n}|G) = \prod_{i=1}^n p_i(y_i|G)$, but in reality classifiers are likely to be correlated, so instead we define the n-dimensional conditional density $p(\mathbf{y^n}|G)$ as a product of $n - 1$ univariate conditional densities:

$$p(\mathbf{y^n}|G) = p(y_n|\mathbf{y^{n-1}}, G)p(y_{n-1}|\mathbf{y^{n-2}}, G) \cdots p(y_2|y_1, G)p(y_1|G). \tag{12.7}$$

Then $p(\mathbf{y^n}|G)$ can be updated sequentially as

$$q(G|\mathbf{y^n}) = \frac{p(y_n|\mathbf{y^{n-1}}, G)q(G|\mathbf{y^{n-1}})}{k(y_n|\mathbf{y^{n-1}})}, \tag{12.8}$$

where $k(y_n|\mathbf{y^{n-1}}) = p(y_n|\mathbf{y^{n-1}}, G)q(G|\mathbf{y^{n-1}}) + p(y_n|\mathbf{y^{n-1}}, B)(1 - q(G|\mathbf{y^{n-1}}))$. Thus all that is needed is to calculate $p(y_n|\mathbf{y^{n-1}}, G)$ and $p(y_n|\mathbf{y^{n-1}}, B)$.

Zhu, Beling, and Overstreet (1999b) suggested calculating these by transforming the y values so that they satisfy a normal distribution applying a regression equation to connect the transformed y_n with the $y_1, \ldots, y_{n-1}$ and then transferring this relationship back to the y_ns. This has the advantage that the regression is done with normal distributions and that it can work whatever the distributions of $p(y_n|\mathbf{y^{n-1}}, G)$ and $p(y_n|\mathbf{y^{n-1}}, B)$.

Formally define z_k^i by $Q(z_k^i) = F(y_k|i)$, $k = 1, \ldots, n$, $i = G$ or B, where Q is the cumulative distribution function of the standard normal distribution. Obtain the conditional distributions $p(z_n^i|\mathbf{z}^{n-1}, i)$ by applying a regression to these transformed values so that

$$z_n^i = r_n^i + s_{n1}^i z_1^i + \cdots + s_{n,n-1}^i z_{n-1}^i + e_{n-1}^i i = G \text{ or } B,$$

where e_n^i is $N(0, (\sigma_n^i)^2)$. This implies that $p(z_n^i|\mathbf{z}^{n-1}, i)$ has a normal distribution with parameters $N(r_n^i + \sum_{j=1}^{n-1} s_{nj}^i z_j, (\sigma_n^i)^2)$. Transforming this back to the original distribution gives

$$p(y_n|y^{n-1}, i) = \frac{p(y_n|i)}{\sigma_n^i \phi(Q^{-1}(F(y_n|i))}\phi\left(\frac{Q^{-1}(f(y_n|i) - r_n^i - \sum_{j=1}^{n-1} s_{nj}^i Q^{-1}(F(y_j|i))}{\sigma_n^i}\right), \tag{12.9}$$

where ϕ is the density function of the standard normal distribution.

This Bayesian procedure produces a sequence of combined classifiers of increasing predictive power but also, unfortunately, increasing complexity. One way to simplify this is to use the idea of sufficiency introduced in the previous section since if S_1 is sufficient for the classifier based on S_1 and S_2, then $p(G|s_1, s_2) = p(G|s_1)$ and one can ignore classifier 2. Zhu, Beling, and Overstreet (1999b) used this approach to combine a linear regression, a logistic regression, and three neural net–based scorecards. The error rates of the individual classifiers were between 12% and 17% on a holdout sample, while the combined classifier, having recognized by the sufficiency tests that two of the classifiers were not needed, had an error rate of 7.5% on that sample.

Another way to combine classifiers would be to average them in some sense—somewhat akin to the basic generic scorecard of section 7.2. A more subtle approach would be to jointly calibrate the classifiers so that each classified the error between the true class and the average of the other classes. This corresponds to saying that if $\mathbf{x}$ are the characteristics and Y_i, $i = 1, \ldots, n$, are the different classifiers, then one defines

$$q(G|\mathbf{x}) = \alpha + \sum_i g(\mathbf{x}, Y_i),$$

where $g(\mathbf{x}, Y_i)$ is the prediction by classifier Y_i of the probability of being good.

Mertens and Hand (1997) looked at a specific example with two classifiers. The first is a linear regression model with parameters $\mathbf{w}$, so $g(\mathbf{x}, \mathbf{w}, LR) = \mathbf{x}^T \cdot \mathbf{w}$. The second is a classification tree with final nodes given by hyperrectangles R_l in the characteristic space and the lth one having a good rate of g_l, so $g(\mathbf{x}, R, CT) = \sum_{l=1}^{L} \chi_{R_l}(\mathbf{x})$. $\chi_{R_l}(\mathbf{x})$ is the indictor function, which is 1 if $\mathbf{x} \in R$ and 0 otherwise. One can then alternate between the tree and the linear regression, keeping the parameters for one fixed and adjusting the parameters of the other to minimize the mean square classification error. Suppose there are n consumers, g of whom are good, in the sample. If the kth one has characteristics $\mathbf{x}_k$ and a probability y_k of being good ($y_k = 0$ or 1), the algorithm first defines $g_0(\mathbf{x}_k, \mathbf{w}, LR) = \frac{g}{n}$ and $g_0(\mathbf{x}_k, R, CT) = \frac{g}{n}$. At iteration i,

(a) calculate $\mathbf{w}_i$ by fitting the model $g_i(\mathbf{x}_k, \mathbf{w}_i, LR)$ as the best linear regressor of the residuals $y_k - g_{i-1}(\mathbf{x}_k, R_{i-1}, CT)$;

(b) grow the tree $g_i(\mathbf{x}_k, CT)$ to fit the residuals $y_k - g_i(\mathbf{x}_k, \mathbf{w}_i, LR)$ and prune the tree under some agreed procedure to obtain the classification tree $g_i(\mathbf{x}_k, R_i, LR)$;

(c) calculate the mean square error and the optimal constant α. If necessary, repeat the procedure.

This is a procedure for combining a regression and a classification tree, but this approach could be used for any number and any combination of classifying approaches. There are several other approaches for combining classifiers, but little work has been done on this area in credit scoring to date.

12.5 Indirect credit scoring

All the credit-scoring methods discussed so far seek to directly predict the future default risk, but one of the current moves in credit scoring is to predict expected profit rather than this default risk. One way to do this is to predict the future values of other variables, like balance outstanding, which affect the profit as well as the default risk. One can then use the estimates of these intermediate variables to forecast the profit or the default risk indirectly. The differences are displayed in Figure 12.2.

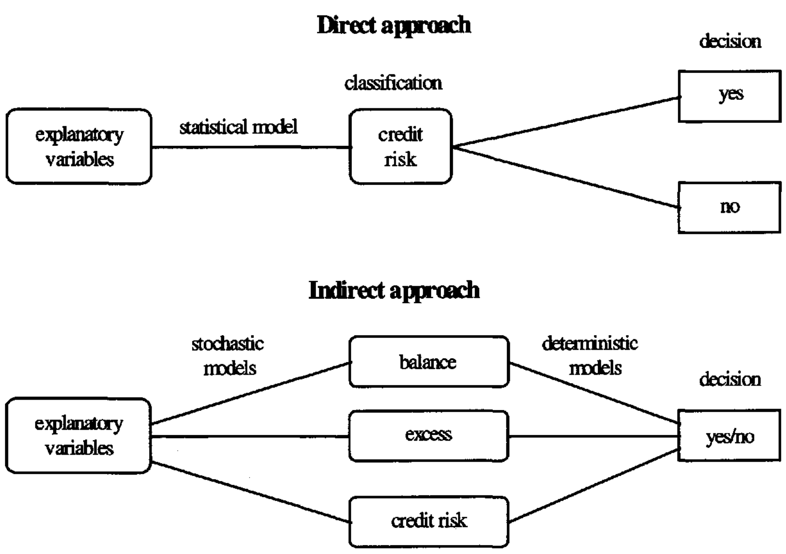


Figure 12.2. *Direct and indirect approaches to credit decisioning.*

There are several advantages to the indirect method. Since one is estimating a number of variables, one can get a better understanding of the way the repayment process develops; one can be more flexible in the definitions used to decide whether to grant credit, without having to recalculate the whole model; and validation of the model is easier to undertake. On the other hand, the more variables one estimates, the more errors that can be introduced into the system.

Li and Hand (1997) investigated this approach on a bank's current account data. They forecasted each component of a multicriterion definition of bad. Thus they used the data over the previous nine months to estimate the month-end balances and maximum balances, credit turnover, and amount the account is over the overdraft limit in months 9 and 10 in the future. These were the variables, $\mathbf{y}$, that they wanted to estimate. There were 29 predictor variables, $\mathbf{x}$, available, and they looked at three types of regression. The full model used lagged values of $\mathbf{x}$ and $\mathbf{y}$ to predict future $\mathbf{y}$; the predictor model used only lagged $\mathbf{x}$ values to predict $\mathbf{y}$, and the lagged model just used lagged values of $\mathbf{y}$ to predict future $\mathbf{y}$. The definition of default taken by the bank involved either being £500 over the overdraft limit if the account was active, £100 over the overdraft limit if the account was inactive for one month, and £100 overdrawn if the account was inactive for two or more months. Thus having the predictions of the $\mathbf{y}$, one could calculate whether the account was bad under this definition. This was compared with the logistic regression approach where each customer was directly classified as good or bad depending on their behavior in the ninth month. This was called the direct approach. The results they obtained are given in Table 12.2, where ER is the error rate and BR is the bad rate among those accepted under each approach.

Table 12.2. *Comparison of errors in the direct and indirect methods.*

Model	Full	Predictor	Lagged	Direct
ER	0.171	0.170	0.167	0.170
BR	0.162	0.160	0.163	0.163

These are very similar results, so Li and Hand (1997) then performed a number of simulation experiments. Although one might expect the performance of the indirect methods to be superior in that they used the exact definition of what is a bad, which is quite complex, this did not turn out to be the case. Only when one was able to classify the intermediate variables very accurately was the indirect method clearly superior, and when the cost of misclassifying a bad as a good was high, the direct method was clearly superior. Thus once again this approach needs some development before it can be used with confidence in practice, although given its flexibility it is clearly worth undertaking the development.

12.6 Graphical models and Bayesian networks applied to credit scoring

The indirect method outlined in the last section suggests that it is useful to model the connection between the variables that describe a consumer's performance and the attributes of the consumer. This is the consequence of moving from credit granting based purely on default risk to credit granting based on the much more complex objective of profit. Graphical modeling is a set of statistical tools that has proved useful in modeling the connections between variables. Recently, several authors (Whittaker 1990, Sewart and Whittaker 1998, Hand, McConway, and Stanghellini 1997, Chang et al. 2000) suggested that graphical models could prove useful in building models in the credit scoring context.

A graphical model of a situation consists of a number of vertices, each of which represents a continuous or discrete variable of interest. Some of these will be the application variables, like residential status or age; others can be outcome variables, like balance after 12 months or number of missed payments; others can be exogenous variables, like unemployment rate or interest rates; while some may be unobservable, like capacity to repay. The vertices are linked by edges, which describe the probabilistic dependency between them. In particular, the lack of an edge between two vertices implies the two variables are conditionally independent given the rest of the variables.

Definition 12.2. *Let X, Y, Z be a set of variables; then X is conditionally independent of Y given Z if when p is either the conditional density function (continuous variable) or the probability function (discrete variable), then $p_{\mathbf{x}|\mathbf{y},\mathbf{z}}(\mathbf{x}|\mathbf{y},\mathbf{z})$ does not depend on y. This is* *equivalent to saying*

$$p_{\mathbf{x}|\mathbf{y},\mathbf{z}}(\mathbf{x}|\mathbf{y},\mathbf{z}) = p_{\mathbf{x},\mathbf{z}}(\mathbf{x},\mathbf{z}) \cdot p_{\mathbf{y},\mathbf{z}}(\mathbf{y},\mathbf{z}). \tag{12.10}$$

With this definition of vertex and edge (or nonedge), the relationship between the variables in the model can be displayed as a graph. Hence the name *graphical models*. If two variables are separated in the graph, i.e., you cannot get from one to the other except through another variable, then they are conditionally independent given this other variable. This is called the global Markov property.

For example, consider four variables W, X, Y, Z in which W is conditionally independent of Z given (X, Y) and Y is conditionally independent on Z given (W, X). This is represented in Figure 12.3.

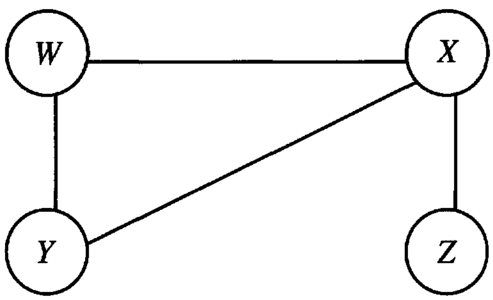


Figure 12.3. *A typical graphical model.*

This means that Z and W are separated by X, so Z is conditionally independent of W given X (and also Z is conditionally independent of W given Y only). Marginal indepedence is the normal definition of independence and says that $p_{xy}(x, y) = p_x(x)p_y(y)$. This is shown in Figure 12.4.

Figure 12.4. *A graphical model of marginal independence.*

We can also bracket variables together in groups called cliques if they are all adjacent to each other in the graph and are the largest group with this property. Thus if we added any other variable, we would not form a clique. Therefore, in Figure 12.3, $\{X, Y, W\}$ and $\{X, Z\}$ are cliques.

There are three ways in which these graphical models can be used in credit scoring. First, they illuminate the relationship between the factors that affect the behavior of borrowers. They model the interrelationships between the variables simultaneously without forcing any particular one to be distinguished as the outcome variable. Thus it may be easier to identify subpopulations with certain behavioral characteristics than by the standard credit-scoring methods, even those based on classification trees. Second, the graphical models can be used to predict risk or profit for each individual. Given the application variables, one can use the relationship to predict the other variables, including unknown ones like default risk. Thus it can be used in the same way as normal credit-scoring systems. Third, since in reality credit granting is a decision process, with different decisions being made at different times—to whom should we mail, who will respond, whom do we accept, who will then use the card, who will change to another lender, who will default—it is useful to understand when information will be known and what insights that information can give. Thus it is a way to relate the possible predictions to the decision-making process and so can be used to ensure that the decision-making process is well designed.

The graph connecting the variables can be constructed either by starting with the complete graph and asking which edges can be left out or by starting with a graph with no edges and adding them. In both cases, one is trying to compare two models, one with an edge in and one with it out. These models correspond to different statistical models depending on whether all the variables are discrete (categorical), all are continuous, or there is a mixture with some variables continuous and others discrete. The discrete variable models correspond to loglinear models; the continuous variable models become multivariate normal models if we assume that the distributions are normal. In this case, conditional independence corresponds to zero partial correlations. Graphical mixed models were introduced by Lauritzen and Wermuth (1989), and a software package, MIM, was developed by Edwards (1995) to estimate the parameters and calculate the graph.

We will sketch the calculations for the discrete case since it is so common in credit and behavioral scoring to translate continuous variables into categorical ones. Suppose there are three categorical variables X, Y, and Z and let $p_{jkl} = P\{X = j, Y = k, Z = l\}$. The most general loglinear model defines these probabilities by

$$\mathrm{Log}(p_{jkl}) = u + u_j^X + u_k^Y + u_l^Z + u_{jk}^{XY} + u_{jl}^{XZ} + u_{kl}^{YZ} + u_{jkl}^{XYZ}. \tag{12.11}$$

There is a relationship between independence of the variables X and Y and whether u^{XY} is zero. If Y is conditionally independent of Z, given X, then that is equivalent to $u_{kl}^{YZ} = u_{jkl}^{XYZ} = 0$. If there are N observations in the sample set and n_{jkl} are the numbers of observations where $X = j, Y = k, Z = l$, then this is a multinomial distribution. Thus the chance of this happening is

$$\left(\frac{N!}{\prod_{jkl} n_{jkl}!}\right) \prod_{jkl} p_{jkl}^{n_{jkl}}.$$

The log of the likelihood of this happening is

$$l_M = \log(L_M(p_{jkl}|n_{jkl})) = \log\left(\frac{N!}{\prod_{jkl} n_{jkl}!}\right) + \sum_{jkl} n_{jkl} \log(p_{jkl}). \tag{12.12}$$

Let l_{full} be the log of the maximum likelihood assuming the full model of (12.11) and let l_M be the log of the maximum likelihood assuming some submodel M (maybe with X conditionally independent of Y). The deviance G^2 is the likelihood ratio test, so $G^2 = 2(l_{\text{full}} - l_M)$. In the full model, the maximum likelihood estimates of p_{jkl} are $\frac{n_{jkl}}{N}$. If $\hat{p}_{jkl}^1$ are the ML estimates, assuming that M_1 and $\hat{m}_{jkl}^1 = N\hat{p}_{jkl}^1$ are the expected cell counts and $\hat{p}_{jkl}^2$ and $\hat{m}_{jkl}^2$ are the corresponding estimates under a different model M_2, $M_1 \subseteq M_2$, then the deviance difference d is given by

$$d = G_2^2 - G_1^2 = 2\left(\sum_{jkl} n_{jkl} \ln \hat{p}_{jkl}^2 - \sum_{jkl} n_{jkl} \ln \hat{p}_{jkl}^1\right) = 2\sum_{jkl} n_{jkl} \ln\left(\frac{\hat{m}_{jkl}^2}{\hat{m}_{jkl}^1}\right). \tag{12.13}$$

Under M_2, d is asymptotically χ_k^2, where k is the difference in the number of free parameters between M_2 and M_1. The alternative is to use the fact that G_1^2 is $\chi_{k'}^2$, where k' is the difference in the number of free parameters between M_1 and M_{full}.

Sewart and Whittaker (1998) gave some examples of the use of these techniques on credit scoring, and the following examples are based on these.

Example 12.1 (Is wealth related to default risk?). A sample of 5000 customers, was compared for wealth (poor or rich) and default status (yes or no). Table 12.3 gives the distributions. M_1 is the model where wealth and default status are independent, so the independent default rate is 2.4% ($\frac{120}{5000}$) and the poor rate is 50.2% ($\frac{2510}{5000}$). This gives expected numbers in each of the four cells of

$$2449.76 \left(2510 \times \frac{4880}{5000}\right), \quad 60.24 \left(2510 \times \frac{120}{5000}\right),$$
$$2430.24 \left(2490 \times \frac{4880}{5000}\right), \quad 59.76 \left(2490 \times \frac{120}{5000}\right).$$

Table 12.3. *Numbers in Example* 12.1.

		X		
		Not defaulted	**Defaulted**	**Total**
Y	**Poor**	2450	60	2510
	Rich	2430	60	2490
	Total	4880	120	5000

M_2 is the full model where wealth and default are connected and the expected numbers there are, of course, the actual numbers. There is one extra parameter in this model (u^{XY}). Hence

$$\begin{aligned} d &= 2\left(2450 \ln\left(\frac{2450}{2449.76}\right) + 60 \ln\left(\frac{60}{60.24}\right) + 2430 \ln\left(\frac{2430}{2430.24}\right) + 60 \ln\left(\frac{60}{59.76}\right)\right) \\ &= .002. \end{aligned}$$

$\chi_{0.95}^2 = .0039$ for one degree of freedom, which suggests that the M_2 model is not a significant improvement over the M_1 model. Hence wealth and default rate are marginally independent.

Example 12.2 (Are wealth and usage related to default risk? (Marginal independence does not imply conditional independence.)). The use made of the credit card by the 5000 customers in Example 12.1 is now added to obtain Table 12.4. Compare this time M_2, which is the full graph with usage, wealth, and default related, and M_1, the graph where default and usage are dependent and usage and wealth are dependent, but default and wealth are conditionally independent. (Remember that they were marginally independent originally.) Again under M_2 with the complete graph, the expected numbers in each cell are the actual numbers in this case. For M_1, we use the fact that estimating someone as a light user is .4844 (heavy .5156) and the estimates of the conditional probabilities are $\hat{p}(\text{default}|\text{light}) = \frac{12}{2422}$, $\hat{p}(\text{default}|\text{heavy}) = \frac{108}{2578}$, $\hat{p}(\text{poor}|\text{light}) = \frac{2220}{2422}$, and $\hat{p}(\text{poor}|\text{heavy}) = \frac{290}{2578}$. This leads to the estimates for the cell entries under M_1 given in Table 12.5.

Table 12.4. *Numbers in Example* 12.2.

		X		
Z	**Y**	**Not defaulted**	**Defaulted**	**Total**
Light use	**Poor**	2210	10	2422
	Rich	200	2	
Heavy use	**Poor**	240	50	2578
	Rich	2230	58	

Table 12.5. *Estimate for* M_1.

		X	
Z	**Y**	**Not defaulted**	**Defaulted**
Light use	**Poor**	2209	11
	Rich	201	1
Heavy use	**Poor**	277.85	12.15
	Rich	2192.15	95.85

There are two extra degrees of freedom between M_1 and M_2, i.e., u^{xy} and u^{xyz}, so the deviance difference (12.13) has a χ^2 distribution with two degrees of freedom. In this case,

$$\begin{aligned} d = 2\Bigg\{&2210\ln\left(\frac{2210}{2209}\right) + 10\ln\left(\frac{10}{11}\right) + 200\ln\left(\frac{200}{201}\right) + 2\ln\left(\frac{2}{1}\right) + 240\ln\left(\frac{240}{277.85}\right) \\ &+ 50\ln\left(\frac{50}{12.15}\right) + 2230\ln\left(\frac{2230}{2192.15}\right) + 58\ln\left(\frac{58}{95.85}\right)\Bigg\} \\ = {}& 90.5. \end{aligned}$$

The 95% significance value of the χ^2 distribution with two degrees of freedom is .103, so this is vastly over that and shows that the M_2 model is significantly better than the M_1 model. Hence the graphical models for Examples 12.1 and 12.2 would be as given in Figure 12.5.

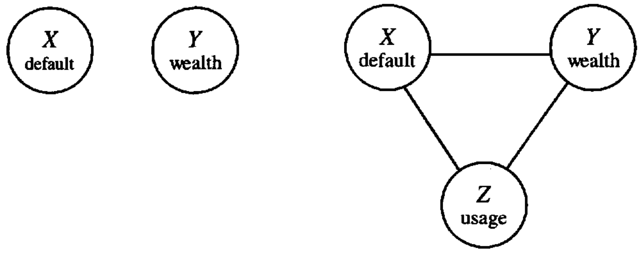


Figure 12.5. *Graphical models of Examples* 12.1 *and* 12.2.

Example 12.3. Another example of graphical models in credit scoring was studied by Hand, McConway, and Stanghellini (1997). They looked at 23,000 unsecured personal loans and investigated the graph of nine variables: AMOUNT (amount of loan); MARRY (marital status); RES (residential status); INCOME (disposable income); AGE (customer's age); INSURANCE (Did they take out insurance on the loan?); CURRAC (current account);

CCDEL (Has the applicant defaulted on a credit card?); and BAD (whether the loan became bad).

With nine variables, even if each variable has only two categories, that gives 512 cells, while if they have three categories each, there are 19,683 cells. Thus even with a sample of 23,000, one has to be very parsimonious in the number of attributes for each variable to avoid too many of the cells being empty. In this example, seven of the variables had two categories and two had three categories—2304 cells in total. The modeling started with the complete graph and tried to drop arcs and ended up with Figure 12.6.

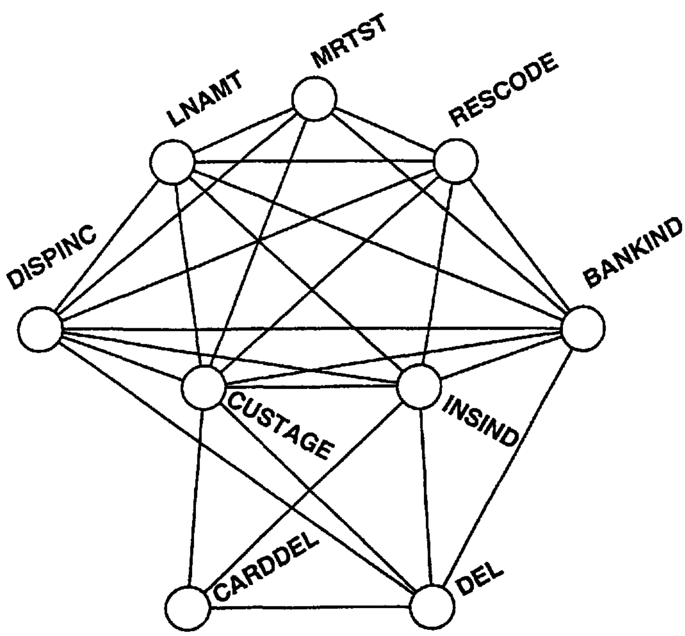


Figure 12.6. *Graphical model of Example* 12.3.

Only nine edges were dropped, but the graph can be decomposed into three groups. One is amount of loan, marital status, and residential position; the second is age, income, current account, and insurance; and the third is default on either credit card loan. What is interesting is that the first and third sets are separated by the second. Therefore, to estimate default likelihood, one does not benefit by knowing loan amount, marital status, and residential position if one knows age, income, insurance, and current account.

Thus far, we have considered graphical models where there is no direction to the edges because they represent conditional dependence between the variables. An undirected graph represents the joint distributions, but one can also define a joint distribution by a sequence of conditional probabilities. If this approach is taken, then one can put a direction on each arc of the graph to represent the way in which the conditioning is undertaken. One can also define a directed graph formally by saying that if $(X_1, \ldots, X_n)$ are a set of ordered variables,

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1)p_{X_2|X_1}(x_2|x_1)\cdots p_{X_n|X_{n-1},\ldots,X_1}(x_n|x_{n-1}, \ldots, x_1). \tag{12.14}$$

For $i < j$, an arrow is drawn from X_i to X_j unless $p_{X_j|X_{j-1},\ldots,X_1}(x_j|x_{j-1}, \ldots, x_1)$ does not depend on X_i; i.e., X_j is conditionally independent of X_i given $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}\}$.

This directed graphical model is also called a Bayesian network because it was introduced to look at probabilistic expert systems where the emphasis is on Bayesian ideas of updating one's beliefs; see (Spiegelhalter et al. 1993) for a survey of the area. In these Bayesian networks, if there is a directed arc from vertex X_i to vertex X_j, then vertex i is a parent (direct predecessor) of node j. Let $P(j)$ be the set of parents of j; i.e., $P(j) = \{i|(i, j)$ is a directed arc$\}$. Similarly, $S(i)$ are the children (direct successors) of node i, so $S(i) = \{j|(i, j)$ is a directed arc$\}$. This enables us to define all the variables in the database that are relevant to the prediction of variable X_i. This is the Markov blanket surrounding node i. A Markov blanket of X_i in a Bayesian network is the subset of nodes

$$M(i) = (P(i) \cup S(i) \cup P(S(i)) - \{i\}; \tag{12.15}$$

i.e., it has the parents, the children, and the parents of the children of the variable X_i.

It follows from these definitions that if $p(.|.)$ is again a conditional probability or a conditional density, then we can show that

$$p(x_i|x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x \ldots) = p(x_i|\mathbf{x}_{M(i)}).$$

Chang et al. (2000) used the idea of a Bayesian network in the credit-scoring context in the following way. They took X_0 to be the variable describing risk of default (i.e., G = good or B = bad) and assumed that it had no predecessors but that the application variables $X_1, \ldots, X_p$ could be its children. They pointed out that with a suitable transformation, the score $S(\underline{X})$ for someone with application details x is equivalent to the log of the odds of being good to being bad. (This was the basis of the logistic regression models.) Applying this gives the relationship

$$s(x) = \log\left(\frac{q(G|x)}{q(B|x)}\right) = \log\left(\frac{p_G p(x|G)}{p_B p(x|B)}\right) = \log\left(\frac{p_G}{p_B}\right) + \log\left(\frac{p(x|G)}{p(x|B)}\right). \tag{12.16}$$

Chang et al. (2000) modified the definition of a clique to define cliques in directed graphs. They defined a successor clique C of the performance node to be a set of nodes at least one of whom is a successor of the performance node, and the set has the property that all children of nodes in the clique lie themselves within the clique; i.e., $S(C) \subseteq C$. This means that if we have two successor cliques C_1 and C_2, then

$$p(\mathbf{x}_{C_1 \cup C_2}|x_0) = p(x_{C_1}|x_0)p(x_{C_2}|x_0)$$

since no arcs reach from one clique to the next. This means that if there are K successor cliques $C_1, C_2, \ldots, C_K$ in the network,

$$p(\mathbf{x}|G) = p(\mathbf{x}_{C_1}|G)p(\mathbf{x}_{C_2}|G)\cdots p(\mathbf{x}_{C_K}|G) \quad \text{and}$$
$$p(\mathbf{x}|B) = p(\mathbf{x}_{C_1}|B)p(\mathbf{x}_{C_2}|B)\cdots p(\mathbf{x}_{C_K}|B).$$

Hence

$$s(x) = \log\left(\frac{p_G}{p_B}\right) + \sum_{k=1}^{K} \log\left(\frac{p(\mathbf{x}_{C_k}|G)}{p(\mathbf{x}_{C_k}|B)}\right) = s_0 + \sum_{k=1}^{K} s(C_k), \tag{12.17}$$

where

$$s(C_k) = \log\left(\frac{p(\mathbf{x}_{C_k}|G)}{p(x_{C_k}|B)}\right).$$

This means that the score splits into the sum of its clique scores. This result is useful in that there may be far fewer elements in a clique than the p original variables and so score calculation is easier. The structure of the cliques might also give insight into the economic and behavioral factors that most influence the score.

However, this decomposition into cliques may be less helpful than is first thought. Chang et al. (2000) applied their results to a sample of 7000 applicants for bank credit with 35 characteristics available on each. Starting with the model where each of the 35 characteristics was assumed to depend on the good-bad performance variable X_0 but to be conditionally independent of each other, a Bayesian network was built to find the Markov blanket of X_0 and the successor cliques. It produced a result where there were 25 nodes in the Markov blanket (so 10 variables were dropped altogether), and these split into 11 one-node cliques, 4 two-node cliques, and 2 three-node cliques. Thus the network got only slightly more complicated than the one that was taken initially.

These results suggest that more research is needed in developing the methodology for using graphical models in credit scoring. At present, it seems if you start with the complete graph, you will not get rid of many arcs, while if you start with a very sparse graph, you will not add many arcs. Clearly, however, the potential for this sort of approach to building profit-scoring models is considerable.

12.7 Survival analysis applied to credit scoring

Credit-scoring systems were built to answer the question, "How likely is a credit applicant to default by a given time in the future?" The methodology is to take a sample of previous customers and classify them into good or bad depending on their repayment performance over a given fixed period. Poor performance just before the end of this fixed period means that customer is classified as bad; poor performance just after the end of the period does not matter and the customer is classified as good. This arbitrary division can lead to less-than-robust scoring systems. Also, if one wants to move from credit scoring to profit scoring, then it matters when a customer defaults. One asks not if an applicant will default but when will they default. This is a more difficult question to answer because there are lots of answers, not just the yes or no of the "if" question, but it is the question that survival analysis tools address when modeling the lifetime of equipment, constructions, and humans.

Using survival analysis to answer the "when" question has several advantages over standard credit scoring. For example,

(a) it deals easily with censored data, where customers cease to be borrowers (either by paying back the loan, death, changing lender) before they default;

(b) it avoids the instability caused by having to choose a fixed period to measure satisfactory performance;

(c) estimating when there is a default is a major step toward calculating the profitability of an applicant;

(d) these estimates will give a forecast of the default levels as a function of time, which is useful in debt provisioning; and

(e) this approach may make it easier to incorporate estimates of changes in the economic climate into the scoring system.

Narain (1992) was one of the first to suggest that survival analysis would be used in credit scoring. Banasik, Crook, and Thomas (1999) make a comparison of the basic survival analysis approach with logistic regression–based scorecards and showed how competing risks can be used in the credit-scoring context. Stepanova and Thomas (1999, 2001) developed the ideas further and introduced tools for building survival analysis scorecards as well as introducing survival analysis ideas into behavioral scoring. This section is based on these last two papers.

Let T be the time until a loan defaults. Then there are three standard ways to describe the randomness of T in survival analysis (Collett 1994):

- survival function: $S(t) = \text{Prob}\{T \geq t\}$, where $F(t) = 1 - S(t)$ is the distribution function;
- density function: $f(t)$, where $\text{Prob}\{t \leq T \leq t + \delta t\} = f(t)\delta t$;
- hazard function: $h(t) = \frac{f(t)}{S(t)}$, so $h(t)\delta t = \text{Prob}\{t \leq T \leq t + \delta t | T \geq t\}$.

Two of the commonest lifetime distributions are the negative exponential, which with parameters λ has $S(t) = e^{-\lambda t}$, $f(t) = \lambda e^{-\lambda t}$, and $h(t) = \lambda$, and the Weibull distribution, which with scale λ and shape k has $S(t) = e^{-(\lambda t)^k}$, $f(t) = k\lambda t^{k-1}e^{-(\lambda t)^k}$, and $h(t) = k\lambda^k t^{k-1}$. The former has no aging effect in that the default rate stays the same over time; the latter is more likely to default early on if $k < 1$, is more likely to default late on if $k > 1$, and becomes the negative exponential distribution if $k = 1$.

In standard credit scoring, one assumes that the application characteristics affect the probability of default. Similarly, in this survival analysis approach, we want models that allow these characteristics to affect the probability of when a customer defaults. Two models have found favor in connecting explanatory variables to failure times in survival analysis: proportional hazard models and accelerated life models. If $\mathbf{x} = (x_1, \ldots, x_p)$ are the application (explanatory) characteristics, then an accelerated life model assumes that

$$S(t) = S_0(e^{\mathbf{w}\cdot\mathbf{x}}t) \quad \text{or} \quad h(t) = e^{\mathbf{w}\cdot\mathbf{x}}h_0(e^{\mathbf{w}\cdot\mathbf{x}}t), \tag{12.18}$$

where h_0 and S_0 are baseline functions, so the $\mathbf{x}$ can speed up or slow down the aging of the account. The proportional hazard assumes that

$$h(t) = e^{\mathbf{w}\cdot\mathbf{x}}h_0(t), \tag{12.19}$$

so the application variables $\mathbf{x}$ have a multiplier effect on the baseline hazard. One can use a parametric approach to both the proportional hazards and acceleration life models by assuming that $h_0(.)$ belongs to a particular family of distributions. It turns out that the negative exponential and the Weibull distributions are the only ones that are both accelerated life and proportional hazard models. The difference between the models is that in proportional hazards, the applicants most at risk for defaulting at any one time remain the ones most at risk for defaulting at any other time.

Cox (1972) pointed out that in proportional hazards one can estimate the weights $\mathbf{w}$ without knowing $h_0(t)$ using the ordering of the failure times and the censored times. If t_i and $\mathbf{x}_i$ are the failure (or censored) times and the application variables for each of the items under test, then the conditional probability that customer i defaults at time t_i given that $R(i)$ are the customers still operating just before t_i is given by

$$\frac{\exp\{\mathbf{w}\cdot\mathbf{x}_i\}h_0(t)}{\sum_{k\in R(i)}\exp\{\mathbf{w}\cdot\mathbf{x}_k\}h_0(t)} = \frac{\exp\{\mathbf{w}\cdot\mathbf{x}_i\}}{\sum_{k\in R(i)}\exp\{\mathbf{w}\cdot\mathbf{x}_k\}}, \tag{12.20}$$

which is independent of h_0.

There are two minor complications when applying these ideas to credit granting. First, the default data tend to be aggregated at a monthly level, so the default time is really discrete, not continuous, and this leads to lots of ties when there are a large number of customers who default or whose data is censored for other reasons in the same month. Cox (1972) suggested a linear log odds model as the discrete time equivalent or proportional hazards, where

$$\frac{h(t)\partial t}{1-h(t)\partial t} = \exp\{\mathbf{w}\cdot\mathbf{x}\}\left(\frac{h_0(t)\partial t}{1-h_0(t)\partial t}\right). \tag{12.21}$$

If d_i are the number of defaults (failures) at time t_i, let $R(t_i, d_i)$ be the set of all subsets of d_i customers taken from the risk set $R(i)$. Let R be any such subset of d_i customers in $R(t_i, d_i)$. Then let $\mathbf{s}_R = \sum_{r\in R}\mathbf{x}_r$ be the sum of the answers over the individuals in set R. Let D_i denote the set of the d_i of customers who actually defaulted at t_i and $\mathbf{s}_{D_i} = \sum_{r\in D_i}\mathbf{x}_r$. The likelihood function arising from model (12.21) is then

$$\prod_{i=1}^{K}\left(\frac{\exp(\mathbf{w}\cdot\mathbf{s}_{D_i})}{\sum_{R\in R(t_i;d_i)}\exp(\mathbf{w}\cdot\mathbf{s}_R)}\right). \tag{12.22}$$

Maximizing this is difficult because of the large number of sets that have to be summed over in the denominator, but Breslow (1974) and Effron (1977) have suggested simplifying approximations.

The results of applying Cox's proportional hazard model to a sample of 50,000 personal loans from a major U.K. financial institution are given in Figure 12.7. Two measures were used for the survival model based on application data:

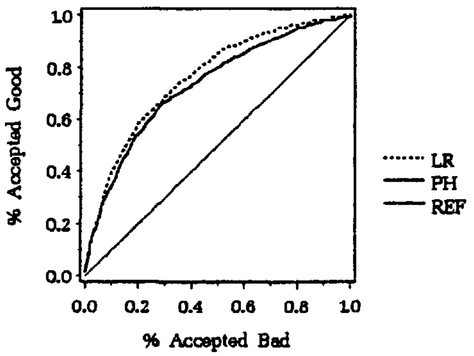
(i) How good was the model at estimating which loans will default in the first 12 months, compared with a logistic regression scorecard with failure in the first 12 months as its definition of bad?

(ii) How good was the model at estimating, among those still repaying after 12 months, which ones will default in the second 12 months? This again is compared with a logistic regression built on those who have survived 12 months, with the definition of bad being default in the second 12 months.

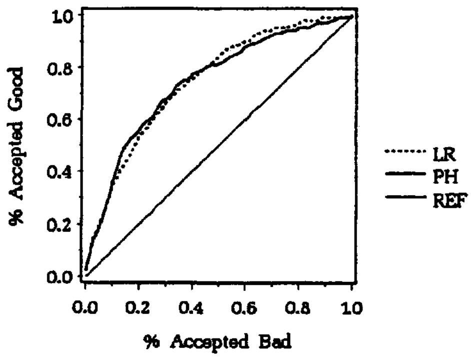
Figure 12.7 shows that the proportional hazard model on each criterion is competitive with the logistic regression model built solely on that criterion. The left-hand graph (a) shows the ROC curves under criterion (i), and the right-hand graph (b) is the ROC curve under criterion (ii). This suggests that even as a measure of default risk on a fixed time horizon, proportional hazard models are very competitive.

The survival analysis approach has a number of extensions that can be used in credit lending modeling. There are other reasons apart from default why a customer may finish before the original intended term—moving to another lender, selling the item that the loan was used to purchase, taking out another loan. All these mean the lender does not make the profit that was expected, and so in their own way these loans are bad for the lender. The competing-risks approach to survival analysis allows us to build survivor function models for each of these risks separately. Consider a sample case where the loan can be defaulted on, paid off early, or paid to term, and let T_d and T_e be the lifetime of the loan until default and until early repayment. If T_m is the term of the loan, the number of months of repayment is $T = \min\{T_d, T_e, T_m\}$.

Just as we were able to estimate T_d previously, we could use the same analytic techniques to estimate T_e, the time until early repayment. One does not need to assume that T_d



(a)



(b)

Figure 12.7. *ROC curves comparing proportional hazards and logistic regression models using default criterion.*

and T_e are independent variables, but as the calculations get complicated, if one does not (Collett 1994), it is usual to assume that is the case. If one repeats the exercise on the 50,000 personal loans and uses Cox's proportional hazard models to estimate early repayment both in the first year and then in the second year for those still repaying after 12 months, the proportional hazard model does not perform as well as the optimal logistic regression models, especially in the second year, where some of the results are worse than chance. Stepanova and Thomas (1999) showed that if one segmented the results on the term of the loan, the results are much better. Figure 12.8 shows the ROC curves for early repayment again in the first year and then in the second year for those still repaying after 12 months. Graph (a) is the ROC curve for this first criterion for loans with terms between months 12 and 24, and graph (b) is the ROC curve for the second criterion for such loans. Graphs (c) and (d) are ROC curves for these two criteria for loans between 24 and 30 months, and graphs (e) and (f) are the same ROC curves for loans with terms of more than three years. In the latter case, the survival approach works well. One of the interesting results of their analysis is that for early payments, it may be better to take t to be the number of months until completion of the term of loan, rather than the length of time since the loan was taken out. This is why segmenting on the term of the loan improved the results so dramatically.

One of the disadvantages of the proportional hazards assumption is that the relative ranking among the applicants of the risk (be it of default or early repayment) does not vary over time. This can be overcome by introducing time-dependent characteristics. So suppose that $x_1 = 1$ if the purpose of the loan is refinancing and 0 otherwise. One can introduce a second characteristic $x_2 = x_1 t$. In the model with just x_1 involved, the corresponding weight was $w_1 = 0.157$, so the hazard rate at time t for refinancing loans was $e^{0.157}h_0(t) = 1.17h_0(t)$, and for other loans, the hazard rate was $h_0(t)$. When the analysis was done with x_1 and x_2 involved, the coefficients of the proportional hazard loans were $w_1 = 0.32$, $w_2 = -0.01$. Thus for refinancing loans, the hazard rate at time t was $e^{0.32-0.01t}h_0(t)$ compared with others $h_0(t)$. Thus in month 1, the hazard from having a refinancing loan was $e^{0.31} = 1.36$ times higher than for a nonrefinancing loan, while after 36 months, the hazard rate for refinancing was $e^{-0.04} = 0.96$ of the hazard rate for not refinancing. Thus time-by-characteristic interactions in proportional hazard models allow the flexibility that the effect of a characteristic can increase or decrease with the age of the loan.

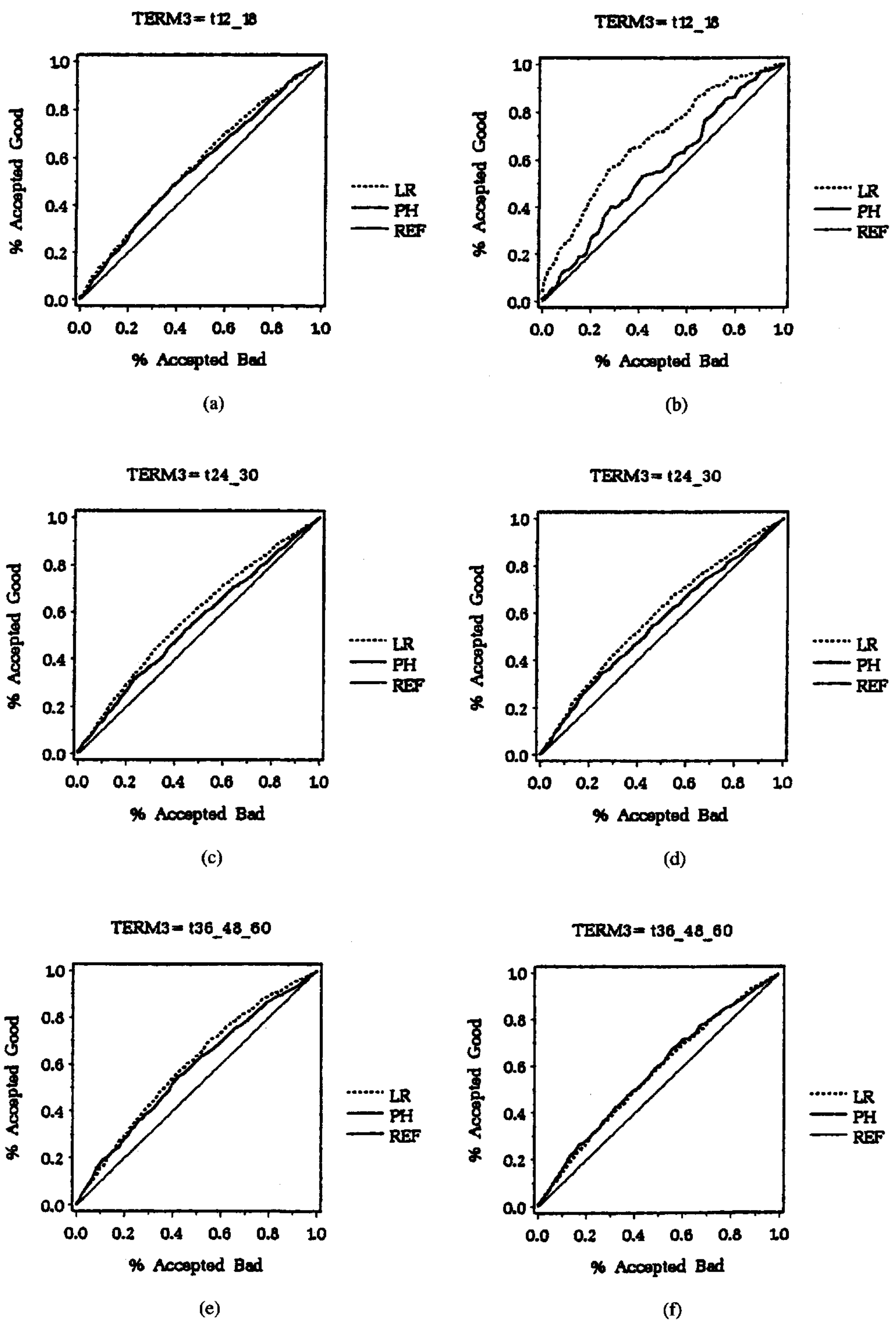


Figure 12.8. *ROC curves for early repayment segmented by term of loan.*

Survival techniques can also be applied in the behavioral scoring context, although a little more care is needed. Suppose that it is u periods since the start of the loan and $\mathbf{b}(u)$ are the behavioral characteristics in period u (which may include application characteristics). Then a proportional hazard model might say the hazard rate for defaulting in another t periods time, i.e., $t+u$ since the start of the loan, is $e^{\mathbf{w}(u)\cdot\mathbf{b}(u)}h_0^u(t)$. At the next period, $u+1$, the comparable hazard rate would be that for $t-1$ more periods to go, i.e., $e^{\mathbf{w}(u+1)\cdot\mathbf{b}(u+1)}h_0^{u+1}(t-1)$. Thus the coefficients $w(u)$ have to be estimated separately for each period u using only the data in the data set that have survived up to period u. As it stands, these coefficients could change significantly from one period to the next. One way to smooth these changes would be to make the behavioral score at the last period one of the characteristics for the current period. So suppose that $\mathbf{x} = (x_1, \ldots, x_p)$ were the application characteristics and $\mathbf{y}(u) = (y(u)_1, \ldots, y(u)_q)$ were the behavioral characteristics at period u. One defines a series of behavioral scores by $s(0) = \mathbf{w}\cdot\mathbf{x}$, where $e^{\mathbf{wx}}h_0^0(t)$ is the default rate hazard function at time 0. At time 1, the hazard function for time from now until default is $e^{\mathbf{w}_0^1 s(0)+\mathbf{w}(1)\cdot\mathbf{y}(1)}h_0^1(t)$, and define $s(1) = w_0(1)s(0) + \mathbf{w}(1)\mathbf{y}(1)$. Repeating this procedure leads to behavior scores at period u satisfying $s(u) = w_0(u)s(u-1) + \mathbf{w}(u)\cdot\mathbf{y}(u)$. Details of this analysis can be found in Stepanova and Thomas (2001).

Chapter 13

International Differences

13.1 Introduction

In this chapter, we explore a number of international differences in the behavior of borrowers and lenders and institutional arrangements affecting markets for credit. We begin by considering differences between countries in the volume of debt that households take on. Then we look at differences in credit bureau information between the U.K. and the U.S. Third, we review studies that have tried to explain differences in the choice of payment methods consumers use to buy goods and services. Fourth, we review some work that examined intercountry differences in scorecards, and finally we examine the rather unique position of the U.S. in the procedures available for debtors who declare themselves bankrupt and the possible affects this might have on some borrowers' behavior.

13.2 Use of credit

13.2.1 Consumer credit

Table 13.1 shows data for the U.S., the U.K., and Canada of consumer credit outstanding relative to GDP, consumer credit per capita of the population, and residential mortgage outstandings relative to GDP. While the U.S. has the largest consumer credit outstanding relative to GDP, Canada has the greatest stock of mortgage debt for its national output. In the U.S., there is $4922 of consumer credit outstanding for every person in the country, but only £1727 ($2862) in the U.K. Table 13.2 shows data for a larger number of countries for household debt outstanding as a percentage of GDP and for the net increase in debt outstanding as a percentage of GDP. These are all average figures for the periods shown in the table, which in most cases is 1992–1996. This period is determined by data availability and the attempt to span a full business cycle. Household debt is the sum of short-term and long-term loans owed by households. It includes consumer credit and mortgage debt. This table reveals great differences between countries. At the top of the list is the U.S., with household debt standing at 68.81% of GDP; Canada is not far behind at around 65%. Then come Spain, Germany, and France at between 54% and 42%, and then finally Italy at around 19% and Japan at a mere 11.98%. In terms of the average annual net increase in debt outstanding, the largest is again the U.S. at 4.4%, with Spain, Canada, and the U.K. not far behind at around 3%. Germany, France, and Italy have considerably lower increases at between 1% and 0.5%, and Japan actually had an average annual decrease, although it was very small. Of course, these figures cannot be regarded as precise because there are minor differences in the definitions used. But the overall impression is reasonably clear. The U.S. had the largest household debt outstanding and the largest annual net increase. Canada is very similar. Spain does not have as sizable a level of debt as the North American countries, but it has such a high rate of increase that its debt relative to GDP will increase further from the levels of comparable major European countries (with the exception of the U.K.). On the other hand, the stocks of consumer debt in Germany, France, Italy, and Japan not only are low relative to GDP but will stay low because of their low growth rates (unless there are dramatic changes in their growth rates). The immediate question arises as to why such differences exist.

Table 13.1. *International differences in the volume of credit,* 1998. (*Sources: Calculated from Financial Statistics, Office for National Statistics* (*U.K.*); *Federal Reserve Board, U.S.; Bank of Canada; and International Financial Statistics Yearbook, Vol.* 11, *International Monetary Fund,* 1999).

Country	Consumer credit outstanding/GDP (%)	Consumer credit outstanding per capita	Residential mortgage debt outstanding/GDP (%)
U.S.	15.65	US$4,922	52.63
U.K.	12.21	£1,727	54.53
Canada	17.67	C$5,223	44.53

Table 13.2. *Intercountry differences in household debt. The numerator consists of the sum of short-term and long-term household debt* (*or net increase in short-term plus long-term household debt*) *divided by the number of years in the period. The denominator is the mean GDP over the years in the period. All prices are current.* (*Sources: Calculated from OECD Financial Statistics, Part* 2: *Financial Accounts of OECD Countries, various issues, and International Monetary Fund, International Financial Statistics, January* 2000.)

Country	Period	Debt outstanding as percentage of GDP	Net increase in debt outstanding as percentage of GDP
U.S.	1992–1996	68.81	4.429
Germany	1992–1996	43.24	0.485
Canada	1992–1995	64.94	3.051
France	1992–1996	42.07	0.903
Spain	1992–1996	53.89	3.441
Italy	1992–1996	18.52	0.831
Japan	1992–1996	11.98	–0.001
U.K.	1992–1995	n.a.	3.323

Remember from Chapter 3 that the volume of credit observed depends on demand, supply, and the extent to which those who demand credit are unable to obtain it, that is, the extent of credit rationing. Therefore, differences in the volume of credit outstanding between countries will depend on differences in demand, in supply, or in the extent to which credit constraints apply between countries. This explanation for intercountry differences in the amount of debt outstanding was investigated by Jappelli and Pagano (1989), and we follow their findings here. Jappelli and Pagano examined the volume of debt as a proportion of consumer expenditure using data from around 1961 to around 1984 for seven countries: Sweden, the U.S., the U.K., Japan, Italy, Spain, and Greece. They used data for consumer credit that was available at that time on a consistent basis for these countries. However, while Jappelli and Pagano explained the rank ordering of debt outstanding in the 1960s, 1970s, and early 1980s, the rank ordering has, according to the figures in Table 13.2, changed. Thus in Jappelli and Pagano's study, the rank ordering was the U.S., the U.K., Japan, Italy, and Spain in descending order in terms of consumer plus mortgage debt outstanding as a percentage of consumption expenditure; in 1992–1996, the rank ordering in terms of short- plus long-term loans as a percentage of GDP was the U.S., Spain, Italy, and Japan (with no figures available for the U.K.). Unfortunately, we know of no other study that tried to rigorously explain intercountry differences in consumer debt. However, the factors that Jappelli and Pagano isolated to explain the ranking in their period may also explain the ranking in 1998; the difference may just be that the conditions that lead to a high ranking in their period now apply to different countries in the 1990s. In any event, we need to be cautious about taking Jappelli and Pagano's reasons as necessarily being applicable today.

First, we might expect that the greater the amount by which the rate of interest on loans exceeds that which lenders have to pay to gain funds (the interest rate wedge), the greater the amount of credit that lenders would wish to supply. However, Jappelli and Pagano found that in mortgage markets, there was no correlation between the wedge and volume of lending. For example, Sweden had a relatively high wedge but relatively low volume of loans, Italy had a high wedge and low volume of loans, and the U.S. had both a high wedge and a high loan loan volume.

Second, Jappelli and Pagano looked at the average percentage of house prices paid as down payment and at the percentage of home owners in younger age groups. They found that, generally speaking, in those countries where the down payments are a small percentage of a house price, home ownership occurs earlier. Thus they found that rationing in the market for mortgages is more common in Japan, Italy, and Spain than in the U.K., the U.S., and Sweden and that this, rather than differences in supply, may contribute to intercountry differences in mortgage debt outstanding.

Finally, they considered differences in demand. One factor that might explain intercountry differences in demand is differences in tax rates, with those countries offering greater tax deductions from interest payments having a greater demand, everything else constant. But this effect will depend on the marginal rate of tax in each country since that is the rate that borrowers would avoid paying. Jappelli and Pagano document the tax incentives to borrow in each sample country and consider a proxy for the marginal tax rate. They conclude that in five of the seven countries, there were no incentives to take consumer loans. However, the greater incentives in Sweden and the U.S. may partially explain the greater debt-to-consumption ratios observed there. In the case of mortgage loans, differences in tax regimens were also too small to explain differences in debt taken.

A second factor which may explain differences in demand is differences between countries in the age distribution and earnings profiles of their populations. To assess the effect of these, for each country, Jappelli and Pagano took the life cycle theory of consumption (see Chapter 3) to predict consumption at each age and the observed earnings profile to simulate the volume of debt-to-consumption ratio. This was then compared with the observed ratios. They concluded that there was little relation between the predicted and actual rankings of debt to consumption. For example, debt to consumption was predicted to be highest in Japan and lowest in the U.K., whereas in Japan the observed ratio was low and in the U.K. high.

Finally, preferences for debt may differ between countries. Unfortunately, in the absence of direct measures of tastes, Jappelli and Pagano can only consider differences in the proportion of consumers' expenditure on durables as indicative of tastes and they find no correlation between this and the debt-to-consumption ratio. Of course, this variable may be a very inaccurate indicator of tastes.

Overall, Jappelli and Pagano conclude that differences in the amount of credit rationing, especially in the markets for mortgages, explains intercountry differences in debt-to-consumption ratios.

13.2.2 Credit cards

Table 13.3 shows the number of credit and charge cards per head issued in Europe in 1997, where the numbers of cards were reported by Datamonitor (1998). The figures suggest that it is possible to distinguish between four groups. The very high users are the U.S. and the U.K. with 1,616 and 661 cards per 1000 people in the population, respectively. Belgium, the Netherlands, Spain, Sweden, and Germany have around 200 cards per 1000 of the population, and finally Italy and France have around 100 cards per 1000 of the population. The numbers of superpremium and premium cards, of corporate cards, and of standard cards per head are in the same rank order as the totals, with one or two exceptions. The exceptions are the very high number of premium and corporate cards per head in Sweden and the relatively high number of premium cards in Germany.

Table 13.3. *European credit and charge cards,* 1997 (*per* 1000 *people in population*).

Country	Superpremium + premium	Corporate	Standard	Total
U.S.	650.4	20.9	945.0	1616.3
U.K.	91.3	22.5	546.7	660.5
Belgium	53.0	6.9	197.4	257.3
Netherlands	38.3	9.4	195.9	243.5
Spain	26.5	4.3	212.0	242.8
Sweden	44.2	46.4	85.8	176.4
Germany	39.7	4.6	127.8	172.0
Italy	18.2	9.7	109.1	137.0
France	25.1	3.1	68.3	96.6

Fianco (1998) discussed some possible reasons for these differences given in the table. She argued that in the U.S., recent intense competition between card issuers has resulted in the acceptance of more risky applicants who previously would not have been accepted. In addition, extensive direct marketing mailshots and preapprovals were issued and interest rates lowered. However, delinquency rates on premium cards increased commensurately. The U.K. followed a pattern similar to that of the U.S., and U.S. card issuers who entered the U.K. market have taken large shares of the premium and superpremium markets.

Fianco also points out some interesting differences in the relative significance of fee and interest income in the premium and corporate sectors. Interest makes up 60% and 80% of income in the U.K. and the U.S., respectively, while in the rest of Europe, fee income makes up around 90% of income. The difference is caused by the greater use of revolving credit in the U.S. and the U.K. and the greater use of corporate cards as charge cards in Europe.

13.3 International differences in credit bureau reports

Details of the information contained in U.K. credit bureau reports are given in Chapter 8. In this chapter, we consider information in bureau reports from other countries.

13.3.1 The U.S.

In the U.S., there are more than 1000 credit bureaus, but almost all are either owned or under contract to one of three main credit reporting companies: Experian, TransUnion, and Equifax. Information contained in a U.S. credit report is similar to that in a U.K. report except that much more information about the characteristics of accounts opened by the person is recorded. Again, the precise items included vary between agencies, but a typical report would include the following information:

- Personal information. Name, address, former address, date of birth, name of current and former employers; Social Security number.

- Public record information. Any details of court judgments against the person, the amount, the balance, the plaintiff, and the current status; details of any declarations of bankruptcy and of tax liens.

- Credit accounts history. For each account, the account number and lender; the date it was opened; the type of account; credit limit; payment in dollars; whether the payments are up to date; the balance outstanding; and how many times the regular payment has been 30 days, 60 days, or 90 days overdue.

- Inquiries. The names of credit grantors that have asked for the credit report and the dates on which they did so.

Most adverse information about an account remains on a report for up to seven years from the date of the first delinquency on the account. The same applies to civil judgments, discharged or dismissed Chapter 13 bankruptcies, and tax liens from the date of payment. Chapters 7, 11, and 12 remain on the report for 10 years. Information on searches of your report remain for one to two years depending on the type of organization that carried out the search.

No credit reference reports in either country contain information on religion, race, or national origin. Credit reports do not include information on checking accounts or savings accounts.

In the U.S., access to a person's credit report is governed by the Federal Fair Credit Reporting Act of 1996, as are a person's rights if he is denied credit. In the U.K., access is governed by the Data Protection Act of 1998, and a person's rights are governed by the Consumer Credit (Credit Reference Agency Regulations) Act of 1999.

13.3.2 Other countries

The content and even availability of credit bureau reports varies greatly between countries. Jappelli and Pagano (2000) carried out a survey to establish the availability of bureau information among 45 countries. Here we briefly summarize some of their results for countries in the European Union (excluding Luxembourg). Jappelli and Pagano found that credit bureau data were available in all countries except France and Greece. In all countries, information on debts outstanding and on defaults was available except for Belgium, Denmark, Finland, and Spain, where only black data were available. They also found that public debt registers are available in all countries except Denmark, Finland, Ireland, the Netherlands, Sweden, and the U.K., although the amount of information varies between countries. Generally, if there is a public credit register, it contains information on defaults, arrears, total loans outstanding, and guarantees.

One would expect that in countries where there is no credit bureau data, then everything else being equal, the statistical discriminatory power would be rather lower than in countries where more information exists. However, there are no empirical studies of this issue.

13.4 Choice of payment vehicle

A small number of papers attempted to predict the method of payment most frequently used for different types of goods and services. Most of these papers relate to consumers in the U.S., although a few relate to consumers in other countries. These studies are difficult to compare because the alternative forms of payment that they analyze differ, as do the goods and services considered and the variables used to predict the customers' chosen methods. It may be that these differences, rather than the differences between the countries in which the consumers studied reside, explain any differences in payment method. This is a further opportunity for research.

The most recent study for the U.S., by Carow and Staten (1999), considers the choice between gas credit cards, cash, general-purpose credit cards, and debt as payment for gas. They found that consumers are more likely to use cash when their income is low, the purchaser is middle-aged, they have relatively little education, and the purchaser has few cards. Other credit and debit cards are more likely to be used when the purchaser is relatively young, is more educated, and holds more cards. Lindley et al. (1989), who considered purchases in 1983, also found that purchases of gas were more likely to be made using a credit card than by cash or checks when the buyer's income was higher. They also found the same result for purchases of furniture, household goods, and clothing, but they did not detect an influence of the buyer's age, nor for years of education, except in the case of clothing. For the purchase of household goods and of clothing, the chance of using a credit card was lower if the buyer owned his own house.

An interesting set of results relating perceived attributes to various characteristics of different forms of payment in 1980 was gained by Hirschman (1982) in his study of U.S. consumers. Hirschman attempted to predict the relative frequency of use of five methods of payment—cards, store cards, travel and entertainment cards, cash, and checks—for different types of goods and services. He found that, for example, the use of bank cards was associated with the perceptions of bank cards' high security, high prestige, speedy transfer time, higher acceptability, lower transaction time, the possession of a transaction record, and the greater ease of facilitating purchases by gaining debt.

For the U.K., Crook, Hamilton, and Thomas (1992b) used a sample of bank credit card holders to distinguish between those who, over a fixed period in the late 1980s, used the card and those who did not. They found that those 30 to 40 years old were more likely to use the card than other age groups and that those of higher income were more likely to use their cards, as were those who had had an account at the issuing bank for four to five years and those who had lived at the same address for under six months or for very many years. Of all the possible categories of residential status, those least likely to use their card were tenants in unfurnished accommodation.

The only published study for Australia is by Volker (1983), who distinguished between those who did not hold a credit card and those who both held a card and used it. The probability of possession and use of credit cards was found to be lower if the person was aged 16 to 19 years, 20 to 24 years, 45 to 54 years, or over 55 years. The probability was also lower for the skilled and unskilled occupational groups than for other occupational groups and for females than for males.

13.5 Differences in scorecards

In section 12.2, we considered generic scorecards versus those specific to a particular credit vendor. In this section, we consider a related issue: differences in scorecards between countries. Little is known about international differences in scorecards. However, with an increasing number of European countries adopting a common currency (the Euro), one might expect to see increasing numbers of applications for loans from overseas residents who are still within the same currency region. If both the lender and the borrower use the same currency, then fluctuations in exchange rates will have no effect on payments. Thus it would be in the interest of lenders for them to advertise for borrowers in other countries, especially if the lender can offer lower interest rates than lenders in the applicant's home country.

This raises the question of whether the same scorecard can be applied to applicants from foreign countries as to home applicants. The only publicly available study of this question is by Platts and Howe (1997), who considered data relating to applicants in the U.K., Germany, Greece, Belgium, and Italy. They found that different variables are collected in different countries, and initially they compared the discriminating power of each variable taken on its own by considering the weights of evidence for each variable in each country. They discovered that the same variable had very different predictive power in different countries. For example, not having a home telephone was very much more predictive of default in the U.K. than in any of the other countries; having a bad credit bureau report was more predictive of default in Italy, Germany, and the U.K. than in Greece; and being a tenant was more predictive in the U.K. than in any of the other countries.

Platts and Howe then built three types of scorecard. First, a separate scorecard for each of the five countries was built using, for each country, data only from that country and using variables that discriminated best for that country: the specific scorecard. The second and third scorecards used the same attributes (variables). The second consisted of a separate scorecard for each country but used data only for the home country of the applicant, thus allowing the scores from each attribute to differ between the countries and for different weights of evidence: the country scorecard. The third used data pooled across all applicants from the five countries: the European scorecard. Platts and Howe then compared the percentage improvement in bad debt when using each scorecard compared with that currently in use when the cutoff scores were chosen to maintain the same acceptance rate. Their results are shown in Figure 13.1.

Figure 13.1 shows that for all five countries, the country-specific scorecard performs better than the country scorecard, which in turn outperforms the European scorecard. In most cases, all the scorecards performed better than existing systems. The differences were smallest for Belgium but largest for the U.K. One possible explanation for the relatively poor performance of the European card in the U.K. is that in the U.K. there is a relatively large amount of credit bureau data. The greatest improvement in performance of the country scorecard over the European scorecard (both using the same attributes) was again in the U.K., with a 12% improvement in bad debt. The benefits to the other countries were very similar at between 4% and 6%. Platts and Howe concluded that these differences were sufficiently large that it was worth the additional cost of building country-specific cards rather than using one single generic card.

13.6 Bankruptcy

If an individual fails to keep up payments of loans, creditors typically pursue debt collection activities. In the U.S., creditors can, subject to federal regulations, garnish the debtor's

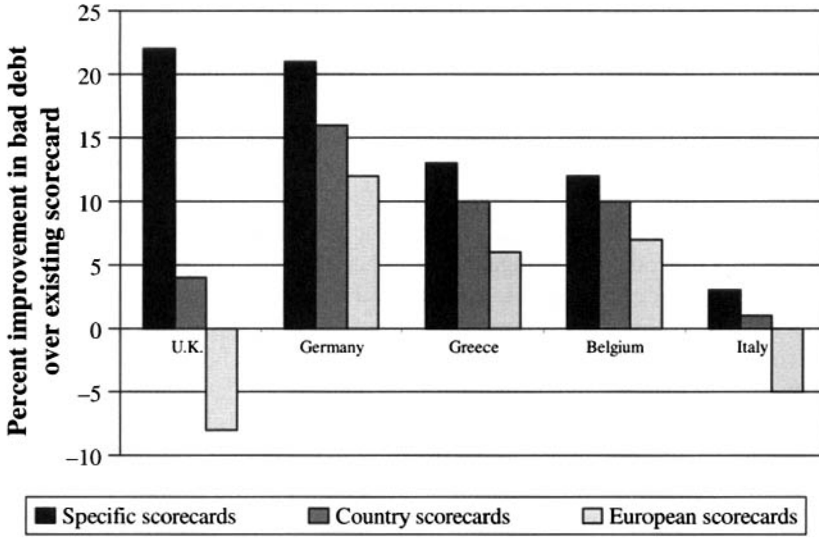


Figure 13.1. *Comparison of European, country, and specific scorecards.*

wages and take actions to repossess the debtor's property. Alternatively, debtors may declare themselves bankrupt. Unlike many other countries, the U.S. allows individuals to choose which of two types of bankruptcy to declare (under the Bankruptcy Reform Act of 1978): Chapter 7 procedures and Chapter 13 procedures.

Under a Chapter 7 filing, a debtor surrenders all of his nonexempt assets to the court. A plan is made to sell the assets and the proceeds to be distributed to creditors. Secured debt would be repaid first, for example, loans where houses or cars were offered as collateral. Unsecured loans would be repaid last. Unsecured include debt where the value of the collateral was less than the loan outstanding and also, usually, debt outstanding on credit cards and installment credit. Once the plan is agreed, the debtor is discharged from the requirement to repay almost all his debts. No further garnishment of wages or repossession of assets is allowed. The debts that are not waived include recent income tax and child support. Note also that if property is collateral for a loan, an exemption cannot prevent the creditor repossessing the property. The nature and value of exempt assets are stated in federal law but also vary considerably between states (see White 1998).

Under a Chapter 13 filing, debtors with regular income offer a plan to reschedule the payments for most of their debt over three to five years. One exception is mortgages, which cannot be rescheduled, although debtors can pay off mortgage arrears to prevent foreclosure. The debtor keeps all of his assets, and creditors are restrained from their efforts to collect debt. Often, debtors who file under Chapter 13 do not have to pay some part of their debt, although they must plan to repay as much unsecured debt as they would under Chapter 7. Chapter 13 plans require the approval only of the judge, not of the creditors, to be implemented.

There has been discussion as to how individuals might—and do—respond to these procedures. White (1998) considered several possible strategies an individual might adopt if he files under Chapter 7. The person may borrow relatively large amounts of unsecured debt (for example, installment credit or credit card debt) and either use it to buy consumption goods (such as holidays) and services that cannot be liquidated (or, if they can be sold, only for a small faction of their original cost) or, alternatively, use it to pay off nondischargeable debt. Debtors may then choose to declare themselves bankrupt. As a result, this recently acquired debt and older unsecured debt will all be written off by lenders. Indeed, Berkowitz and Hynes (1999) argue that "many debtors" file for bankruptcy so that they can discharge their nonmortgage debt to reduce their payments on these so as to have sufficient funds to keep repaying their mortgage. As evidence, Berkowitz and Hynes cite Sullivan, Warren, and Westerbrook (1989), who found that in a sample of 1529 bankruptcy cases in 1981, 40% of debtors had secured debt and 10% owned their own homes but did not declare a mortgage debt.

A second strategy White considered is for individuals to liquidate assets that are not exempt from seizure by debtors and use the proceeds to reduce the mortgage on their main home, which in many states under state law is at least partially exempt from seizure. This would enable the individual to retain a greater amount of assets after bankruptcy. A third strategy White considered exists if, having adopted the second strategy, the individual's mortgage is completely repaid but he still has some nonexempt assets. In this case, the individual would be able to retain even more assets after bankruptcy if he uses the proceeds to improve his residence or move to a more valuable residence.

Using data from the Survey of Consumer Finance (1992), White estimated the percentage of households in the U.S. as a whole and in eleven selected states who would benefit from these three strategies rather than declaring bankruptcy and not following any of these strategies. Her findings for the U.S. as a whole are that by not following one of the three strategies, 17% of households would benefit, whereas the percentage who would benefit from pursuing each of the strategies is 24% in each case.

The recent empirical evidence that examined the choice of chapters an individual makes is not so clear-cut. For example, Domowitz and Sartain (1999) found that the choice of chapter depended more on location, whether the person was married, employment, auto equity, home equity, and personal assets than on home exemption relative to debt, other exemptions relative to debt, the volume of credit card debt, or different types of secured or unsecured debt.

This page intentionally left blank

Chapter 14

Profit Scoring, Risk-Based Pricing, and Securitization

14.1 Introduction

Thus far, with a few exceptions, credit and behavioral scoring have been used to estimate the risk of a consumer defaulting on a loan. This information is used in the decision on whether to grant the loan (credit scoring) or extend the credit limit (behavioral scoring). In this chapter, we discuss how scoring can be used to make more sophisticated financial decisions, so we extend some of the topics outlined in Chapter 9. In section 14.2, we discuss how the accept-reject decision can be made explicitly in terms of maximizing the profit, even with a default-based credit score. We point out that inconsistencies in the scoring system can then lead to more complicated decisions than just one cutoff score. Moreover, the objective might be not just profit but could also involve the chance of losses being not too great and that the volume of loans accepted meets the lender's requirements. Section 14.3 describes a profit-based measure of scorecard accuracy that allows comparison of scorecards or cutoff levels purely in profit terms.

Section 14.4 presents a brief review of the ways one could try to build a profit-scoring system as opposed to a default-risk system. Although many organizations seek to develop such scorecards, it is fair to say that no one has yet perfected an approach to this problem. If profit is one's objective, then of course the lender's decision is not just, "Shall I accept the consumer or not?" but also, "What interest rate should I charge?" This leads to the idea of risk-based pricing, which has been very slow to come in credit cards or other forms of consumer lending. The reasons behind this hesitancy are discussed in section 14.5, which contains a simple model of risk-based pricing where the score can be used to set the interest rate.

Another area where the score could be used as a measure of general risk rather than just the default risk of the individual consumer is securitization. Groups of loans are packaged together, underwritten, and distributed to investors in the form of securities. The price at which the package of loans is sold must reflect the risk of nonrepayment, i.e., should be related to the credit scores of the individual loans in the package. In practice this link is not very transparent, and section 14.6 discusses the issue of securitization. The one area where securitization has been an overwhelming success is mortgage-backed securities in the U.S. This is the second-biggest bond market in the U.S. after Treasury bonds, so in the last section we look at how this market functions and why it has proved so much more successful in the U.S. than anywhere else.

14.2 Profit-maximizing decisions and default-based scores

Recall the model of the credit-granting process developed in section 4.2. The application characteristics $\mathbf{x}$ give rise to a score $s(\mathbf{x})$. (We drop the x dependency of the score at times for ease of notation.) p_G and p_B are the proportion of goods and bads in the population as a whole and so $o_p = \frac{p_G}{p_B}$ is the population odds. $q(G|s)$ ($q(B|s)$) is the conditional probability that a consumer with score s will be good (bad) and $q(G|s) + q(B|s) = 1$. Let $p(s)$ be the proportion of the population that has score s. Assume that the profit from a consumer is a random variable R, where

$$R = \begin{cases} 0 & \text{if the account is rejected,} \\ L & \text{if the account is accepted and becomes good (L is lost profit from ruling out} \\ & \text{a good),} \\ -D & \text{if the account is accepted and becomes bad (so D is the default amount).} \end{cases} \tag{14.1}$$

The expected profit per consumer if one accepts those with score s is

$$E\{R|s\} = Lq(G|s) - D(1 - q(G|s)) = (L + D)q(G|s) - D. \tag{14.2}$$

Thus to maximize profit, one should accept those with scores s if $q(G|s) \geq \frac{D}{D+L}$. Let A be the set of scores where the inequality holds; then the expected profit per consumer from the whole population is

$$E^*\{R\} = \sum_{s \in A} ((L + D)q(G|s) - D)p(s). \tag{14.3}$$

This analysis ignored fixed costs. As section 9.2 suggests, fixed costs encourage the lender to set lower cutoff scores than do those that maximize profits.

Notice that if the profits and default losses $L(s)$ and $D(s)$ were score dependent, the decision rule would be to accept if $q(G|s) \geq \frac{D(s)}{D(s)+L(s)}$. In that case, A may consist of several regions of scores, i.e., accept if the score is between 300 and 400, accept if between 500 and 550, and accept if over 750; reject otherwise. However, we assume that the profits and losses are independent of the score and that $q(G|s)$ is monotonically increasing in s. This means that $A = \{s|s \geq c\}$, where $q(G|s) \geq \frac{D}{D+L}$, so c is the cutoff score. In this case, define $F(s|G)$, $F(s|B)$ to be the probabilities a good or a bad has a score less than s:

$$\begin{aligned} E^*\{R\} &= \sum_{s \geq c} ((L + D)q(G|s) - D)p(s) = \sum_{s \geq c} (Lp_G p(s|G) - Dp_B p(s|B)) \\ &= Lp_G(1 - F(c|G)) - Dp_B(1 - F(c|B)) \\ &= Lp_G - Dp_B + (Dp_B F(c|B) - Lp_G F(c|G)). \end{aligned} \tag{14.4}$$

The first term on the right-hand side, $Lp_G - Dp_B$, is the profit if we accept everyone, so the second term is the profit that the scorecard brings. One can rewrite (14.4) in a different way to say how far away $E^*\{R\}$ is from the expected profit if there were perfect information $E\{PI\}$. With perfect information, one would only accept goods, so $E\{PI\} = Lp_G$. Hence

$$\begin{aligned} E^*\{R\} &= Lp_G - (Dp_B(1 - F(c|B)) + Lp_G F(c|G)) \\ &= E\{PI\} - (Dp_B(1 - F(c|B)) + Lp_G F(c|G)). \end{aligned} \tag{14.5}$$

This analysis follows the approach of Oliver (1993) very closely, and one of the points he makes is that terms like $q(G|s)$ and $F(c|G)$ have two interpretations. They can be the forecasted probabilities and they can be the actual proportions in the portfolio. Thus (14.4) and (14.5) give both the forecast performance profit and the actual performance profit, but these are two very different quantities although they have the same expression. It brings us back to some of the questions discussed in Chapter 7 on measuring the performance of a scorecard. If one thinks of a scorecard as a forecast, then the questions are how discriminating it is (how far from 0.5 $q(G|s)$ is) and how well calibrated it is (how different the forecast $q(G|s)$ is from the actual $q(G|s)$). Usually, the most discriminating forecasts are not well calibrated, and the well-calibrated forecasts are not usually the most discriminating.

In a subsequent paper, Oliver and Wells (2001) developed these models further to describe the different trade-offs that credit granters make in choosing their portfolios of loans by setting a cutoff. Let $F(s)$ be the proportion of scores below s; i.e., $F(s) = F(s|G)p_G + F(s|B)p_B$. Many lenders look at the trade-off of the bad acceptance rate (the percentage of the total bad population accepted) against the acceptance rate, i.e., $(1 - F(s|B))p_B$ against $1 - F(s)$. The actual bad rate, which is the percentage of those accepted who are bad, is the ratio of these two numbers, so

$$\text{Actual bad rate} = \frac{(1 - F(s|B))p_B}{1 - F(s)}. \tag{14.6}$$

This is called the strategy curve, and Figure 14.1 shows the strategy curve for a typical scorecard. The bold line shows the strategy curve for the scorecard with perfect information and hence perfect discrimination. Clearly, the nearer the scorecard strategy curve gets to that, the better it is.

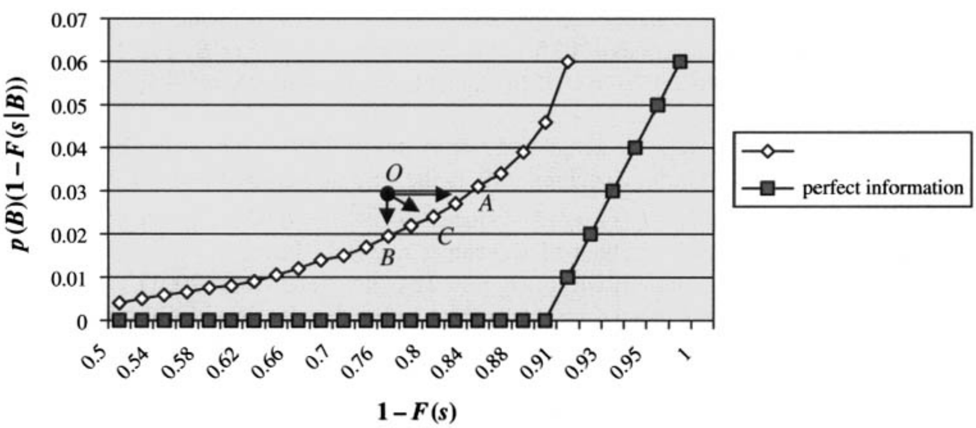


Figure 14.1. *Strategy curve.*

What often happens on introducing a new scorecard is that the existing operating policy gives a point O that is above the new strategy curve. The question then is where on the strategy curve one wants to go by choosing the appropriate cutoff. If one moves to A, then one keeps the bad acceptance rate the same but accepts more people, while moving to B would accept the same numbers but lower the bad acceptance rate and hence the bad rate. Moving to C would keep the bad rate the same and again increase the numbers accepted. The following example, based on Oliver and Wells (2001), investigates this.

Example 14.1. A bank's current policy O has an acceptance rate of 72%, a bad rate of 2.6%, and the population odds ($o = \frac{p_G}{p_B}$) of 10.3 to 1. The data for a new scorecard are as given in Table 14.1. It follows from the existing statistics that $p_G = \frac{o}{1+o} = \frac{10.3}{11.3} = 0.912$ and so $p_B = 0.088$. Since $1 - F(s) = 0.72$ (the acceptance rate), from (14.6) the bad acceptance rate

$$p_B(1 - F(s|B) = (\text{bad rate}) \cdot (\text{acceptance rate}) = 0.026 \cdot 0.72 = 0.0187.$$

Hence $1 - F(s|B) = \frac{0.0187}{0.088} = 0.213$. Since $1 - F(s) = (1 - F(s|G))p_G + (1 - F(s|B))p_B$, we can write

$$1 - F(s|G) = \frac{(1 - F(s)) - p_B(1 - F(s|B))}{p_G} = \frac{0.72 - 0.0187}{0.912} = 0.769.$$

Table 14.1. *Strategy curve details for new scorecard in Example* 14.1.

Marginal odds	7.2:1	7.8:1	8.3:1	8.9:1	9.6:1	11.0:1	12.5:1	15.1:1	20.6:1	30.2:1
$1 - F(s\|G)$	0.848	0.841	0.833	0.827	0.818	0.803	0.776	0.746	0.693	0.631
$1 - F(s\|B)$	0.233	0.218	0.213	0.193	0.181	0.166	0.144	0.126	0.097	0.069
$1 - F(s)$	0.793	0.786	0.778	0.771	0.762	0.747	0.720	0.691	0.640	0.581

We can use this table to calculate the statistics for the new possible points A, B, and C. At point A, the bad acceptance rate is maintained at 1.87%, so $1 - F(s|B) = \frac{0.0187}{0.088} = 0.213$. This is the point with marginal odds 8.3:1 in the table and corresponds to an acceptance rate of 77.8% and hence an actual bad rate of $\frac{0.0187}{0.778} = 0.024$, i.e., 2.4%.

At B, we keep the acceptance rate the same, so $1 - F(s) = 0.72$, which corresponds to the point with marginal cutoff odds of 12.5:1. Since for this point $1 - F(s|B) = 0.144$, the bad acceptance rate is $(0.144)(0.088) = 0.012672$, and hence the bad rate is $\frac{0.012672}{0.72} = 0.0176$, i.e., 1.76%.

Finally, if at C we want the actual bad rate to stay at 2.6%, we need to find the point where $\frac{0.088(1-F(s|B))}{1-F(s)} = 0.026$. Checking, this is the case at marginal odds of 7.2:1 when $1 - F(s|B) = 0.232$ and $1 - F(s) = 0.794$ since $\frac{0.088(0.232)}{0.794} = 0.0257$. Hence at this point the acceptance rate is 79.4% and the bad acceptance rate is 2.04%.

An alternative to the strategy curve, suggested by Oliver and Wells (2001), is to plot the expected losses against the expected acceptance rate, but all that this does is multiply the scale of the Y-axis in Figure 14.1 by D, the cost of a default. More interesting is to plot the expected losses against the expected profit, where the profit is $L(1 - F(s)) - (L + D)F(s|B)p_B$, so the new X-axis is at an angle to the original one in Figure 14.1. This leads to Figure 14.2.

This is intriguing in that one can get the same expected profit in at least two ways. The points on the lower part of the curve have higher cutoff scores, so less bads are accepted; those on the higher part of the curve correspond to lower cutoff scores with higher numbers of bads accepted. The efficient frontier of this curve is the lower part from C to D. These give points with an expected profit and an expected loss, in which the former cannot be raised without the latter also being raised.

If a lender is at present operating at point O, then again he can move onto the new scorecard curve either by keeping the bad acceptance rate the same, i.e., A, or by keeping the acceptance rate the same (which would move to point B on the curve). In this case, one would suggest that the move to A is less sensible because it is not an efficient point and one could have the same expected profit with lower expected losses.

One can obtain the efficient frontier in a more formal way as follows. One is seeking to minimize the expected loss with lower bound P^* on the expected profit:

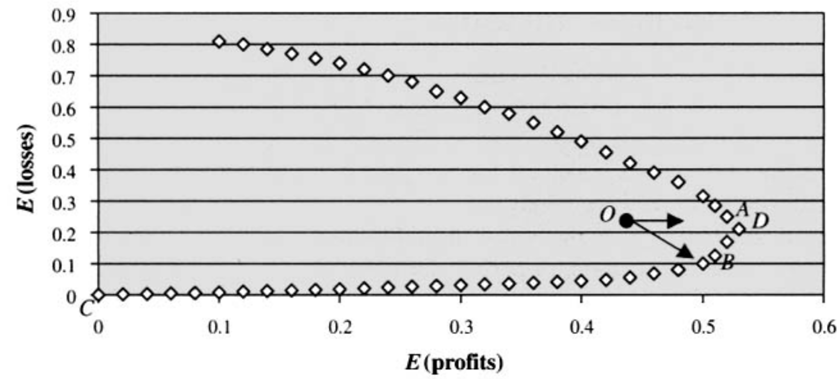


Figure 14.2. *Expected losses against expected profits.*

$$\begin{array}{ll}\text{Min}_s & Dp_B(1-F(s|B)) \\ \text{subject to} & Lp_G(1-F(s|G)) - Dp_B(1-F(s|B)) \geq P^*.\end{array} \tag{14.7}$$

One can solve this using the Kuhn–Tucker optimality conditions for nonlinear optimization, i.e., the derivative of (objective + λ(constraint)) = 0 at a minimum, to get the condition

$$\begin{array}{r}Dp_B f(s|B) - \lambda(Lp_G(1-f(s|G)) - Dp_B(1-f(s|B))) = 0 \quad \text{and} \\ \lambda(Lp_G(1-F(s|G)) - Dp_B(1-F(s|B) - P^*)) = 0, \quad \lambda \geq 0,\end{array} \tag{14.8}$$

where $f(s|B)$ is the derivative of $F(s|B)$ with respect to s. Then the shadow price λ satisfies

$$\lambda = \frac{Dp_G f(s^*|B)}{Lp_G f(s^*|G) - Dp_B f(s^*|B)} = \frac{1}{\frac{o^*}{\bar{o}} - 1} > 0, \tag{14.9}$$

where $o^* = \frac{p_G f(s^*|G)}{p_B f(s^*|B)}$ is the marginal odds at the cutoff score and $\bar{o} = \frac{D}{L}$ is the optimal odds for the unconstrained problem. Notice that this implies that $o^* > \bar{o}$—the odds for the constrained problem should be higher than those for the unconstrained problem.

Some lenders add another constraint, namely, that there should be a lower bound on the numbers accepted. This leads to the following extension of (14.7), where the total profit is being maximized, N is the number in the population, and N_0 is the minimum number that must be accepted:

$$\begin{array}{ll}\text{Min}_s & DNp_B(1-F(s|B)) \\ \text{subject to} & N(Lp_G(1-F(s|G)) - Dp_B(1-F(s|B))) \geq P^*, \quad N(1-F(s)) \geq N_0.\end{array} \tag{14.10}$$

This splits into three cases depending on what happens to the multipliers λ and μ of the profit constraint and the size constraint at the solution point. If the profit constraint is exactly satisfied and the size constraint is a strict inequality ($\mu = 0$), then the result is similar to (14.9). If the profit constraint is a strict inequality ($\lambda = 0$) and the size constraint is an equality, then the number N_0 accepted by this constraint gives the cutoff limit by $1 - F(c) = \frac{N_0}{N}$. Third, if N_0 is very large, then there will be no feasible cutoff that meets both constraints.

14.3 Holistic profit measure

The criteria described in Chapter 7 to measure the performance of scoring systems were all discrimination or risk based. Hoadley and Oliver (1998) followed the profit approach of the previous section to define a profit measure. The clever thing about this measure is that it is independent of D and L, which are the two parameters of the profit calculations that are always difficult to estimate accurately.

Recall that (14.3) gives the expected profit under a scorecard with cutoff c, and the profit under perfect information (or perfect discrimination) is Lp_G. Then the holistic profit is defined as

$$\text{HP} = \frac{\text{expected profit under scorecard with cutoff } c}{\text{expected profit under perfect discrimination}} = 1 - F(c|G) - \frac{Dp_B}{Lp_G}(1 - F(c|B). \tag{14.11}$$

If instead of defining the cutoff c we define the marginal odds at the cutoff, $o^* = \frac{q(G|s)}{q(B|s)} = \frac{D}{L}$ and the population odds $o_0 = \frac{p_G}{p_B}$. Then we can rewrite (14.11) as the holistic profit scan $h(o^*)$, which is the holistic profit in terms of the marginal cutoff odds o^* by

$$h(o^*) = 1 - F(s(o^*)|G)| \left(\frac{o^*}{o_0}\right)(1 - F(s(o^*)|B)). \tag{14.12}$$

Using the data of Table 14.1, we get the holistic profit scan in Figure 14.3.

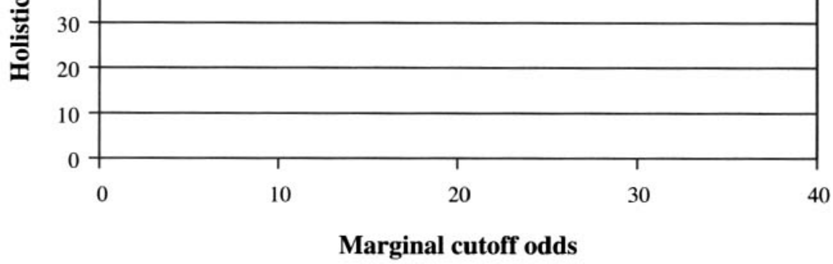


Figure 14.3. *Holistic profit scans.*

One has to be careful not to read too much into this graph. It is not the case that the best cutoff is where the curve is a maximum. It depends on the values of L and D. What the graph does is give a view of how the scorecard will perform compared with the perfect discriminating scorecard as the ratio $\frac{D}{L} = o^*$ varies.

One can also use holistic profit functions to assess how sensitive this profit is to choice of suboptimal cutoffs. Suppose that $o^* = \frac{D}{L}$; then one can draw the holistic profit function as

This is a basic solution; it contains only m positive components. The matrix in the preliminary problem (8) has the $n + m$ columns $a^1, \ldots, a^n, e^1, \ldots, e^m$; the first n columns come from A, and the last m columns come from the $m \times m$ identity matrix. Using composite matrices and vectors, we could write (8) in the form

$$\begin{gathered}[A \quad I]\begin{bmatrix} x \\ z \end{bmatrix} = b \\ x \geqslant 0, \quad z \geqslant 0, \quad \sum z_i = \text{min.}\end{gathered} \tag{8'}$$

Our basic feasible solution (9) depends on the m independent columns $e^1, \ldots, e^m$. The cost of this solution is $\sum z_i = \sum b_i > 0$.

Now carry out a Phase II calculation on the preliminary minimum problem (8). When you're done, there are two possibilities:

Case 1: $\min \sum z_i = 0$. In this case the final z equals zero, and so the basic optimal solution of the preliminary problem satisfies

$$Ax = b, \quad x \geqslant 0. \tag{10}$$

Then x is a basic feasible solution of the original problem.

Case 2: $\min \sum z_i > 0$. In this case the original has no feasible solution. For if x were a feasible solution of the original problem, then x along with $z = 0$ would give a *zero*-cost solution of the preliminary problem (8).

EXAMPLE 2. Suppose the original feasibility problem is

$$-x_1 - 2x_2 = 3, \quad x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{11}$$

Then the preliminary minimum problem is

$$\begin{gathered} -x_1 - 2x_2 + z_1 = 3 \\ x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad z_1 \geqslant 0 \\ z_1 = \text{minimum.} \end{gathered} \tag{12}$$

The unique solution of this problem is $x_1 = 0$, $x_2 = 0$, $z_1 = 3$. Since the preliminary minimum cost is positive, the original problem (11) has no feasible solution.

EXAMPLE 3. Suppose the original feasibility problem is

$$x_1 + 2x_2 = 3, \quad x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{13}$$

Then the preliminary minimum problem is

$$\begin{gathered} x_1 + 2x_2 + z_1 = 3 \\ x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad z_1 \geqslant 0 \\ z_1 = \text{minimum.} \end{gathered} \tag{14}$$

The first feasible solution is $x_1 = 0$, $x_2 = 0$, $z_1 = 3$; a basic optimal solution is $x_1 = 3$, $x_2 = 0$, $z_3 = 0$. This gives a feasible solution for (13).

suggest that profit scoring requires a fully integrated information system in the organization. One needs the information on all the customers' transactions (and maybe a whole family's transactions) and accounts collated together to calculate the customers' profitability to the firm; hence the push to data warehousing by companies so that all this information is kept and is easily accessible. This could lead to legal problems because the use of personal information for reasons other than those for which it was originally collected is frowned on by legislators in many countries.

The advent of data-mining techniques—see (Jost 1998) for its uses in credit scoring—means that the technical problems of analyzing such vast amounts of data are being addressed. However, there are still major problems in developing models for profit scoring. Over what time horizon should profit be considered, given that we do not wish to alienate customers by charging high prices now and then losing future profit as they move to another lender? Also, profit is a function of economic conditions as well as the individual consumer's characteristics. (See Crook et al. (1992a) for the effect economic conditions have between one year and the next.) Thus it is even more important to include economic variables into profit scoring than it was in credit scoring. Profit is dependent on how long a customer stays with a lender, and so one wants to know how long customers stay and whether they default or just move their custom elsewhere. Thus one needs to estimate attrition rates as part of profit scoring.

Last, two difficulties affect which methodology to choose. Should one look at the profit on each product in isolation or look at the total profit over all possible products? The former means that one could decide not to offer a customer a new credit card because he does not use it enough. This refusal may offend a customer so much that his profitable home loan is moved to another lender. Going for total profit, on the other hand, ignores the fact that the decision on which products a customer takes is the customer's decision. He can cherry-pick as well and may refuse the product where the firm felt it would make the most profit from him. Second, there is the problem of censored data. In a sample of past transactions, the total profit for current customers will not be known but only the profit up to the date that the sample history finished.

So what approaches are being tried? They can be classified into four groups. One approach is to build on the existing scorecards, which estimate default, usage, acceptance, and attrition, and try to define the profit for groups of the population segmented according to their scores under these different measures. Oliver (1993) was one of the first to suggest this and looked at what decision rules should be used if one has a transaction profit score and a default score. Fishelson-Holstine (1998) described a case study where one tried to segment according to two types of profit. A bank runs a private-label credit card for a retailer. The retailer wants to increase profits by sending cards to people who will use it to buy more in their stores, while the bank wants the credit card operations of the customer to be profitable. By using a demographically based segmentation tool, details of the retailers sales, and the credit card transaction database, groups were identified who were profitable for both. The idea of estimating intermediate variables, outlined in section 12.5, is an obvious step toward estimating profit. One should try to estimate intermediate variables like balance outstanding and purchases and use these to estimate the final outcome. This approach is related to the graphical networks and the Bayesian learning models of section 12.6.

A second approach is to mimic the regression approach of credit scoring by trying to describe profit as a linear function of the categorical application form variables. Almost all the data will be censored in that the total profit is not known, but there is a body of literature on regression with censored data (Buckley and James 1979), and although the censoring that occurs in credit scoring has not yet been dealt with, research in this area is continuing (Lai and Ying 1994).

The third approach is to build on the Markov chain approaches to behavioral scoring outlined in section 6.4 and develop more precise stochastic models of customer behavior. Cyert, Davidson, and Thompson's (1962) original model was used to model profit in a one-product case, and these approaches have proved very successful in estimating debt provisioning for portfolios of customers with the same product. If one extends the ideas to the profit a customer brings when one has to look at the several products he has or may possibly have with the lender, the problems become one of data availability and computational power rather than of modeling difficulty. One runs up against the curse of dimensionality that arises when one seeks to use Markov chains to model complex real situations. However, a number of techniques are proving very successful on other areas to overcome this problem, including aggregation-disaggregation and parallel computing (Thomas 1994).

Fourth, one could recognize that the survival analysis approach of section 12.7 is another way to begin to address profit by estimating how long the loan will run before unexpected—and hence usually unprofitable—events occur. The techniques of survival analysis—proportional hazard models and accelerated life models—could then be used to estimate the long-term profit from a customer given only the experience in the first few months or years. Narain (1992) was the first to suggest that one could use this analysis on credit-scoring data, while the paper by Banasik, Crook, and Thomas (1999) showed that one could also use the idea of competing risks from reliability to get good estimates of when borrowers will default and when they will pay off early, thus incorporating default and attrition in the same analysis. Proportional hazards and accelerated life approaches are useful ways to think about how economic affects can be introduced into profit models. By taking the characteristic variables in proportional hazards and accelerated life to describe the economic conditions as well as the characteristics of the borrower, one can build a model that allows for the speeding up in default rates that occurs in poor economic conditions.

Profit-scoring systems seem more difficult to obtain than might have been first thought, but the prize for success is enormous. A profit-scoring system would provide a decision support aid that has the same focus throughout the different decision-making areas of the organization. It provides an excellent way to take advantage of all the new data on consumer behavior that has become available in the last few years with electronic point-of-sales equipment and loyalty cards.

14.5 Risk-based pricing

One of the real surprises in consumer credit over the 50 years of its history is how long it has taken lenders to start to price the cost of the loan to the risk involved. For most economic goods, the supplier produces a good at a price to clear the market, yet in the credit industry there are usually funds available for consumer credit—it being so much less risky than other forms of lending—and yet there are whole segments of the population who cannot get the credit they want. This is because normally instead of adjusting the price of credit, which is the interest rate they charge, lenders decide on a fixed interest rate (price) and use credit scoring to decide to whom to lend at this price. There have always been different groups of lenders targeting different parts of the credit market: naïvely, one could say gold cards are for the least risky group, normal credit cards for the next group, store cards for a slightlier risky group, and door-to-door collectors (tallymen in Old English) for the group even more at risk. Even so, there are groups who fall between or outside these bands and who find it hard to get the credit they want. Within each product group, the price tends to be fixed across most lenders. For example, it is only in the past decade that one has had subprime lending coming into the credit card market or some financial institutions offering very attractive rates to their own best customers; see Crook (1996) and Jappelli (1990), who identify the characteristics of those unable to gain the credit they wish.

This slowness in trying to relate price to risk is surprising in consumer lending because credit scoring is an ideal technique for setting risk-based prices or interest rates, as the following example shows.

Suppose one has a scoring system with p_G, p_B, $p(s|G)$, $p(s|B)$, $p(s) = p(s|G)p_G + p(s|B)p_B$, $q(G|s) = \frac{p(s|G)}{p(s)}$, $q(B|s) = \frac{p(s|B)}{p(s)}$ defined as in section 14.2. Assume that the interest rate charged, i, is a function of the credit score s, so $i(s)$. Let the cost of default be D, and we assume this is independent of the interest rate but that the profit on a good customer, $L(i)$, does depend monotonically on the interest rate i charged. Then for each score s, one has to decide whether to accept consumers at that score and, if so, what interest rate to charge. One is interested for a score s in maximizing

$$\max_i\{(L(i)q(G|s) - Dq(B|s))a_s(i), 0\}. \tag{14.14}$$

This is clearly maximized when $L(i)$ is maximized, and this silly result shows what is missing in that not all the potential customers with a credit score s will accept a loan or credit card when the interest rate being charges is i. So let $a_s(i)$ be the fraction of those with credit score s who will accept interest rate i, while $q(G|s,i)$, $q(B|s,i)$ are the fraction of acceptors with score s when the interest rate is i who are good or bad. Adverse selection suggests that these conditional probabilities do depend on i because as the interest rate increases, it is only the people who cannot get credit elsewhere who are accepting, and these will increasingly be poorer risks even if their credit score has not changed. Clearly, $a_s(i)$ is decreasing in i, and what we want is to maximize

$$\max_i\{(L(i)q(G|s,i) - Dq(B|s,i))a_s(i), 0\}. \tag{14.15}$$

Differentiating with respect to i and setting the derivative to 0 to find the maximum gives that the optimal interest rate for a score s satisfies

$$\begin{aligned} &-L'(i)q(G|s)a_s(i) + (L(i)q'(G|s,i) - Dq'(B|s,i))a_s(i) \\ &= (L(i)q(G|s) - Dq(B|s))a'_s(i). \end{aligned} \tag{14.16}$$

In a specific case, suppose that $a_s(i) = e^{-\alpha(s)(i-i^*)}$, i.e., everyone accepts an interest rate i^* and the subsequent dropoff is exponential, and that there is no effect of interest rate on the fraction of goods who accept. Let $L(i) = \frac{R}{(1+i)^T} - \frac{R}{(1+i^*)^T}$; i.e., there is one payment of R at time T with the interest charged being i, while the real cost of capital is i^*. Then (14.14) becomes

$$\begin{aligned} &\frac{-R}{(1+i)^{T+1}}q(G|S)e^{-\alpha(s)(i-i*)} = -\left(\frac{R}{(1+i)^T}q(G|s) - Dq(B|s)\right)\alpha(s)e^{-\alpha(s)(i-i*)} \\ \text{or}\quad &\frac{q(G|s)}{(1+i)^T}\left(\alpha(s) + \frac{1}{1+i}\right) = Dq(B|s)\alpha(s). \end{aligned} \tag{14.17}$$

Solving this gives the interest rate to be charged for those with credit score s. One can do similar calculations for all types of loan products.

One reason that risk-based pricing will become more prevalent for individual loans is that the idea of risk-based pricing of portfolios of loans is already with us in the concept of securitization. The last two sections of this chapter look at this issue.

14.6 Securitization

Securitization is the issuing of securities in the finance market backed not like bonds by the expected capacity of a corporation to repay but by the expected cash flows from specific assets. The term *securitization* is usually applied to transactions in which the underlying assets are removed from the original owner's balance sheet, called "off-balance sheet securitization," although in some European countries there is a system of "on-balance sheet securitization." The most widely used assets in securitization are residential mortgages, credit card receivables, automobile loans, and other consumer purchases—the very areas where credit scoring is used to assess the credit risk of the individual loan.

Securitization begins with a legal entity known as a special-purpose vehicle (SPV) set up by three parties—the sponsor, the servicer, and the trustee. The sponsor, who is often the original owner of the assets (the originator), but need not be, sells the receivables to the SPV, which is often called the issuer. The only business activity of the issuer is to acquire and hold the assets and issue securities backed by the assets, and the servicer is the administrator of this securitization. Often the sponsor is also the servicer, but there are third parties who also specialize in this area. To properly administer the securities, the servicer must be able to collect on delinquent loans and recover on defaulting loans. It must be able to repossess houses in the case of mortgage assets, dispose of collateral in the case of car loans, and generate new receivables in accordance with underwriting standards in the case of credit card loans. The trustee acts as an intermediary between the issuer and the investors in the event of default and ensures the orderly payment of the interest and the principal by the issuer to the investors. For the publicly offered securities, credit-rating agencies will assign a rating to the SPV based on the possible credit problems with the assets and the legal structure of the operation.

Securitization began in the 1970s to promote residential mortgage finance in the U.S., which is still one of the largest securitization markets. In the 1980s, the techniques developed were applied to an increasingly wide range of assets in the U.S. One would have expected these financial instruments to become worldwide very quickly. There has been considerable progress in Canada, the U.K., and Australia, which have legal and regulatory systems similar to the U.S. However, in many countries, the existing laws have prevented securitization, although some countries have started to change their rules to allow wider securitization. Even today, the most startling feature of securitization is the difference between the size of the asset-backed security market and the mortgage-backed security markets in the U.S. and those in the rest of the world.

It is vital for securitization that one can separate the assets from the originating firm so that the investors bear the risk only on a clearly defined existing pool of loans that meet specific criterion. They are not exposed to other risks that the originator might have, like the geographic concentration of its total portfolio (loans from several originators are put into the same SPV to overcome that), changes in the default rate of future customers, or losses on other of the originators assets. The assets can be transferred from the sponsor to the issuer in three ways:

(a) Novation or clean transfer. The rights and the obligations of the assets are transferred to the issuer. This is how most retail store cards are securitized. However, it requires the consent of all parties, including the borrower of the original loan, since consumer protection laws may be violated otherwise. In effect, a new contract is written between the original borrower and the issuer of the securities.

(b) Assignment. The original borrower keeps paying the originator, who in turn pays the money to the issuer. This is the norm for finance loans for automobile purchase. In this case, the issuer cannot change the rules of the original agreement with the borrower.

(c) Subparticipation. There is a separate loan between the issuer (the subparticipant) and the sponsor which the latter pays back out of the money repaid by the borrowers. The issuer has no rights over the original borrowers and so has a double credit risk of whether the original borrower and the sponsor will default.

The advantages of securitization are manifold. Borrowers can either borrow from financial institutions or raise funds directly in the capital market. It is usually cheaper to raise money by the sale of securities on the capital market than to borrow from banks. The cost of securitization must be less than the difference in the interest rates paid on the securities and on their loan equivalents. Therefore, securitization is typically used for loans where the credit risk is easy to assess (perhaps because there is lots of historic information) and where monitoring is adequate. Investors are willing to pay for the greater liquidity and credit transparency that the securities bring over loans. It is also the case that securitization allows originators who are not highly credit rated to get access to funds at better rates. Essentially, the originators' loans may be less risky than they are, and so by separating out their loans, the originators can create SPVs that have higher credit ratings than they do and so borrow money more cheaply. Banks also like securitization because it increases liquidity and because originating and then securitizing new loans increases their profit by raising the turnover rather than the volume of their assets. The biggest boost to securitization in the U.S. was the lack of liquidity in the banking sector associated with the asset quality problems of the savings and loan associations in the crisis of the mid 1980s. Securitization is also a way banks avoid having to hold capital to cover the risk of the loans defaulting, as they are required to do under the Basle rules.

There are reasons for not securitizing as well. Clearly, if a ready supply of cheap money is available from normal methods, there is less need for the hassle of securitization. There is also reluctance by some institutions to damage their special relationships with the original borrowers that creating an SPV involves. One has to be careful when securitizing the quality part of a loan portfolio to get the high credit rating that what remains is not so high risk and unbalanced that the whole organization is at risk. However, the main problem to securitization is still the financial cost and the need for skilled people to undertake the project. This is more so in Europe, where there has been a development in one-off exotic types of securitization rather than the standard securitizations that are the bulk of the U.S. market.

Another problem with securitization is that the credit risk of the portfolio of loans may be too great to be easily saleable. To reduce that risk, issuers will offer credit enhancement, which is usually of one of four types:

(a) Third-party enhancement. An external party like an insurance company will guarantee against the risk of nonpayment.

(b) Subordination. There are different sorts of investors, some of whom get priority in the case of difficulties in repayment in exchange for not getting such good rates of return.

(c) Overcollateralization. The assets are of greater value than is needed to support the agreed payments to the investors.

(d) Cash collateral accounts. Cash is held in a deposit account to be paid out if there a shortfall in the cash received from the receivables.

The second and even more vital part of reducing risk is the role of the ratings agencies, who will provide ratings of the credit (i.e., default) risk of the SPV. The ratings agency examines the historical performance of the receivables and uses simulations based on difficult economic conditions in the past to see how likely the loans are to default. They also examine the external credit enhancements and usually apply the weak-link principle, meaning that a security cannot be rated higher than the rating of the provider of the enhancement. The ratings given can be one of about a dozen levels going from AAA to C, for example.

So what is the relationship between credit scoring and securitization? There are two areas in which credit scoring can be used. The first is deciding what portfolio of loans to put together. Sometimes the portfolio is composed of loans all of the same type, for example, automobile purchasers of the same type of automobile, but one could also put together loans with the same likelihood of default, i.e., the same good:bad odds. Notice that we did not say the same score band because the loans could have come from a number of lenders with very different scorecards, but the odds ratio is the connection between them all. There is very little published work describing how sponsors put together their loan portfolios (probably because it is a very lucrative area), but one would assume that using the score is a sensible way to proceed.

There are some caveats, however, in that the credit score of two different accounts does not describe how those accounts are correlated in terms of risk. It could be that the borrowers have very similar characteristics and so might be vulnerable to the same economic or personal pressures that lead to defaulting, or they could be quite different, and although they have the same overall default risk, the underlying pressures are independent. Clearly, there is need for much more research to identify how credit-scoring techniques can be extended to help identify correlations.

The second and related area is how the ratings agencies can use the credit scores to come up with their overall ratings. One would expect a credit score giving the default risk of each loan in the portfolio would be of considerable help in determining the overall risk of default in the portfolio. However, the problem that the credit score does not describe correlations between the risks to the individuals still holds, and one may need to develop more creative approaches to get the credit score of a portfolio of loans. The methods the ratings agencies use are not widely disseminated, but again one feels there is room for considerable developments in this area. For more details of the whole of the securitization issues, there are good reviews by Lumpkin (1999), and for a view of the international scene, see an article in *Financial Market Trends* (1995).

14.7 Mortgage-backed securities

Mortgage-backed securities were the first of the consumer loans to be securitized and are still far and away the biggest market. In the mid 1990s, almost three-quarters of the home mortgage lending in the U.S. was financed through securitization. As of the end of 1998, the volume of outstanding mortgage-backed securities in the U.S. was $2.6 trillion, accounting for roughly half of all outstanding mortgage credit. The amount of other asset-backed securities in the U.S. was $284 billion. Compared with this, the mortgage-backed securities market in other countries is much smaller. Even in the U.K., where activity has grown considerably, nowhere near this percentage of the market is covered by securitized products.

The reasons for this unevenness are regulatory, historical, and related to the financial markets in the different countries. Historically, the U.S. housing market has long fixed-rate mortgages with financing provided by banks and savings and loan associations. In the 1970s, interest rate volatility heightened the mismatch between the short-term deposit rates and the long-term repayment rates. Thus federal agencies started to guarantee residential mortgages and then issued them as collateral securities to investors. These agencies—the Government National Mortgage Association (Ginnie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), and the Federal National Mortgage Association (Fannie Mae)—were responsible for most mortgage-backed securitization. Also, the legal framework in the U.S. allowed for SPVs to be set up as trusts, while other countries have had to change their laws to allow them. It is also worth noting that in the late 1980s, U.S. banks started shifting away from traditional lending to fee-based sources of income. Securitization was the key component in this strategy, and only recently have banks in other countries started moving more to this strategic approach.

One of the major differences between mortgage-backed securities and asset-backed securities is that the chance of default is probably even smaller in the former than in the latter. However, the chance of prepayment or early repayment is much higher since the average mortgage loan term is more than 25 years but the average duration of a mortgage is less than 7 years. Prepayment has a negative effect on the investor as they were expecting to have their money tied up for longer maturities than turns out to be the case. So in terms of scoring, this means that scoring for early repayment is as important as scoring for default. Quite sophisticated models have been built to estimate the risk of prepayment of mortgage-backed securities in terms of changes in the interest rate, but these tend not to involve the early repayment scores or the default scores of the underlying accounts. In this area, the survival analysis models of section 12.7 as well as the Markov chain modeling of section 6.4 may prove alternatives to the interest-based models that are currently most commonly used.

References

H. T. ALBRIGHT (1994), *Construction of a Polynomial Classifier for Consumer Loan Applications Using Genetic Algorithms*, Working Paper, Department of Systems Engineering, University of Virginia, Charlottesville, VA.

E. I. ALTMAN (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *J. Finance*, 23, 589–609.

T. W. ANDERSON AND L. A. GOODMAN (1957), Statistical inference about Markov chains, *Ann. Math Statist.*, 28, 89–109.

P. BACCHETTA AND F. BALLABRIGA (1995), *The Impact of Monetary Policy and Bank Lending: Some International Evidence*, Working Paper 95.08, Studienzentrum Gerzensee, Gerzensee, Switzerland.

S. M. BAJGIER AND A. V. HILL (1982), An experimental comparison of statistical and linear programming approaches to the discriminant problem, *Decision Sci.*, 13, 604–611.

J. BANASIK, J. N. CROOK, AND L. C. THOMAS (1996), Does scoring a subpopulation make a difference?, *Internat. Rev. Retail Distribution Consumer Res.*, 6, 180–195.

J. BANASIK, J. N. CROOK, AND L. C. THOMAS (1999), Not if but when borrowers default, *J. Oper. Res. Soc.*, 50, 1185–1190.

M. S. BARTLETT (1951), The frequency goodness of fit test for probability chains, *Proc. Cambridge Philos. Soc.*, 47, 86–95.

G. BENNETT, G. PLATTS, AND J. CROSSLEY (1996), Inferring the inferred, *IMA J. Math. Appl. Business Indust.*, 7, 271–338.

J. BERKOWITZ AND R. HYNES (1999), Bankruptcy exemptions and the market for mortgage loans, *J. Law Econom.*, 42, 809–830.

B. S. BERNANKE (1986), Alternative explanations of money-income correlation, *Carnegie-Rochester Ser. Public Policy*, 25, 49–100.

B. S. BERNANKE AND A. S. BLINDER (1988), Is it money or credit, or both, or neither?, *Amer. Econom. Rev. Papers Proc.*, 78, 435–439.

B. S. BERNANKE AND A. S. BLINDER (1992), The federal funds rate and the channels of monetary transmission, *Amer. Econom. Rev.*, September, 901–921.

B. S. BERNANKE AND M. GERTLER (1995), Inside the black box: The credit channel of monetary policy transmission, *J. Econom. Perspectives*, 9, 27–48.

H. Bierman and W. H. Hausman (1970), The credit granting decision, *Management Sci.*, 16, 519–532.

D. Biggs, B. De Ville, and E. Suen (1991), A method of choosing multiway partitions for classification and decision trees, *J. Appl. Statist.*, 18, 49–62.

C. M. Bishop (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford.

M. Blackwell (1993), *Measuring the Discriminatory Power of Characteristics, Proceedings of Credit Scoring and Credit Control* III, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

Board of Governors of the Federal Reserve System (1992), *Survey of Consumer Finance*, Board of Governors of the Federal Reserve System, Washington, DC.

M. Boyle, J. N. Crook, R. Hamilton, and L. C. Thomas (1992), Methods for credit scoring applied to slow payers, in *Credit Scoring and Credit Control*, L. C. Thomas, J. N. Crook, and D. B. Edelman, eds., Oxford University Press, Oxford, 75–90.

L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

N. Breslow (1974), Covariance analysis of censored survival data, *Biometrics*, 30, 89–99.

J. Buckley and I. James (1979), Linear regression with censored data, *Biometrika*, 66, 429–436.

N. Capon (1982), Credit scoring systems: A critical analysis, *J. Marketing*, 46, 82–91.

K. A. Carow and M. Staten (1999), Debit, credit or cash: Survey evidence on gasoline purchases, *J. Econom. Business*, 51, 409–421.

G. G. Chandler and D. C. Ewert (1976), *Discrimination on Basis of Sex and the Equal Credit Opportunity Act*, Credit Research Center, Purdue University, West Lafayette, IN.

K. C. Chang, R. Fung, A. Lucas, R. Oliver, and N. Shikaloff (2000), Bayesian networks applied to credit scoring, *IMA J. Math. Appl. Business Indust.*, 11, 1–18.

S. Chatterjee and S. Barcun (1970), A nonparametric approach to credit screening, *J. Amer. Statist. Assoc.*, 65, 150–154.

G. A. Churchill, J. R. Nevin, and R. R. Watson (1977), The role of credit scoring in the loan decision, *Credit World*, March, 6–10,

R. T. Clemen, A. H. Murphy, and R. L. Winkler (1995), Screening probability forecasts: Contrasts between choosing and combining, *Internat. J. Forecasting*, 11, 133–146.

J. Y. Coffman (1986), *The Proper Role of Tree Analysis in Forecasting the Risk Behavior of Borrowers*, MDS Reports, Management Decision Systems, Atlanta, GA, 3, 4, 7, 9.

D. Collett (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, London.

A. W. Corcoran (1978), The use of exponentially smoothed transition matrices to improve forecasting of cash flows from accounts receivable, *Management Sci.*, 24, 732–739.

D. R. Cox (1972), Regression models and life-tables (with discussion), *J. Roy. Statist. Soc. Ser. B*, 74, 187–220.

J. N. Crook (1989), The demand for retailer financed instalment credit: An econometric analysis, *Managerial Decision Econom.*, 10, 311–319.

J. N. Crook (1996), Credit constraints and U.S. households, *Appl. Financial Econom.*, 6, 477–485.

J. N. Crook (1998), Consumer credit and business cycles, in *Statistics in Finance*, D. J. Hand and S. D. Jacka, eds., Arnold, London.

J. N. Crook (2000), The demand for household credit in the U.S.: Evidence from the 1995 Survey of Consumer Finance, *Appl. Financial Econom.*, to appear.

J. N. Crook, R. Hamilton, and L. C. Thomas (1992a), The degradation of the scorecard over the business cycle, *IMA J. Math. Appl. Business Indust.*, 4, 111–123.

J. N. Crook, R. Hamilton, and L. C. Thomas (1992b), Credit card holders: Users and nonusers, *Service Industries J.*, 12 (2), 251–262.

R. M. Cyert, H. J. Davidson, and G. L. Thompson (1962), Estimation of allowance for doubtful accounts by Markov chains, *Management Sci.*, 8, 287–303.

S. Dale and A. Haldane (1995), Interest rates and the channels of monetary transmission: Some sectoral estimates, *European Econom. Rev.*, 39, 1611–1626.

C. Darwin (1859), *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, London.

Datamonitor (1998), *Analysis Profitability in Prestige Sectors*, Report DMFS0376, Datamonitor, London.

R. H. Davis, D. B. Edelman, and A. J. Gammerman (1992), Machine-learning algorithms for credit-card applications, *IMA J. Math. Appl. Business Indust.*, 4, 43–52.

M. H. De Groot and E. A. Ericksson (1985), Probability forecasting, stochastic dominance and the Lorenz curve, in *Bayesian Statistics*, J. M. Bernado, M. H. De Groot, D. V. Lindley, and A. F. M. Smith, eds., North-Holland, Amsterdam, 99–118.

V. S. Desai, D. G. Convay, J. N. Crook, and G. A. Overstreet (1997), Credit scoring models in the credit union environment using neural networks and genetic algorithms, *IMA J. Math. Appl. Business Indust.*, 8, 323–346.

V. S. Desai, J. N. Crook, G. A. Overstreet (1996), A comparison of neural networks and linear scoring models in the credit environment, *European J. Oper. Res.*, 95, 24–37.

Y. M. I. Dirickx and L. Wakeman (1976), An extension of the Bierman-Hausman model for credit granting, *Management Sci.*, 22, 1229–1237.

I. Domowitz and R. L. Sartain (1999), Incentives and bankruptcy chapter choice: Evidence from the Reform Act 1978, *J. Legal Stud.*, 28, 461–487.

L. M. Drake and M. J. Holmes (1995), Adverse selection and the market for consumer credit, *Appl. Financial Econom.*, 5, 161–167.

L. M. Drake and M. J. Holmes (1997), Adverse selection and the market for building society mortgage finance, *Manchester School*, 65 (1), 58–70.

D. Durand (1941), *Risk Elements in Consumer Instalment Financing*, National Bureau of Economic Research, New York.

D. B. Edelman (1988), Some thoughts on the coding of employment categories, *Viewpoints*, 12 (4), San Rafael, CA.

D. B. Edelman (1997), Credit scoring for lending to small businesses, in *Proceedings of Credit Scoring and Credit Control* V, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

D. B. Edelman (1999), Building a model to forecast arrears and provisions, in *Proceedings of Credit Scoring and Credit Control* VI, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

D. Edwards (1995), *Introduction to Graphical Modelling*, Springer-Verlag, New York.

B. Efron (1977), The efficiency of Cox's likelihood function for censored data, *J. Amer. Statist. Assoc.*, 72, 557–565.

R. A. Eisenbeis (1977), Pitfalls in the application of discriminant analysis in business, finance and economics, *J. Finance*, 32, 875–900.

R. A. Eisenbeis (1978), Problems in applying discriminant analysis in credit scoring models, *J. Banking Finance*, 2, 205–219.

R. A. Eisenbeis (1996), Recent developments in the application of credit scoring techniques to the evaluation of commercial loans, *IMA J. Math. Appl. Business Indust.*, 7, 271–290.

Equal Credit Opportunity Act (1975), *United States Code*, Title 15, Section 1691 et seq.

Equal Credit Opportunity Act Amendments of 1976 (1976), *Report of the Committee on Banking Housing and Urban Affairs, 94th Congress*, U.S. Government Printing Office, Washington, DC.

S. S. Erenguc and G. J. Koehler (1990), Survey of mathematical programming models and experimental results for linear discriminant analysis, *Managerial Decision Econom.*, 11, 215–225.

S. Fianco (1988), Platinum players, *European Card Rev.*, October, 26–28.

Finance and Leasing Association (2000), *Guide to Credit Scoring* 2000, Finance and Leasing Association, London.

Financial Market Trends (1995), Securitisation: An international perspective, *Financial Market Trends*, 61, 33–53.

H. Fishelson-Holstine (1998), Case studies in credit risk model development, in *Credit Risk Modeling*, E. Mays, ed., Glenlake Publishing, Chicago, 169–180.

R. A. Fisher (1936), The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, 7, 179–188.

E. Fix and J. Hodges (1952), *Discriminatory Analysis, Nonparametric Discrimination, Consistency Properties*, Report 4, Project 21-49-004, School of Aviation Medicine, Randolph Field, TX.

C. Forgy (1982), Rete: A fast algorithm for the many pattern/many object pattern match problem, *Artif. Intell.*, 19, 17–37.

N. Freed and F. Glover (1981a), A linear programming approach to the discriminant problem, *Decision Sci.*, 12, 68–74.

N. Freed and F. Glover (1981b), Simple but powerful goal programming formulations for the discriminant problem, *European J. Oper. Res.*, 7, 44–60.

N. Freed and F. Glover (1986a), Evaluating alternative linear programming models to solve the two-group discriminant problem, *Decision Sci.*, 17, 151–162.

N. Freed and F. Glover (1986b), Resolving certain difficulties and improving classification power of LP discriminant analysis formulations, *Decision Sci.*, 17, 589–595.

M. Friedman and A. J. Schwartz (1963), *A Monetary History of the United States*, 1867–1960, Princeton University Press, Princeton, NJ.

H. Frydman (1984), Maximum likelihood estimation in the Mover-Stayer model, *J. Amer. Statist. Assoc.*, 79, 632–638.

H. Frydman, J. G. Kallberg, and D.-L. Kao (1985), Testing the adequacy of Markov chains and Mover-Stayer models as representations of credit behaviour, *Oper. Res.*, 33, 1203–1214.

K. Fukanaga and T. E. Flick (1984), An optimal global nearest neighbour metric, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-1, 25–37.

N. C. Garganas (1975), An analysis of consumer credit and its effects on purchases of consumer durables, in *Modelling the Economy*, G. Renton, ed., Heinemann Educational Books, London.

G. D. Garson (1998), *Neural Networks: An Introductory Guide for Social Scientists*, Sage, London.

M. Gertler and S. Gilchrist (1993), The role of credit market imperfections in the transmission of monetary policy: Arguments and evidence, *Scand. J. Econom.*, 95 (1), 43–64.

J. J. Glen (1999), Integer programming models for normalisation and variable selection in mathematical programming models for discriminant analysis, *J. Oper. Res. Soc.*, 50, 1043–1053.

F. Glover (1990), Improved linear programming models for discriminant analysis, *Decision Sci.*, 21, 771–785.

D. E. Goldberg (1989), *Genetic Algorithms in Search Optimization and Machine Learning*, Addison–Wesley, Reading, MA.

B. J. Grablowsky and W. K. Talley (1981), Probit and discriminant functions for classifying credit applicants: A comparison, *J. Econom. Business*, 33, 254–261.

D. J. HAND (1981), *Discrimination and Classification*, John Wiley, Chichester, U.K.

D. J. HAND (1997), *Construction and Assessment of Classification Rules*, John Wiley, Chichester, U.K.

D. J. HAND AND W. E. HENLEY (1993), Can reject inference ever work?, *IMA J. Math. Appl. Business Indust.*, 5, 45–55.

D. J. HAND AND W. E. HENLEY (1997), Statistical classification methods in consumer credit, *J. Roy. Statist. Soc. Ser. A*, 160, 523–541.

D. J. HAND AND S. D. JACKA (1998), *Statistics in Finance*, Arnold, London.

D. J. HAND, K. J. MCCONWAY, AND E. STANGHELLINI (1997), Graphical models of applications for credit, *IMA J. Math. Appl. Business Indust.*, 8, 143–155.

A. HARTROPP (1992), Demand for consumer borrowing in the U.K., 1969–90, *Appl. Financial Econom.*, 2, 11–20.

S. HAYKIN (1999), *Neural Networks: A Comprehensive Foundation*, Prentice–Hall International, London.

W. E. HENLEY (1995), *Statistical Aspects of Credit Scoring*, Ph.D. thesis, Open University, Milton Keynes, U.K.

W. E. HENLEY AND D. J. HAND (1996), A k-NN classifier for assessing consumer credit risk, *Statistician*, 65, 77–95.

E. C. HIRSCHMAN (1982), Consumer payment systems: The relationship of attribute structure to preference and usage, *J. Business*, 55, 531–545.

B. HOADLEY AND R. M. OLIVER (1998), Business measures of scorecard benefit, *IMA J. Math. Appl. Business Indust.*, 9, 55–64.

P. G. HOEL (1954), A test for Markov chains, *Biometrika*, 41, 430–433.

K. HOFFMANN (2000), Stein estimation: A review, *Statist. Papers*, 41, 127–158.

J. H. HOLLAND (1968), *Hierarchical Description of Universal and Adaptive Systems*, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor.

J. H. HOLLAND (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

M. A. HOPPER AND E. M. LEWIS (1991), Development and use of credit profit measures for account management, *IMA J. Math. Appl. Business Indust.*, 4, 3–17.

M. A. HOPPER AND E. M. LEWIS (1992), Behaviour scoring and adaptive control systems, in *Credit Scoring and Credit Control*, L. C. Thomas, J. N. Crook, and D. B. Edelman, eds., Oxford University Press, Oxford, 257–276.

D. C. HSIA (1978), Credit scoring and the Equal Credit Opportunity Act, *Hastings Law J.*, 30, 371–448.

R. JAMBER, S. JAV, D. MEDLER, AND P. KLAHR (1991), The credit clearing house expert system, in, *Innovative Applications of Artificial Intelligence* 3, R. Smith and C. Scott, eds., AAAI Press, Menlo Park, CA.

T. JAPPELLI (1990), Who is credit constrained in the U.S.?, *Quart. J. Econom.*, 105, 219–234.

T. JAPPELLI AND M. PAGANO (1989), Consumption and capital market imperfections, *Amer. Econom. Rev.*, 79 (5), 1088–1105.

T. JAPPELLI AND M. PAGANO (2000), *Information Sharing, Lending and Defaults: Cross-Country Evidence*, Working Paper 22, Centre for Studies in Economics and Finance, Dipartimento di Scienze Economiche, Università degli Studi di Salerno, Salerno, Italy.

E. A. JOACHIMSTHALER AND A. STAM (1990), Mathematical programming approaches for the classification problem in two-group discriminant analysis, *Multivariate Behavioural Res.*, 25, 427–454.

R. W. JOHNSON (1992), Legal, social and economic issues implementing scoring in the U.S., in *Credit Scoring and Credit Control*, L. C. Thomas, J. N. Crook, and D. B. Edelman, eds., Oxford University Press, Oxford, 19–32.

A. JOST (1998), Data mining, in *Credit Risk Modeling*, E. Mays, ed., Glenlake Publishing, Chicago, 129–154.

J. G. KALLBERG AND A. SAUNDERS (1983), Markov chain approach to the analysis of payment behaviour of retail credit customers, *Financial Management*, 12, 5–14.

A. K. KASHYAP, J. C. STEIN, AND D. W. WILCOX (1993), Monetary policy and credit conditions: Evidence from the composition of external finance, *Amer. Econom. Rev.*, 83 (1), 78–98.

S. R. KING (1986), Monetary transmission through bank loans or bank liabilities?, *J. Money Credit Banking*, 18 (3), 290–303.

G. J. KOEHLER AND S. S. ERENGUC (1990), Minimising misclassification in linear discriminant analysis, *Decision Sci.*, 21, 63–85.

P. KOLESAR AND J. L. SHOWERS (1985), A robust credit screening model using categorical data, *Management Sci.*, 31, 123–133.

T. L. LAI AND Z. L. YING (1994), A Missing information principle and M-estimators in regression analysis with censored and truncated data, *Ann. Statist.*, 22, 1222–1255.

S. L. LAURITZEN AND N. WERMUTH (1989), Graphical models for association between variables some of which are qualitative and some quantitative, *Ann. Statist.*, 17, 51–57.

Y. LE CUN (1985), Une procedure d'apprentissage pour resau a seuil assymetrique, in *Proceedings of Cognitiva* 85: *À la Frontiere de l'Intelligence Artificielle, des Sciences de la Connaissance et des Neurosciences*, CESTA, Paris, 500–604.

K. J. LEONARD (1993A), Detecting credit card fraud using expert systems, *Comput. Indust. Engrg.*, 25, 103–106.

K. J. LEONARD (1993B), A fraud alert model for credit cards during the authorization process, *IMA J. Math. Appl. Business Indust.*, 5, 57–62.

K. J. LEONARD (2000), *The Development of a Just-in-Time Risk Scoring System for "Early Delinquent Account" Management*, Working Paper, Department of Health Administration, University of Toronto, Toronto, ON, Canada.

E. M. LEWIS (1992), *An Introduction to Credit Scoring*, Athena Press, San Rafael, CA.

H. G. LI AND D. J. HAND (1997), Direct versus indirect credit scoring classification, in *Proceedings of Credit Scoring and Credit Control* V, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

J. T. LINDLEY, P. RUDOLPH, AND E. B. SELBY JR. (1989), Credit card possession and use: Changes over time. *J. Econom. Business*, 127–142.

A. D. LOVIE AND P. LOVIE (1986), The flat maximum effect and linear scoring models for prediction, *J. Forecasting*, 5, 159–186.

A. LUCAS AND J. POWELL (1997), Small sample scoring, in *Proceedings of Credit Scoring and Credit Control* V, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

S. LUDVIGSON (1998), The channel of monetary transmission to demand: Evidence from the market for automobile credit, *J. Money Credit Banking*, 30 (3), 365–383.

S. LUMPKIN (1999), Trends and developments in securitisation, *Financial Market Trends*, 74, 25–59.

P. MAKOWSKI (1985), Credit scoring branches out, *Credit World*, 75, 30–37.

W. M. MAKUCH (1999), The basics of a better application score, in *Credit Risk Modeling, Design and Applications*, E. Mays, ed., Fitzroy Dearborn Publishers, Chicago, 59–80.

O. L. MANGASARIAN (1965), Linear and nonlinear separation of patterns by linear programming, *Oper. Res.*, 13, 444–452.

R. E. MARTIN AND D. J. SMYTH (1991), Adverse selection and moral hazard effects in the mortgage market: An empirical analysis, *Southern Econom. J.*, 1071–1084.

E. MAYS (1998), *Credit Risk Modeling*, Glenlake Publishing, Chicago.

B. J. A. MERTENS AND D. J. HAND (1997), *Adjusted Estimation for the Combination of Classifiers*, Department of Statistics, Trinity College, Dublin.

Z. MICHALEWICZ (1996), *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin.

M. L. MINSKY AND S. A. PAPERT (1969), *Perceptrons*, MIT Press, Cambridge, MA.

F. MODIGLIANI (1986), Life cycle, individual thrift and the wealth of nations, *Amer. Econom. Rev.*, 76, 297–313.

J. H. MYERS AND E. W. FORGY (1963), The development of numerical credit evaluation systems, *J. Amer. Statist. Assoc.*, 58, 799–806.

B. NARAIN (1992), Survival analysis and the credit granting decision, in *Credit Scoring and Credit Control*, L. C. Thomas, J. N. Crook, D. B. Edelman, eds., Oxford University Press, Oxford, 109–122.

R. Nath, W. M. Jackson, and T. W. Jones (1992), A comparison of the classical and the linear programming approaches to the classification problem in discriminant analysis, *J. Statist. Comput. Simul.*, 41, 73–93.

R. Nath and T. W. Jones (1988), A variable selection criterion in the linear programming approaches to discriminant analysis, *Decision Sci.*, 19, 554–563.

J. R. Nevin and G. A. Churchill (1979), The Equal Credit Opportunity Act: An evaluation, *J. Marketing*, 42, 95–104.

R. M. Oliver (1993), Effects of calibration and discrimination on profitability scoring, in *Proceedings of Credit Scoring and Credit Control* III, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

R. M. Oliver and E. Wells (2001), Efficient frontier cutoff policies in credit problems, *J. Oper. Res. Soc.*, 52, 1025–1033.

G. A. Overstreet and E. L. Bradley (1996), Applicability of generic linear scoring models in the U.S. credit union environment, *IMA J. Math. Appl. Business Indust.*, 7, 271–338.

G. A. Overstreet, E. L. Bradley, and R. S. Kemp (1992), The flat maximum effect and generic linear scoring models: A test, *IMA J. Math. Appl. Business Indust.*, 4, 97–110.

S. Park (1993), The determinants of consumer instalment credit, *Federal Reserve Bank of St. Louis Economic Rev.*, November/December, 23–38.

D. B. Parker (1982), *Learning Logic*, Invention Report S81-64, File 1, Office of Technology Licencing, Stanford University, Stanford, CA.

R. Pavar, P. Wanart, and C. Loucopoulos (1997), Examination of the classification performance of MIP models with secondary goals for the two group discriminant problem, *Ann. Oper. Res.*, 74, 173–189.

A. P. Duarte Silva and A. Stam (1997), A mixed integer programming algorithm for minimising the training sample misclassification cost in two-group classifications, *Ann. Oper. Res.*, 74, 129–157.

G. Platts and I. Howe (1997), A single European scorecard, in *Proceedings of Credit Scoring and Credit Control* V, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

R. Pollin (1988), The growth of U.S. household debt: Demand side influences, *J. Macroeconomics*, 10 (2), 231–248.

M. L. Puterman (1994), *Markov Decision Processes*, John Wiley, New York.

J. R. Quinlan (1993), *C4.5: Programs for Machine Learning*, Morgan–Kaufman, San Mateo, CA.

A. K. Reichert, C.-C. Cho, G. M. Wagner (1983), An examination of the conceptual issues involved in developing credit scoring models, *J. Business Econom. Statist.*, 1, 101–114.

M. D. Richard and R. P. Lipman (1991), Neural network classifiers estimate Bayesian a posteriori probabilities, *Neural Comput.*, 3, 461–483.

C. D. Romer and D. H. Romer (1990), New evidence on the monetary transmission mechanism, in *Brookings Papers on Economic Activity*, Brookings Institution Press, Washington, DC.

E. Rosenberg and A. Gleit (1994), Quantitative methods in credit management: A survey, *Oper. Res.*, 42, 589–613.

F. Rosenblatt (1958), The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Rev.*, 65, 386–408.

F. Rosenblatt (1960), *On the Convergence of Reinforcement Procedures in Simple Perceptrons*, Report VG-1196-G-4, Cornell Aeronautical Laboratory, Buffalo, NY.

P. A. Rubin (1990), Heuristic solution procedures for a mixed-integer programming discriminant model, *Managerial Decision Econom.*, 11, 255–266.

P. A. Rubin (1997), Solving mixed integer classification problems by decomposition, *Ann. Oper. Res.*, 74, 51–64.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams (1986a), Learning representation by back-propagating errors, *Nature*, 323, 533–536.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams (1986b), Learning internal representations by error backpropagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, MA.

D. E. Rumelhart and J. L. McClelland, eds. (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press, Cambridge, MA.

S. R. Safavian and D. Landgrebe (1991), A survey of decision tree classifier methodology, *IEEE Trans. Systems Man Cybernetics*, 21, 660–674

B. S. Sangha (1998), A systematic approach for managing credit score overrides, in *Credit Risk Modeling*, E. Mays, ed., Glenlake Publishing, Chicago, 221–244

J. Saunders (1985), This is credit scoring, *Credit Management*, September, 23–26.

P. Sewart and J. Whittaker (1998), Fitting graphical models to credit scoring data, *IMA J. Math. Appl. Business Indust.*, 9, 241–266.

G. M. Shepherd and C. Koch (1990), Introduction to synaptic circuits, in *The Synaptic Organization of the Brain*, G. M. Shepherd, ed., Oxford University Press, New York.

J. L. Showers and L. M. Chakrin (1981), Reducing revenue from residential telephone customers, *Interfaces*, 11, 21–31.

O. A. Smalley and F. D. Sturdivant (1973), *The Credit Merchants: A History of Spiegel Inc.*, Southern Illinois University Press, Carbondale, IL.

D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen, and R. G. Cowell (1993), Bayesian analysis in expert systems, *Statist. Sci.*, 8, 219–283.

V. Srinivasan and Y. H. Kim (1987a), The Bierman-Hausman credit granting model: A note, *Management Sci.*, 33, 1361–1362.

V. SRINIVASAN AND Y. H. KIM (1987B), Credit granting: A comparative analysis of classification procedures, *J. Finance*, 42, 665–683.

M. STEPANOVA AND L. C. THOMAS (1999), Survival analysis methods for personal loan data, in *Proceedings of Credit Scoring and Credit Control* VI, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

M. STEPANOVA AND L. C. THOMAS (2001), PHAB scores: Proportional hazards analysis behavioural scores, *J. Oper. Res. Soc.*, 52, 1007–1016.

J. STIGLITZ AND A. WEISS (1981), Credit rationing in markets with imperfect information, *Amer. Econom. Rev.*, 71, 393–410.

A. C. SULLIVAN (1987), *Economic Factors Associated with Delinquency Rates on Consumer Installment Debt*, Working Paper 55, Credit Research Center, Krannert Graduate School of Management, Purdue University, West Lafayette, IN.

T. A. SULLIVAN, E. WARREN, AND J. L. WESTERBROOK (1989), *As We Forgive Our Debtors*, Oxford University Press, New York.

H. TALEBZADEH, S. MANDUTIANU, AND C. WINNER (1994), Countrywide loan underwriting expert system, in *Proceedings of the Sixth Innovative Applications of Artificial Intelligence Conference*, AAAI Press, Menlo Park, CA.

L. C. THOMAS (1992), Financial risk management models, in *Risk Analysis, Assessment and Management*, J. Ansell and F. Wharton, eds., John Wiley, Chichester, U.K., 55–70.

L. C. THOMAS (1994), Applications and solution algorithms for dynamic programming, *Bull. IMA*, 30, 116–122.

L.C. THOMAS (1998), Methodologies for classifying applicants for credit, in *Statistics in Finance*, D. J. Hand and S. D. Jacka, eds., Arnold, London, 83–103.

L. C. THOMAS, J. BANASIK, AND J. N. CROOK (2001), Recalibrating scorecards, *J. Oper. Res. Soc.*, 52, 981–988.

L. C. THOMAS, J. N. CROOK, AND D. B. EDELMAN (1992), *Credit Scoring and Credit Control*, Oxford University Press, Oxford.

H.-T. TSAI, L. C. THOMAS, AND H.-C. YEH (1999), Using screening variables to control default rate in credit assessment problems, in *Proceedings of Credit Scoring and Credit Control* VI, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

H.-T. TSAI AND H.-C. YEH (1999), *A Two Stage Screening Procedure and Mailing Credit Assessment*, Working Paper, Department of Business Management, National Sun Yat-Sen University, Kaohsiung, Taiwan.

J. A. M. VAN KUELEN, J. SPRONK, AND A. W. CORCORAN (1981), Note on the Cyert-Davidson-Thompson doubtful accounts model, *Management Sci.*, 27, 108–112.

P. A. VOLKER (1983), A note on factors influencing the utilization of bankcard, *Econom. Record*, September, 281–289.

H. M. WEINGARTNER (1966), Concepts and utilization of credit scoring techniques, *Banking*, 58, 51–53.

H. White (1989), Learning in artificial neural networks: A statistical perspective, *Neural Comput.*, 1, 425–464.

M. J. White (1998), Why it pays to file for bankruptcy: A critical look at the incentives under U.S. personal bankruptcy law and a proposal for change, *Univ. Chicago Law Rev.*, 65 (3), 685–732.

J. Whittaker (1990), *Graphical Models in Applied Multivariate Statistics*, John Wiley, New York.

J. C. Wiginton (1980), A note on the comparison of logit and discriminant models of consumer credit behaviour, *J. Financial Quantitative Anal.*, 15, 757–770.

E. F. Wonderlic (1952), An analysis of factors in granting credit, *Indiana Univ. Bull.*, 50, 163–176.

M. B. Yobas, J. N. Crook, and P. Ross (1997), *Credit Scoring Using Neural and Evolutionary Techniques*, Working Paper 97/2, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

H. A. Ziari, D. J. Leatham, and P. N. Ellinger (1997), Development of statistical discriminant mathematical programming model via resampling estimation techniques, *Amer. J. Agricultural Econom.*, 79, 1352–1362.

H. Zhu, P. A. Beling, and G. A. Overstreet (1999), A Bayesian framework for the combination of classifier outputs, in *Proceedings of Credit Scoring and Credit Control* VI, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland.

H. Zhu, P. A. Beling, and G. A. Overstreet (2001), A study in the combination of two consumer credit scores, *J. Oper. Res. Soc.*, 52, 974–980.

Index