Banff Credit Risk Conference 2003

Tom Åstebro (University of Toronto),
Peter A. Beling (University of Virginia),
David Hand (Imperial College),
Robert Oliver (University of California at Berkeley),
L.C. Thomas (Southampton University)
October 11 – 16, 2003

1 Workshop Objectives

The objective of this workshop was to bring together a small group of people interested in the foundations and underlying theory of the mathematical and statistical prediction and decision-making models in Retail Credit Risk, an area whose most important components include the behaviors, actions and preferences of individuals for financial products. We were particularly interested in attempting to understand the similarities, differences, and interactions between retail credit risk and corporate financial risk. Modern portfolio theory, which now has an enormous scientific literature (both theoretical and experimental), plays a central role in investments, trading and valuation of assets through option pricing formulas and arbitrage models. It is worth noting that the Merton [8] and Black-Scholes [2] papers on option pricing and asset valuation are now some of the most widely cited scientific papers and have served as the foundation for the development of a rich theory of corporate risk. Although there are many similarities between retail and corporate risk, the differences are greater still - as the atomic building block of retail credit risk appears to be the account of a single individual with behavioral preferences, whereas in the corporate world the building block for a large portfolio of assets is comprised of shares of stocks or bonds of publicly traded and priced corporations.

The meetings and workshops allowed us to identify and frame the most important unsolved problems, discover linkages with modern financial theory and attempt to show where the similarities with corporate finance are meaningful and realistic. About half of the conference was devoted to four special interest workshops whose topics and special problems are reported in greater detail in the body of this report. In the first three days, a small number of papers were presented that posed questions and issues; in the afternoon of the first meeting day we organized a preliminary taxonomy of unsolved problems and asked each participant to add to or modify the list. The topics of interest were then subdivided into four major workshops, which from that point forward met during breaks and periods when no papers were presented. In several cases the workshops broke down into even smaller sub-groups and by the third and fourth day of the conference individuals and groups were making presentations of preliminary findings and recommendations to all participants. Having small individual meeting rooms with whiteboards and projectors was an invaluable asset. The findings of the workshops are reported in the sections that follow.

2 Default Models for Portfolios of Consumer Loans

The problem we addressed was whether there exist mathematical or statistical models describing the credit risk of portfolios of consumer loans. This is an important topic because the Basel New Accord assumes specific formulae for this which are taken directly from Merton's models of credit risk in corporate loans [8]. One also wants to develop such loan models for other reasons, such as portfolio management and securitization of loans. Thus we considered problems in three related areas:

- *New Models:* Can we develop new models that describe credit risk of consumer loans and portfolios of consumer loans ?
- *Ties to Corporate Risk:* Can we construct models of portfolio credit risk for consumer loans that link directly with the corporate credit risk models?

- *Existing Models:* Are there models of the credit risk of consumer loans and portfolios of consumer loans that lead to the formulae appropriate for current and future versions of the Basel regulations?

2.1 New Models

Three possible approaches were considered – models that segment on the dynamics and level of the default losses; models exploiting conditional independence once economic and age-related factors had been removed; and models which sought to mimic the reduced form and survival analysis approaches to credit risk for corporate loans.

The most advanced 'dynamics of losses segmentation model' for the credit risk of portfolios of consumer loans was that presented by Wedling [13]. This is an empirically based approach which requires considerable data so is probably most appropriately developed by credit bureaus or regulators. It begins by defining a large number of segments on socio-demographic variables (say more than 300) which are the basic building blocks of the portfolio. These are used to estimate a loss index function, such as actual payments compared with expected payments. One then merges segments, creating larger groupings with homogeneous dynamics of loss and similar correlations between loss indices (calendar time is used by Wedling, so this gives the response to economic effects, but one could use duration since lending start as well which would pick up financial naivety and fraud). Typically there will be around five such segments. One then does a regression of loss over given time interval against applicant and loan variables and segments according to expected loss levels. One chooses only a few, say three (high, medium and low), such segments. The resulting subsegments which come from combining the dynamics and the levels of loss (15 segments for the indicative numbers given)are then considered separately and a loss distribution is built on each by first deciding empirically on distribution function and then fitting. Correlations between the subsegments are obtained by using correlation coefficients in segment definitions and a chosen copula. One then does Monte Carlo simulation to get final results for loss distribution. It would seem very unlikely that one could obtain an analytic expression for this final distribution.

The independence approach derives from the belief that a model of loan loss should be developed which is based on the structure of conditional independence of default losses given uncertain economic factors such as interest rate, unemployment etc. Marginalization of these economic factors would induce a mathematical structure to the observed correlations between loan losses which may be fundamental to understanding the underlying dependencies, and may, in addition, help distinguish the retail credit process from the corporate one. A theory developed along these lines would not only make direct use of available data on defaults and default losses but would also give a theoretical rationale for the underlying common cause correlations between individual accounts.

The last class of models is considered in Section 2.2.

2.2 Ties to Corporate Risk

We identified six approaches to modelling the credit risk of portfolios of corporate loans:

1. Creditmetrics (J.P.Morgan 1997, see [3] a mark-to-market, ratings-based approach.

2. KMV (Kealhofer, McQuown, Vasicek, see [5] a mark-to market Merton-style model.

3. CreditRisk+ (Credit Suisse 1997, see [4] which is a default mode actuarial style model.

4. Credit Portfolio View (Wilson 1999, see [14] which is logistic regression model using lagged and correlated macroeconomic variables.

5. Markov Chain Reduced Form Models (Jarrow Lando Turnbull 1997, see [6] where default is exogenous but one estimates transition between ratings.

6. Intensity based reduced form models (Lando 1998, see [7] which use survival analysis ideas to estimate directly time to default as a function of macro-economic variables.

As was argued above there are problems with translating the models based on the Merton approach to the consumer loans unless they incorporate jump effects and even then there are problems in identifying what is a default level of a consumer's assets or his credit worthiness let alone how to estimate the correlation in these. Thus it would seem that if one wishes to try and relate models for the credit risk in both corporate and consumer loans the following are the better options.

2.2.1 CreditRisk+ and Related Models

This is a default model with two states – default, not default - not unlike the approach in Section 2.3. It calculates capital requirement based on actuarial approaches found in the property insurance literature. It has minimal data input but only gives loss rates, not loan value changes. It assumes each loan has a small probability of default that is independent of default of other loans. So the distribution is Binomial; one usually takes the Poisson approximation to get analytic expressions. The severities of losses are put into bands; combining frequency of default and severity of losses gives distribution of losses for each exposure band which are then summed across exposure bands. Most of these results can be applied to consumer loans, but it was pointed out the model proves difficult if default probabilities are high (above 4%) since the Poisson approximation is no longer valid and one would need to simulate using the Binomial distribution.

2.2.2 Markov chain reduced form type models

These are mark-to-market models (so there are several states the loan can be in). The state space is a ratings agencies rating of the bonds. Default occurs when the rating hits level D. One can build either continuous or discrete time Markov process of the change in the bonds rating. The transition matrices are estimated as a mixture of the historical process and some limiting risk processes (i.e. with transition matrices I, $p(j,j) = 1$ so no default or $p(j,D) = 1$ so all default) to get correlations. To deal with economic cycles one can let the transition matrices $p(j,k) = f($ macro variables, shock factors $)$ where the former are obtained from the data and the latter are simulated.

In a consumer loan related model one could envisage the ratings being behavioral score buckets plus a bucket for default. One could use historical transition matrices (roll out rates) and then follow the rest of the Markov chain reduced form approach. There would be lots of parameters to estimate but it has the advantage that as one gets ratings distributions at each period, one can check early and often that the model is tracking reality.

2.2.3 Intensity based reduced form type models

These are default mode models and are similar to survival analysis in which we estimate the hazard rate (intensity function) as a function of economic variables and loan dependent variables)

$$h(t) = \exp\{a + b\,I(t) + cW(t)\}, \qquad E[I(t)W(t)] \neq 0$$

under the assumption that $(I(t), W(t))$ is multivariate normal with assumed correlation structure derived theoretically or experimentally from the joint distributions of I and W.

Lando (Lando 1998 [7]) uses a Cox process whose structure depends on the identification of different state variables.

A consumer version of the model could be based on the survival analysis approach to behavioral scoring developed by Stepanova et al. [11].

2.3 Existing Models

There is no real clarity in how the Basel formulae were developed save that they are based on a model for the credit risk of a portfolio of corporate loans in which the portfolio consists of infinitely granular one year loans with one risk factor and value of borrowers' assets being log normally distributed. Although it is difficult to obtain mathematical derivations and references in the published literature, such models are usually attributed to proprietary models developed by the companies Credit Metrics, KMV, and CreditRisk. The assumption is that companies default if debts exceed assets and that the correlation between companies' share prices describes the correlation between their asset movements.

It was suggested that the following simple common factor model can be used to derive the Basel formulae (see, e.g., [10]). In this it is assumed that the distribution of n defaults in a portfolio of N firms is given by the well-known Binomial probability mass function

$$\mathrm{P}\left\{n \text{ defaults in } (0,T)\right\} = \binom{N}{n} p^n (1-p)^{N-n} \qquad 0 \le n \le N,$$

where p is the (common) probability of default of one firm. For large N, this distribution is approximated by the normal density. Under specialized assumptions for the losses of the firm given default, it is usually a straightforward exercise to assess the distribution of losses (risk) to the portfolio under independence assumptions. If we assume that $V_j(t)$ denotes the value of firm j at time t, common factor models assume that default of the firm occurs when this value drops below a pre-specified barrier, say K. The distribution of losses to a portfolio composed of shares of these firms are usually assumed to be proportional to the product of a fractional loss given default with the credit exposure of the firm.

The next assumption that is usually made is that the value of the firm is composed of a term with a common cause factor and a noise term structured in such a way that, given the common factor, firm defaults are independent of one another but correlation between firms exists because of the common cause factor. It is a straightforward but often difficult probabilistic calculation to determine the effect of removing the condition of the common cause factor. In general, if there are more than two common cause factors, analytical solutions are not easy to obtain. Under restrictive assumptions it is sometimes possible to derive analytic results.

A common assumption is that the $V_n(T)$ are jointly normally distributed random variables with covariance matrix Σ, i.e., the firm values are jointly dependent (the Credit metric model), and the barrier is determined by the probability of default $K = \Phi^{-1}(p)$. For example, if

$$V_n(T) = \sqrt{\rho}\, Y + \sqrt{1-\rho}\, \epsilon_n \qquad n = 1, 2, \ldots, N$$

and one assumes that Y and ϵ_n are i.i.d. standard normal random variables, then the probability that the firm's value $V_n(T)$ falls below the barrier K, given that the common factor Y takes on the value y, is

$$p(y) = \Phi\left(\frac{K - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right),$$

which means that the probability of m or more defaults is given by

$$\sum_{n=m}^{N} \binom{N}{n} \int_{-\infty}^{+\infty} \left(\Phi\left(\frac{K-\sqrt{\rho}\, y}{\sqrt{1-\rho}}\right)\right)^{n} \left(1 - \Phi\left(\frac{K-\sqrt{\rho}\, y}{\sqrt{1-\rho}}\right)\right)^{N-n} \phi(y)dy.$$

It can be shown that as the number of firms in the portfolio goes to infinity the continuous distribution function for the number of defaults exceeding x is $1 - F(x)$ where

$$F(x) = \Phi\left(\frac{1}{\sqrt{\rho}}\left(\sqrt{1-\rho}\, \Phi^{-1}(x) - \Phi^{-1}(p)\right)\right)$$

Generalizations have also been made to include cases where each firm has a probability of default p_n rather than a common p as described above, to include assumptions of multiple factors, relaxation of the i.i.d. normality assumption for Y and noise terms as well as the introduction of stochastic time-dependent volatilities.

If one seeks to consider this in the consumer loan context there are a number of problems. Do consumers default when the value of their assets fall below a prespecified barrier? More fundamentally, do consumers even know the value of their assets and if they do can they realize them? What would be a suitable K for consumers and how does one estimate the covariance matrix Σ in the consumer case when there is no equivalent of share price? These are definitional problems but there is also a structural one, as we shall discuss below.

Suppose we develop a model for distribution of loan losses of a consumer loan inspired by the corporate models of Merton [8] and Vasicek [12]. This would lead to

$$\frac{dV}{V} = Adt + BdW + CdY$$

Percentage Change in Value = Drift + Gaussian Diffusion + Stochastic Jump Process

a diffusion equation describes the change of the value of a single firm (the borrower) over time. Debt of the firm is assumed to be constant over time. When the asset value drops below the debt line (or some function of it) default is assumed to occur. Under strong assumptions distributions of loan losses for portfolio of such loans can be derived as in the way described above. More recently Zhou [15] extended the model by adding jump processes (also, see [9]).

The analogy for retail credit is that if we define a new variable, say U, to be the creditworthiness of a single consumer, an analogous diffusion equation can be derived in which the consumer (borrower) has a put option on his credit worthiness with a fixed strike price, R. Parameters are denoted by A', B', C'', R and a starting point U_0. The A term may be zero or can take into account such factors as the ageing of the consumer. R is analogous to debt. Some felt that behavioural score might be a proxy (a noisy signal) for U and one could apply the model on behavioural score.

$$\begin{aligned}
\frac{dU}{U} &= A'dt + B'dW + C''dY \\
\frac{dU}{U} &= \text{Drift + Diffusion(Normal) + Jump Income Shocks(Poisson)}
\end{aligned}$$

The parameters of the diffusion equation can possibly be estimated from microeconomic data such as credit bureau data. There seems to be some preliminary evidence that over short time periods credit scores may follow a simple geometric Brownian motion.

The problem is that it is not possible to obtain the Basel formulae if one has the jump term in the model and yet it was felt by all workshop participants that the jump was possibly the most important feature for a good model of the retail credit process – the reason being that events such as divorce, termination of employment, etc., must be accounted for.

3 Improved Models

This workshop addressed three main problems:

1. Reject Inference
2. Bayesian Marginalization, and
3. Dropout/Withdrawal Inference.

3.1 Reject inference

Financial institutions build models to predict creditworthiness, u, from variables, x, available at the time of application. Such models are based on a retrospective database of customers for whom the x variables are known. The outcome, y, which for convenience we will take to be a binary variable (1=good outcome, such as 'repay the loan'; 0=bad outcome, such as 'default'), will also be known for those customers who were previously accepted, but it is meaningless to speak of a good/bad outcome variable for customers who were rejected. Let a be a random variable taking the value 1 if a customer was accepted, and 0 if they were rejected.

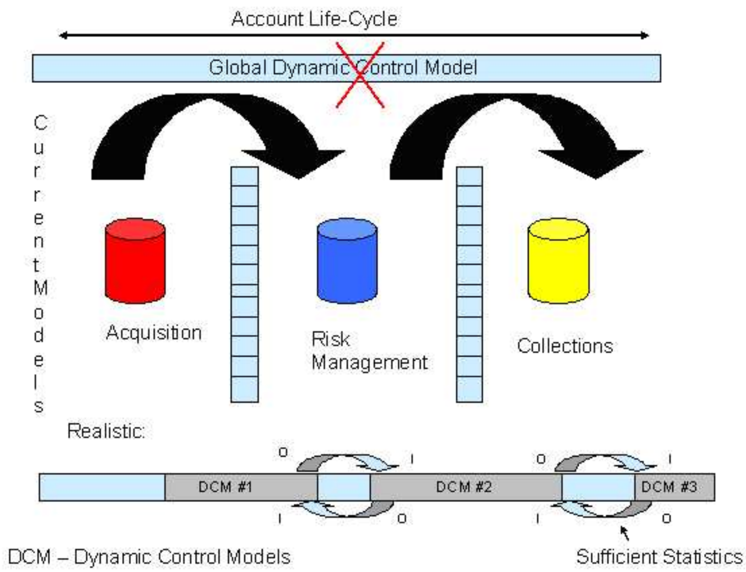

Figure 1: State-space modeling for retail credit.

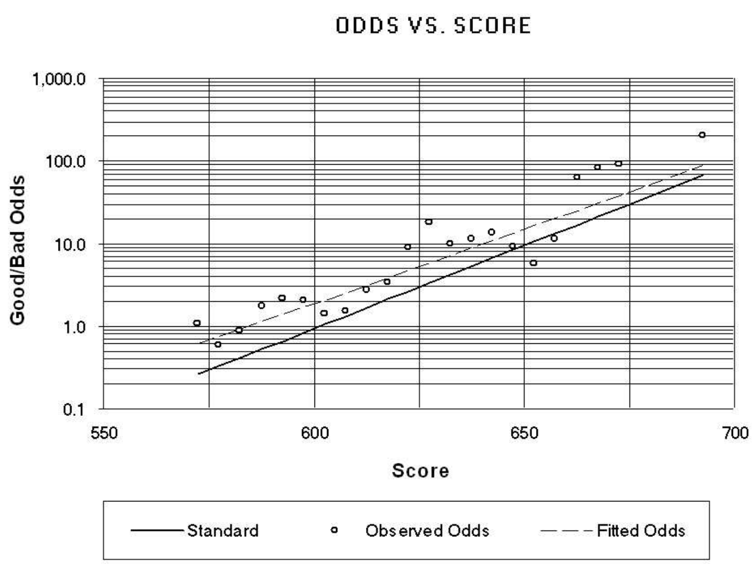

Figure 2: Odds versus score.

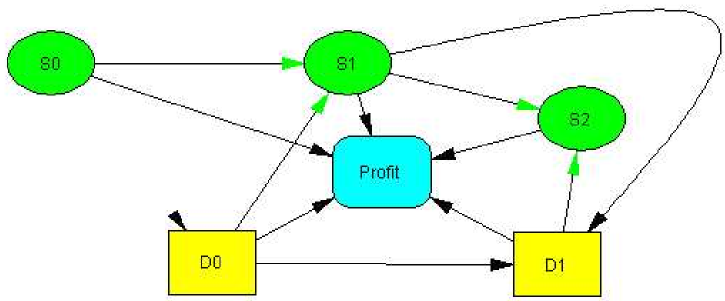

Figure 3: Influence Diagram.

Our aim, then, is to estimate $f(u \mid x)$, the distribution of creditworthiness, u, given the variables, x, available at the time of application. Future accept/reject decisions will be based on a comparison of some summary of our estimate of $f(u \mid x)$ with a threshold, with this threshold being determined from operational considerations. For example, we might compare the median of $f(u \mid x)$ with a threshold t, accepting an applicant with vector x if this median is above t, since we estimate that more than 50% of such people have a creditworthiness exceeding t. A difficulty arises, however, because $f(u \mid x)$ is the distribution across *applicants*, whereas we have outcome information y (related to u as described below) only for those *accepted*: we have a biased sample. *Reject inference* is the term used for strategies aimed at overcoming this problem. The problem is a universal one in the retail credit industry, and has been the focus of much interest. Several papers at the workshop described particular issues of and approaches to reject inference, and a workshop was devoted to it.

We have the decomposition

$$f(u \mid x) = f(u \mid x, a = 0)P(a = 0 \mid x) + f(u \mid x, a = 1)P(a = 1 \mid x)$$

and we can immediately distinguish between two situations.

Case 1: when the accept/reject decision was based solely on variables included amongst those now available in x, so that $a = a(x)$. That is, x may include extra variables which were not used in the accept/reject decision, but certainly includes all variables which were used in that decision. In this case, we have outcome information, y, on all of the applicants in region $A = \{x : a(x) = 1\}$ and outcome information, y, on none of the applicants in region $R = \{x : a(x) = 0\}$, where R is the complement of A. In such a situation, the only possible strategy is to build a model for the data in region A and extrapolate it over region R. The model in region A will be unbiased by the rejection process (assuming it is a properly specified model). If the extrapolation is not into parts of R far from the surface separating A from R then one might expect the model to perform reasonably well.

Case 2: when information, v, additional to that in x, was used in making the accept/reject decision, so that $a = a(x, v)$. This is the case if policy overrides were used (though it may then be difficult to articulate v explicitly). It is also the case if variables are no longer collected. In this case, the distribution of creditworthiness, u, amongst those accepted, is likely to differ from the distribution of creditworthiness amongst those rejected, for a given x.

A simple model often used in the industry assumes that

$$f(u \mid x, a = 0) = f(u \mid x, a = 1),$$

so that

$$P(a = 0 \mid x, u) = P(a = 0 \mid x).$$

That is, this model assumes that the creditworthiness distribution is the same (at given x) for applicants who are accepted and applicants who are rejected. This is the missing-at-random assumption. It assumes that, conditional on x, the actual value of the creditworthiness, u, does not influence the accept/reject decision.

In Case 1, since $a = a(x)$, we see that this applies, so that this case involves data which are missing at random. In Case 1 the values of u are missing at random. Unfortunately, since complete outcome data are observed in region A and no outcome data in region R, this does not expedite the analysis: extrapolation is necessary.

In Case 2, since the additional information in v is likely to be related to the creditworthiness variable u (else why was v used?), u is non-ignorably missing.

Of course, we only observe y, not u. The variables u and y are related. A simple model takes

$$\begin{cases} y = 1 & if\ u > s \\ y = 0 & otherwise \end{cases}$$

where s is a threshold. A more sophisticated model acknowledges that there are additional random influences on outcome, even given someone's creditworthiness, and takes $y = 1$ if $u + \delta > s$, where δ is a random variable, but we will not describe this in this summary. The simple model assumes that 'creditworthiness' is the sole determinant of outcome. Hence

$$P(y = 1 \mid x) = \int_s^\infty f(u \mid x)du = P(a = 0 \mid x) \int_s^\infty f(u \mid x, a = 0)du + P(a = 1 \mid x) \int_s^\infty f(u \mid x, a = 1)du$$

$$= P(a = 0 \mid x) \int_s^\infty f(u \mid x, a = 0)du + P(a = 1 \mid x)P(y = 1 \mid x, a = 1)$$

In the right hand side of this expression, the probabilities $P(a = 0 \mid x)$ and $P(a = 1 \mid x)$ can be estimated immediately from the retrospective database as the proportions of customers rejected and accepted. Similarly, the probability $P(y = 1 \mid x, a = 1)$ can be estimated immediately from the retrospective database as the proportion of accepted customers who have good outcomes. Unfortunately, as explained above, we cannot estimate $\int_s^\infty f(u \mid x, a = 0)$ from the data. Reject inference describes attempts to infer the creditworthiness status of the rejected applicants, so that $\int_s^\infty f(u \mid x, a = 0)$ and hence $P(y = 1 \mid x)$ may be estimated.

In order to tackle Case 2, it is necessary to obtain extra information. This can take various forms including (i) assumptions about the forms of the distributions involved, and (ii) information from other suppliers on outcomes of rejected applicants.

There are two distinct estimation strategies for the non-ignorably missing case. The *selection model* postulates an explicit model for the missing data probabilities, so that

$$f(u, a \mid x, \theta, \phi) = f(u \mid x, \theta)P(a \mid x, u, \phi).$$

In contrast, the *pattern-mixture model* describes the marginal distribution of u as a mixture over the missing data patterns:

$$f(u, a \mid x, \phi, \pi) = f(u \mid x, a, \phi)P(a \mid x, \phi),$$

where θ and ϕ are parameters of the respective models.

The most famous selection model is that due to Heckman. This is based on making assumptions about the forms of the distributions, in particular, explaining the influence of u on the probability that $a = 0$ via the relationship between u and an unobserved variable v by assuming that u and v have a bivariate normal distribution:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N(\begin{pmatrix} x^T\beta_u \\ x^T\beta_v \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix})$$

with

$$P(a = 0 \mid x, u, v, \psi) = \begin{cases} 1 & v > 0 \\ 0 & v \leq 0 \end{cases}$$

From this it follows that

$$P(a = 0 \mid x, u, \psi) = \Phi(\frac{x^T\beta_v + \rho(u - x^T\beta_u)/\sigma}{\sqrt{1 - \rho^2}})$$

We see immediately from this that, in the special case of no correlation between u and v, the probability that u will be missing is independent of u, given x - we have the MAR case.

In general, we then have

$$f(u \mid x, \theta, \psi) = \frac{f(u \mid x, a=1, \theta) P(a=1 \mid x)}{1 - P(a=0 \mid x, u, \psi)},$$

with all of the components on the right hand side being estimable. From this, of course, $P(y=1)$ is estimated by integration.

Unfortunately, it seems that selection models are sensitive to the assumptions made.

3.2 Bayesian Marginalization

In designing retail credit risk models there appears to be widespread belief that models should recognize major changes or disruptions in 'lifestyle' variables - examples include such events as divorce, termination of employment, heart attack. In many cases, the lifestyle data is only available in small samples at critical points of time and may be buried or hidden from view in the data on characteristics of individuals that is normally available. Nevertheless, the effects of lifestyle data may influence some of the characteristics in the standard datasets. Given a dataset D_n we want to make the prediction

$$p(y_{n+1}|D_n) \qquad D_n = \{y_i, \mathbf{x}_i\} \qquad 1 \le i \le n$$

by using information from a 'latent' or unobservable variable θ, and model parameter estimates, β. Note that the index on the y (say default, fraud, "creditworthiness") to be predicted is $(n+1)$ whereas the dataset has a subscript of n. Assuming that posterior densities can be obtained for all parameters the predictive density for y given the data is

$$p(y_{n+1}|D_n) = \int p(y_{n+1}|D_n, \beta, \theta) p(\beta, \theta | D_n) d\beta d\theta.$$

The first term inside the integral is independent of the data D_n given the parameters β and θ, i.e.

$$p(y_{n+1}|D_n, \beta, \theta) = p(y_{n+1}|\beta, \theta);$$

thus, the conditioning on the dataset used to develop the model has been removed. The second term is the posterior distribution of the unobservable parameters and latent variable conditioned on D_n. We have

$$p(\beta, \theta \mid D_n) = p(\theta \mid \beta, D_n) p(\beta \mid D_n),$$

in which $p(\theta \mid \beta, D_n) = p(\theta \mid \beta)$, the parametric distribution adopted for θ. For example, $p(\theta \mid \beta) = \lambda(\beta) \mathrm{e}^{-\lambda(\beta)\theta}$, where $\lambda(\beta)$ is a given function of β.

In practice, accurate estimates of the βs are derived from the data D_n: we denote these estimates by $\hat{\beta} = \hat{\beta}(D_n)$ and assume that the posterior $p(\beta \mid D_n)$ is effectively concentrated at $\hat{\beta}$. Then, the predictive equation can be replaced by the approximation

$$p(y_{n+1}|D_n) \approx \int p(y_{n+1} \mid \hat{\beta}, \theta) p(\theta \mid \hat{\beta}) d\theta.$$

Use of the method rests on trying out various models for the effect of the latent variable, assessing the effect of the specified conditional density function $p(\theta \mid \beta)$.

It is possible that the latent variable can be observed and measured in a separate experiment unrelated to the data used for deriving the parameter estimates of β. To obtain $p(y_{n+1}|D_n)$ it is suggested to use a two-stage procedure where the form of $p(\theta|\beta, z)$ is first estimated from a small-scale survey where information on θ can be obtained from additional observations of z at some significant cost. (In general, z can also be a vector although, for simplicity, we only consider a single variable). Then one can obtain a new (approximate) predictive density

$$p(y_{n+1}|D_n, z) = \int p(y_{n+1}|\hat{\beta}, \theta) p(\theta|\hat{\beta}, z)$$

in which we have assumed that

$$p(y \mid \beta, \theta, D_n, z) = p(y \mid \beta, \theta) \quad \text{and} \quad p(\beta, \theta \mid D_n, z) = p(\theta \mid \beta, z)p(\beta \mid D_n).$$

The steps recommended to obtain $p(y_{n+1}|D_n, z)$ are:

1. In a small-scale (possibly expensive) survey, relate θ to z;
2. In that survey test that y does not significantly depend on z;
3. Infer estimate of θ or distribution of θ from β and z;
4. Calculate the predictive distribution of y_{n+1} given D_n and z.

Steps 1 through 4 are similar to those performed in 2-stage least squares estimation with latent variables (see, e.g., [21]) except that here we switch from one (small-scale) data set when estimating θ, another when estimating $p(y|\beta, \theta)$.

Another approach that may be more efficient is to use maximum likelihood estimation where the first $p(\theta|\beta, z)$ and second $p(y|\beta, \theta)$ stage equations are estimated simultaneously (iteratively back and forth) using for example the EM algorithm developed by Dempster et al [23]. The point of a simultaneous estimation is that the first-pass estimate of stage 1 in the 2-stage approach above might not be the most efficient. The simultaneous estimation approach poses a technical challenge as one is toggling between two different data-sets to converge on global parameter estimates.

3.3 Dropout/Withdrawal Inference

This problem can be formulated in a similar way to the reject inference problem where we define

$$a = \begin{cases} 1 & \text{if applicant declines the offer made} \\ 0 & \text{otherwise} \end{cases}$$

$$y = \begin{cases} 1 & \text{if applicant is creditworthy} \\ 0 & \text{otherwise,} \end{cases}$$

and where x is a vector of applicant descriptors.

In this case we know even less about the selection mechanism than in the reject inference case. This means there are likely to be even more unobserved variables in the selection equation than in the reject inference case. This makes the bias from using the drop-in sample even more likely. The same analysis as above holds with MNAR more likely. The magnitude of the bias depends crucially on how able we are to specify the true selection equation. Our hypothesis, following Åstebro and Bernhard (2003, see [17]) is that the better credit worthy applicants are likely to self-select out from the pool of applicant given a single price offer. Indicators of that self-selection bias are possibly, education, work experience and other variables that makes the applicant go to suppliers of credit that are more sensitive to this information, typically friends family and angel investors. In a credit market where other lenders are also not able to evaluate human capital, and in addition there is little or no informal market for credit, then this problem diminishes. In the U.S the informal market is rather large and so the problem is probably of some significance.

Selection will also depend on offers made by competing credit suppliers. These offers will be dependent on applicants characteristics and on anticipated offers by other suppliers.

4 Decision models and dynamic control actions

4.1 Issue

Credit management can be seen as a control problem. The aim of the exercise is for managers to take actions to achieve business objectives. The process by which actions are chosen is the focus this workshop. The decision process can be thought of as a function mapping from states describing the lender, the borrower, and any relevant context to the actions. What form should such a function take?

The motivation behind this workshop was the belief that, currently, decision functions constitute a very small and specialized subset of the space of all possible decision functions. The concern is that the subset which is commonly used, is used for historical or other accidental reasons, rather than fitness to purpose. Current decision functions typically do not choose actions to optimise some objective function; they usually only consider the effect of the current action and not possible future actions; they usually do not take into account the sophisticated analysis of the interplay between the lender and the borrower; and they typically do not consider the possibility of a wide range of actions at one time. Are there classes of decision functions which are particularly well-suited to the types of problems facing the credit manager?

Often, there exists the well-known stove-pipe situation, where different analytical models in the life of an account do not effectively communicate in terms of integrating decision models. Acquisition models, for example, often do not allow for the fact that there are numerous actions that can be taken by risk management to control accounts and improve profitability. Thus, an average case is assumed for the behavior of the account. It is unlikely that a global dynamic control model, i.e., one that covers the entire life-cycle of an account or portfolio of accounts, can or will be constructed in the near future. This is a result of the complexity and data requirements for such a model. It is our belief that dynamic control models can be built in the near future that cross the individual traditional stovepipes, for example, from acquisition to risk management.

4.2 Aims

The aim of this work was to provide a conceptual framework for modelling individually and collectively the decisions involved in the consumer lending process. This would go from prospect mailing, at one end, to longevity bonuses, attrition modelling, collection scoring, and rebranding at the other. One would not expect to have one model to cover the whole process, but one could expect coherence between the models where coherence would mean consistency of the objective functions and descriptions of the answers that are consistent when the models of different phases of the process overlap. This would allow the scoring community to identify the issues that need to be modelled: risk-based pricing, adaptive adjustments in product features at the individual customer level, appropriate segmenting of the population, and overall customer-profit optimisation.

4.3 A dynamic programming approach

The generalised approach that was beginning to be formulated during the workshop has much in common with a dynamic programming or optimal control approach. This is not really surprising, since the idea of using this approach on a very simple consumer credit model was suggested forty years ago by Cyert and Davidson (1962) and Liebman [1]. What is more surprising is that the ideas have not been taken up for practical implementation by the industry. We speculate that this is partly because of the division into the distinct stovepipes mentioned above: those concerned with acquisitions have different optimality criteria from those concerned with risk, and there is relatively little communication between them.

The problem has various critical dimensions which would need to be taken account of in any complete model, including: the number of periods of time, the number of players (a single lender; a lender and a borrower; a lender and two types of borrower; more). The model can be made as complex as one wants, and the trick, as in all scientific modelling, will be to establish a model which is simple enough to be implementable in practice, but sophisticated enough to be useful. It was too much to hope that such a model could be developed at the meeting itself.

5 Scorecard Alignments

5.1 Problem:

Model Updating. During the development phase much effort is made to ensure that the scorecard is as predictive as possible on the holdout sample. However, the performance of the scorecard soon begins to deviate from the theoretical performance when applied to new data in the live environment.

5.2 Questions:

What level of performance deterioration warrants a scorecard rebuild?
When to realign versus when to rebuild?

In the case of deteriorating performance three options are considered:

1. Scorecard realignment

   It is common for a scorecard to undergo realignment at regular periods (usually 12 months). The process involves assessing previously scored data for which the full outcome period is observed. The observed score to odds relationship will be determined using linear regression and adjustments made to the slope and intercept of the regression so that the scores will be aligned with the standard score to odds relationship as shown in Fig. 6.

2. Scorecard weights re-estimation

   The cost associated with compiling a data set would be the same whether a full re-development or a weight re-estimation were being carried out. Therefore this approach is quite uncommon and would only happen if other system issues were pertinent.

3. Redevelop scorecard

   Due to the desirability of a stable live system and the prohibitive cost of a new rebuild, scorecards often are used for long periods of time. Regular model developments do not happen.

5.3 When to rebuild

During the development phase much effort is expended to achieve high predictability of a new scorecard on the training sample. Much of this effort goes towards defining suitable x-variables derived from the application information. However, the performance of a scorecard deteriorates as it continues to be applied to new applications in the live environment. This loss of discrimination carries a cost. Thus, a decision has to be made on when to renew the scorecard, i.e. to find the optimal balance between the costs of renewal and non-renewal.

An applicant for credit is required to supply various details such as age, employment status, residence status, etc. On the basis of this information, summarized as a vector $x = (x_1, ..., x_p)$ of scores, a decision is made as to whether to issue credit or not. A common vehicle for this decision is a logistic regression function that has been developed on a training sample. Thus, the training data have been used to estimate the parameters $\beta = (\beta_1, ..., \beta_p)$ in the model

$$\mathrm{P}(G \mid x) = 1/(1 + \mathrm{e}^{-x^{\mathrm{T}}\beta}),$$

where $\mathrm{P}(G \mid x)$ denotes the probability that an applicant with score vector x will be a 'Good' (defined as one who will always repay a loan). The model may be re-expressed in the form

$$\log[\mathrm{P}(G \mid x)/\{1 - \mathrm{P}(G \mid x)\}] = x^{\mathrm{T}}\beta.$$

The left-hand side here is the *log-odds*, odds being the ratio of Good to Bad probabilities, and the right-hand side is the score combination. In practice, the score combination is subjected to a linear transformation, say as $a + b(x^{\mathrm{T}}\beta)$, in order to make the resulting *score* take values in a standardized range. The standard straight-line relationship of log-odds to score is shown as the bold line in Fig. 1. A score threshold is set and only applicants with a score above the threshold will be accepted, i.e offered credit.

There are various levels of renewal of a scorecard, each incurring a significant cost. However, we are here concerned only with the cost of non-renewal, and for this we need some notation. Let

$\gamma_G(t_0, t)$ be the probability of accepting a Good at time t using a scorecard developed at time t_0, where $t_0 \leq t$. Likewise, let $\gamma_B(t_0, t)$ be the probability of accepting a Bad. One common practice is to set the score threshold to achieve a given ratio of accepted Goods and Bads; for illustration, we will take this ratio to be 15/1.

Among applicants accepted at time t, using a scorecard developed at time t_0, the odds of Goods to Bads is

$$\begin{aligned}\phi(t_0, t) &= \mathrm{P}(Good \mid accepted)/\mathrm{P}(Bad \mid accepted) \\ &= \mathrm{P}(accepted \mid Good)\mathrm{P}(Good)/\mathrm{P}(accepted \mid Bad)\mathrm{P}(Bad) \\ &= \{\gamma_G(t_0, t)\pi_G\}/\{\gamma_B(t_0, t)\pi_B\},\end{aligned}$$

where $\pi_G = 1 - \pi_B$ is the overall proportion of Good applicants.

(For simplicity, it is assumed here that π_G is constant over time.)

For illustration, suppose that, at time t, there are N_t applicants for a loan of 5000 (pounds) at interest rate r, so the amount to be repaid is $5000(1 + r)$. On average, there will be $\pi_G N_t$ Good applicants, of whom $\gamma_G(t_0, t)\pi_G N_t$ are accepted on the basis of the t_0-vintage scorecard. Likewise, the expected number of Bads accepted is $\gamma_B(t_0, t)\pi_B N_t$. Each accepted Good will yield a profit of $5000r$ and each accepted Bad will lead to a loss of 5000. (For simplicity, we assume that the whole amount is lost to a Bad, otherwise a specified fraction of 5000 is to be applied.) Thus, the expected net profit from the N_t applicants is

$$5000r\gamma_G(t_0, t)\pi_G N_t - 5000\gamma_B(t_0, t)\pi_B N_t = 5000N_t\gamma_B(t_0, t)\pi_B\{r\phi(t_0, t) - 1\}.$$

Thus, the expected profit difference between using an up-to-date scorecard and one developed at time $t_0 < t$, is

$$5000N_t\pi_B[\gamma_B(t, t)\{r\phi(t, t) - 1\} - \gamma_B(t_0, t)\{r\phi(t_0, t) - 1\}].$$

Suppose, for example, that $\gamma_G(t, t) = 0.90$, $\gamma_B(t, t) = 0.06$, $\gamma_G(t_0, t) = 0.84$ and $\gamma_B(t_0, t) = 0.07$; then $\phi(t, t) = 15$ and $\phi(t_0, t) = 12$. Also, suppose that $r = 0.1$. Then, the expected difference is

$$5000N_t\pi_B[0.06\{1.5 - 1\} - 0.07\{1.2 - 1\}] = 80N_t\pi_B.$$

This cost can be set against that of renewing the scorecard and thereby contribute to the decision process.

5.4 How to determine when a rebuild is needed

When the cut-off score is chosen in the scorecard development process this is defining $\gamma_G(t_0, t_0)$ and $\gamma_B(t_0, t_0)$ - they are the coordinates of the point on the ROC curve chosen as the cut-off. When one starts implementing the scorecard one is actually calculating $\gamma_G(t_0, t_0 + d)$ and $\gamma_B(t_0, t_0 + d)$, where d is the lead time between the development sample observation point and the scorecard implementation point. It seems reasonable to assume that, if a scorecard was left for ever, it would lose all its discrimination, i.e. $\gamma_G(t_0, \infty) = \gamma_B(t_0, \infty)$, which is a point on a diagonal ROC curve. If nothing is done to the scorecard, and if one assumes that the deterioration is the same for goods as bads then it moves on the straight line on the ROC curve between these two points. If there are regular recalibrations, so that one keeps the same accept rate, then one moves on a line where $a(t) = \gamma_G(t_0, t)(1 - \pi_B) + \gamma_B(t_0, t)\pi_B$ is a constant and ends up at a different point on the ROC curve diagonal. In either case one also has to estimate the speed of movement along these curves. The simplest reasonable movement would be negative exponential so that, after time t, if the length of the curve is a, the scorecard would have moved $a(1 - \mathrm{e}^{-bt})$ along it. One can estimate b, and also check the validity of the negative exponential assumption by looking at the delinquency reports, which are segmented by start date and duration of account. When one has these estimates of γ_G and γ_B one can apply the decision structure described in Section 5.3 to determine when a rebuild is likely to be advantageous, but this has yet to be done in practice.

6 Summary and Conclusions

The BIRS meetings and workshops gave attendees the opportunity to informally discuss and describe some of the underlying mathematical and statistical risk problems associated with retail credit. We concluded that the published literature on these types of problems at both the account and portfolio level is sparse.

One needs to build on the success of consumer credit scorings ability to make assessments of relative likelihoods of default and other risk outcomes to make good time-dependent probability and categorical forecasts. There is considerable confusion in the credit literature between probability forecasts and categorical forecasts and how the former, in conjunction with business and financial assumptions lead to decision rules (e.g. an Accept/Reject rule) that can then be used to derive the latter.

There appear to be important difficulties in understanding the theoretical and experimental ways in which retail risk scores can be transformed and converted into absolute default probabilities largely because traditional scoring has been used to establish relative improvement in simple statistical and business measures that depend on the probability of default. To the best of our knowledge there are no documented academic or industry attempts to incorporate multiple economic factors into the population odds of default or into the stochastic process describing the evolution of future scores.

We discussed the important behavioral risks associated with consumer lending, including the formulation of the consumer as an active participant in the borrowing/lending negotiation process. These negotiations depend on the preference functions of both parties. We presented a preliminary borrowing/lending model that incorporated many aspects of negotiation and exchange between lender and borrower and concluded that experiments to evaluate preferences are a prerequisite to developing models and strategies for customization of financial products.

Simulation techniques in part depend on relevant performance data but it was by no means clear that we have common agreement on probability models that capture the essential features of the prediction and decision-making structure as an integral component of the lending process.

There was some discussion as to how artificial neural networks could and could not be used to predict (on the basis of past data fits) the outcomes of random quantities that depend on future controls and actions not recorded in the historical fitting exercises.

Work was undertaken to develop models based on the monitoring tools in consumer credit risk systems to identify when such systems needed recalibration, re-estimation and replacement.

A part of one workshop discussed risk-based pricing models that included the propensity of the customer to take a loan offer at a specified price and term; there was considerable debate as to the form of the elasticity of response to price of risk. Some specific models were discussed and there was general agreement that this important area needed considerable future study and experimentation.

There was unanimous agreement that there are enormous differences between retail and wholesale credit financial markets because retail risk is affected by social behavior as well as by business cycles and economic factors. A further complication is that there is scant pricing information available for the purchase and sale of retail loan portfolios in either a primary or secondary market. These additional complexities offer important research challenges to academics and practitioners alike.

There appeared to be near-unanimous agreement that the parameters and models that are widely used in measuring, assessing and predicting default risk in wholesale commercial loan markets cannot be applied to retail loan portfolios.

In one workshop we discussed an approach that could be used to cope with the highly dimensional models that are traditionally used to study correlated returns (which are used in the standard value at risk models for liquid corporate securities). The approach would be to correctly formulate a portfolio loan model that, at its core, has a strong conditional independence structure in which correlations are induced by marginalizing over one or more common-cause factors.

A set of problems and models that attracted considerable interest was the inclusion of the proposed Basel II capital accords in setting required levels of regulatory and equity capital. These requirements would recognize the risk contributions of different loan types as well as the composition of individual/behavioral risk profiles. Several models and theoretical studies were proposed.

References

[1] L.H. Liebman, A Markov decision rule for selecting optimal credit control policies. *Management Science*, **18**, B519-B525, 1972.

[2] F. Black and M. Scholes, The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–654, 1973.

[3] *CreditMetrics*, Technical Document, JP Morgan, Inc., 1997.

[4] *Credit Risk+: A Credit Management Framework*, Credit Suisse Financial Products, 1997.

[5] M. Crouhy, D. Galai, and R. Mark, A comparative analysis of current risk models. *Journal of Banking and Finance*, **24**, 59-117, 2000.

[6] R.A. Jarrow, D. Lando, and S.M. Turnbull, A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**, 481-523, 1997.

[7] D. Lando, On Cox processes and credit risky securities. *Review of Derivatives Research* **2**, 99-120, 1998.

[8] R. Merton, On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **28**, 449-470, 1994.

[9] M. Saunders, *Credit Risk Measurement*, Wiley, New York, 1999.

[10] Schonbucher, J. Philipp, *Factor Models for Portfolio Credit Risk*, Department of Statistics, Bonn University, 2000.

[11] M. Stepanova, L.C. Thomas, PHAB scores: Proportional hazards behaviour scores. *Journal of Operational Research Society* **52**, 1007-1016, 2001.

[12] O. Vasicek, *Credit valuation Net Exposure 1*, 1997.

[13] F. Wedling, Ph.D. thesis, University of Sao Paulo, Brazil, 2003.

[14] T.C. Wilson, Portfolio Credit Risk in Credit Risk Models and Management. *Risk Books, (D. Shimko, ed.)*, London, 23-42, 1999.

[15] C. Zhou, A jump diffusion approach to modelling credit risk and valuing defaultable securities, Working Paper, Federal Reserve Board of Governors, 1997.

[16] T. Åstebro and I. Bernhardt, Start-Up Financing, Owner Characteristics and Survival. *Journal of Economics and Business*, **55(4)**, 303-320, 2003.

[17] T. Åstebro and G. Chen, Bound and Collapse Bayesian Reject Inference When Data are Missing not at Random. paper presented at the *BANFF Credit Risk Conference*, October 11-16, 2003, Banff, Canada.

[18] J.B. Banasik, J.N. Crook, and L.C. Thomas, *Sample selection bias in credit scoring models.* Working paper 01/5, Credit Research Centre, University of Edinburg, England, 2001.

[19] A.J. Feelders, Credit scoring and reject inference with mixture models. *International Journal of Intelligent System in Accounting, Finance and Management*, **9**, 1-8, 2000.

[20] W.R. Gilks, S. Richardson and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall, 1996.

[21] W. Greene, *Econometric Analysis, 4th edition* , London, Prentice Hall International, 2000.

[22] R.J.A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, 1987.

[23] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Statist. Soc. B*, **39**,1-38, 1977.