Application of Robust Portfolio Modeling to Credit Scoring Problem

Helsinki University of Technology
System Analysis Laboratory
Marko Tuominen 55179D
Department of Engineering Physics and Mathematics
Contents

1 Introduction 2

2 Credit Scores and Cutoff Policies 4
   2.1 Credit Scoring Problem 4
   2.2 Credit Scoring Models 5
   2.3 Problem of setting optimal cut-off 7

3 Robust Portfolio Modeling 11
   3.1 The framework for optimizing cutoffs 11
   3.2 Validation of results 13

4 Results 15

5 Conclusions 19

References
List of Figures

1  An example of efficient frontier for setting optimal cutoffs. . . . . . . . . . . . . . . . 7
2  A second example of efficient frontier for setting optimal cutoffs. . . . . . . . . . . . . . . 8
3  A third example of efficient frontier for setting optimal cutoffs. . . . . . . . . . . . . . . 9
4  Setting cutoffs when multiple scores are available. . . . . . . . . . . . . . . . . . . . . . 10
5  Robust Portfolio Modelling as a process(from [LM805], https://www.ri.dtu.dk/ ) . . . 12
6  \textit{minCore} – \textit{maxExterior} classification rule. . . . . . . . . . . . . . . . . . . . . . 14
7  Histograms of scores and core index in validation sample. a) Left above, logistic-
   regression score, b) Right above, linear-regression score, c) Left below, k-NN score, d)
   Right below, MC-RPM core index. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
8  Scorevice scatter plots. a) Left, linear vs. logistic score. b) Right, linear vs. k-NN score. 17
9  ROC-curves of scores in validation sample. a) Red, logistic-regression score, b) Green,
   linear-regression score, c) Blue, k-NN score. d) Straight line, ROC curve, when decision
   making process is random. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

List of Tables

1  Performance of MC-RPM classified sample by Core Index range. . . . . . . . . . . . . . . 16
2  Performance of \textit{minCore} – \textit{maxExterior} classified validation sample. . . . . . . 16
3  Misclassification rates. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
1 Introduction

Lending institutions employ in their credit decisions, both in consumer lending and in retail banking, credit scoring models [LS02]. Automated decision-making and credit modelling systems had speeded up loan processing considerably in last few decades [May04]. These credit scoring models have enlarged from so called acquisition scoring models, used only in loan acquisition phase, to the most important management tools in the industry [LS02] [Tho00].

Many credit scoring systems are based on single score, which seeks to estimate the creditworthiness from historical data and present it in an one figure. Creditworthiness of the applicant must be defined and it may vary greatly among different institutions and sometimes among products. Combination of creditworthiness estimates from different models has been under examination [ZBO01]. For most models, the best cut-off level, that is, the level below which applicants score are denied for credit must be set.

For LP, ILP [JS90] [Tho00] as well as vector support machine approaches [BVVS03] to credit discrimination problem the optimal cut-off will be parameter itself. For other models, the cut-off must be set after the model has been estimated. Traditionally cut-off for scorecard is set using one criterion only - expected loss or bad rate [May04]. Approaches have evolved towards multi-objective frameworks in cut-off setting [W001]. These are referred as efficient frontier methods. Objectives may be technical, such as, minimum number of misclassified applications and maximum number of correctly classified or economical, such as minimum expected losses or maximum return on equity [May04]. In the case of many scores the problem of multi-criteria optimization must be solved.

This paper presents an application of Robust Portfolio Modelling [LMS05] to the credit scoring problem, in particular, setting optimal cut-offs, when scores from several models are available. The scores have been calculated using three different models for classification (logistic regression, linear regression and k-Nearest-Neighbour method). The optimal cut-off is found as a solution to multicriteria optimization problem, where information on crite-
rion weights and true scores are allowed to be incomplete. This approach of RPM to credit scoring context allows one to seek cut-off policies that are robust over different conditions (economic cycles and management preferences), which will have influence on parameters (scores and weights).

The paper is organized as follows, section 1 is an introduction, section 2 introduces credit scoring problem as a classification problem, and gives a short literature review over recent developments of models proposed to solve the problem as well as policies for setting optimal cut-offs. Section 3 illustrates the Robust Portfolio Modelling framework and measures to validate its ability to set cut-offs for multiscore applications. RPM’s classification ability with regard to loan applications, is also validated. Section 4 presents results. Section 5 concludes.
2 Credit Scores and Cutoff Policies

2.1 Credit Scoring Problem

The description of the problem is simple; to whom to give credit. The decision must involve consideration of the risk associated in the lending situation, i.e., positive probability of not getting receivables back. This is essentially the same as classifying applicants good and bad credit risks. Traditionally this is done by human judgment based on things like, applicant’s character, employment situation, wealth etc. [May04].

Today the problem faced by credit industry is the same, although more sophisticated questions must also be answered. Problems may involve loss forecasting, choosing right operating level, that is, profit, risk, volume etc. combination for loan portfolio manager’s risk attitude. These problems ask for integration of classification models (i.e. loan level risk forecasting) and portfolio management models [TW04]. Scoring has also moved from credit classification towards loan level profitability forecasting, referred as profitability scoring [Tho00].

Practical problems arise when the model is to be implemented. These are significant and should be paid attention in model selection phase. One of the problems, when building the scorecard is so called sample selection bias. This refers to the fact that usually only information on previously approved applications is available [May04]. This is usually due improper information such as missing values or false information in rejected applications, cost of storing the data and legislation [May04]. For instance, in some countries lenders are allowed to keep information on rejected applicants only a limited period of time, after which data must be removed, if the particular applicant has not sent another approved application mean while. An excellent study on the effects of sample selection bias, see [Par03].
2.2 Credit Scoring Models

The first statistical classification rule is from Fisher [Fis34]. If c is the applicant’s characteristics, $\mu_G$ and $\mu_B$ are the expected values of populations characteristics (good and bad credit risks) and $\Sigma$ is pooled covariance matrix of these. Then c belongs with group of goods if:

$$c^T\Sigma^{-1}(\mu_G - \mu_B) \geq \frac{1}{2}(\mu_G + \mu_B)^T\Sigma^{-1}(\mu_G - \mu_B),$$  \hspace{1cm} (1)

else c belongs with group bads. This is so called Fisher’s linear classification function (FLC) and is based on assumption of multinormally distributed applicant characteristics with equal covariance matrices. Another classical rule is from Smith [Smi47]. Here c is classified to group goods if,

$$(c - \mu_B)^T\Sigma_B^{-1}(c - \mu_B) - (c - \mu_G)^T\Sigma_G^{-1}(c - \mu_G) \geq \ln |\Sigma_G| - \ln |\Sigma_B|.$$ \hspace{1cm} (2)

else c is classified to group bads. Here $\Sigma_G$ and $\Sigma_B$ are unpooleed covariance matrices of populations. This rule is referred as quadratic classification function (QDF). These both (LCF and QCF) have worked relatively well in comparison to more modern classification methods, see [HH97] and [JS90].

LP and ILP models are presented in [JS90] [Wil96] [Tho00] [TEC02] [BVVSV03]. Equation (3) below, shows an example of classification problem formulated as mixed-ILP. This formulation was proposed by [Geh86].

$$\min_{w, c} \left\{ \sum_i \delta_{i1} + \sum_h \delta_{h2} \right\},$$ \hspace{1cm} (3)

$\sum_j x_{ij}w_j - M\delta_{i1} \leq c$, for all cases i in group 1
$\sum_j x_{hj}w_j + M\delta_{h2} > c$, for all cases h in group 2
$\delta_{i1}, \delta_{h2} = \begin{cases} 1 & \text{if case k is misclassified} \\ 0 & \text{if case k is correctly classified} \end{cases}$
$w_j$, unrestricted

Here, $x_{ij}$ and $x_{hj}$ denote j th characteristics of i th and h th applicant, w contains weights for characteristics, c is optimal cut-off and $\Sigma_B$ is large constant, which penalizes for misclassifications. The objective is to minimize
number of misclassifications. This formulation is computationally simple, but
gives very good classification results [JS90].

Other optimization approaches such as combinatorial optimization solved
via genetic algorithms are presented in [TEC02] and vector support machines
(reduces to solving convex QP) are presented in [BVVSV03]. Statistical parametric approaches such as linear- and logistic regression are clearly presented
in [May04], [TEC02] and [Sun05]. Non-parametric statistical approaches such
as k-nearest-neighbour [HH96] and neural networks are illustrated in [TEC02]
and [Sun05].

K-nearest-neighbour method works as follows. An estimate of applicants
creditworthiness is the proportion of bads/goods within k most similar applicants. Similarity is measured using some metric based on applicant characteristics. Before calculating distances among applicants, the characteristics
are WoE\(^1\) transformed. Characteristics are replaced by,

\[ w_{ij} = \ln \left( \frac{p_{ij}}{q_{ij}} \right), \tag{4} \]

where \( p_{ij} \) is the proportion of goods and \( q_{ij} \) proportion of bads in \( j \) th
value of \( i \) th characteristics. The distance between applicants \( j \) and \( h \) can be
calculated e.g. as,

\[ d(c_j, c_h) = \sqrt{(c_j - c_h)(I + De_p e^T_p)(c_j - c_h)}, \tag{5} \]

where \( e_p \) is an estimate for the direction of equiprobability contours and
\( c_i \)'s are applicants’ characteristics, see more [HH96].

Sometimes scoring models are divided to direct and indirect methods
[HL02]. Direct model refers to methods discussed above, that is, directly
classifying applicants as good or bad, whereas indirect model refers to an
approach where applicant characteristics are estimated. Score is derived after
wards, see more [HL02].

Good reviews on scoring models are [Tho00] and [HH97]. An essential
book on subject is [TEC02].

\(^1\text{Weight of Evidence}\)
2.3 Problem of setting optimal cut-off

The problem of setting optimal cut-off is a general multi-objective optimization problem. Criteria are usually, minimum percentage losses, maximum profit to sales and maximum sales [May04] [OW01]. Maximum sales objective tries to capture the effects of possibly involved real options. For instance, it will make sense to sacrifice profitability on short-term to get more market share, and thus build larger distribution channel, where more profitable products can be marketed later. For setting optimal cut-off the most traditional approach is to look approval rate versus expected losses. This is presented in figure 1 below.

![Graph](image)

Figure 1: An example of efficient frontier for setting optimal cutoffs.

Horizontal axis represents expected acceptance rate in percents, while vertical axis represents expected percentage bad rate. Each point in frontier is uniquely linked to particular score cut-off, such that the lower the cut-off the higher the expected acceptance rate. Efficient frontier methods concentrate on finding out management’s risk attitude and the selecting of so called efficient (pareto-optimal) operating point for the institution. The problem here is to choose a desirable operating point for the lender from given frontier
(blue points). Before new scoring system is introduced the operating point of the firm might be the red point in figure 1. Possibilities of the use of new scoring system can be presented as two alternatives. Acceptance rate can be kept as it is and expected losses reduced or the previous loss level can be kept while sales is increased (acceptance rate).

The second method for setting optimal cut-off is illustrated in figure 2 above. Here expected losses are compared to expected profits. This is very common way of setting optimal cut-offs [OW01], that is, to compare risk versus profit. The lower part of the frontier (green curve) is so called efficient frontier whereas the above part is clearly inefficient. Current operating point might be the red circle in figure. The same profitability can be achieved with far lower level of expected losses, however. This is the case when new operating point of the lender is moved to filled red point in figure 2. It should be noted that operating point at the red circle might still be desirable in some situations since it yields higher sales. This is considered in figure 3 next page.

Here all, expected losses, expected profits (axis on plane) and expected sales (vertical axis) are considered when optimizing the cut-off. In the case
Figure 3: A third example of efficient frontier for setting optimal cutoffs.

of single score the problem can be formulated as follows,

$$\max_c \left\{ w_1 A(c) - w_2 B(c) + w_3 C(c) \right\}. \quad (6)$$

$$\sum w_i = 1$$

$$A(c) = \int_{c}^{UB} (1 + r) \ell(x)f(x)dx$$

$$B(c) = \int_{c}^{UB} (1 + r)\ell(x)f(x)\lambda(x)dx$$

$$C(c) = \int_{c}^{UB} r(1 - \lambda(x))\ell(x)f(x)dx$$

$$c = \text{Optimal cutoff}$$

$$UB = \text{Upper Bound of score}$$

$$r = \text{Interest rate}$$

$$\ell(x) = \text{Distribution of sales in score range}$$

$$\lambda(x) = \text{Bad rate}$$

$$f(x) = \text{Score distribution}$$

The solution to this is one red point in figure 3 above. When one wants to use more than one score this problem becomes somewhat more complicated. If one replaces optimal cut-off c by vector $c = (c_1, \ldots, c_n)$ and integrates over all possible $c$s, then the solution is one fixed cut-off for each score $c_i$. 
This is illustrated in figure 4 below in the case of two scores.

![Figure 4](image)

**Figure 4:** Setting cutoffs when multiple scores are available.

One can accept all applications that score above cut-off level at every score, this corresponds to the white area in figure 4, and rejects applications that score below the cut-off at every score, and this policy corresponds to the black area in figure 4. This approval strategy would leave the majority of potential business to the gray area. More information should be gathered to derive lending policy towards these intermediate loans. Cut-off can be optimized dynamically, however. This approach is introduced in figure 5.

It should be noted that sometimes misunderstandings of the use of more than one score can result very inefficient lending policies, indeed [LS02, p. 68]. Economical optimization criteria are desirable with models, where cutoffs are not set during the parameter estimation procedure. Thus this cut-off setting framework excludes the use of some LP, ILP [JS90] and vector support machines [BVVSSV03] scoring models. Also some accounting information on cash flows etc. can be used to construct economical performance measures such as ROA, ROI and ROE to be optimized. Good illustration of this is given in [May04].
3 Robust Portfolio Modeling

The information on criterion weights as well as scores in cutoff optimization problem (6) may not be complete. Scores may vary for many reasons. Variates whose effect on applicant’s score cannot be controlled for and possible errors in estimation procedures cause some uncertainties about the true score. For instance, scores for applicant’s profitability and creditworthiness change over economic situations [TEC02, Chapter 3: Economic Cycles and Lending and Debt Patterns]. The management will also have varying preferences over time, which asks for solutions that perform well under these uncertainties. Robust Portfolio Modelling [LMS06] offers a methodology to solve problems formulated in this way. It seeks to discriminate applications to three distinct groups, namely core applications, borderline applications and exterior applications. Core loans will be robust in the sense that they will score high (and thus will be accepted) under different conditions, such as economic situations and management preferences. To decide for borderline investments more information need to be gathered thus narrowing the set of borderline loans. The process is illustrated in figure 6.

3.1 The framework for optimizing cutoffs

The RPM works as follows. First, information on score ranges and weights is gathered. The uncertainties are modelled through set inclusion, that is, true parameter value is within given range. Formally, let \( \ell_i \) denote loan application \( i \) and \( s_{ij} \) the \( j \) th score of \( i \) th application. Total score of the given application is modelled as additive value, \( V_i = \sum_{j=1}^{n} w_j s_{ij} \), where \( \mathbf{w} = (w_1, \ldots, w_n) \) are weights. The weights must belong to following set,

\[
\{ \mathbf{w} \in \mathbb{R} | w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \}.
\]

Uncertainties about scores will be taken in to account by giving an interval \([\bar{s}_j, \tilde{s}_j]\), where true score in supposed to lie.

The score of the whole portfolio is,
V(p, w, s) = \sum_{\ell_j \in p} V(\ell_j) = \sum_{\ell_j \in p} \sum_{i=1}^{n} w_i s_{ij}.

A particular portfolio \( p \) is feasible if, \( \{ p \in P \mid C(p) \leq B \} \), where \( P \) is a set of all possible portfolios, \( C(p) \) is the cost of the portfolio and \( B \) is the resource constraint vector. A non-dominated set of portfolios is the set, where all portfolios are within resource constraints (feasible) and whose overall score (value) will always be greater or equal to the overall score of any other portfolio within given uncertainties about weights and loan level scores and will be greater at least for one set of weights and loan level scores. For each loan a core index is specified (here \( S \) denotes the information set on weights and score intervals),

\[
CI(\ell_j, S) = \frac{\#\{ p \in P_N(S) | \ell_j \in p \}}{\#P_N(S)}.
\]
Loans are classified using core indexes as follows,

\[
\begin{align*}
\text{Core loans} & : CI(\ell_j, S) = 1 \\
\text{Border loans} & : 0 < CI(\ell_j, S) < 1 \\
\text{Exterior loans} & : CI(\ell_j, S) = 0.
\end{align*}
\]

For each loan the core index will be calculated, that is, all the non-dominated portfolios in which particular loan is contained.

### 3.2 Validation of results

Results are validated using validation sample, which contains 2335 loan applications. Typically scorecard training sample contains loans from few thousands to several millions [May04]. The size of the sample faces some restrictions within exact robust portfolio selection algorithms. This leads to the use of approximative Monte Carlo method, referred MC-RPM hereafter. The following simulation scheme is used,

1. Calculate default probability from development sample.
2. Estimate linear-, logistic regression and k-NN scoring models from development sample.
3. Calculate scores for validation sample.
4. Use few percents as variation range for scores and draw random score matrix and weights.
5. Use default probability from development sample as a budget constraint and select corresponding number of loans and sort them (descending) by overall score.
6. Repeat until core indexes stabilize

The variation for scores \( p\% \) is relative. This is desirable, since error type 1 (bad credit risk is accepted) is greater cost of risk to the lender than error
type 2 (good credit risk is rejected), that is, the greater the scores the more uncertainty is accepted, which is modelled as relative variations. Classified loans are followed by core index over validation sample. Furthermore cut-offs for each score are set using the following approach; if for loan application \( j \) in validation sample,

\[
\begin{align*}
  s_{ij} &> s_{i,\text{minCore}}, \forall i \\
  s_{ij} &< s_{i,\text{maxExterior}}, \forall i
\end{align*}
\]

If (11) holds, application in validation sample is classified as core loan, if (12) holds it is classified as exterior, and if none of these holds it is left as borderline. Here \( \text{minCore} \) and \( \text{maxExterior} \) refer to minimum core loan and maximum exterior loan (according to \( \sum w_is_{ij} \) with equal weights) in classified validation sample. Figure 7 below illustrates this method in two dimensions. Furthermore, if \( s_{i,\text{minCore}} < s_{i,\text{maxExterior}} \) scores are swapped correspondingly.

![Diagram](image)

**Figure 6:** \( \text{minCore} - \text{maxExterior} \) classification rule.
4 Results

The methods of linear-, logistic regression and \( k \)-nearest-neighbour were used to construct three different scores for applications. These scores were used as criteria for RPM. Scores from linear- and logistic regression models were highly correlated with each other, but not that much with the score from \( k \)-NN method.

In \( k \)-NN method \( D \) was set to 1.5 (see equation (5)) and \( k \) to 250. Size of the development sample for each method was 15000 applications. Direction of equiprobability contours was derived from linear regression of variable Weight of Evidences (WoE) as suggested in [HH96]. Calculations were done by simulations scheme described in last section. The loop was repeated 250 times, and score variation was set to 0.25\%. Score weights were random, such that, \( w_1 > w_2 > w_3 \) and \( \sum_i w_i = 1 \) held true. Without these settings i.e. when weights are as loose as possible and variations are greater than few percent, the variation in core indexes become very small. As a result neither core nor exterior loans can be identified.

In \( \text{minCore} - \text{maxExterior} \) setting scorewise comparisons were made for each loan e.g. if loan’s scores are greater than \( \text{minCore} \), for each score it is core index is 1. Otherwise each score vise dominance with regard to \( \text{minCore} \) increases the core index and scorewise non-dominance with regard to \( \text{maxExterior} \) decreases it. Finally index is scaled from 0 to 1. Next page shows classification results.

Table 1 shows results of a sample classified using MC-RPM. Core Indexes do not vary greatly in validation sample. This is visualized in figure 7 on page 17. Core Indexes are also skewed to right in comparison to other scores thus falsely suggesting that riskiness of validation sample is somehow lesser that it actually is. This results from two facts. Firstly, logistic- and linear-regression scores are correlated with correlation coefficient 0.9712 and \( k \)-NN is correlated with linear- and logistic models with coefficient 0.7518 and 0.7848 correspondingly (all correlations are significant with 5\% risk level). Thus loans with two high score and one low or mediocre or vice versa exist.
Table 1: Performance of MC-RPM classified sample by Core Index range.

<table>
<thead>
<tr>
<th>CoreIndex</th>
<th>0</th>
<th>0-0.2</th>
<th>0.2-0.4</th>
<th>0.4-0.6</th>
<th>0.6-0.8</th>
<th>0.8-1.0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.39%</td>
<td>42.18%</td>
<td>56.79%</td>
<td>0.64%</td>
</tr>
<tr>
<td>True</td>
<td>21.41%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>78.59%</td>
<td></td>
</tr>
<tr>
<td>Goods-%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>77.78%</td>
<td>77.06%</td>
<td>79.56%</td>
<td>93.33%</td>
</tr>
<tr>
<td>Bads-%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>22.22%</td>
<td>22.94%</td>
<td>20.44%</td>
<td>6.67%</td>
</tr>
<tr>
<td>Odds</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>3.5</td>
<td>3.36</td>
<td>3.89</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2: Performance of minCore – maxExterior classified validation sample.

<table>
<thead>
<tr>
<th>CoreIndex</th>
<th>0</th>
<th>0-0.2</th>
<th>0.2-0.4</th>
<th>0.4-0.6</th>
<th>0.6-0.8</th>
<th>0.8-1.0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.09%</td>
<td>26.38%</td>
<td>3.13%</td>
<td>55.50%</td>
<td>14.90%</td>
</tr>
<tr>
<td>True</td>
<td>21.41%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>78.59%</td>
<td></td>
</tr>
<tr>
<td>Goods-%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>50.00%</td>
<td>65.75%</td>
<td>75.34%</td>
<td>82.10%</td>
<td>89.08%</td>
</tr>
<tr>
<td>Bads-%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>50.00%</td>
<td>34.25%</td>
<td>24.66%</td>
<td>17.90%</td>
<td>10.92%</td>
</tr>
<tr>
<td>Odds</td>
<td>Na</td>
<td>Na</td>
<td>1</td>
<td>1.92</td>
<td>3.06</td>
<td>4.59</td>
<td>8.16</td>
</tr>
</tbody>
</table>

Table 3: Misclassification rates.

<table>
<thead>
<tr>
<th>Sample</th>
<th>MC-RPM</th>
<th>minCore-maxExt</th>
<th>Log.Reg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Rate-%</td>
<td>17.52%</td>
<td>18.32%</td>
<td>20.87%</td>
</tr>
<tr>
<td>Bads as Goods</td>
<td>99.00%</td>
<td>99.80%</td>
<td>89.00%</td>
</tr>
<tr>
<td>Goods as Bads</td>
<td>0.60%</td>
<td>0.05%</td>
<td>2.34%</td>
</tr>
</tbody>
</table>

This is illustrated in figure 8 next page. Secondly, risk ranking ability of these scores are no good. KS (Kolmogov-Smirnov) statistics (maximum difference between cumulative goods and bads distributions in score range) for linear, logistic and k-NN models were, 19.61%, 20.26% and 17.12% respectively. These are all lower end of the acceptable scorecard [May04]. ROC curves (Lorenz curves) of the different models are shown in figure 8 next page. These are not great, although nor extremely bad either. Gini-coefficients (two
Figure 7: Histograms of scores and core index in validation sample. a) Left above, logistic-regression score, b) Right above, linear-regression score, c) Left below, k-NN score, d) Right below, MC-RPM core index.

Figure 8: Scorewise scatter plots. a) Left, linear vs. logistic score. b) Right, linear vs. k-NN score.

Figure 9: ROC-curves of scores in validation sample. a) Red, logistic-regression score, b) Green, linear-regression score, c) Blue, k-NN score. d) Straight line, ROC-curve, when decision making process is random.
times the area between ROC curve and straight line; varies between 0 and 1) were, 0.415 for linear model, 0.422 for logistic model and 0.360 k-NN model. Few percent of variation is added to these scores to give a suitable loan level score range for MC-RPM, such that, the range seeks to model uncertainties about the true score. Here is the problem; the poorer the initial scoring models behind the scores the more variation should be added to account for uncertainties. When this is done to poor initial scores finding robust solutions become impossible. Thus variation was set small and preferences about scores were used as described above. MC-RPM expects majority of applications between $0.6 < CI < 1$. A positive observation is that Good/Bad odds increase with core index, as it should.

In table 2 performance of the validation sample, when $minCore – maxExterior$ cut-offs are used to classification, are shown. Results have same characteristics as, when MC-RPM was used. Here core loans are greater in number, but perform somewhat poorer than in MC-RPM setting. Good / Bad odds also increase with core index.

Table 3 shows misclassification rates. Here the whole validation sample is classified Goods and Bads using a particular cut-off (the one that minimizes overall misclassification rate) thus yielding more realistic results. If misclassification rates were calculated using core and exterior loans only, the results would be promising. That would not be justifiable, however. As mentioned above only small minority of loans were classified as Core or Exteriors and in the end; an application must be accepted or rejected, that is, classified as good or bad independent of scoring model. RPM yields better overall misclassification rate than $minCore – maxExterior$ and pure logistic regression. With used cut-off, results for error type 1 look no good at the first glance e.g. for MC-RPM classification algorithm fully 99.00% of bad credit risks are accepted. This is 1% lesser if no scoring model were used. A swap set is 0.60% of rejected good loans, that is, decreased sales. Other cut-off yield different figures for reduced risk and decreased sales. Misclassification rates for $minCore – maxExterior$ are about the same. As suggested by Thomas
et al. [TEC02] cut-offs should be optimized with regard to expected misclassification costs. When this is considered simple logistic regression ranks risk better than MC-RPM and minCore – maxExterior classification methods.

5 Conclusions

This paper presents a study of an application of PRM to credit scoring problem. Cut-offs are set in multiscore situation using minCore – maxExterior rule. The rule needs core indexes from PRM. Furthermore RPM was used to classification problem. Both results were validated. The sample sizes were so large that no exact algorithm could find robust solutions. Thus approximative MC-RPM simulation scheme was used. Both MC-RPM and minCore- maxExterior methods give relatively good results in the terms of misclassification rates in comparison to traditional logistic regression model. Still logistic model remain better and much simpler. The combination of three scores to one core index using RPM did not yield better classifications nor did the combination using logistic regression (for three scores, from which one was already calculated in means of logistic regression). The approach took utilizes the strengths of the methodology, namely finding robust solutions under given uncertainties. Underlying scores were allowed to vary few tens of percents and weights were selected such that given preferences on score importance’s held.

It was noted that in order to have the solution to classification problem and reliable core indexes using RPM, underlying scores should be good classifiers of the phenomenon themselves, coarse and strongly interdependent rather that relatively bad classifiers (as individuals), abundant and independent. First conditions are fulfilled, when traditional models are used to calculate scores for strongly co-dependent phenomena (such as risk and return) from good set of data. If one wishes to use RPM by using applicant’s characteristics as scores (bad classifiers as individuals, numerous and weakly correlated), then great results are not expected.
In literature there exists some approaches to the modelling of uncertainties in consumer loan portfolio selection [Par05], but strengths of RPM are evident. The ability to model uncertainties about model parameters is unique to any other approach in portfolio selection [LMS05]. It can also capture synergies among investments once estimated. Synergies might be quite hard to estimate at the loan level, however. RPM can also take uncertainties about true loan amount (project costs) into account [LMS05], when finding robust solutions. When this is done RPM could also be used to solve the problem of setting policy limits. Reliable core index can be used for that purpose as well as for risk-based pricing of loans, indeed. If one wants to use RPM methodology as a primary decision tool in lending business a sort of reference sample or portfolio might be useful. This is because core index itself is relative to the sample. A good candidate for reference sample would be current outstanding loans. This would have several benefits over traditional scoring models. First, each new application is added to the current portfolio (this is a must in RPM setting) and the decision is based on the risk level of the entire portfolio rather than the risk estimate of single loan. Especially, when loan level co-dependencies are successfully modelled RPM methodology will ensure that the lender will be provided with more efficient operating point (e.g. higher sales since bad individual credit risks will be accepted if the riskiness of the portfolio remains acceptable). Budget constraint for RPM follows from portfolio level risk restrictions (absolute or relative amount of expected credit losses). Current difficulties include, calculation times for exact algorithms and modelling of loan co-dependencies – more development on these areas would be welcome.
References


