VALIDATION OF CREDIT RATING SYSTEMS AND SCORECARDS: A GENERALISED ROC APPROACH

S. Z. Xanthopoulos & C.T. Nakas
Credit rating or scoring systems

attach a score or rate to a bank’s customers attempting to rank order them with respect to their credit worthiness (or even with respect to some other aspect of their credit relation with the bank)
Credit rating or scoring systems are important tools for:

- decision making for loan granting / pricing
- loan monitoring
- business efficiency improvement
- performance measurement
- offer a common language framework
- *can serve as a platform on which to base the credit policy of the bank*
The discriminatory power of a credit scoring system

Traditionally refers to its ex ante capability to distinguish adequately between “bad” and “good” clients

(E.g. between potential defaulters and non-defaulters)
Example

- **Suppose:**
  - A system gives scores to the clients according to how likely they are to *default* within a certain time horizon.
  - Default means that some due payment is delayed for more than 90 days.

- **Then:**
  - “Good” = non-defaulters, “Bad” = defaulters.
  - The discriminatory power of the system is reflected on the probability that whenever it is presented with a good and a bad client the system will score them in the right order.
Validation of the discriminatory power of a credit scoring system

Receiver Operating Characteristic (ROC) curves are among the most popular techniques for validating the power of models that try to discriminate between two classes.
Given a definition of “Good” and “Bad”

- the Bank can have histograms (or probability distributions) of the scores of these two populations and then ...

- introduce a decision threshold $c$ (accept those who score better than $c$ and reject the others)
Score histograms of “good” and “bad” customers
Each decision threshold $c$ gives rise to a classification table.

<table>
<thead>
<tr>
<th></th>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S(X) \leq C$</td>
<td>$H_1(C)$ (hit)</td>
<td>$M_{12}(C)$ (miss)</td>
</tr>
<tr>
<td>$S(X) &gt; C$</td>
<td>$M_{21}(C)$ (miss)</td>
<td>$H_2(C)$ (hit)</td>
</tr>
</tbody>
</table>
Where...

\[ H_1(c) = \text{prob} \left( s(x) \leq c \mid x \text{ is } "\text{Good}" \right) \]
\[ M_{12}(c) = \text{prob} \left( s(x) \leq c \mid x \text{ is } "\text{Bad}" \right) \]
\[ M_{21}(c) = \text{prob} \left( c < s(x) \mid x \text{ is } "\text{Good}" \right) \]
\[ H_2(c) = \text{prob} \left( c < s(x) \mid x \text{ is } "\text{Bad}" \right) \]
The ROC curve

is the plot of $H_2(c)$ against $M_{21}(c)$ for all $c$. 

Bad Suffered = $M(12)$

Bad Avoided = $H(2)$

Good Sacrificed = $M(21)$

Good Accepted = $H(1)$
The Area under the ROC curve

\[ AUC = \text{prob} \ (s(x_1) \leq s(x_2)) \]

where \( x_1 \) "Good" and \( x_2 \) "Bad"

Therefore \( AUC \) is a “good” measure of the diagnostic accuracy of the scoring system

(\textit{Bamber, 1975; Hand \& Till, 2001})
Banks try to accommodate all aspects of their credit relation of the customer with the bank:

- pre-application stage
- application stage
- performance monitoring
- bad debt management

In this context classification into more than two classes will appear more and more frequently in the management of the credit relation of a bank with its customers at each of these stages.
Smoothed histograms of scores for three classes of default
Each decision pair of thresholds $c_1 < c_2$ gives rise to a classification table

<table>
<thead>
<tr>
<th></th>
<th>$x \in D_1$</th>
<th>$x \in D_2$</th>
<th>$x \in D_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s(x) \leq c_1$</td>
<td>$H_1$ (hit)</td>
<td>$M_{12}$ (miss)</td>
<td>$M_{13}$ (miss)</td>
</tr>
<tr>
<td>$c_1 &lt; s(x) \leq c_2$</td>
<td>$M_{21}$ (miss)</td>
<td>$H_2$ (hit)</td>
<td>$M_{23}$ (miss)</td>
</tr>
<tr>
<td>$c_2 &lt; s(x)$</td>
<td>$M_{31}$ (miss)</td>
<td>$M_{32}$ (miss)</td>
<td>$H_3$ (hit)</td>
</tr>
</tbody>
</table>
Where...

\[ H_1 = H_1 (c_1, c_2) = \text{prob} \ (s(x) \leq c_1 \mid x \in D_1) \]

\[ H_2 = H_2 (c_1, c_2) = \text{prob} \ (c_1 < s(x) \leq c_2 \mid x \in D_2) \]

\[ H_3 = H_3 (c_1, c_2) = \text{prob} \ (c_2 < s(x) \mid x \in D_3) \]

\[ M_{12} = M_{12} (c_1, c_2) = \text{prob} \ (s(x) \leq c_1 \mid x \in D_2) \]

\[ M_{21} = M_{21} (c_1, c_2) = \text{prob} \ (c_1 < s(x) \leq c_2 \mid x \in D_1) \]

\[ M_{13} = M_{13} (c_1, c_2) = \text{prob} \ (s(x) \leq c_1 \mid x \in D_3) \]

\[ M_{31} = M_{31} (c_1, c_2) = \text{prob} \ (c_2 < s(x) \mid x \in D_1) \]

\[ M_{23} = M_{23} (c_1, c_2) = \text{prob} \ (c_1 < s(x) \leq c_2 \mid x \in D_3) \]

\[ M_{32} = M_{32} (c_1, c_2) = \text{prob} \ (c_2 < s(x) \mid x \in D_2) \]
The ROC surface is the set of all points with coordinates

\[(H_1(c_1, c_2), H_2(c_1, c_2), H_3(c_1, c_2))\]
The volume under the ROC surface

\[ VUS = \text{prob}(s(x_1) < s(x_2) < s(x_3)) \]

\[ x_1 \in D_1, \ x_2 \in D_2, \ x_3 \in D_3 \]

(Scurfield, 1996)
Application

- Data from the loans book of a Greek bank.
- Grades of their rating system from 1 to 8 (rates 7 and 8 normally prohibiting loan granting)
- 6 rating classes left, appropriate for evaluation purposes
- Unsatisfactory granularity, so we merged two couples of these rating classes to end with 4 credit rates
Next we consider the following four well defined classes of delinquency status at the end of the year under consideration.

A: No delay of payments on the last day of the year under consideration.

B: Maximum delay of some payment on the last day of the year under consideration: between 1 and 90 days.

C: Maximum delay of some payment on the last day of the year under consideration: between 90 and 180 days.

D: Maximum delay of some payment on the last day of the year under consideration: more than 180 days.
## Application Results

<table>
<thead>
<tr>
<th>Categorization</th>
<th>Null value</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) &lt; (B+C+D)</td>
<td>0.5</td>
<td>0.5195</td>
<td>(0.4877; 0.5513)</td>
</tr>
<tr>
<td>(A+B) &lt; (C+D)</td>
<td>0.5</td>
<td>0.5268</td>
<td>(0.4903; 0.5633)</td>
</tr>
<tr>
<td>(A+B+C) &lt; (D)</td>
<td>0.5</td>
<td>0.6382</td>
<td>(0.5680; 0.7084)</td>
</tr>
<tr>
<td>(A+B) &lt; (C) &lt; (D)</td>
<td>0.167</td>
<td>0.2556</td>
<td>(0.2435; 0.2685)</td>
</tr>
<tr>
<td>(A) &lt; (B+C) &lt; (D)</td>
<td>0.167</td>
<td>0.2569</td>
<td>(0.2448; 0.2690)</td>
</tr>
<tr>
<td>(A) &lt; (B) &lt; (C+D)</td>
<td>0.167</td>
<td>0.1858</td>
<td>(0.1750; 0.1966)</td>
</tr>
<tr>
<td>(A) &lt; (B) &lt; (C) &lt; (D)</td>
<td>0.042</td>
<td>0.0743</td>
<td>(0.0670; 0.0816)</td>
</tr>
</tbody>
</table>
A < (B+C) < D
Conclusions

- Work on the multiple-class classification can also be found in [Hand et al. ’01] where a measure generalising AUC is presented.
- Our approach differs significantly in that a whole geometric object is proposed (the ROC surface or hypersurface) rather than a single measure.
- This is important since it allows for geometric and probabilistic interpretations and supports intuition.
- Furthermore, our approach sets the ground to adapt and extend techniques used in [Stein ’05] for an integration of ROC surfaces and hypersurfaces with loan pricing.