CREDIT
SUISSE

INTRODUCTION TO RATING MODELS

Dr. Daniel Straumann

Credit Suisse
Credit Portfolio Analytics

Zurich, May 26, 2005

Motivation

Due to the Basle II Accord every bank has to assess the default probability (DP) of each obligor (firm, private client, consumer) in its credit portfolio (IRB Approach).

This calls for internal rating systems. A rating system is a rational and replicable procedure, which assigns to each obligor a rating (one–year DP) based on the following information:

- market information (equity prices, interest rates, credit spreads, derivatives prices, credit agencies ratings, macroeconomic variables, . . .),
- balance sheet information resp. income data and assets (for individuals),
- qualitative data (marital status, domicile, industry sector, . . .),
- qualitative assessment (by credit analysts),
- client's individual financial behavior,
- . . .

Examples of Ratings

Standard & Poor's		Moody's	
Rating	DP	Rating	DP
AAA	0.003 %	Aaa	0.003 %
AA+	0.006 %	Aa1	0.006 %
AA	0.01 %	Aa2	0.01 %
AA-	0.02 %	Aa3	0.02 %
A+	0.03 %	A1	0.03 %
A	0.05 %	A2	0.05 %
A-	0.09 %	A3	0.09 %
BBB+	0.15 %	Baa1	0.15 %
BBB	0.25 %	Baa2	0.27 %
BBB-	0.44 %	Baa3	0.46 %
BB+	0.75 %	Ba1	0.79 %
BB	1.28 %	Ba2	1.35 %
BB-	2.20 %	Ba3	2.33 %
B+	3.77 %	B1	4.00 %
B	6.45%	B2	6.89 %
B-	11.06 %	B3	11.86 %
CCC	18.95 %	Caa & $<$	20.42 %

Table 1: Masterscales

Brief History of Rating Systems

- Starts with scorecard models, which are used for evaluating creditworthiness of customers or firms. First proposals during Second World War.

  Example: Consumer loans.

Residential Status		Age		Loan Purpose	
Owner	36	18–25	22	New car	41
Tenant	10	26–35	25	Second–hand car	33
Living with parents	14	36–43	34	Home improvement	36
Other specified	20	44–52	39	Holiday	19
No response	16	53 +	49	Other	25

Table 2: Simple scorecard for consumer credit.

This means: a 20–year old person, living with its parents, wishing to borrow money for a second–hand car, has a score of 14+22+33= 69.

Example: Altman's Z–score (1968)

$$Z = 0.012\,X_1 + 0.014\,X_2 + 0.033\,X_3 + 0.006\,X_4 + 0.999\,X_5, \tag{1}$$

where

$$\begin{aligned}
X_1 &= \text{Working capital/Total assets,} \\
X_2 &= \text{Retained earnings/Total assets,} \\
X_3 &= \text{Earnings before interest and taxes/Total assets,} \\
X_4 &= \text{Market value of equity/Book value of total debt,} \\
X_5 &= \text{Sales/Total assets.}
\end{aligned}$$

Classification rule:

$$\begin{aligned}
Z \leq 1.81 &\qquad \text{bankruptcy,} \\
1.81 < Z \leq 2.99 &\qquad \text{"zone of ignorance",} \\
Z > 2.99 &\qquad \text{non–bankruptcy.}
\end{aligned}$$

Remarks:

- Coefficients are optimized for a particular data sample.
- Nevertheless Altman's Z–score is still used (can e.g. be downloaded from Bloomberg).

- Scorecard models are often applications of statistical classification methods.

- Beginning of the 90's: Focus on default probabilities.

- Rating systems in practice
  - RiskCalc (Moody's): Scorecard model
  - CreditEdge (KMV): Structural default model
  - ...

The Classification Problem

Group of non–defaulters ($Y = 0$) and defaulters ($Y = 1$).
Each obligor has certain (observed) random attributes $\mathbf{X} \in \mathbb{R}^m$.

We *assume*:

non–defaulters have density $f_0(\mathbf{x}) = f(\mathbf{x} \mid Y = 0)$,

defaulters have density $f_1(\mathbf{x}) = f(\mathbf{x} \mid Y = 1)$.

Let $p = \mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0)$ be the default probability in the combined sample.
Assume classification rule of the following kind for a subset $A \subset \mathbb{R}^m$:

$$\widehat{Y}(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \mathbf{x} \in A.$$

Classification errors:

$$\alpha = \mathbb{P}(\widehat{Y} = 0 \,|\, Y = 1) = \int_{\mathbb{R}^m \setminus A} f_1(\mathbf{x}) d\mathbf{x} \qquad \text{(1-(hit rate))},$$

$$\beta = \mathbb{P}(\widehat{Y} = 1 \,|\, Y = 0) = \int_{A} f_0(\mathbf{x}) d\mathbf{x} \qquad \text{(false alarm rate)}.$$

Total Misclassification Error

$$\begin{aligned}\text{TME} &= \mathbb{P}(\text{classification error}) = p\alpha + (1-p)\beta \\ &= \int_{\mathbb{R}^m \setminus A} p f_1(\mathbf{x}) d\mathbf{x} + \int_A (1-p) f_0(\mathbf{x}) d\mathbf{x}.\end{aligned}$$

Aim: Minimize TME. Since $\int_{\mathbb{R}^m} f_1(\mathbf{x}) d\mathbf{x} = 1$, we have that

$$\int_{\mathbb{R}^m \setminus A} p f_1(\mathbf{x}) d\mathbf{x} = p - \int_A p f_1(\mathbf{x}) d\mathbf{x},$$

and can therefore write

$$\text{TME} = p + \int_A \left((1-p) f_0(\mathbf{x}) - p f_1(\mathbf{x})\right) d\mathbf{x}.$$

Hence TME is minimized for

$$\mathbf{x} \in A \quad \Longleftrightarrow \quad (1-p) f_0(\mathbf{x}) - p f_1(\mathbf{x}) < 0 \quad \Longleftrightarrow \quad S(\mathbf{x}) = \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} < \frac{p}{1-p}. \tag{2}$$

Remarks

- We can regard $S(\mathbf{x})$ as a score.
- Costs of misclassification can easily be incorporated.
- Minimal TME Rule (2) is equivalent to

$$\mathbf{x} \in A \quad \Longleftrightarrow \quad \mathbb{P}(Y = 0 \mid \mathbf{x}) < \mathbb{P}(Y = 1 \mid \mathbf{x}).$$

Examples

Assume that $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 0, 1$. Then

- $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1$:

$$\mathbf{x} \in A \quad \Longleftrightarrow \quad \log S(\mathbf{x}) = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - 0.5(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)) < \log \left(\frac{p}{1-p} \right).$$

  The score $\log S(\mathbf{x})$ is a *linear* function.

- $\boldsymbol{\Sigma}_0 \neq \boldsymbol{\Sigma}_1$:

$$\mathbf{x} \in A \quad \Longleftrightarrow \quad \log S(\mathbf{x}) = -\frac{1}{2}\log\left(\frac{\det(\boldsymbol{\Sigma}_0)}{\det(\boldsymbol{\Sigma}_1)}\right) - \frac{1}{2}\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1^T$$
$$-\frac{1}{2}\mathbf{x}^T(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1})\mathbf{x} - 2\mathbf{x}^T(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1) < \log\left(\frac{p}{1-p}\right).$$

The score $\log S(\mathbf{x})$ is a *quadratic* function.

Alternative classification methods

- Classification trees
- Nearest–neighbor methods
- Neural networks
- Expert systems

Logistic Regression

For the derivation, assume that $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 0, 1$. Then

$$\begin{aligned} \log(\text{"posterior odds"}) &= \log\left(\frac{\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})}{\mathbb{P}(Y = 0 \mid \mathbf{X} = \mathbf{x})}\right) = \log\left(\frac{f_1(\mathbf{x})p}{f_0(\mathbf{x})(1-p)}\right) \\ &= \text{const.} + \boldsymbol{\alpha}^T\mathbf{x}, \qquad \boldsymbol{\alpha} \in \mathbb{R}^m. \end{aligned}$$

Definition

Let Y_i be 0–1 – response variables and $\mathbf{x}_i \in \mathbb{R}^m$ vectors of explanatory variables. Then we say that Y_i obeys a logistic regression if:

1. The Y_i's given $\mathbf{x}_i$ are conditionally independent.

2. There are $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $c \in \mathbb{R}$ with

$$\mathbb{P}(Y_i = 1 \mid \mathbf{x}_i) = F(c + \boldsymbol{\alpha}^T\mathbf{x}_i), \tag{3}$$

   where $F(u) = 1/(1 + \exp(-u))$.

Remarks

- The above model is a so–called generalized linear model with link function F^{-1}.
- The choice $F(u) = \Phi(u) = (2\pi)^{-1/2} \int_{-\infty}^{u} \exp(-s^2/2) ds$ would correspond to a probit model.
- The logistic regression is the workhorse for the development of new rating models. Other areas of application: medical and natural sciences.

Practical Issues of Calibration

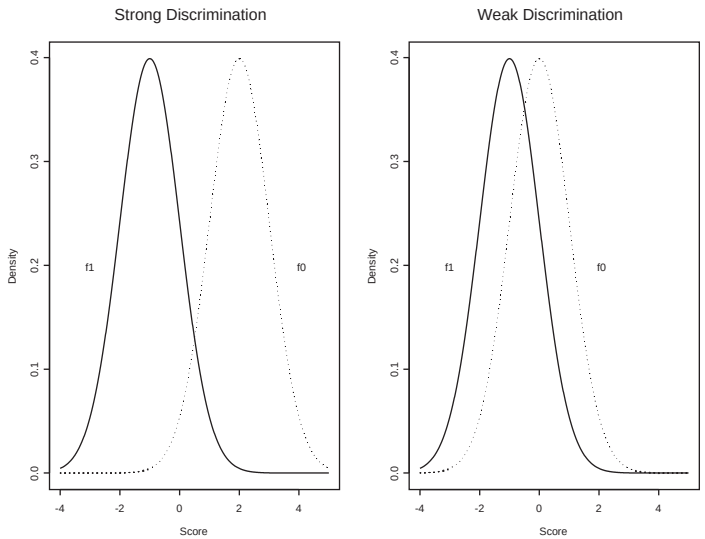
- Estimation by MLE.
- Data cleaning, construction of meaningful financial ratios.
- Variable selection methods (Stepwise forward, stepwise backward).
- Goodness–of–fit checks.

Validation of Rating Models

Suppose we have constructed a score $S = S(\mathbf{X})$, where $\mathbf{X}$ is an observed random vector. We want to predict Y in the following way:

$$\widehat{Y}(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad S(\mathbf{x}) \leq s. \tag{4}$$

The number $s \in \mathbb{R}$ is called cut–off point.

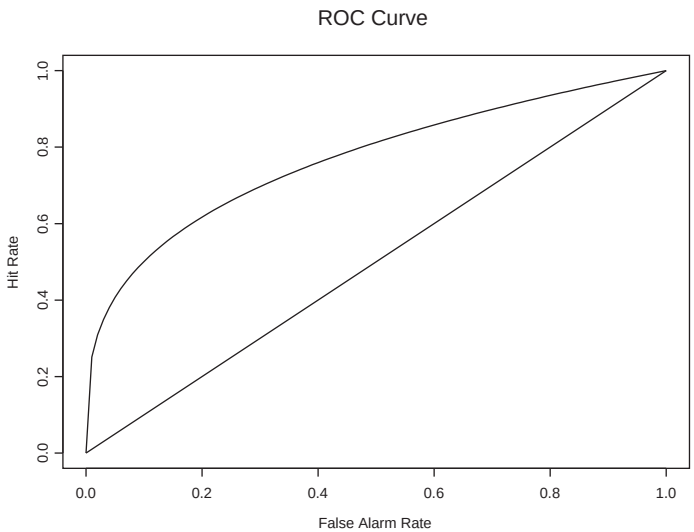


We need to measure the distance between f_0 and f_1 in a meaningful way.

The misclassification errors for (4) are given by

$$\begin{aligned} \alpha(s) &= \mathbb{P}(S > s \mid Y = 1) && \text{(1-hit rate)}, \\ \beta(s) &= \mathbb{P}(S \leq s \mid Y = 0) && \text{(false alarm rate)}. \end{aligned}$$

The curve $(\beta(s), 1 - \alpha(s))$ is called receiver operating characteristics (ROC–Curve).



AUROC

The area under the ROC–Curve (AUROC) serves as a measure for the discriminatory power of the score S.

Properties of AUROC

- $0 \leq \text{AUROC} \leq 1$.
- Perfect discrimination:

$$\begin{aligned} \text{AUROC} = 1 \quad &\Longleftrightarrow \quad \text{There is } c \in \mathbb{R} \text{ such that: } Y = 1 \quad \Longleftrightarrow \quad S \leq c, \\ \text{AUROC} = 0 \quad &\Longleftrightarrow \quad \text{There is } c \in \mathbb{R} \text{ such that: } Y = 0 \quad \Longleftrightarrow \quad S \leq c. \end{aligned}$$

- If Y and S are independent, then $\text{AUROC} = 1/2$.
- AUROC is invariant under strictly increasing transforms of S.

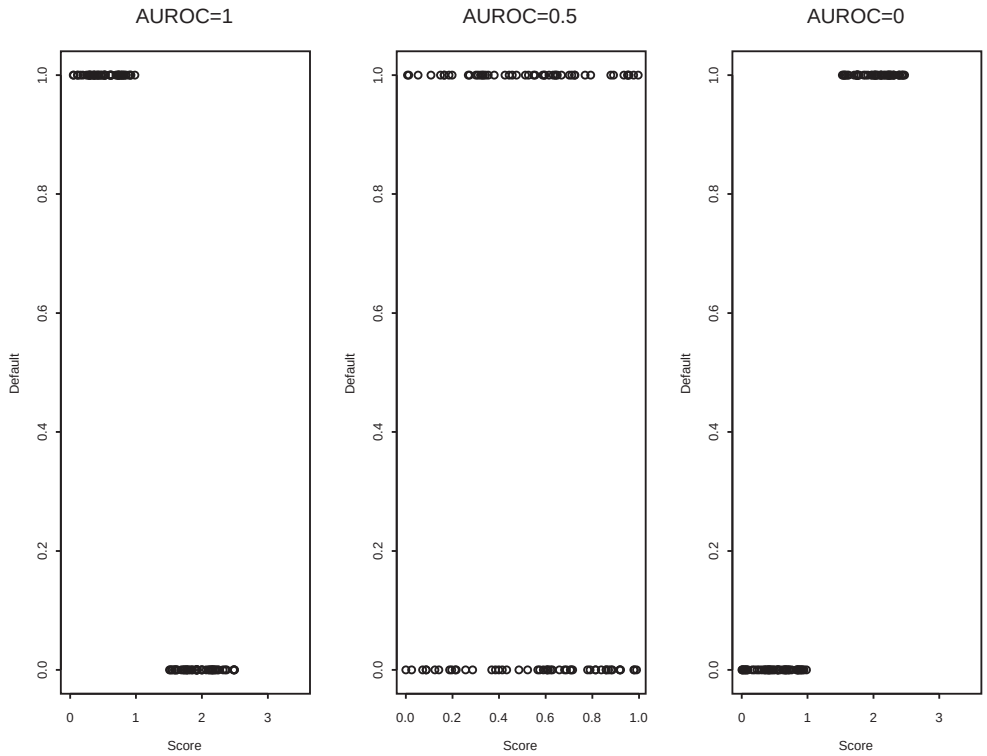


Figure 1: Representation of joint distributions for AUROC$= 1, 0.5, 0$.

Practical Applications of AUROC

- AUROC–values of 0.8 and more are often considered as "good".

- Similar to linear correlation, AUROC values are per se not very meaningful and rather difficult to interpret.

- Danger of misuse:
  - AUROC is not a measure of goodness–of–fit: add a constant to the logistic regression score, and the AUROC will not change.
  - By the Neyman–Pearson lemma, AUROC is bounded from above by a functional depending on the densities $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$.