Default Prediction Accuracy

Jorge A. Chan-Lau[1] Jin-Chuan Duan[2] Wei Sun[3]

[1]Institute for Capacity and Development
International Monetary Fund

[2]NUS Business School and Credit Research Initiative
National University of Singapore

[3]Credit Research Initiative
National University of Singapore

International Monetary Fund, Washington DC, July 17-18, 2017

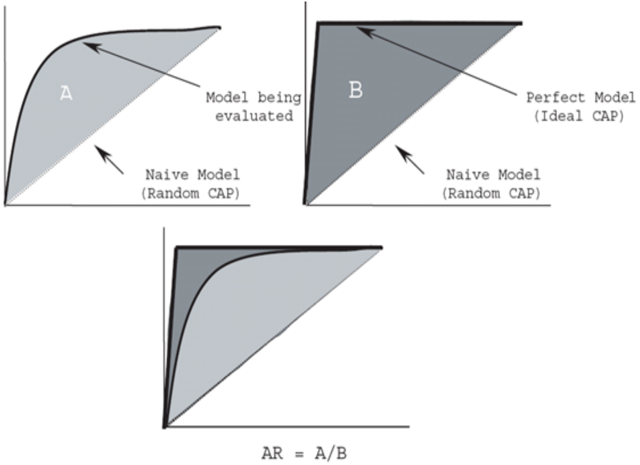
Cumulative accuracy profile

- To assess a credit rating model's out-of-sample performance, we can employ the "cumulative accuracy profile" (CAP).
- To plot CAP, companies are first ordered by model-implied default likelihood. For the top x% riskiest companies, y(x) is the percentage of the actual defaulters whose default likelihood is equal to or greater than the one for fraction x%.

- If a model were totally uninformative and assigned default likelihood randomly, we would generate a 45-degree straight line, which is referred to as “Random CAP”.
- A perfect model would produce the “Ideal CAP”, which is a straight line capturing 100% of the actual defaults within a fraction of the population being considered.

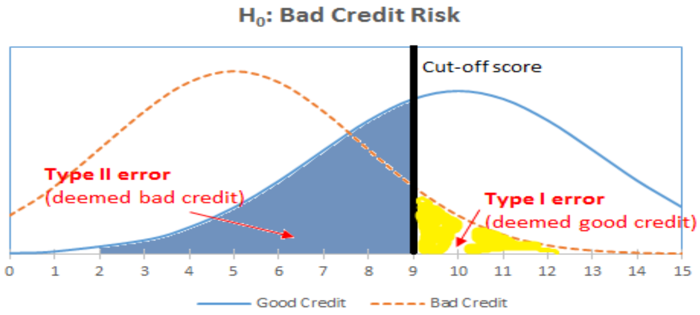
Accuracy ratio

Accuracy ratio (like the Gini coefficient) is calculated as A/B as shown in the following figure:



Type I and Type II errors

Credit score distribution produced by a model can be conceptually divided into two.



Type I error if the null is true but (wrongly) rejected. A credit risk model's null is that potential obligor is bad risk. Models are developed to find sufficient evidence to reject the null, i.e., to approve credit application. More stringent model => lower p(Type I error). Low Type I error leads to high Type II error where the alternative (risk is good) is true but not accepted.

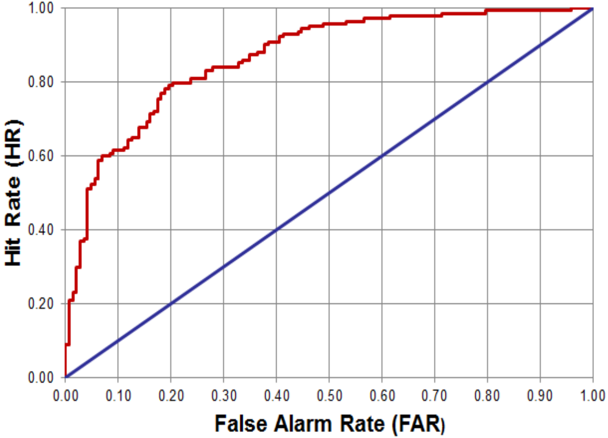
Receiver operating characteristic

Receiver operating characteristic (ROC) is a performance measure equivalent to CAP. Using any cut-off value for a classification decision based on a scoring system produces the following table:

	Classification Decision	
True State	Non-default	Default
Non- default	Score > C (Correct Prediction)	Score ≤ C (False Alarm: Type II Error)
Default	Score >C (Miss: Type I Error)	Score ≤ C (Hit)

ROC curve

Varying the cut-off value C and plotting the hit rate versus the false alarm rate produces the ROC curve.



AR-AUROC equivalence

Denote the area under the ROC curve by *AUROC*.

By Engelmann and Tasche (2003, *RISK*), $AR = 2 \times (AUROC - 0.5)$

Thus, a totally uninformative rating/scoring system has $AR = 0$ and $AUROC = 0.5$. For a perfect model, both have a value of 1.

AR-AUROC: Users beware

AR or AUROC is intuitive, but actually provides rather limited information on credit risk because

- AR or AUROC is based on ordinal rankings, and is invariant to location-scale adjustment to the scoring system; that is, adding and/or multiplying all PDs by a constant leads to the same result.
- AR or AUROC offers marginal information, meaning that joint default likelihood plays no role as long as marginal default likelihoods are the same. As a result, AR or AUROC may be insensitive to different time series patterns of defaults (i.e., default clustering).

An example of scale invariance

Three models differ significantly in their PD magnitude, but vary very little in AR or AUROC.