Credit Risk in Banking

CLIENT SCORING PROCEDURE

Sebastiano Vitali, 2017/2018

Determination of Good and Bad Clients

- The input set of clients entering the scoring function model consists of clients having information

$$X_k^r, k = 1, \dots, n$$

during period r and who are able to be determined as **good** or **bad** at the end of the period.

Determination of Good and Bad Clients

- The definition of a good and a bad client differs across models and it is very important to take the definition into account and examine it thoroughly when constructing models.
- A commonly used bound to define a good and a bad client is the inability to pay bank installments within an interval of thirty to ninety days.

Determination of Good and Bad Clients

- **Too strict definition** of a bad client makes too many borderline clients being included amongst them: we are discarding clients as bad too soon.
- **Less strict classification** means we categorize even borderline clients as good during scoring function development, with the resulting good scores carrying the risk of making deals with these borderline clients: we are accepting clients as good too soon.

Good Client at the Start of the Period

- We have a set of clients $K = 1, \dots, n$ with information X_k^r about them **at the beginning** of period r, and we have information Y_k^r **at the end** of period, with $Y_k^r, k = 1, \dots, n$ being alternatively distributed random variables with parameter $\pi(X_k^r)$, i.e. Y_k^r is the defaulted/not defaulted random variables.
- It is obvious, yet important, that while information $\boldsymbol{X_k^r}$ **is tied to the start** of the examined period, client's quality $\boldsymbol{Y_k^r}$ **is evaluated at the end** of the examined period. But the definition of a good and a bad client can differ between the start and the end of the examined period.

Good Client at the Start of the Period

- Additionally, the goodness of a client is further assured nowadays by the **credit register systems** supplying online information about the good and bad definitions based on mutually shared bank information. Thus, client's affordability to pay installments as well as total credit amount of all loans shall be considered to prevent **over-indebtedness**.

Logistic Regression Tool for Scoring

To recap the main steps:

- We construct a logistic regression:
  - estimate a suitable model and its parameter values
    - recursive or gradually widening regression model
  - perform statistical checks for the model significance
  - outlying observations
  - calculate Hosmer-Lemeshow statistics
  - calculate R^2
- We estimate the values of
  - the distribution functions F^G and F^B,
  - Gini coefficient
  - Lift

It is always a good idea to develop multiple models to compare their results and improve the safety of the final selection.
It is always a good idea to consult the final model's parameters with financial analysis experts.

Development and Reference Sample

- A very important issue in the development of a model is:
  - to develop on a "development sample"
  - to test on a "reference sample"
- (we are talking mainly about selection of the variables)

Development and Reference Sample

- If we want to use the scoring function to predict clients' quality, we need to test its **forecasting function**. This is why we use the method of gradual model widening or a recursive regression to select the parameters on a randomly selected subset of clients (its size is usually half of all the clients).
- We estimate the model parameters and test the Gini coefficient values and other statistics for both the development and reference samples. In case of choosing between alternate models, we need to favor the models with high enough Gini coefficients on the reference sample.
- Moreover, out-of-time samples should be considered to evaluate scoring model quality evolution as time progresses forward.

Determining Risk Grades

Based on an arbitrarily constructed function *s* we need to set the bank's strategy towards closing deals. Available strategies are:

- **We will offer deals to everyone, but good clients will cover the defaults of bad clients.** This results in a reduction of the bank's competitiveness towards good clients. The bank offers too high rate and the competition steals those clients. This is why this model is not suitable even though it is very simple.

Determining Risk Grades

Based on an arbitrarily constructed function *s* we need to set the bank's strategy towards closing deals. Available strategies are:

- **We set a limit for the scoring function value, and if the examined client surpasses it, we will not offer him a credit.** For granted credits we again assume that good clients have to cover the losses from bad deals. This results in the same problem as above, but with the difference that the final rate does not have to be as markedly unfavorable as in the previous model. This model is sometimes accepted for private persons loans.

Determining Risk Grades

Based on an arbitrarily constructed function s we need to set the bank's strategy towards closing deals. Available strategies are:

- **We set multiple risk grades.** We will offer deals in all the grades, but within each the good clients have to cover for the bad clients of the same grade.

Determining Risk Grades
multiple risk grades

- **We set multiple risk grades.** We will offer deals in all the grades, but within each the good clients have to cover for the bad clients of the same grade.

This model sufficiently covers the problem of a competitive rate, especially towards the creditworthy clients. On the other hand, the worst zone has a problem that the risk rate meant to cover the losses is too high. This leads to the better clients of this zone either not accepting the rate or having problems with repayment, while the worse clients in this zone accept the rate without a second thought, because they know they will have problems with paying the installments anyway, and the bank loan will only help them to postpone a problem and potentially carry out a credit fraud. This leads then to a so-called adverse selection, where the credits are more likely to be accepted by the worse clients and the real risk rate is therefore higher than the estimates based on the past.

Determining Risk Grades
multiple risk grades - evolution

- **We set multiple risk grades with an assumption that we will not offer credits in the worst one.**

In the rest we set the risk premiums in a way so that the installments of the good clients cover the losses from the bad clients. This model is the most correct one and when being careful with setting the risk grades, it can offer the bank a substantial competitive edge while respecting the minimization of losses. Even in this model though we need to take into account the fact that the bank is losing a large percentage of potentially good clients in the worst zone, but their probability of being unable to pay installments is so high that the bank cannot risk this insolvency. This is a psychological barrier that the banks and especially the management of their business departments have to be able to overcome and respect.

Determining Risk Grades
multiple risk grades - evolution

- When we are setting more than one risk grade, **we need to specify their boundaries.** Even though we can determine the probability of not paying the installments from the scoring function values, another approach is often used. All the clients that are **good at the start of the period** (i.e. they meet the condition of not being overdue at the start of the period) enter the estimation. We proceed with the **increasing score of the individual clients** and calculate the probability of **not paying the installments** $p(x)$ on intervals $[0, x], x \in [0,1]$. $p(x)$ should be an increasing function in x, which is why the estimate is constructed using the so-called isotonic regression. We will now formalize this base thought. In accordance with the previous designations let us recall that we have K clients with score $s(X_k), k = 1, \dots, K$. Without loss of generality, we will assume that $s(X_1) \leq s(X_2) \leq \cdots \leq s(X_K)$. The probability estimate of default for clients with score lower than x is reached based on:

$$\operatorname{argmin}_{p(x):[0,1]\to[0,1], non-decreasing} \sum_{k=1}^{K} \Big(Y_k - p\big(s(X_k)\big)\Big)^2$$

Determining Risk Grades
multiple risk grades - evolution

$$\text{argmin}_{p(x):[0,1]\to[0,1],non-decreasing} \sum_{k=1}^{K} \left(Y_k - p\left(s(X_k)\right)\right)^2$$

We assume $p(x)$ is linear between points $s(X_k)$. The values of x where the most significant turns of the function p happen are the subjective criterion for setting the **risk grade bounds**. At the same time we have to keep in mind that the **number of clients in each grade should be roughly equal**. The boundary of the last grade should be set in a way so that the probability of not paying installments in the last active grade **does not lead to adverse selection**. Here we always have to take into account the current **credit market situation**, **marketability of rates** and other business criteria. In case of applying the grade boundaries as an extension to already existing boundaries, we also need to think about the continuity of the probability of not paying installments, so that the probability in the respective grade does not change substantially, and the risk cost with it.