Automatic modelling of credit risk through internal ratings: an application of advanced statistical models and machine learning techniques

Raquel Florez-Lopez
Department of Economics and Business Administration
University of Leon
SPAIN
e-mail: dderfl@unileon.es

In last years, financial regulations as Basel II or Solvency II have pointed the utility of the credit risk measurement through internal rating systems (IRB approach), which employ own internal estimates of risk components to categorize exposures. There are four risk components to be evaluated: probability of default (PD), loss given default (LGD), exposure at default (EAD) and effective maturity (M). From them, the probability of default is perhaps the most critical and firm-specific variable, such that other internal risks components could be directly provided by supervisors but not PD.

Firms must provide clear definitions of their internal ratings, and they may to use all data and techniques that take into account of the long-run experience when estimating the PD associated to each rating grade. To get it, it is recommended to associate internal grades to the scale used by an external credit assessment institution, and then attribute the default rate observed for this entity to the internal grades. Nevertheless, these mappings must be based on an extensive comparison of internal rating criteria to the criteria used by the external institution, thus firms which want to estimate internal PD need to understand the rating models developed by external rating agencies.

In this paper, we propose to build an internal rating model for credit risk management, which will be map to the Standard & Poor's financial strength ratings scale. To do it, a sample of European insurance firms being rated by S&P is studied, and public financial information is analysed to discover the main attributes and relationships which define the final rating for each firm.

This process is developed in four main stages. Firstly, a data mining feature selection has been done, combining statistical methods, Bayesian techniques and machine learning algorithms (filter, embedded and wrapper approaches).

Secondly, multivariate classification models are developed, considering statistical approaches (MDA, logit and ordinal logit), together to multiple decision trees strategies.

Then, results are evaluated and discussed between models, in terms of robustness, performance and comprehensibility; particularly, some decision trees strategies are found to provide very valuable solutions.

Finally, a Bayesian hybrid proposal is provided in order to achieve synergies among previous techniques.