

Anatomy of credit scoring models

Matías Alfredo Gutiérrez Girault

Central Bank of Argentina

Summary

Introduced in the 70's, the use of credit scoring techniques became widespread in the 90's thanks to the development of better statistical and computational resources. Nowadays almost all the financial intermediaries use these techniques, at least to originate the credits they grant.

Credit scoring models are algorithms that in a mechanical way assess the credit risk of a loan applicant or an existing bank client, by means of statistical, mathematic, econometric or artificial intelligence developments. They are focused on the borrower's creditworthiness or credit risk, regardless of his interaction with the rest of the portfolio. Although all of them yield fairly similar results, those most commonly used are probit and discriminant analysis models, lineal and logistic regressions, and decision trees.

Credit scoring models can be used to evaluate retail and corporate obligors. However, in general they are used to evaluate the retail portfolio, whereas corporate obligors are assessed with rating systems. Besides using different explanatory variables, the assessment of corporate borrowers implies revising qualitative aspects of their business that are difficult to standardize. Therefore the result of their assessment is better expressed with a rating.

When financial institutions use these models to grant credits, they are referred to as application scorings. Banks use behavioral scorings to manage their loan portfolio, for example in setting credit limits, marketing products and evaluating the risk and risk adjusted profitability of existing clients. Financial

intermediaries commonly use generic credit scores, which evaluate the overall credit risk of a loan applicant, without taking into consideration the characteristics of the requested credit. However, some credit scores are aimed at getting more precision in the risk estimates: they are designed to forecast the risk of individuals applying for specific credits, such as residential mortgages. Credit scores can be estimated with external or pooled data, such as bureau scores, or with internal data by the banks themselves or external consultants.

Regardless of how credit scores are constructed, their result is condensed in a credit risk measure that allows to compare and rank order the individuals according to their perceived risk, as well as to quantify it. In general, they assign a score, classification or rating.

To clarify how credit scores are constructed and used, with the information contained in the BCRA's public credit registry (*Central de Deudores del Sistema Financiero* (CENDEU)) we estimate a sample credit score and show how it operates with a probit model. The CENDEU has detailed information of all borrowers in the financial system, such as their ID, business sector, type of borrower, risk rating, outstanding debt with each bank, type of credits, amount of financial and real collateral, etc. This credit score predicts the repayment behavior of retail borrowers in the financial system: individuals and small and medium enterprises. The only purpose of this model is to show some stylized facts of credit scores, and by no means seeks to establish or indicate what are the best practices in their use, construction or interpretation.

To estimate the model, for each of the years in the 2000-2006 period we chose those retail borrowers that at the beginning of each year were not in default. Having defined our target population, comprised by slightly more than 32.600.000 borrowers, we draw a random sample of 20%. Although it is not the purpose of this paper to discuss how these models can be validated, the remaining 80% of the data can be used to perform *out-of-sample* tests to evaluate the reliability of the model. The variable to be explained is the situation of default (or not) of the borrowers by the end of each year, which gives the probability of default (*PD*) an annual dimension. Some explanatory variables are: risk rating at the outset of each year, credit history, worst rating in the financial system, type of creditor, type of borrower, number of banks with which the borrower has debts, GDP growth, outstanding debt with the bank, outstanding debt with the financial

system, degree of collateralization of the exposures. All these variables were computed with the information at the CENDEU.

In general, all the estimated coefficients have the expected sign. For example, borrowers that at the beginning of each year were risk classified as 2 are riskier than those that were rated as 1: their *PD* is 22,7% higher. Estimates also show that the worse the borrower's rating in the financial system, the higher his *PD*. The dummy for the existence of real or financial collateral indicates that these borrowers have a lower *PD*. Finally, estimates for GDP growth indicate that on average higher economic growth reduces the *PD* in approximately 36%.

JEL: C25, G32.

Key words: credit risk, credit scoring, binary probit.