What is at stake in the construction and use of credit score?

Mireille Bardos

Banque de France
Companies Observatory
(email: mireille.bardos@banque-France.fr)

Abstract. "Statistical inference techniques, if not applied to the real world, will lose their import and appear to be deductive exercises. Furthermore, it is my belief that a statistical course emphasis should be given to both mathematical theory of statistics and to application of the theory to practical problems. A detailed discussion on the application of a statistical technique facilitates better understanding of the theory behind the technique." C. Radhakrishna RAO in Linear Statistical Inference and Its Applications

The following summary presents important topics currently being debated for companies risk assessment and the main problems to be solved in the construction and use of credit scoring. Many examples and statistical issues will be presented during my presentation.
Keywords: discriminant analysis, credit risk forecasting, accuracy of probability of failure, stability of risk classes, transition matrices, credit risk models.

1 Introduction

The need for better control of credit risk by banks has led to a stepping-up of research concerning credit scoring. Several types of technique make possible the early detection of payment defaults by companies. These techniques fall within the field of discriminant analysis.

One of banks' major objectives is to estimate expected loss and, using an extreme quantile, unexpected loss, for a population of companies, for example the customers of a bank. In order to do so, it is necessary to know the probability of default for each company for a given horizon. It is then possible to determine homogenous classes of risk.

Such an objective gives rise to several questions about the properties of the score made available. These relate to:

- the accuracy of estimates of probability and homogeneity of risk classes
- the stability over time of risk classes and their properties
- the dependence of the risk measurement on the business cycle
- the stability of transition matrices
- the correlation of risks

In order to get to grips with these questions, it is necessary to investigate the process of constructing the score and to examine the sensitive stages of the process. The practice of constructing and using scores leads to a second set of questions regarding:

- the type of variables
- the historical period of the files used to construct the scores
- the process of selecting variables
- the choice of discriminant analysis technique
- the forecasting horizon
- the stability of companies within risk classes
- the frequency with which the tool is revised
- the interaction between the business cycle, forecasting and revision

These issues are important for the quality of risk forecasting. They have increasingly been the subject of research and it appears that they are highly interdependent. The successive stages of a score's construction have an impact on the robustness and effectiveness of the tool eventually developed. In examining them, we shall set out the choices made at the Banque de France. Various different uses of the tool will be looked at.

2 Construction

2.1 How appropriate is the model for the data

2.1.1 The data Defining the event to be detected constitutes the first difficulty: Should this be legal proceedings or payment default? How serious does the payment default need to be? For the statistician, the criterion chosen is dependent on the information available. The question then arises of the correlation between these events for the same company. The population of target companies. For the statistical work to be of good quality, it requires: homogeneity of the population, representativeness of samples and their possible adjustment.

The forecasting horizon is determined by the needs of the banking system, but is dependent on how recent the data are and the impact of the economic cycle. The way in which the data files are organised is the result of a compromise. The choice of explanatory variables is also determined by availability and reliability. Qualitative variables are especially fragile and often better suited to expert assessment. Among quantitative variables, monitoring companies' bank accounts is probably very revealing over the short term, but this option is not available to the Banque de France. Economic and financial ratios constructed using accounting variables are widely available and relatively homogenous thanks to the existence of a chart of accounts. They are based on an underlying theory: financial analysis.

They are tricky and time-consuming to prepare. Abnormal, extreme or "bizarre" values are examined, as well as their law of probability, discriminant capacity, correlations and linearity or non-linearity with respect to the problem in question. This last characteristic may only be known via an examination of ratio distributions. It also determines the choice of technique (linear or non-linear). Once the correct ratios have been identified, the discriminant capacity is assessed by tests on quantiles [Vessereau, 1987]. Some statisticians use the stochastic dominance test [Davidson and Duclos, 1999].

2.1.2 The models The aim in constructing a score may be confined to the desire to identify risk signals. However, if one wishes to obtain an operational tool, its construction needs to be based on a decision rule, but its practical use requires knowing the probability of failure at a given horizon. The methods that result in linear combination of ratios are by far the most robust and are easy to interpret.

Indeed, corporate failure is a complex phenomenon for which the actual causal variables are difficult to access and to identify. The score functions therefore make use of symptoms such as descriptors of the company's situation before its failure. In other words, it is impossible to accurately define the phenomenon of company failure, contrary to what occurs in other fields of application of discriminant analysis that are closer to physical science, such as shape recognition, where overlearning is easier to master and techniques such as neural networks are successfully applied. It is, therefore, the very traditional linear discriminant analysis (LDA) of Fisher that is used at the Banque de France.

Estimating probability may be associated with the theoretical model used or may be done on the basis of empirical distributions and Bayes' theorem. The choice between the two will depend on how representative the files are and the extent to which the data correspond to the assumptions in the model.

2.2 Some thoughts on models

With detailed theoretical comparisons having been made in several studies [Baesens *et al.*, 2003], here we suggest some thoughts about suitability for companies' economic data and robustness over time. The main models will be looked at: Fisher's linear or quadratic discriminant analysis; logistical regression; Disqual [Lelogeais, 2003]; decision tree; neural networks; and other non-parametric methods.

The much-debated comparison between Fisher LDA and LOGIT warrants some further investigation – in terms of the theoretical properties [Amemiya and Powell, 1983], interpretability (the great advantage of Fisher's LDA: contributions of variables to the value of the score), sensitivity to the sampling plan of the logistical regression [Celeux and Nakache, 1994], estimation of the probability of failure (either via a theoretical formula or by use of Bayes' theorem on empirical distributions).

2.2.1 Some arguments regarding the choice between models The model's appropriateness for the data derives from the following properties and the corresponding choices:

- linearity or non-linearity,
- robustness to loss of parametric assumptions
- sensitivity to extreme values
- robustness over time (problem of thresholds for economic variables)
- interpretability of results for the user.

2.3 Probability of failure

The probability of failure provides a measure of the intensity of risk. It is much more informative than a decision threshold. Several crucial issues determine the quality of the tool:

1. The forecasting horizon must be consistent with the nature of the data. There is by definition a lag of a few months between balance sheet variables and the time at which the company is assessed, and these variables describe what has occurred over the course of the past year; they are consequently better suited to a medium-term forecast than to a short-term one. Balance sheets undoubtedly provide useful and robust information, provided that the assessment and the forecasting horizon are well matched. With a one-year horizon, it might be thought that it would be possible to create a short-term indicator, which, if it were re-estimated sufficiently often, would allow us to track the conditions under which companies are operating. But such an indicator would then follow the business cycle closely. However, this kind of perspective is very difficult to work with as frequent re-estimation in a changing environment is liable to lead to functions that always lag the current situation. It was therefore decided to work on a medium-term horizon with quantitative variables based on balance sheets and which are submitted to a method of financial analysis whose quality is long established. Given that balance sheet structures are related to the sector to which a company belongs, scores are created according to the major sectors (industry, wholesale trade, retail trade, transport, construction, business services).
2. An estimate of posterior failure probabilities well suited to the empirical data using Bayes' theorem is closely associated with the determination of risk classes. Robustness over time must be ensured for the average probability per risk class. The confidence interval of this average indicates the accuracy and provides a measure of what can happen in the worst case scenario.
3. The stability of companies in risk classes is studied using transition matrices. This paper is participating in the currently heated debate about "through the cycle" vs. "point in time" estimates.

3 Use

3.1 Individual diagnosis

Credit scoring is the first stage in the analysis of individual cases. A whole range of tools is made available. The scores are accompanied by aids to interpretation for the user, who is not a statistician but rather a financial analyst: failure probabilities, contributions of ratios to the score, and the company situation relative to the sector as a whole.

It is a great asset for the statistician to be able to identify ways in which the tool is unsuitable thanks to the analyst users who point out concrete cases where there are measurement difficulties. Their observations make it possible to improve the statistical measure of concepts of financial analysis and understanding of corporate failure processes.

3.1.1 Risk assessment for a given population Progress reports for a given group of customers are recommended by the Basel Committee. The Banque de France has produced some examples aimed at monitoring a particular population: IRISK method; economic impact of corporate failure.

3.2 Research under way

The scores constructed at the Banque de France cover a wide range of sectors. Applied to a representative sample of firms whose turnover exceeds EUR 0.75 million, they allow us to study many questions related to credit risk.

Risk contagion [Stili, 2003] [Stili, 2005] can be studied using the Banque de France's database of payment incidents involving trade bills.

Risk correlation [Foulcher *et al.*, 2003] between companies has a substantial impact on the assessment of potential losses.

If the **link between risk and the business cycle** [Bataille *et al.*, 2005] can be clarified, it would make it possible to better anticipate the risk of future failures in the light of macroeconomic variables or specific factors.

Looking at the **paths followed by companies** [Bardos, 1998b] makes possible the dynamic study of risk.

The **transition matrices** between classes of risk allow us to study the Markovian character or otherwise of failure processes.

Concentration of debt [Bardos and Plihon, 1999] is a source of major risk for banks.

Furthermore, investigation of credit risk models requires comparisons between company rating systems. Statistical research on the simulation of **distributions of default rates by rating** [Tiomo, 2002] make it possible to establish scales of reference for these comparisons.

References

[Altman and Saunders, 1998]E. I. Altman and A. Saunders. Credit risk measurement: developments over the last 20 years. *Journal of Banking and Finance*, 21:1721–1742, 1998.

[Amemiya and Powell, 1983]T. Amemiya and J. Powell. *Karlin, Amemiya, Goodman, Studies in econometrics, time series and multivariate statistics.* Academic Press New York, 1983.

[Anderson, 1984]T. W. Anderson. *An introduction to multivariate statistical analysis*, chapter 6: Classification of observations. Wiley, 1984.

[Baesens *et al.*, 2003]Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. Benchmarking state-of-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 2003.

[Bardos and Plihon, 1999]M. Bardos and D. Plihon. Détection des secteurs risqués - la méthode IRISK. *Bulletin de la Banque de France*, 69, 1999.

[Bardos and Zhu, 1998]M. Bardos and W. H. Zhu. *Bio-mimetic approaches in management science*, chapter Comparison of linear discriminant analysis and neural networks, application for the detection of company failures, pages 77–100. Kluwer Academic Publishes, 1998.

[Bardos *et al.*, 2004]M. Bardos, S. Foulcher, and E. Bataille. Banque de France scores: method, results, practice. Technical report, Banque de France, 2004.

[Bardos, 1998a]M. Bardos. Detecting the risk of company failure. *The Journal of Banking and Finance*, 22, 1998.

[Bardos, 1998b]M. Bardos. Le score BDFI: du diagnostic individuel à l'analyse de portefeuille, les études de l'observatoire des entreprises. Technical report, 1998.

[Bardos, 2001]M. Bardos. *Analyse discriminante: application au risque et scoring financier.* Dunod, 2001.

[Barret and Donald, 2002]G. Barret and S. G. Donald. Consistent tests for stochastic dominance. 2002.

[Bataille *et al.*, 2005]E. Bataille, C. Bruneau, and F. Michaud. Use of principal components method to follow the link between business cycle and risk of companies' failure. Technical report, 2005.

[Blockwitz and Hohl, 2001]S. Blockwitz and S. Hohl. Reconciling ratings. *Risk Magazine*, June 2001.

[Breiman *et al.*, 1984]L. Breiman, J. H. Freidman, R. A. Ohlson, and C. J. Stone. *Classification and regression trees.* Edition Wadswoth International Group, California, 1984.

[Celeux and Nakache, 1994]G. Celeux and J. P. Nakache. *Analyse discriminante sur variables qualitatives.* Polytechnica, 1994.

[Davidson and Duclos, 1999]R. Davidson and J.Y. Duclos. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. 1999.

[Efron and Tibshirani, 1993]B. Efron and R. J. Tibshirani. *An introduction to the Bootstrap.* Chapman & Hall, 1993.

[Efron, 1975]B. Efron. The efficiency of logistic regression compared to normal discriminant analysis. *Journal American Statistical Society*, 70:892–898, 1975.

[Foulcher *et al.*, 2003]S. Foulcher, C. Gouriéroux, and A. Tiomo. La structure par terme des taux de défauts et ratings, études et recherches de l'observatoire des entreprises. Technical report, 2003.

[Foulcher *et al.*, 2004]S. Foulcher, C. Gouriéroux, and A. Tiomo. études et recherches de l'observatoire des entreprises :la corrélation de migration: méthode d'estimation et application aux historiques de notation des entreprises françaises. Technical report, 2004.

[Gnanadesikan and *al*, 1989]R. Gnanadesikan and *al.* Discriminant analysis and clustering. *Statistical Science*, 4 (1):34–69, 1989.

[Gouriéroux, 1989]C. Gouriéroux. économétrie des variables qualitatives. *Économica*, 1989.

[Hand, 1981]D. J. Hand. Discrimination and classification. *Wiley series in probability and mathematical statistics*, 1981.

[Hristache *et al.*, 2004]Hristache, Delcroix, and Patiléa. On semi parametric m-estimation. *Journal of Statistic companies' economic data and robustess over time*, 2004.

[Kendall and Stuart, 1961]M. G. Kendall and A. Stuart. *The advanced Theory of Statistics.* Griffin, London, 1961.

[Lebart *et al.*, 1998]L. Lebart, A. Morineau, and M. Piron. *Statistique exploratoire multidimensionnelle.* Dunod, 1998.

[Lelogeais, 2003]L. Lelogeais. Un score sur variables qualitatives pour la détection précoce des défaillances d'entreprises. *Bulletin de la Banque de France*, 114, 2003.

[Lo, 1986]A. W. Lo. Logit versus discriminant analysis. *Journal of Econometrics*, 31:151–178, 1986.

[Maddala, 1999]G. S. Maddala. *Limited-dependent and qualitative variables in econometrics.* Cambridge University Press, 1999.

[Mc Lachlan, 1992]J. Mc Lachlan. *Discriminant Analysis and Statistical Pattern Recognition.* Wiley, New York, 1992.

[Michaud, 2004]F. Michaud. Mémoire de stage. Technical report, 2004.

[Rao, 1973]C. R. Rao. *Linear statistical inference and its applications.* Wiley, 1973.

[Saporta, 1990]G. Saporta. *Probabilités, analyse des données et statistique.* Technip, 1990.

[Stili, 2003]D. Stili. Etudes et recherches de l'observatoire des entreprises : Les incidents de paiement sur effets de commerce. Technical report, 2003.

[Stili, 2005]D. Stili. *Détection précoce des défaillances d'entreprises et contagion du risque.* PhD thesis, Université Paris I, 2005.

[Thiria *et al.*, 1997]S. Thiria, Y. Lechevallier, O. Gascuel, and S. Canu. *Statistique et méthodes neuronales.* Dunod, 1997.

[Tiomo, 2002]A. Tiomo. Risque de crédit et variabilité des taux de défaut: une analyse empirique par simulation, études et recherches de l'observatoire des entreprises. Technical report, 2002.

[Vessereau, 1987]Vessereau. Une propriété peu connue: l'intervalle de confiance de la médiane. *Revue de Statistique Appliquée*, XXXV (1):5–8, 1987.