Ordinal Models for Financial Evaluation

Paola Cerchiello

Abstract In this contribution we propose to estimate the probability of financial default of companies using efficiently the information contained in different databases. In this respect, we propose a novel approach, based on the recursive usage of Bayes theorem, that can be very helpful in integrating default estimates obtained from different sets of covariates. Our approach is ordinal: on one hand, the default response variable is binary; on the other hand, covariates that induce partitioning of companies are measured on an ordinal scale. We use our approach not only in a Bayesian variable averaging perspective but also to binarize ordinal variables in the most predictive way. The method is evaluated from two perspectives: one non parametric, in which we consider a functional of the probability of default and a second, fully parametric, in which we address directly the probability of default by using a mixture of Binomial and Beta random variables. The application of our proposal to an Italian credit risk database shows that it performs quite efficiently, allowing to predict for each company the probability of default by averaging the covariates contribute.

Key words: credit risk, Bayesian variable averaging, rating classes

1 Background

The financial meltdown of 2008-2009 questioned the validity of risk models and their practical implications. Within the framework of existing regulatory models, Basel II and III, banks have a tendency to uniform their models of risk evaluation and enterprise funding generating a pro-cyclical approach that, in the current economic and financial environment, highlights and exacerbates the difficult conditions in which the firms operate. Over the last 25 years, technological advances and infor-

Paola Cerchiello
Department of Economics and Management, University of Pavia, Via S. Felice 5, e-mail: paola.cerchiello@unipv.it

mation sharing have increased the use of credit scoring in almost all forms of loan origination (Altman and Saunders, 1998). However, the use of credit scoring is not without limitations (Mester et al. 1997). Theoretical studies have demonstrated the importance of information sharing in mitigating the problems of adverse selection (Jappelli and Pagano, 1993), moral hazard (Padilla and Pagano, 2000). However, the high dimensional data available from public financial statements make credit analysis difficult, and the problems are exacerbated by the necessity to account jointly for qualitative and quantitative data. In order to improve empirical results, and obtain credit scores that are more predictive and less procyclical, research is needed in the area of "scoritisation" of ordinal variables and in variable selection, preliminary to the inclusion in a full Bayesian model averaging perspective.

2 Proposal

In this contribution we propose to estimate the probability of financial default of companies using efficiently the information contained in different databases.

We want to classify companies in groups (i.e. rating classes) in a supervised way. Such groups to comply with BASEL II requirements, have to be: homogeneous with regard to target variable (i.e. default- not default), order preserving (i.e. ordering ability) and stable with regard to horizon time. In this context we are typically provided with databases of various origin, often not transparent and made of qualitative and quantitative variables. Our proposal is to build, effective but easy to explain, ordinal rating models integrated by means of Bayes theorem.

Given the available data $\underline{X}$, the model we propose can be essentially described as follows:

$$E(\theta_i \mid \underline{X}) = \sum_{k=1}^{K} E(\theta_j|\underline{X}, g_k) \cdot p(g_k|\underline{X}) \tag{1}$$

where g_k is a partition induced by a covariate k (to be combined) that classifies each unit i into one and only one level j. Equation (1) is obtained using a Bayesian model based on a mixture of Binomial and Beta random variables, corresponding respectively to the sample and the a priori distributions of the default frequency.

More precisely, our aim is to predict the target event, default or not default of a company, by averaging the most relevant available variables with the employment of Bayes theorem. We follow a factorial approach: each covariate is dichotomized into two classes that respect the binary nature of the target variable. The best dichotomization and consequently the most relevant variables are chosen by maximizing the marginal likelihoods. We also assume the probability of default of each company to be constant within the same level j of the covariate. We consider two different formulation of expression 1. On one hand, we borrow the general idea from Giudici et al. (2003) that proposed a mixtures of products of Dirichlet process in the survival analysis context in order to compare the explanatory power of each

available covariate. Thus, in the context of a non parametric framework, instead of modeling directly the probability of default, we consider an appropriate functional of the parameter of interest (i.e. the cumulative distribution function). On the other hand we evaluate a fully parametric context assuming directly that θ_j parameters are independent Beta random variables with parameters α β.

In the non parametric framework we consider the following marginal likelihood:

$$p(\underline{X}|g_k) = \prod_{j=1}^{J} \frac{M^{d_j}}{M^{[d]}} \prod_{i=1}^{r} (n_{j(i)} - 1)! \times \left[\frac{\beta_j}{\alpha_j + \beta_j} I_{[0,1)}(x_j) + 1 I_{[1]}(x_j) \right] \tag{2}$$

where M is a precision measure a priori set, d_j is the number of default companies in level j of the considered covariate, r is the number of levels of the target variable and $n_{j(i)}$ is the number of distinct units in non descending order in each level of the target variable.

In the parametric context, the marginal likelihood distribution turns out to be:

$$p(\underline{X}|g_k) = \int \prod_{j=1}^{J} \theta_j^{d_j} (1-\theta)^{nd_j} \theta_j^{\alpha-1} (1-\theta_j)^{(\beta-1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \tag{3}$$

$$\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \right]^{J} \prod_{j=1}^{J} \frac{\Gamma(\alpha+d_j)\Gamma(\beta+nd_j)}{\Gamma(\alpha+\beta+nd_j)} \tag{4}$$

In order to evaluate the performance of the proposed models, we were provided from an Italian bank a small database containing a list of variables on a set of corporate companies. The dataset is made up of 1000 companies and 13 variables: a target one describing the occurrence of the default or not default event; the remaining 12 on an ordinal measurement. These 12 variables can be divided in two subsets: the first containing 3 independent external rating evaluations and the second containing 9 items from the internal rating questionnaire.

Preliminary analysis based on the above described sample produces interesting results. Especially for the parametric approach, the weights of the 12 covariates are all comparable and assign particular importance to internal rating variables that seems to be the most important in influencing the probability of default. On the other hand, the non parametric approach since comprises an higher level of uncertainty given by the modeling of a functional of the probability of default, produces less stable results in terms of comparability of covariates weights.

3 Conclusions

Our objective is to develop an harmonization method within credit risk assessment, which typically uses multiple sources of information: balance sheets, assessment questionnaires, bank account flows, rating from credit agencies. We therefore need a more structured approach, that fully employs the potential of Bayesian modeling, a natural way to merge different information. We address the problem of integrate

several scoring, i.e. databases, by homogenizing the variables level measurement and importance. In particular we binarize the ordinal variables, typically arising from assessment questionnaires and ratings, to be matched with the target variable (default or not default). We have also proposed a novel Bayesian variable averaging procedure that leads to produce efficient estimation of probabilities of default and relative rating classes when dealing with several variables of different nature on a set of companies. Such variables averaging approach will also be able to select the most important covariates and the relative best binarization in terms of predicting the target event: default or not default. We developed such procedure by considering two contexts: a first non parametric based on an appropriate functional of the probability of default, and a second fully parametric based directly on such probability.

Future research development include a thorough predictive performance tests of the model and the extension to longitudinal data. In view of such extensions we may need a methodology able to summarize internal rating variables into summary indicators. One possibility is to follow what recently proposed by Cerchiello et al (2010) that suggest to employ stochastic dominance and quantile-based indicators. A further need in modelling of microeconomic credit risk data is to take into account interdependencies between risk variables, and their causal factors; one possibility is to employ bayesian network modelling as suggested in Cornalba and Giudici (2004) or Bonafede et al.(2007) or, alternatively, Bayesian graphical models as suggested in Giudici (2001).

Finally we remark that what proposed here can be applied to other assessment context such as quality and reputational risk as introduced in Cerchiello (2012).

Acknowledgements The author acknowledges Prof. Paolo Giudici for useful suggestions and comments.

References

1. Altman, E. I., Saunders, A.: Credit Risk Measurement: Developments over the last 20 Years. Journal of Banking & Finance, **21(11-12)**, 1721–1742 (1997)
2. Bonafede, E., Cerchiello, P., Giudici, P.: Statistical models for Business Continuity Management. Journal of Operational Risk, **2(4)**, 79–96 (2007)
3. Cerchiello, P.: Statistical Model to measure corporate reputation. Forthcoming in Journal of Applied Quantitative Methods, (2012)
4. Cerchiello, P., Dequarti, E., Giudici, P., Magni, C.: Scorecard models to evaluate perceived quality of academic teaching. Statistica & Applicazioni, **2**, 145-156 (2010)
5. Cornalba, C., Giudici, P.: Statistical models for operational risk management. Physica A: Statistical Mechanics and its applications, **338**, 166–172 (2004)
6. Giudici, P.: Bayesian data mining, with application to credit scoring and benchmarking. Applied Stochastic Models in Business and Industry, **17**, 69–81 (2001)
7. Jappelli, T., Marco Pagano, M.: Information Sharing in Credit Markets. Journal of Finance, **48(5)**, 1693–1718 (1993)
8. Mester, L. J.: What is the Point of Credit Scoring? Federal Reserve Bank of Philadelphia Business Review, 3–16 (1997)
9. Padilla, A. J., Pagano, M.: Sharing Default Information as a Borrower Discipline Device. European Economic Review, **44**, 1951–1980 (2000)