Estimation of Default Probabilities Using Incomplete Contracts Data*

J. M. C. Santos Silva
Department of Economics, University of Essex and CEMAPRE

J. M. R. Murteira
Faculdade de Economia, Universidade de Coimbra and CEMAPRE

This version: March 16, 2007

Abstract

The typical situation dealt with in the study of credit scoring models is the case in which data on previous clients of a lending institution are used to define a set of rules that permits the classification of prospective clients as credit worthy or not. However, models constructed using this type of data may suffer from *population drift* problems caused by the continuous changes in the distribution of the characteristics of the clients. That is, the sample used to estimate these models may not be representative of the population of current bank clients and credit applicants. To mitigate this problem, current clients are sometimes included in the sample and are classified according to their present status. However, this procedure will inevitably induce some degree of missclassification because some clients currently classified as non-defaulters may actually default before the end of the contract.

*Address for correspondence: João Santos Silva, Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. E-mail: jmcss@essex.ac.uk.

We address this issue by considering a count data model for credit scoring, which allows the estimation of default probabilities using data on incomplete contracts, and does not require the classification of the clients as defaulters or not. The advantage of this approach is that it allows the use of data which is both up-to-date and readily available to the lending institution. Moreover, conditional on the characteristics both of the client and the loan, it is possible to see how the probability that a client will default varies with the time horizon considered.

The model is based on the beta-binomial distribution. Although this model is rarely used in a regression context, it is particularly attractive for the problem considered here because it can account for the specific characteristics of the data and its estimation is as easy as that of the logit, which is regularly used in the construction of credit scoring models.

A well known data set on personal loans granted by a Spanish bank is used to illustrate the application of the proposed model.

JEL classification code: C21, C51, G21.

Key Words: Beta-binomial distribution; Credit scoring; Population drift.