Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 4 (2007) 251–276
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

Credit risk modeling based on survival analysis with immunes

Jan Beran[a,*], Abdel-Yazid Karim Djaïdja[b]

[a] *Department of Mathematics and Statistics, University of Konstanz, Germany*
[b] *Quantitative Research, Swiss Union of Raiffeisen Banks, Switzerland*

Received 5 October 2005; received in revised form 22 August 2006; accepted 9 September 2006

Abstract

Statistical modeling of credit risk for retail clients is considered. Due to the lack of detailed updated information about the counterparty, traditional approaches such as Merton's firm-value model, are not applicable. Moreover, the credit default data for retail clients typically exhibit a very small percentage of default rates. This motivates a statistical model based on survival analysis under extreme censoring for the time-to-default variable. The model incorporates the stochastic nature of default and is based on incomplete information. Consistency and asymptotic normality of maximum likelihood estimates of the parameters characterizing the time-to-default distribution are derived. A criterion for constructing confidence ellipsoids for the parameters is obtained from the asymptotic results. An extended model with explanatory variables is also discussed. The results are illustrated by a data example with 670 mortgages.

Keywords: Credit risk modeling; Default probability; Survival analysis; Extreme censoring; Mixture distribution; Time to default; Maximum likelihood estimation

0. Introduction

There is an increasing demand by regulatory bodies, such as the Basel Committee on Banking Supervision, to implement methodically sound credit risk measurement and management systems. In this article, the aim is to present a new methodology to evaluate default risk of retail clients.

* Corresponding author.
E-mail addresses: jan.beran@uni-konstanz.de (J. Beran), karim.djaidja@raiffeisen.ch (A.-Y.K. Djaïdja).

1572-3127/$ - see front matter
doi:10.1016/j.stamet.2006.09.001

A common definition of the risk of default is the risk that an obligor is unable to meet a specific financial obligation. Mathematically this may be quantified as a probability that a certain event occurs. Let i be an obligor and D_i the default indicator at time t of the obligor i, defined by

$$D_i(t) = \begin{cases} 1 & \text{if the obligor goes default at time } t, \\ 0 & \text{else.} \end{cases}$$

The risk of default at time t of obligor i is the probability $P(D_i(t) = 1)$. The time $\tau = \inf\{t : D_i(t) = 1\}$ is called time to default, or failure time.

In the last ten years, risks of default and more general credit risks have received much attention from both practitioners and researchers. From a practitioner's point of view, this interest is mainly due to the following factors. The rate of bankruptcy has grown to a level that is no longer negligible. This has a considerable impact on portfolio losses. Moreover, the classical model derived from Merton's seminal article (Merton [11], also see Schönbucher [13] for a review of related literature in the context of credit risks) and its extensions are difficult to calibrate for retail clients. In particular, essential information needed for calibration is not available. Finally, the New Basel Capital Accord edited by the Basel Committee on Banking Supervision allows banks and supervisors to evaluate the various risks and the adequate capital requirements by using an internal model.

In credit risk modeling, the random variable of interest is the time-to-default variable T. This is a non negative random variable which indicates the time where the default occurs. In the context of credit risk analysis, T may be infinite, namely if a default does not occur. Looking at historical statistics of default rates shows that the event of default is rare. It is thus very likely that the survival time T of a subject is very large (or even infinite) compared to the length of the study. As a result, most survival times are right-censored. Classical models and statistical methods for censored data are typically geared at the case where censoring is not too extreme (see e.g. [7,8,4,14]). An approach to survival analysis with extreme censoring, or immunity, based on exponential mixture distributions is proposed e.g. in Maller and Zhou [10].

In this paper, we propose to use survival models with immunity to model credit default rates. An extension of the mixture model by Maller and Zhou is considered. Consistency and asymptotic normality is derived in a general setting. The results are extended to the case where explanatory variables are available. Asymptotic normality leads to asymptotic tests and confidence intervals. The approach is illustrated by modeling default rates for a random sample of 670 mortgages of the Swiss Union of Raiffeisenbanks.

The outline of the paper is as follows. The model without explanatory variables is defined in Section 1. Maximum likelihood estimation and asymptotic results for this model are discussed in Section 2. An extension of the model to include explanatory variables and asymptotic results are presented in Sections 3 and 4. Finally, a data example is discussed in Section 5. Proofs are given in the Appendix.

1. Model without explanatory variables

The following notation will be used. The observed time-to-default variable T (or observed survival time) is defined by $T = \min(Y, C)$ where $Y \geq 0$ is the true survival time Y and C the censoring time. For an individual i, the corresponding random variables are denoted by T_i, Y_i and C_i, and their observed values by t_i, y_i and c_i respectively. Thus, a sample of n individuals consists of the vectors $z_i = (t_i, c_i)$ $(1 \leq i \leq n)$ where $t_i = \min(y_i, c_i)$. Moreover, the censoring