Kalpa Publications in Computing

Volume XXX, 2022, Pages 124–127

Proceedings of V XoveTIC Conference. XoveTIC 2022

Nonparametric probability of default estimation in presence of cure fraction

Peláez, Rebeca[1]*, Van Keilegom, Ingrid[2], Cao, Ricardo[1], and Vilar, Juan M.[1]

[1] Research Group MODES, Department of Mathematics, CITIC, University of A Coruña, A Coruña, Spain

[2] Research Centre for Operations Research and Statistics (ORSTAT), KU Leuven, Leuven, Belgium

Abstract

An estimator of the probability of default (PD) in credit risk is proposed. It is based on a nonparametric conditional survival function estimator for mixture cure models. Asymptotic properties of the proposed estimator are proved. A simulation study shows the performance of the nonparametric estimator compared with other semiparametric methods. A real data analysis illustrates the practical behaviour.

Acknowledgements

This research has been supported by MICINN Grant PID2020-113578RB-100, by the Xunta de Galicia (Grupo de Referencia Competitiva ED431C-2020-14 and Centro Singular de Investigación de Galicia ED431G 2019/01), all of them through the ERDF and by the European Research Council (2016-2022, Horizon 2020 / ERC grant agreement No. 694409). Peláez, R. was sponsored by inMOTION Programme of grants for pre-doctoral stays Inditex-UDC 2021.

1 Introduction

The probability of default (PD) measures the probability of a borrower to run into arrears on his/her credit obligation after a certain period of time since the granting of a credit. To estimate the probability of default, one commonly faces the problem that the time to default is censored to the right, because some customers will not have defaulted at the end of the study period or might be lost to follow up. Since the work by [5], abundant literature has been developed using survival analysis in credit risk. Nonparametric approaches were proposed by [6, 7, 2]. The possibility that some customers could never default should also be considered. Survival models that take this feature into account are called cure models. In this work we propose and study a nonparametric method to estimate the probability of default (PD) in a time horizon $t + b$ from a maturity time t based on mixture cure models.

*Corresponding author: rebeca.pelaez@udc.es

2 Nonparametric Cure Model Estimator

Let $\{(X_i, Z_i, \delta_i)\}_{i=1}^n$ be a random sample of (X, Z, δ) where X is the credit scoring, $Z = \min\{T, C\}$ is the observed maturity, T is the time to default, C is the time until the end of the study or the time until the anticipated cancellation on the credit and $\delta = I(T \leq C)$ is the uncensoring indicator. Let x be a fixed value of the credit scoring X and b a horizon time, then the probability of default is defined as follows

$$PD(t|x) \quad = \quad P(T \leq t + b | T > t, X = x) = 1 - \frac{S(t+b|x)}{S(t|x)}. \tag{1}$$

Mixture cure models consider the following useful decomposition of the conditional survival function $S(t|x) = 1 - p(x) + p(x)S_0(t|x)$, where $1 - p(x)$ is the probability of being cured (nonsusceptible to default) and $S_0(t|x)$ the conditional survival function of the uncured population, those who will never default. The functions $p(x)$ and $S_0(t|x)$ are called the incidence and the latency, respectively. The nonparametric cure model estimator of the conditional survival function proposed by [3] and [4] is given by

$$\widehat{S}_{h,g}^{NPCM}(t|x) = 1 - \widehat{p}_h(x) + \widehat{p}_h(x)\widehat{S}_{0,g}(t|x), \tag{2}$$

where $\widehat{p}_h(x) = 1 - \widehat{S}_h^B\big(\max\{T_i : i = 1, ..., n, \delta_i = 1\}|x\big)$ and $\widehat{S}_{0,g}(t|x) = \dfrac{\widehat{S}_g^B(t|x) - \big(1 - \widehat{p}_g(x)\big)}{\widehat{p}_g(x)}$, with $\widehat{S}_g^B(t|x)$ being the Beran's estimator of the conditional survival function ([1]) and $h = h_n$ and $g = g_n$ smoothing parameters. Replacing (2) in (1), we obtain the nonparametric cure model estimator (NPCM) of the probability of default.

Asymptotic properties of the nonparametric cure model estimators of conditional survival function and the probability of default have been deeply studied. Expressions for the asymptotic bias and variance were found and the asymptotic normality was proved. More details on the asymptotic properties can be found at [8].

3 Simulation study

A simulation study was conducted in order to analyse the performance of the proposed estimator of the probability of default. The study is focused on two different models.

In Model 1, a uniform distribution $U(0, 1)$ is considered for the credit scoring variable, X. The probability of cure $1 - p(x)$ is a logistic function and the latency is given by $S_0(t|x) = e^{-(1+5x)t^2}$. Model 1 fits Cox and AFT cure models.

In Model 2, a uniform distribution $U(0, 1)$ is considered for the credit scoring variable, X. The incidence is defined by $p(x) = \dfrac{\exp(15 - 190/3x + 88x^2 - 128/3x^3)}{1 + \exp(15 - 190/3x + 88x^2 - 128/3x^3)}$ and the latency is given by $S_0(t|x) = e^{-(2+58x-160x^2+107x^3)t}$. Model 2 moves away from Cox and AFT cure models.

The probability of default is estimated in a time grid of size $n_t = 100$, $0 < t_1 < \cdots < t_{n_t}$, where t_{n_t} is about the 95th percentile of the time distribution and b is about 20% of the time grid. The sample size is $n = 400$.

The optimal bivariate bandwidth (h, g) involved in the NPCM estimator is chosen as the pair that minimises a Monte Carlo approximation of the mean integrated squared error. The values of the square root of the mean integrated squared error (RMISE) are shown in Table 1.

The proportional hazards cure model estimator (PHCM) and the accelerated failure time cure model estimator (AFTCM) are considered in this analysis as benchmark methods. The simulation study carried out shows that the NPCM estimator is a very reasonable choice for estimating the probability of default, since it provides smaller estimation errors than classical methods, even in semiparametric models. See Table 1.

		Model 1			Model 2		
		NPCM	PHCM	AFTCM	NPCM	PHCM	AFTCM
RMISE	$x = 0.2$	0.1349	0.1391	0.0970	0.0766	0.0939	0.1026
	$x = 0.8$	0.0376	0.0457	0.0452	0.0551	0.0519	0.0521

Table 1: RMISE of the probability of default estimators when $x = 0.2$ and $x = 0.8$ in Models 1 and 2.

4 Application to real data

The estimation method presented in the previous section is now applied to a real data set. The data consists of a sample of 10000 consumer credits from a Spanish bank registered between July 2004 and November 2006. They were previously used in [2, 7, 6]. To preserve confidentiality, the proportion of defaulted credits has been modified. The sample censoring percentage is 92.8%. The probability of default is estimated at $x = 0.8$ for a horizon time $b = 5$ months. The results in Figure 1 show its decreasing tendency. Then, the probability of falling into default is reduced while the debt maturity is increasing. Moreover, it is close to zero at all points, due to the considered high value of the covariate that indicates a greater solvency of the borrower. The PHCM and AFTCM estimations are overlapping.

Figure 1: Estimation of $PD(t|x)$ at horizon $b = 5$ for $x = 0.8$ by means of the NPCM estimator (solid line), the PHCM estimator (dashed line) and the AFTCM estimator (dotted line) on the consumer credits dataset.

References

[1] R. Beran. Nonparametric regression with randomly censored survival data. *Technical report, University of California*, 1981.

[2] R. Cao, J. M. Vilar, and A. Devia. Modelling consumer credit risk via survival analysis (with discussion). *Statistics and Operations Research Transactions*, 33(1):3–30, 2009.

[3] A. López-Cheda, R. Cao, and M. A. Jácome. Nonparametric latency estimation for mixture cure models. *TEST*, 26(2):353–376, 2017.

[4] A. López-Cheda, R. Cao, M. A. Jácome, and I. Van Keilegom. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics and Data Analysis*, 105(12):144–165, 2017.

[5] B. Naraim. Survival analysis and the credit granting decision. In L. C. Thomas, J. N. Crook, and D. B. Edelman, editors, *Credit Scoring and Credit Control*, pages 109–121. Oxford University Press, Oxford, 1992.

[6] R. Peláez, R. Cao, and J. M. Vilar. Nonparametric estimation of probability of default with double smoothing. *SORT*, 45(2):93–120, 2021.

[7] R. Peláez, R. Cao, and J. M. Vilar. Probability of default estimation in credit risk using a non-parametric approach. *TEST*, 30(2):383–405, 2021.

[8] R. Peláez, I. Van Keilegom, R. Cao, and J. M. Vilar. Probability of default estimation in credit risk using mixture cure models. *Technical Report, Universidade da Coruña*, 2022.