Sampling and validation methods
for hazard estimation

Madormo F., Pafundi P., Artaria A.
Università degli Studi Milano Bicocca
f.madormo@campus.unimib.it — +39 02 6448 - 5889

Introduction

In this poster we compare in terms of predictive performance sampling and validation methods for hazard estimation in credit risk assessment. More precisely, we compare proportional hazard models based on the Cox model (Cox, 1972) with non parametric survival approaches based on random survival forest (Ishwaran et al., 2008) under different validation settings. On the basis of predictive performance measures (i.e. Breir Score and Prediction Error), we compare validation techniques based on the Bootstrap: Bootstrap without replacement (B=200), BOOT 632 (see e.g. Davison and Hinkley, 1997) and BOOT 632 PLUS (Efron and Tibshirani, 1997). Empirical evidence are given on a real data set provided by a credit rating agency composed of 742 small and medium enterprises, 9 financial ratios, a binary dependent variable which express a solvency indicator and a duration indicator.

Main Objectives

1. Explore how validation techniques affect model assessment in predictive models with a special focus on time-dependent survival models.
2. Compare in terms of predictive performance measures (i. e. Prediction Error and Brier score) semi-parametric proportional hazard models with respect to non parametric techniques based on random survival forest .

Models and methods

Models

- Semi-parametric model based on the proportional hazard model (Cox, 1972): the Cox model defines the cumulative hazard function dependent on the vector of predictor variables:

$$\Lambda(t|\underline{X}) = \Lambda_0(t)\exp(\beta\underline{X})$$

  where $\Lambda_0(t)$ is the baseline hazard function.
- Non parametric model based on the Random Survival Forests (Ishwaran et al., 2008): is a non parametric predictive dynamic method for the analysis of right-censored survival data that recognize the interaction between pairs of variables.

Selected validation methods (see e.g. Gerds and van de Wiel, 2011, and the references therein)

- Bootstrap cross validation without replacement (Efron and Tibshirani, 1993)
- Bootstrap 632 (Davison and Hinkley, 1997)
- Bootstrap 632+ (Efron and Tibshirani, 1997)

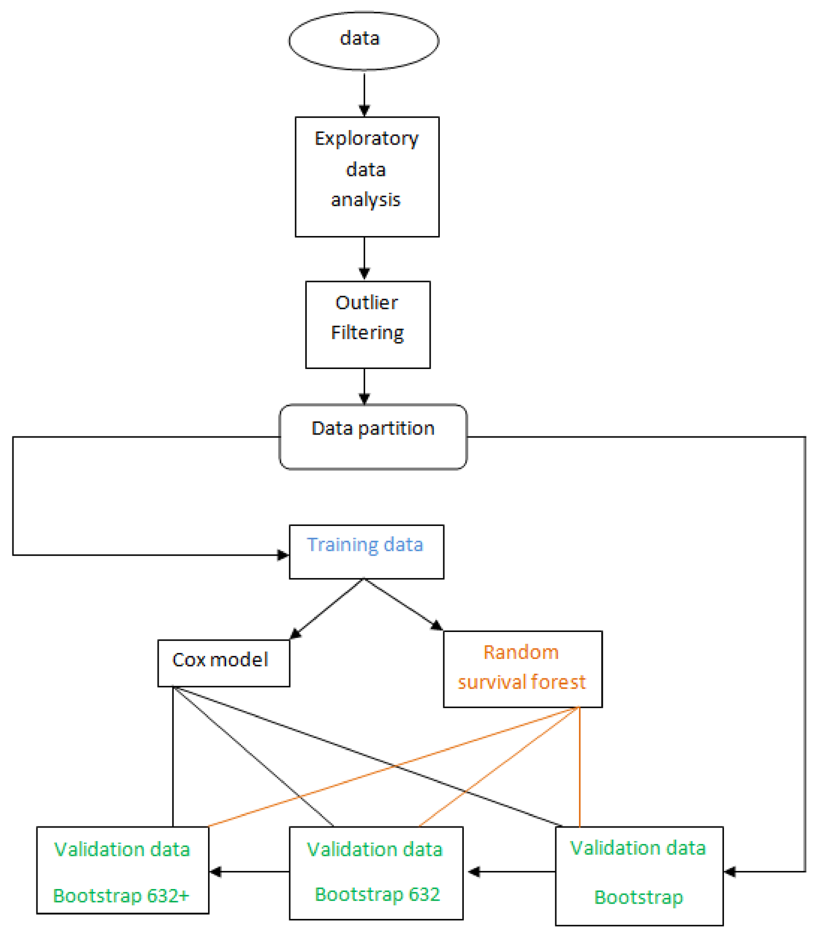
Performance Measures

- Brier score (see e.g. Gneiting and Raftery, 2007)
- Prediction error (see e.g. Mogensen et al., 2012)

Data and flow analysis

We analyzed real data coming from a credit risk dataset composed of 742 observations for a binary **target variable** with its duration (Solvency,Duration) and **9 covariates**, i.e. financial ratios:

- Supplier target days
- Outside capital structure
- Cash ratio
- Capital tied up
- Equity ratio
- Cost income ratio
- Trade payable ratio
- Liabilities ratio
- Liquidity ratio



Results

The following tables show the results in terms of prediction error (Tab. 1) and Brier score (Tab. 2). We note that the best model in terms of Prediction error is the Random survival forest which presents lower values for these measures. If we consider the validation method, the method which realizes the best performance in terms of Prediction error is the Bootstrap 632.

Prediction error	T=2	T=3	T=4	T=5	T=6
Bootstrap cross validation (B=200)					
Cox	1.3	8.3	8.3	8.3	8.3
	(1,06; 1,54)	(8,06; 8,54)	(8,06; 8,54)	(8,06; 8,54)	(8,06; 8,54)
RSF	1.3	8.2	8.1	8	7.4
	(1,06; 1,54)	(7,96; 8,44)	(7,86; 8,34)	(7,76; 8,24)	(7,16; 7,63)
Bootstrap 632 (B=200)					
Cox	1.2	8.1	8.1	8.2	8.2
	(0,96;1,44)	(7,86; 8,34)	(7,86; 8,34)	(7,96; 8,43)	(7,96; 8,43)
RSF	1	6	5.9	5.9	5.4
	(0,83; 1,17)	(5,83; 6,17)	(5,73; 6,07)	(5,73; 6,07)	(5,23; 5,56)
Bootstrap 632+ (B=200)					
Cox	1.2	8.2	8.2	8.2	7.6
	(0,90; 1,44)	(7,95; 8,44)	(7,95; 8,44)	(7,95; 8,44)	(7,36; 7,84)
RSF	1.2	7	6.9	6.8	6.3
	(0,99; 1,40)	(6,79; 7,20)	(6,69; 7,10)	(6,60; 7,00)	(6,09; 6,50)

Tab.2 shows the results for Brier score. Specifically, in terms of model selection and of choice of validation techniques the results confirm that the best model is Random survival forest and the validation technique that performs better than the others is Bootstrap 632.

Brier score	T=2	T=3	T=4	T=5	T=6
Bootstrap cross validation (B=200)					
Cox	0.1	0.5	2.4	3.6	4.4
	(0,06; 0,13)	(0,46; 0,54)	(2,36; 2,44)	(3,56; 3,64)	(4,36; 4,44)
RSF	0.1	0.5	2.4	3.5	4.3
	(0,06; 0,14)	(0,46; 0,54)	(2,36; 2,44)	(3,46; 3,54)	(4,26; 4,34)
Bootstrap 632 (B=200)					
Cox	0.1	0.5	2.4	3.5	4.4
	(0,06; 0,14)	(0,46; 0,54)	(2,36; 2,44)	(3,46; 3,54)	(4,26; 4,34)
RSF	0.1	0.5	1.8	2.6	3.2
	(0,07; 0,13)	(0,37; 0,43)	(1,77; 1,83)	(2,57; 2,63)	(3,17; 3,23)
Bootstrap 632+ (B=200)					
Cox	0.1	0.4	2.4	3.6	4.3
	(0,06; 0,14)	(0,36; 0,44)	(2,36; 2,44)	(3,56; 3,64)	(4,26; 4,34)
RSF	0.1	0.5	2.1	3.1	3.7
	(0,07; 0,13)	(0,37; 0,43)	(2,07; 2,13)	(3,07; 3,13)	(3,67; 3,73)

We observe that the classic Bootstrap performs worse than the other techniques. It would seem that the results affirm that the two generalizations of Bootstrap (Boostrap 632, Bootstrap 632+) perform better than the classical bootstrap.

In the simulation study we controlled the Monte Carlo error through the number of Monte Carlo runs (5000 Monte Carlo runs). So, we control the variation of the estimated value. For all methods this error is lower than 5%.

Conclusions

1. In this poster we have compared in terms of different sampling techniques for validation assessment the performance of Cox model and random survival forest.
2. In terms of the data, at the hand, the best model in terms of prediction performance is the random survival forest and the lowest prediction errors are achieved using the Bootstrap 632 as validation sampling technique. In terms of Brier score the results show that the lowest Brier scores are obtained by Bootstrap 632 validation technique. We note that in both cases the selected models perform better than the reference model.

Forthcoming Research

Improve the comparison considering several techniques of validation as done in (Madormo et al., 2013).

References

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC, 1993.

Bradley Efron and Robert Tibshirani. "improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

Thomas A Gerds and Mark A van de Wiel. Confidence scores for prediction models. *Biometrical Journal*, 53(2):259–274, 2011.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. *The Annals of Applied Statistics*, pages 841–860, 2008.

Filomena Madormo, Fulvia Mecatti, and Silvia Figini. Survival models for credit risk estimation. *Electronic book "Advanced in Latent Variables"*, 2013.

Ulla B Mogensen, Hemant Ishwaran, and Thomas A Gerds. Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, 50(11), 2012.