Competing Risks in Credit Scoring Using Survival Analysis and Economical Modelling

Cristián Bravo R. (1), Lyn C. Thomas (2), and Richard Weber (1)
(1) Department of Industrial Engineering (2) Centre for Risk Management
Universidad de Chile
{cbravo,rweber}@dii.uchile.cl
University of Southampton
L.Thomas@soton.ac.uk

About Us!

- Cooperation between the University of Chile and the University of Southampton since 2008.
- Support:

Doctorado en Sistemas de Ingeniería

CONICYT Ministerio de Educación Gobierno de Chile

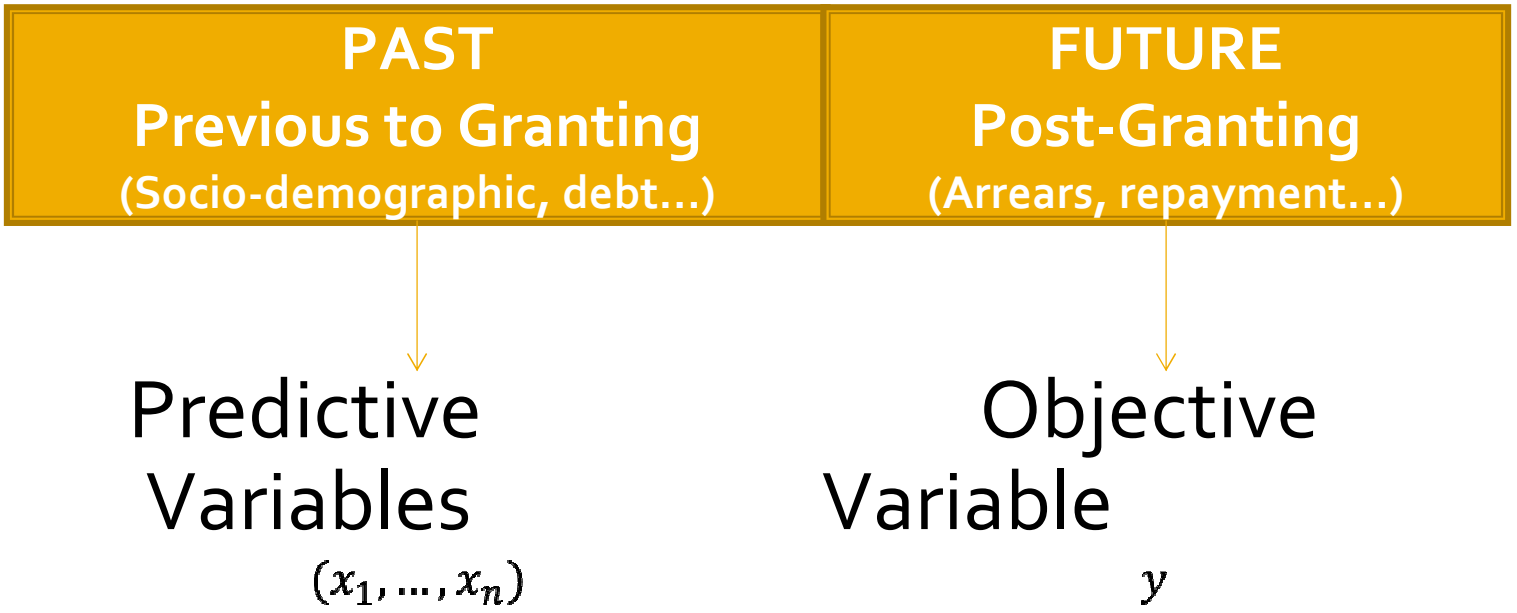
MILLENNIUM INSTITUTE COMPLEX ENGINEERING SYSTEMS

INGENIERIA INDUSTRIAL UNIVERSIDAD DE CHILE

CF CENTRO DE FINANZAS

The Information We Have

- Consumer credit scoring is based on the principle that historical patterns describe present events.
- For any given loan:



The Information We Use

- For past information, a lot of work has been done:
  - Evolution of debt.
  - Socio-economic profiling.
  - Context-specific variables.
- For future information…
  - Default: One (or more) installment was in arrears for more than 90 days in the first year of the loan.
  - **What about all the other information?**

Competing Risks

- Is it possible to enrich the objective variable using more of the available future information?
- In 2007, Thomas[1] identified that survival analysis could be used in the credit scoring context to differentiate **competing risks**:
  - Risk of attrition.
  - Risk of defaulting.
  - Risk of prepayment.
- Considering defaulting, which are the risks?

[1]Consumer Credit Models. 2007. Oxford University Press.

Competing Risks when Defaulting

- Is it possible to enrich the objective variable using more of the available future information?
- In practice, predictors are divided between two types:
  - Variables that describe **capacity of repayment**.
    - These variables describe how much money a customer has.
    - If a customer cannot pay because of lack of capacity is colloquially referred to as a "Can't pay".
  - Variables that describe **willingness to repay**.
    - These variables describe how prone a customer is to pay back the loan.
    - Customers with problems in this respect are usually referred to as a "Won't pay".

Modelling Propensity of Repayment

- Basic principles:
  - A loan is **desirable**.
  - Everyone is **rational**.
- If we assume that are three classes: good payers (P), clients with problems in capacity of repayment (C), and clients with problem in their willingness of repayment (W) then:
  - Customers of classes C and P **desire future loans**.
  - Customer of class W **do not**.
  - In the context of this model, the only difference between a person in P and one in C is that an external shock did not allow them to pay.

...Formally

- We create a procedure to grant the loans, considering the class of the customers. Considering that:
  - Customers request an amount $X_{C,t}$ or X_W, with $t = \{present, future\}$.
  - Customers may have an income given by I, and are subject to an external shock (for example being out of a job) with probability q.
  - The company can grant the loan, given by $y_t = 1$, or reject it ($y_t = 0$), and request a collateral valued at C_t for it.

Utilities for the Players

- All the members discount their flows with discount factors $\delta_C, \delta_W, \delta_L$, with "$L$" describing the lender.
- The utilities for the customers are:

$$\begin{aligned} u_C(s) =& y_1(x_{C1} - q\delta_C C_1 - (1-q)\delta_C(1+r)x_{C1}) + (1-q)\delta_C I \\ &+ (1-q)\delta_C y_1 y_2(x_{C2} - q\delta_C C_2 - (1-q)\delta_C(1+r)x_{C2}) + (1-q)^2\delta_C^2 I \\ u_W(s) =& y_1 x_{W1} - y_1\delta_W C_1 + (1-q)\delta_W I + (1-q)^2\delta_W^2 I \end{aligned}$$

- The difference between the customers is just that in the first case, he/she desires a second loan.

Utility of the Company and Rationallity

- The company receives the loan (if paid back) with an interest rate r, believes that θ percent of customers are of class C, and takes a "haircut" of α in case of default.
- The company has an utility of:

$$\begin{aligned} u_L(S) =& y_1(1-\theta)(-x_{W1}+\delta_L \alpha C_1) + y_1\theta(-x_{C1}+q\delta_L\alpha C_1+(1-q)\delta_L(1+r)x_{C1}) \\ &+\theta(1-q)\delta_L y_1 y_2(-x_{C2}+q\delta_L\alpha C_2+(1-q)\delta_L(1+r)x_{C2}) \end{aligned}$$

- Now we can derive a set of conditions on the customers.

Constraints

- The constraints will be based on the following principles:
  - The loan process is **individually rational**.
    - Each player receives a utility greater than zero.
  - Requesting (and receiving) two loans is the best strategy for class C and the lender.
- We will not search for equilibriums.
  - It is not expected to exist (Coase, 1996).
  - It is **too restrictive**.

Improving the Objective Variable

- Now it is possible to refine the objective variable.
- Given a dataset D of customers to construct a credit scoring model, then:
  - Select only the defaulters.
  - Group them into two classes, such that each class is homogeneous, and satisfies the constraints using only **variables derived after granting the loan**.
  - Construct a model using the new objective (multiclass!) variable using only **predictive variables**.

Constrained Clustering

- To obtain the clusters, the semi-supervised data mining technique called **"Constrained Clustering"** can be used.
  - The method creates the minimum variance cluster that simultaneously satisfy the constraints.
  - We would like that **"most"** cases satisfy the constraints, since there is a large variance.
    - The constraints define "profiles" of customers, which may have variance, or even be outliers. The latter must be filtered.
- We used the CCF algorithm (Bravo and Weber, 2011)
  - Constrained clustering with Outlier Filtering.

Experimental Results

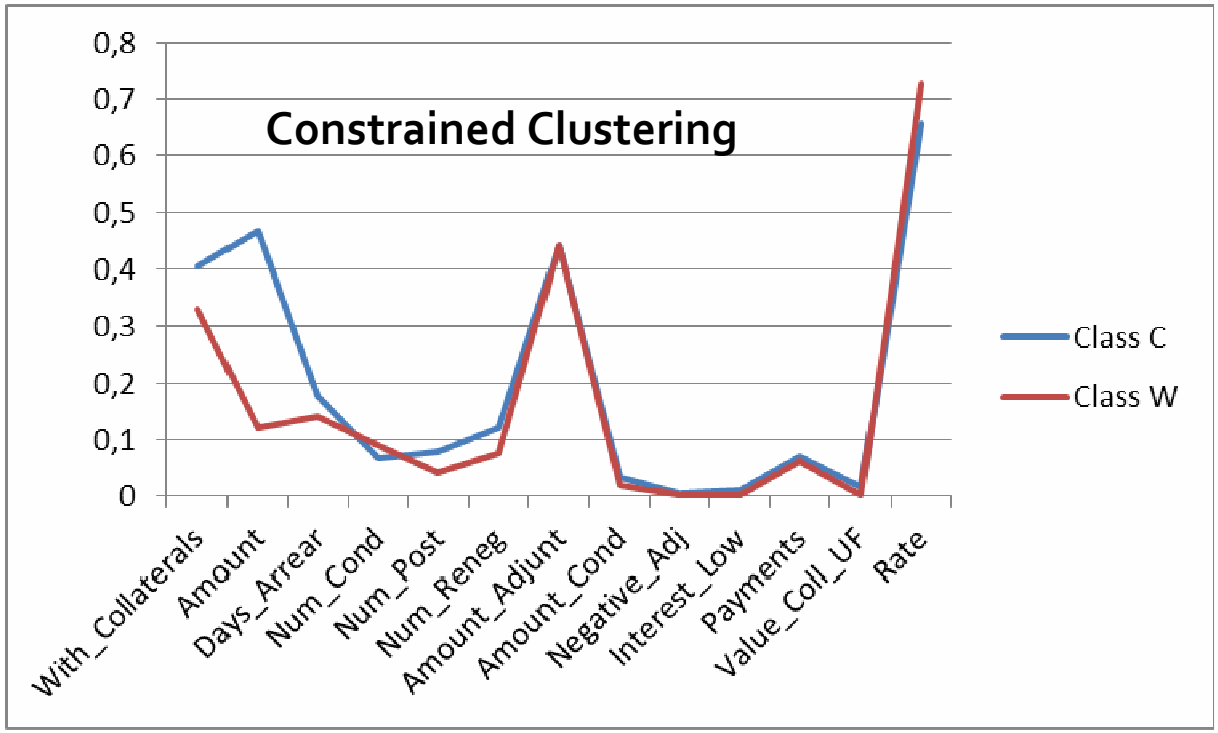
- The proposed methodology was tested using two datasets of approx. 25.000 loans to Small and Medium Enterprises (SME) during the years 2000 – 2007.
  - The predictors included socio-demographic variables, debt, sales, age, etc.
  - The variables that describe the evolution of the loan once granted are renegotiations, amount, value of the collateral, requested amount, term, etc.
  - Roughly half the loans are secured in the first dataset, and 100% of the second dataset.
  - Parameters for the model are obtained using metrics from the institution (repayment rates, "haircuts", etc…)

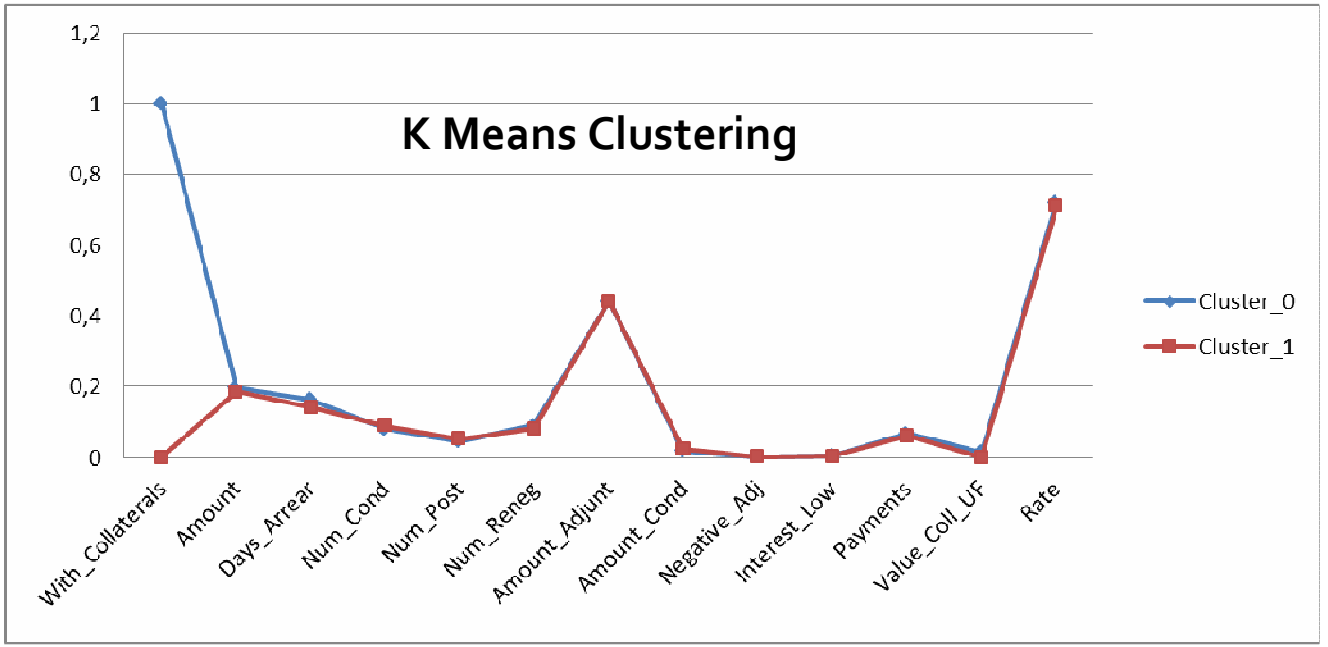
Constrained Clustering Results

Variable	Class C	Class W
With Collaterals	0,41	0,33
Amount	52,94	14,82
Days in Arrears	895,78	719,82
Condonation	0,27	0,36
Postponments	0,85	0,46
Renegotiations	0,72	0,45
Amount Adjusted	-0,27	-0,08
Amount Condoned	4,23	2,44
Negative Adjustment	0,41	0,09
Interest Lowered	1,32	0,31
Paid	2,81	2,42
Collateral (Value)	71,37	7,93
Rate	1,10	1,10

Constrained Clustering Results

Variable	Class C	Class W
Amount	530.238	338.118
Term	15,27	12,82
Grace period (month)	2,21	2,13
Grace period (day)	61,41	58,46
Installment Val.	43.829	32.656
Collateral Val.	1.576.543	435.231
Rate	2,25	2,14
Dif. In Rate	-0,62	-0,40
Prepay req.	0,85	0,18
Prepaid	0,44	0,07
Renegotiated	0,50	0,21
Term Req.	16,65	13,81
Installment Req.	55.788	43.707
Amount Paid	42.206	33.455
Avg. Arrear	73,14	111,41
Amount Req.	614.072	388.277


Constrained Clustering
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
With_Collaterals
Amount
Days_Arrear
Num_Cond
Num_Post
Num_Reneg
Amount_Adjunt
Amount_Cond
Negative_Adj
Interest_Low
Payments
Value_Coll_UF
Rate
Class C
Class W

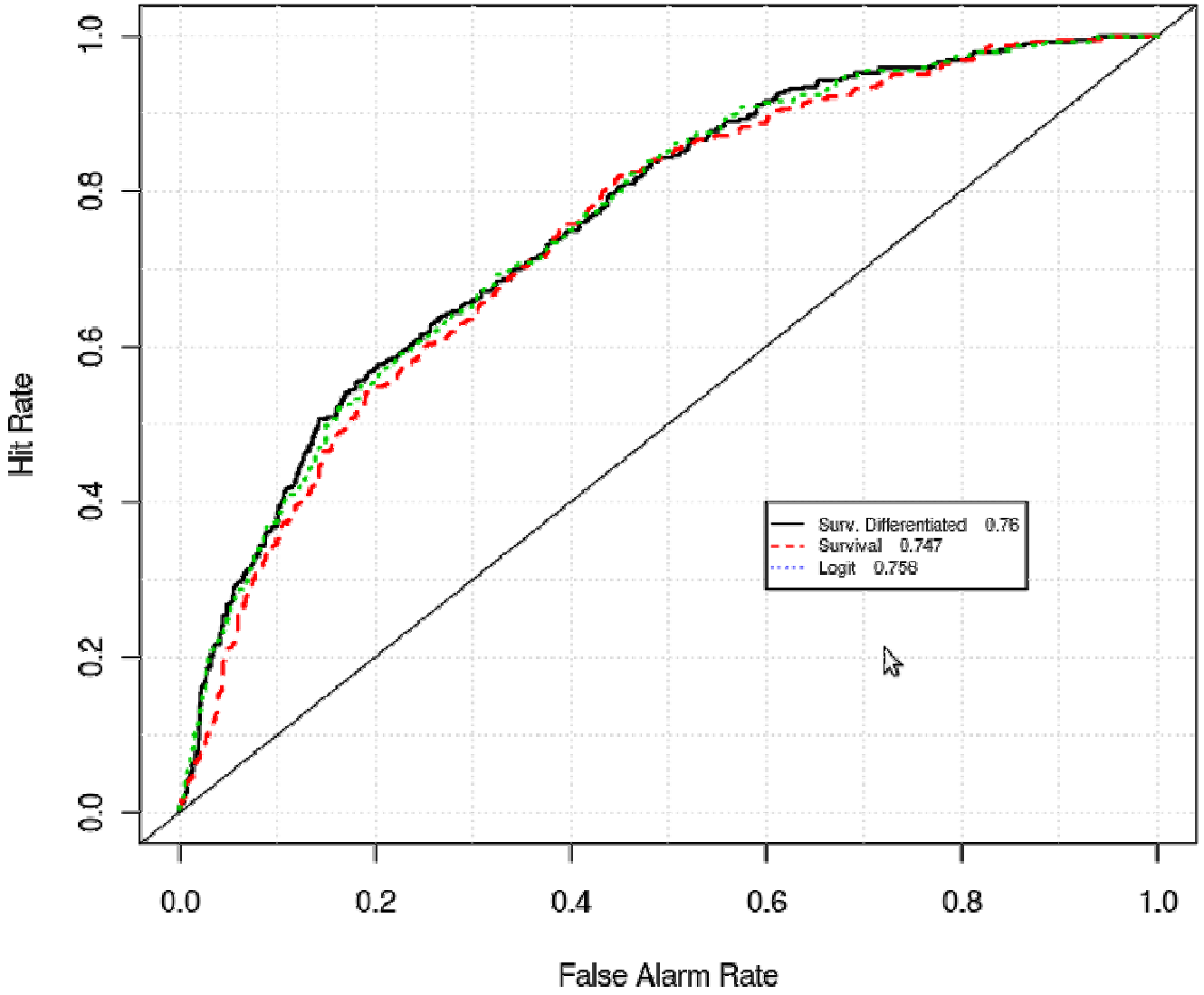

K Means Clustering
1,2
1
0,8
0,6
0,4
0,2
0
With_Collaterals
Amount
Days_Arrear
Num_Cond
Num_Post
Num_Reneg
Amount_Adjunt
Amount_Cond
Negative_Adj
Interest_Low
Payments
Value_Coll_UF
Rate
Cluster_0
Cluster_1

Classification Results

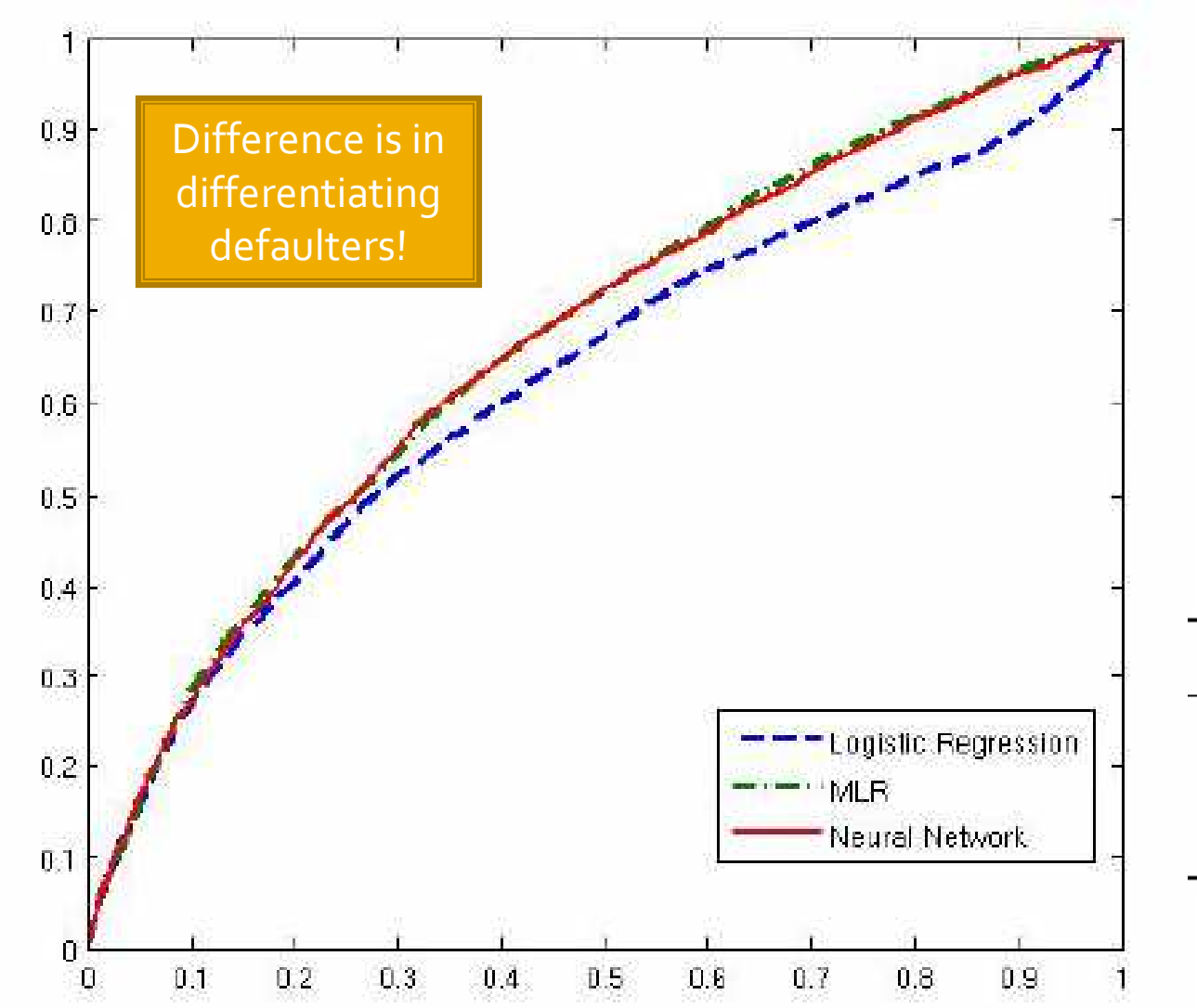
- We compared four models: Survival Analysis, Neural Networks (ANN), Multinomial Logistic Regression (MLR) and Binary Logistic Regression (LR).
  - LR model was trained with class Defaulter, and used as a benchmark , as well as regular (two class, single risk) survival analysis.
  - MLR and ANN were trained as a three class model (P, C & W).
  - Survival analysis using the three classes (competing risks) was used.

Classifications Results: Accuracy

ROC Curve



Classifications Results: Accuracy (II)



AUC (defaulters)
0.6275±0.017
0.6678±0.004
0.6660±0.004

Classification Results: Differences in Variables

LOGISTIC REGRESSION

Variable	Defaulter
Activity_A	0.33±0.04
Activity_B	-0.29±0.02
Owner	-0.27±0.02
Let	0±0.02
Middle_Own	0.39±0.03
No_Property	0.39±0.03
One_Property	0.69±0.02
Region_1	0.2±0.02
Region_2	0.73±0.03
Guarantor	-0.56±0.03
Term	-0.25±0.03
Age	0.19±0.04

NEURAL NETWORKS

Variable	P	C	W
Activity_A	0.06±0.01	-0.29±0.03	0.16±0.02
Activity_B	0.05±0.01	-0.16±0.02	0.09±0.01
Owner	0±0	-0.1±0.62	-0.07±0.01
Let	-0.09±0.01	0.28±0.03	-0.21±0.02
Middle_Own	-0.09±0.01	0.21±0.03	-0.12±0.02
No_Property	-0.15±0.02	0.07±0.03	0.11±0.01
One_Property	-0.05±0.01	0.01±0.01	0.04±0.01
Region_1	-0.17±0.02	0.16±0.03	-0.03±0.01
Region_2	0.12±0.02	-0.21±0.03	0±0
Guarantor	0.06±0.01	0.09±0.01	-0.15±0.02
Term	-0.44±0.18	-0.45±0.46	1.12±0.18
Age	7.31±1.13	3.91±1.88	-11.99±1.93

Conclusions

- An experimental method to construct a scorecard using the risk of being a defaulter for willingness or capacity was presented.
  - It takes advantage of information not used in classifying loans.
  - It is based on simple assumptions on rationality.
- The method provides interesting insights about the customers.
  - Future information allows to profile defaulters.
  - Patterns are enriched, since some variables that were not useful now are.

Conclusions (II)

- Results on a test database show the usefulness of the approach.
  - There are gains in classification.
    - Potential for reducing costs.
  - The clustering method allows us to refine behaviour.
    - Potential for better origination policies.
- Future work:
  - How to better model behaviour?
  - Which future variables are more descriptive?
  - Which past predictors describe best the new situation?

Thank you!! Questions?

Competing Risks in Credit Scoring Using Survival Analysis and Economical Modelling

Cristián Bravo R. (1), Lyn C. Thomas (2), and Richard Weber (1)
(1) Department of Industrial Engineering (2) Centre for Risk Management
Universidad de Chile
{cbravo,rweber}@dii.uchile.cl
University of Southampton
L.Thomas@soton.ac.uk