Analyzing prepayment and default under changing credit market conditions for SMEs by applying advanced analytics on Credit Bureau data

Quirini Lorenzo[1], Vannucci Luigi[2], Bagli Enrico[3], Capuzzo Davide[3], Poluzzi Alessandro[3], Fucile Assunta[3]

[1]Monte dei Paschi di Siena Group, Piazza Salimbeni, 3 53100 Siena SI
[2]Università degli Studi di Firenze, Dipartimento di Scienze per l'Economia e l'Impresa, Via delle Pandette, 32 - 50127 Firenze
[3]CRIF S.p.A., Via Mario Fantin, 1, 40131 Bologna BO

- Introduction
- SMEs in Italy
- Hidden Markov Chain
- CRIF Studio
- Data Analysis
- Results & Conclusions

Introduction

Goal

- compute the **random net present value (RNPV)** of a portfolio of loans under **changing credit market conditions (CCMC)** for SMEs

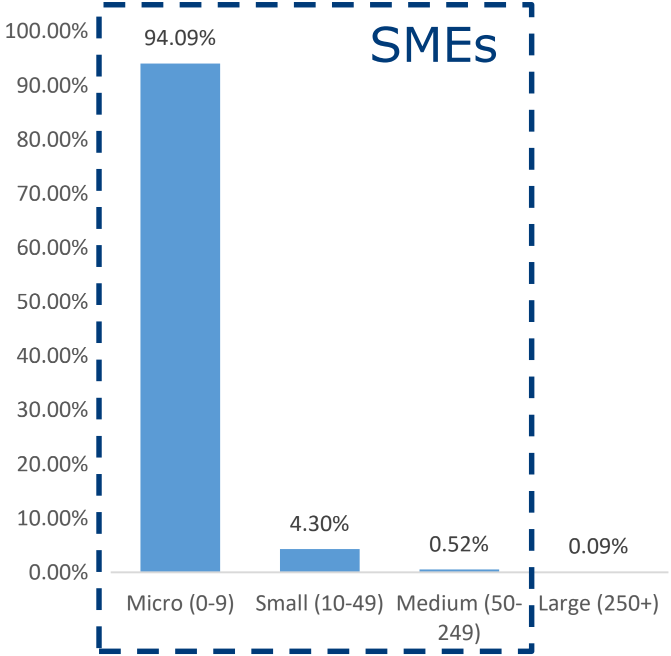
Method

- implementation of **Hidden Markov Chain (HMC)** in the CRIF **CRIF Studio** on Bureau Data

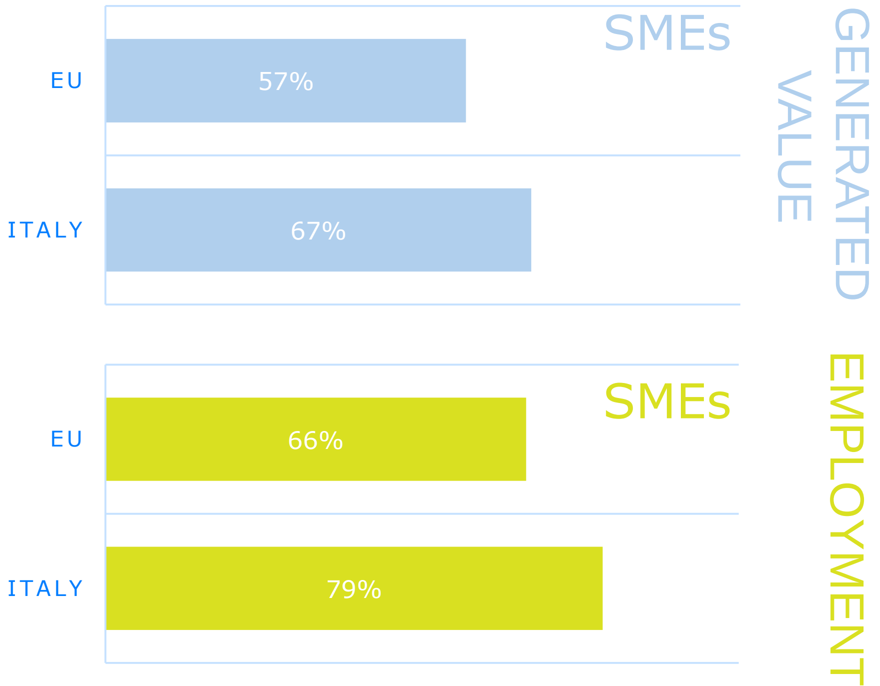
Results

- computation of the RNPV conditioned by CCMC **analysing Credit Bureau time series**

SMEs in Italy



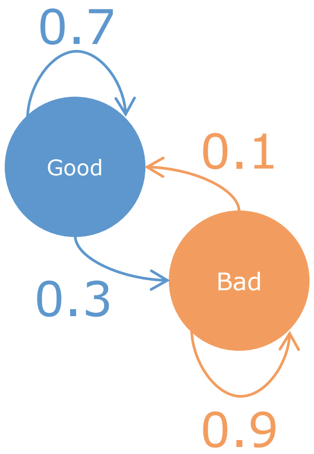
Italian Companies Distribution by Number of Employees
Data Source: Istat Statistics 2017, ISTAT



Data Source: 2018 SBA Fact Sheet, European Commission

Markov Chain

A Markov chain is a **stochastic model** that describes a sequence of possible events where the probability of each event **depends** only on the state attained in the **previous event**



States

$$S = \{S_1, S_2, \dots, S_N\}$$

Sequence of States

$$Q = \{q_1, q_2, q_3, \dots\}$$

Initial Probabilities

$$\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$$

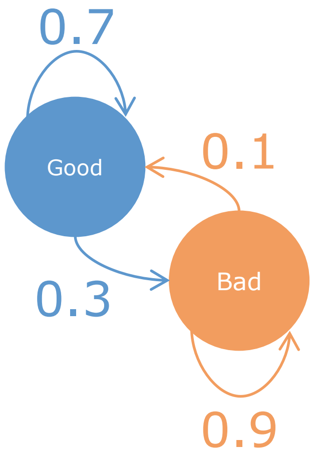
Transition Matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}$$

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$$

Markov Chain

A Markov chain is a **stochastic model** that describes a sequence of possible events where the probability of each event **depends** only on the state attained in the **previous event**



States

Good | Bad

Sequence of States

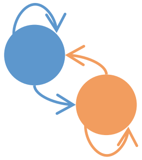
$q_1 - q_2 - q_3 - q_4 - q_5 - q_6 - q_7 - q_8 \rightarrow$

Initial Probabilities

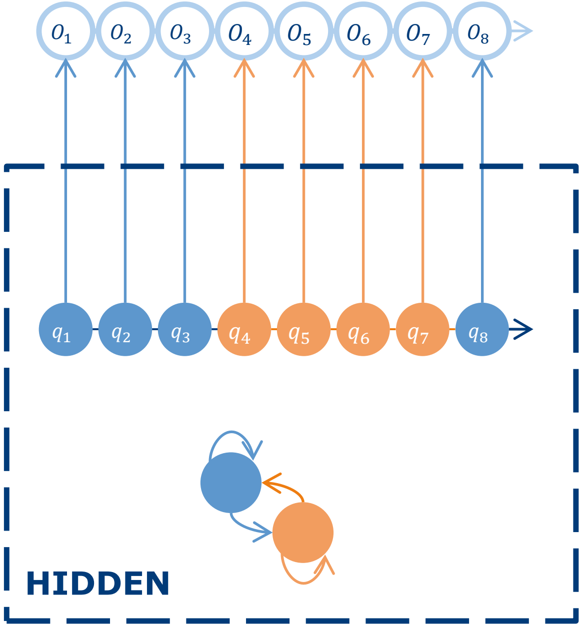
0.9 | 0.1

Transition Matrix

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}$$



Hidden Markov Chain



Observation

$$O = \{O_1, O_2, \dots, O_N\}$$

Probability Distribution

$$B = \{b_i(v)\}$$

$$b_i(v) = P(O_t = v | q_t = S_i)$$

States

$$S = \{S_1, S_2, \dots, S_N\}$$

Sequence of States

$$Q = \{q_1, q_2, q_3, \dots\}$$

Initial Probabilities

$$\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$$

Transition Matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}$$

HIDDEN

Hidden Markov Chain

Three Main Usages

$$\vartheta = \{A, B, \pi\}$$

Decoding

- Finding the most likely pattern of a sequence, given the parameters of the model.
- Viterbi algorithm

$$O, \vartheta \rightarrow P(O|\vartheta)$$

Evaluation

- Computing the likelihood that a specific sequence is generated by the model.
- Forward-Backward algorithm

$$O, \vartheta \rightarrow \operatorname{argmax}_Q P(Q|O, \vartheta)$$

Learning

- Finding the model parameters that maximize the likelihood of the observed sequence.
- Baum–Welch algorithm

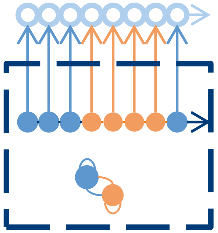
$$O \rightarrow \operatorname{argmax}_\vartheta P(O|\vartheta)$$

Random Net Present Value

Net Present Value

$$v_0 = \sum_{j=1}^{n} r_j (1+y)^{-j} - w$$

$$w = \sum_{j=1}^{n} r_j (1+x)^{-j}$$



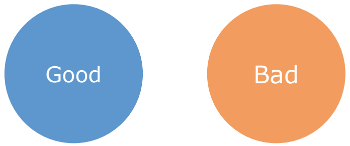
$r_1, r_2, \dots, r_n$ contractual sequence of instalments

x contractual internal rate of return

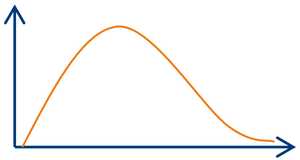
y discount monthly rate

L. Quirini and L. Vannucci, *"Modelling Prepayment and Default under Changing Credit Market Conditions for a Net Present Value Analysis", under publication (2019)*

Market Condition



Poisson Distribution



Event $A_{s,B}(h), A_{s,G}(h)$

The loan, started in s at time 0, is in the B,G state in the h-th period, and it defaults exactly at time h.

Event $C_s(h)$

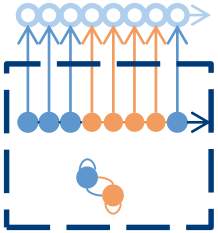
The loan, started in s at time 0, prepays exactly at time h.

Random Net Present Value

Net Present Value

$$v_0 = \sum_{j=1}^{n} r_j (1+y)^{-j} - w$$

$$w = \sum_{j=1}^{n} r_j (1+x)^{-j}$$



$r_1, r_2, \dots, r_n$ contractual sequence of instalments

x contractual internal rate of return

y discount monthly rate

L. Quirini and L. Vannucci, *"Modelling Prepayment and Default under Changing Credit Market Conditions for a Net Present Value Analysis", under publication (2019)*

Random Net Present Value

$$V_s(0) = \begin{cases} \sum_{j=1}^{n} r_j (1+y)^{-j} - w\,, & \mathrm{P}(E_S) \\ \sum_{j=1}^{h-1} r_j (1+y)^{-j} + (1+y)^{-h} Z_B(h)\varphi(h) - w\,, & \mathrm{P}\left(A_{S,B}(h)\right) \\ \sum_{j=1}^{h-1} r_j (1+y)^{-j} + (1+y)^{-h} Z_G(h)\varphi(h) - w\,, & \mathrm{P}\left(A_{S,G}(h)\right) \\ \sum_{j=1}^{h-1} r_j (1+y)^{-j} + (1+y)^{-h} \gamma(h)\varphi(h) - w\,, & \mathrm{P}\left(C_S(h)\right) \end{cases}$$

$Z_B(h), Z_G(h)$ random recovery rates of the exposure at default

$\varphi(h)$ exposure at default at time h

$\gamma(h)$ charged prepayment ratio

$A_{s,B}(h), A_{s,G}(h)$ The loan, started in *s* at time 0, is in the *B*,*G* state at the *h*-th time, and it defaults exactly at time *h*.

$C_s(h)$ The loan, started in *s* at time 0, prepays exactly at time *h*.

Credit Scoring & Credit Control Conference XVI

CRIF Studio

- Ready-to-use environment in which data from different sources are mixed and analysed through advanced analytic techniques, e.g. ML, to optimize specific business processes
- Experienced support to data preparation activities leveraging on CRIF risk expertise and marketing knowledge
- The client is enabled to design, run and conduct the analysis up to the deployment phase

Customer value maximization

Continuous improvement

Ready to use

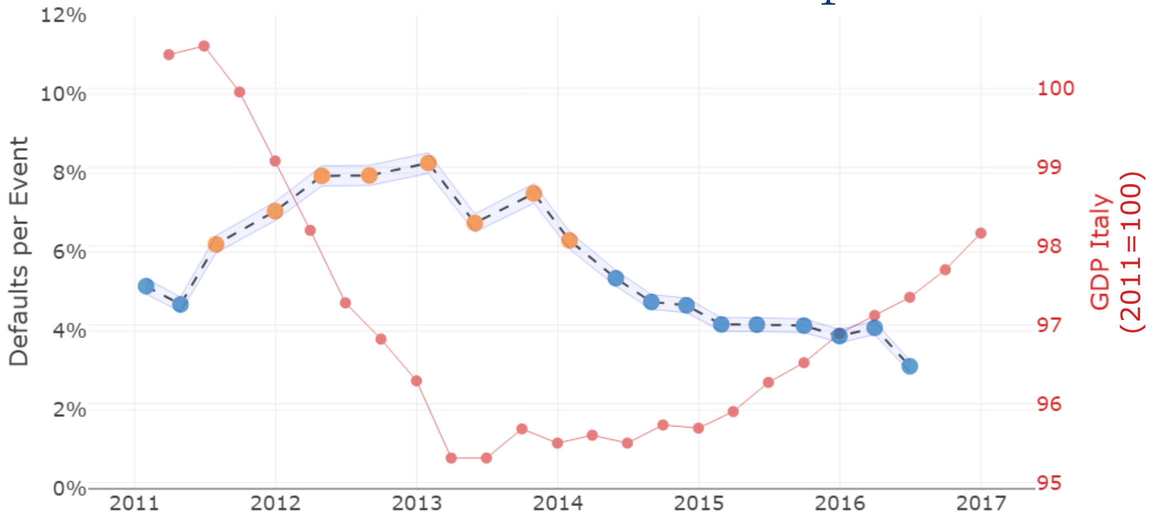
Data Analysis

CRIF Bureau data can be segmented to retrieve HMC parameters for specific clients

- Event
  - Default
  - Prepayment
- CRIF Credit Bureau Score
  - A,B,C,D
  - E,F,G,H,I
  - L,M,N,O,P
- Economy Activity
  - Buildings
  - Wholesale
  - …
- Type
  - Mortgage
  - Loans
  - …
- Amount
  - < 1000 €
  - > 1000 € & < 10000 €
  - …
- …

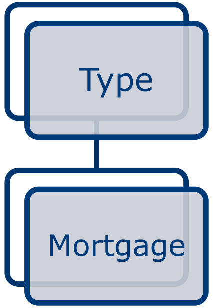
Data Analysis

$$A = \begin{pmatrix} 0.93 & 0.07 \\ 0.06 & 0.94 \end{pmatrix} \qquad \lambda_0 = 4.4 \quad \lambda_1 = 7.2$$



Data Source: elaboration on ISTAT data

Default

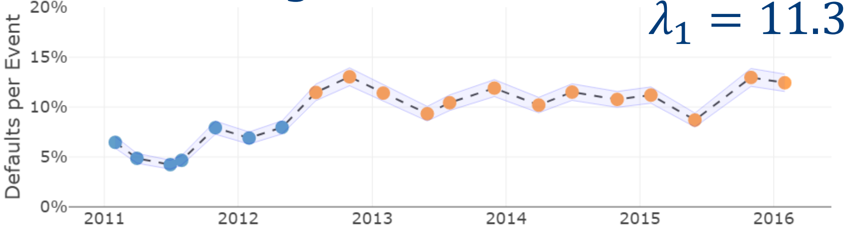


Correlation between GDP and mortgages default

Data Analysis

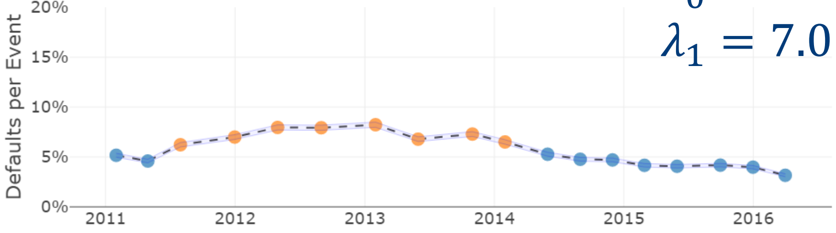
Buildings

$\lambda_0 = 6.1$
$\lambda_1 = 11.3$

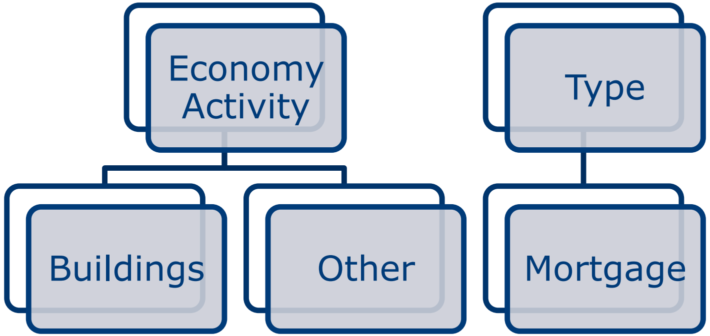


Other

$\lambda_0 = 4.2$
$\lambda_1 = 7.0$



Default

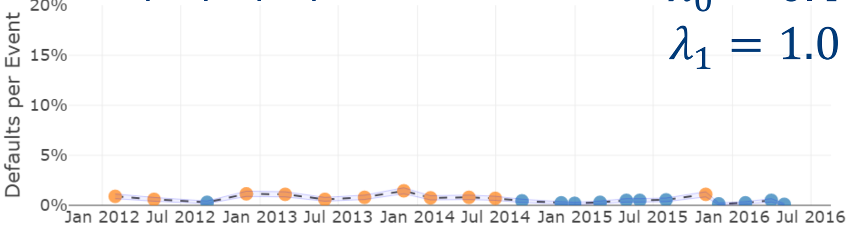


Different default rates for buildings sector and other fields

Data Analysis

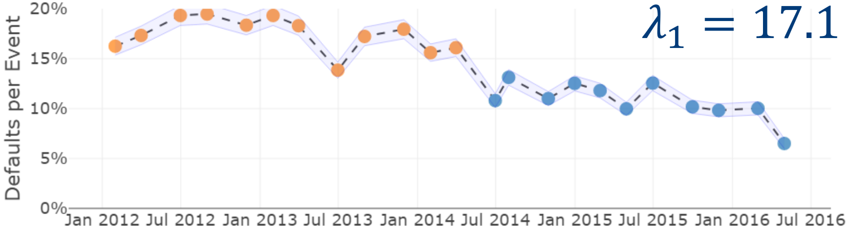
L,M,N,O,P

$$\lambda_0 = 0.4$$
$$\lambda_1 = 1.0$$

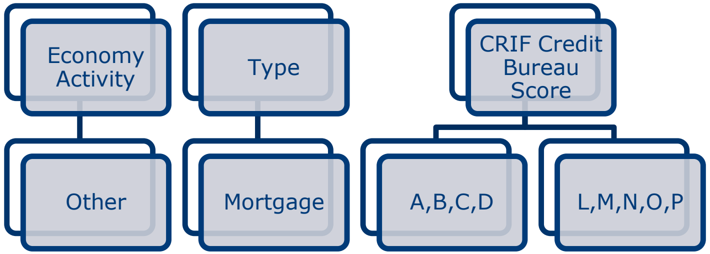


A,B,C,D

$$\lambda_0 = 10.6$$
$$\lambda_1 = 17.1$$



Default

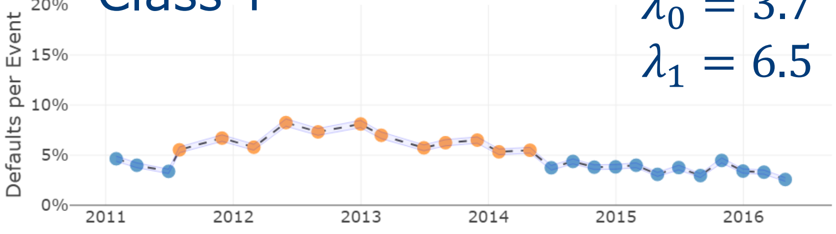


Default rates vary as a function of CRIF Credit Bureau Score

Data Analysis

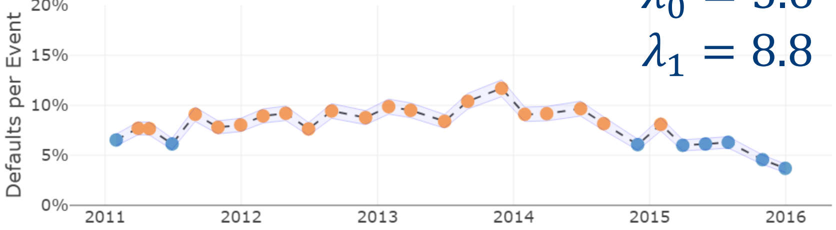
Class Y

$$\lambda_0 = 3.7$$
$$\lambda_1 = 6.5$$

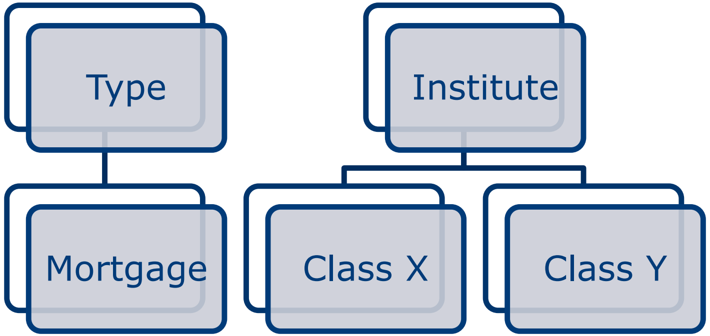


Class X

$$\lambda_0 = 5.6$$
$$\lambda_1 = 8.8$$



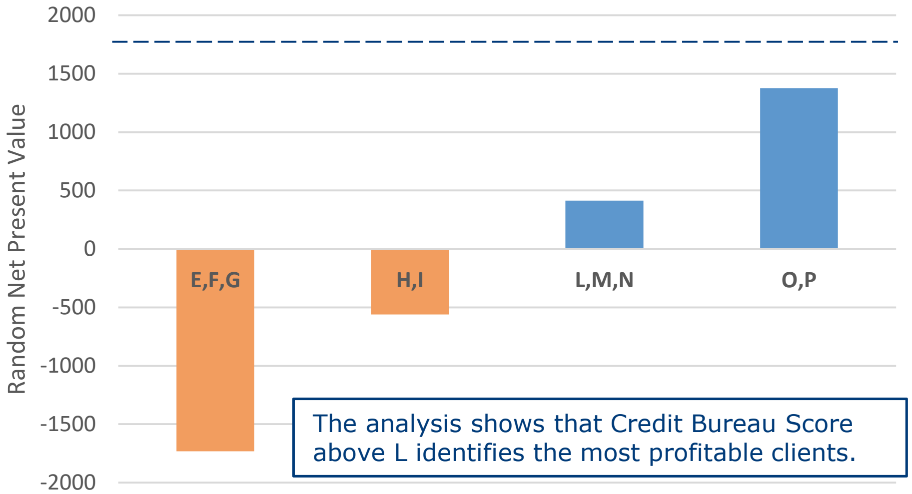
Default



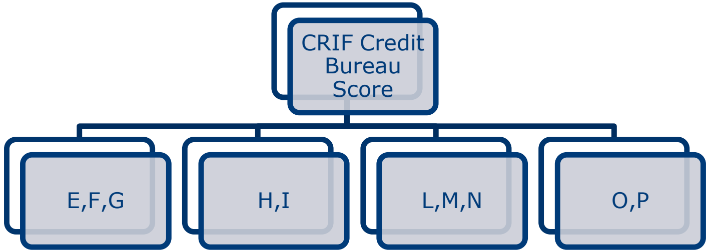
Not all banks have had the same default rate

Data Analysis

Random Net Present Value (1 loan)



The analysis shows that Credit Bureau Score above L identifies the most profitable clients.



Loan Condition

$$NPV_{contr} = 1835\ €, \quad n_{months} = 60$$
$$r_h = 190\ €, \quad w = 8500\ €$$
$$d_{rate} = 0.04, \quad x = 0.010179$$

$$\lambda_{A,0} = \cdots$$
$$\lambda_{A,1} = \cdots$$
$$A_A = (\ldots)$$
$$\pi$$

Results & Conclusions

- Description of the evolution over time of the credit environment in terms of a Hidden Markov Chain
- Estimation of HMC parameters on CRIF Credit Bureau data using Poisson distribution in multiple scenarios
- Calculation of Random Net Present Value based on the resulting parameters for a specific case and applicability to a portfolio

Thank you for the attention

CRIF Together to the next level

MONTE DEI PASCHI DI SIENA BANCA DAL 1472

UNIVERSITÀ DEGLI STUDI FIRENZE