DEPARTMENT OF ECONOMICS AND STATISTICS

SCHOOL OF ECONOMICS AND MANAGEMENT

MASTER OF SCIENCE IN

FINANCE

MANAGING PREPAYMENT RISK

Supervisor:

Prof. Giampaolo Gabbi

Co-supervisor:

Fabio Massimo Valente

Author:

Federica Buonocore

Academic Year 2017-2018

Abstract

This thesis wants to present the main problems which arise when dealing with prepayment risk on mortgages. The topic is not new but an agreed methodology to deal with prepayment risk has not been reached yet. This thesis proposes a solution exploiting a Survival Analysis log-logistic AFT model and testing it on real data. The aim is to predict the probability to prepay on the basis of mortgage/householder characteristics and economic/financial data. The model found a range of survival probabilities depending on the values of the explanatory variables. This will give the bank a starting point instrument, compliant to the Basel Committee IRRBB standards, able to provide competitive advantages supporting risk management and commercial choices.

Acknowledgements

I would like to express my deepest appreciation to my supervisor Professor Giampaolo Gabbi, who guided me in the elaboration of this project, and to my co-supervisor Fabio Massimo Valente, Head of Liquidity and Interest Rate Risk department at Montepaschi Group, who gave me all the support I needed for the analysis and supervised me together with his colleague Letizia De Giorgi. I also would like to show gratitude to Professor Arsenio Stabile for his statistical advices and to my MPS colleagues for their assistance in the interpretation of data. Finally, a special thank must be expressed to all my family, who encouraged and sustained me during the course of my studies and especially in this very complex year.

Table of contents

Introduction: The Behavioural Finance

The classical finance theory has been questioned after the 2008 financial crisis. Modern Portfolio Theory (Markowitz, 1952) and the Capital Asset Pricing Model (Sharpe, 1964), which are based on the notion that investors act rationally and consider all available information in the decision making process, could not explain the empirical evidence. Moreover, the Efficient Market Hypothesis (Fama, 1970) which asserts that financial markets always get asset prices right given the available information, was deeply criticized. As Krugman (2009) states, "Discussion of investor irrationality, of bubbles, of destructive speculation had virtually disappeared from academic discourse. (...) In short, the belief in efficient financial markets blinded many if not most economists to the emergence of the biggest financial bubble in history''.

Practice has shown that there are many cases where emotions and psychology influence the decision making process, and in which decision makers behave unpredictably, sometimes even irrationally. Analyzing this issue is the aim of behavioural finance.

"Behavioural finance deals with the influence of psychological factors on the behaviour of individuals who make decisions in the field of finance and consequences such a behaviour causes on the market" (Sewell, 2010).

Behavioural finance is not a recent theory. It was born in 1979 by the work of psychologists Daniel Kahneman and Amos Tversky, who developed "prospect theory" to understand how the framing of risk influences economic decision-making. Their initial studies, focused only on risk attitudes, mental accounting, and overconfidence issues, were extended later on by the Nobel prize winner Richard Thaler, who connected many psychological and economics principles. The importance of behavioural finance has grown after the financial crisis, because it challenges the classical theories of efficient markets and rational investors, aiming at explaining "why and how people make seemingly irrational or illogical decisions, why they save, invest, spend and borrow money"(Belsky and Gilowich (1999)) combining economics and psychology. The issue is not if there is information on the market but the ways in which investors interpret information at their disposal.

According to behavioural finance theory[1], financial agents form their expectations follow some principles:

- Heuristics Representativeness: people tend to infer that a single observation is representative of the entire population, thus a well-run company represents a good investment.
- Overconfidence: people overestimate their precision of the knowledge and ability to do well and usually recall more their past successes than their failures.
- Anchoring: people will tend to hang on to losing investments by waiting for the investment to reach the value it once had, even though it has no relevance to its current valuation.
- Framing: people perceive a loss as less painful when it is an increment to a larger loss than when it is considered alone.
- Herding: people are influenced in their decisions by what the others are doing.
- Prospect theory: people are risk adverse with gains but often risk seeking with losses, trying to recover.
- Loss Aversion: people weigh losses significantly more heavily than gains. Thus people are more likely to sell a winning stock rather than a losing stock (disposition effect)
- Regret Aversion: people are unable to decide because they feel sorrow after making an error in judgement.
- Mental Accounting: people prefer buying dividend paying stock because they don't have to sell stock to pay for life's necessities.
- Self-control: people believe they can influence uncontrollable events, so investors like to look at charts, although charts are theoretically not helpful in predicting the future prospects for a stock.

Far from being a closed research field, behavioural finance nowadays is extending its reach from micro to macroeconomic themes. Models with behavioural variables are seen to provide a better fit to the data of the housing bubble[2] while the influence of behavioural finance can be used to explain the mechanisms of monetary policy with a model that can explain asymmetries in monetary transmission (Santoro et al., 2014)

1 As explained in the paper Behavioural Finance:a challenge to market efficiency. Mulla Parveen Yusuf. (2015)

2 Housing and behavioural factors. John H. Huston* and Roger W. Spencer Applied Economics Letters, 2014 Vol. 21, No. 3, 215–219

In recent years regulators have recognized that people are not rational in their decisions and that the customer behaviour can influence the stability of the banking system. This lead to the generation of the new regulation frontier called “nudge”. The term “nudge” represents regulatory action that aims at encouraging virtuous customer behaviours without enforcing rules but taking advantage of behavioural factors. The “nudging regulation”, firstly developed by Richard Thaler and Cass Sunstein[3], today is considered more effective by many experts. “Due to cognitive limitations and because laws and regulations lack an adequate ergonomic approach, consumers are unable to make appropriate choices. Nudging can be an alternative: regulators willing to create a pro-consumer environment can design rules that favour choices that are considered better for the public.”[4]

On the other side, commercial banks are deeply interested in understanding, modelling and predicting customer behaviour and it is the consequences on their business. Indeed, bank’s activity is founded on trust and history showed how trust can be weak and easily removed. Moreover, banks’ core business of lending and borrowing is strictly connected to the choices of their clients: if all the clients decide to withdraw money at the same time, the bank will probably face liquidity problems. This is an extreme example, but many common customer behaviours can deeply affect bank’s balance sheets and generate high risks. The problem is that in their day by day activity banks give the customer many options embedded in the product they buy, that will have important consequences on liquidity, funding and interest rate risk. Some examples of this kind of behavioural options are:

- accelerate the speed of redemption of a term deposits with a contractual maturity term or with step-up clauses
- extend the debt maturity (restructure)
- withdraw a stochastic amount of money (with a maximum level) using credit cards
- bank overdrafts

3 Nudge. Improving decisions about health, wealth and happiness. Richard Thaler and Cass Sunstein (2008)

4 European Banking 3.0. Bank Industry and Supervision in the Behavioural Finance Revolution. Fondazione Rosselli (2015)

- pipeline exposures (fixed rate loan commitments)
- unexpectedly close bank accounts at any time
- switch the rate type or modify the payment conditions of a loan or a mortgage
- unexpectedly repay a mortgage residual debt before the contractual maturity

This thesis will be focused particularly on the last topic, which represents one of the biggest risk. Mortgages are one of the main assets in the balance sheets of commercial banks, a great source of profit but also of potential risks. Standard mortgages give the mortgagor an implicit option of full prepayment flexibility, that is to say, the borrower is allowed to pay off the entire principal outstanding at any time. Due to prepayment, the actual maturity of a mortgage contract is stochastic and generally shorter than its contractual maturity. This causes to financial institutions several problems, not only in terms of missing earnings but also for the impacts on liquidity and interest rate risks due to the stochastic cash flow schedule and the arising mismatch between interest rates paid and received that can lead to over-hedging costs. In literature this problem is called “prepayment risk” or “early termination risk”.

Therefore it is important for banks to understand the factors driving this choice.

Behavioural theory states that people are not rational in their financial decision, and this is particularly true even when they have to decide whether or not they should exercise the implicit prepayment option embedded in their mortgage contract. Behavioural risk are those risks that can’t be predicted using economic or financial factors but arise irrationally, driven by human aspects. Prepayment belongs to behavioural risks because the prepayment option is not always exercised optimally: the mortgagor doesn’t exercise the option when it is profitable while it happens that he exercises the option when it is out of the money. The problem is that in the real world the behaviour of a mortgagor can not be easily predicted, because his financial decisions are guided by many aspects which include financial considerations but may go beyond them. Personal specific characteristics of the mortgagor, like his age, his job position, or the geographic area where he is living can impact on his decision to renegotiate the contract conditions (or sell the property) as well as the decreasing market rates.

The irrational behaviour of customers leads to misleading results when trying to quantify and price the prepayment option using the common risk neutral models. As a consequence, banks should use a prepayment model that takes into account also specific behavioural factors, in order to properly charge the mortgagor of the implicit option cost. Moreover, this will aid banks Asset and Liability management, with significant effects on their balance sheets.

Before 2007 in Italy this problem was very limited because banks were free to charge a prepayment penalty on the mortgage contract and usually this penalty was so high that prepayment events were very limited. But after 2007 the situation is deeply changed because of a new law that allows the mortgagor to prepay his residential mortgage without any penalty.

The aim of this thesis is to propose a prepayment model for residential mortgages, focusing in particular on fixed rate mortgages. The resulting model wants to predict the probability of mortgagor‘s prepayment using economic and behavioural factors. In order to do this, a preliminary analysis to discriminate the main characteristics driving the prepayment event is made.

More in detail, this thesis is structured as follows:

Chapter 1 illustrates the general background and the main problems arising when banks are facing prepayment on mortgage loans, presenting risk and commercial problems.

Chapter 2 describes the European situation and the best practices in modelling prepayment risk on mortgage loans, exploiting the results of the 2017 ECB stress test on sensitivity analysis of IRRBB. It will present the new regulatory standards on the topic. Then it will illustrate the principal solutions suggested in literature for modelling prepayment risk, analysing pros and cons.

Chapter 3 is focused on the model application to mortgages real data: it shows the resulting model and the main covariates.

The last chapter sums up the findings and goes on illustrating the possible commercial applications of the described model.

Chapter 1. Causes and effects of prepayment

1.1 The increasing relevance of prepayment risk

The global financial and economic crisis caused the drastic drop of European inflation to a level well below the ECB's definition of price stability (inflation close but below 2%). The ECB counteracted, adopting expansive measures that were mainly directed at ensuring the provision of liquidity and repairing the bank-lending channel, in order to bring inflation back to 2% in the medium term. Among the many unconventional measures adopted, it is worth to mention that the Main Refinancing Operations (MRO) and Longer-Term Refinancing Operations (LTRO) started to be allocated at a fixed rate and full-allotment basis. This essentially meant that banks had unlimited access to central bank liquidity, on the basis of pledged collateral. In particular, the ECB official interest rate level on MROs, which had climbed up until 4.25 on July 2008, had a drastic drop on October 2008 and continued to fall in the subsequent years until reaching the minimum level of zero on March 2016, which is still on. The low ECB rate level encouraged the reduction of the main reference rates used for mortgages, the European Interest Rate Swap rate and the Euribor rate. The Graph 1 shows, for example, the last 10 years evolution of the ECB rate, together with the 30 years European Interest Rate Swap rate and the Euribor 6 months rate, which are on July 2018 respectively at the level of 0%, 1.48% and -0.27% :[5]

Graph1: 10 years evolution of the ECB rate, Euribor 6M rate, EurIrs 30y rate

[5] Source: https://www.mutuionline.it/guide-mutui/osservatorio-tassi-mutui.asp

The low interest rates level positively influenced the recovery of the European housing market. Focusing on the Italian current situation, the residential market is slowly recovering after the financial crisis, with a positive trend that is going on since 2013. Northern regions have the best performances, but the trend concerns the overall area. Mortgages demand is increasing as well, following the same geographical distribution.

Moreover, the low level of interest rates enhanced banking competition. As the April 2018 Bank of Italy "Bank Lending Survey" shows, banks are easing credit supply. Banks margins are reducing mostly on new fixed-rate mortgages and therefore families are keeping on moving their preferences on fixed-rate loans.

However, the recovery of the mortgage demand hides some important risks for banks. Indeed, when signing a mortgage contract, the bank expects to receive a (fixed or floating) regular payments for the borrower for a predetermined period of time, while giving the mortgagor the loan he needs to purchase a home (or using the property as collateral).

However, it often happens that the cash flow deriving from the mortgage deviates from the contractual schedule, with big consequences for the lender. This deviation is essentially due to the exercise by the mortgagor of the implicit options embedded in the mortgage contract, which he can exercise whenever he wants during the life of the mortgage contract. The range of possible embedded options is huge but it is worth to summarize the main cases:

- switch the rate type from fixed to floating
- switch the rate type from floating (or capped floating) to fixed

- modify the amortizing plan
- extend the contractual maturity
- partially or totally reimburse the outstanding debt before the contractual maturity

The last option is the focus of this analysis, and it is known in literature as “early termination risk” or “prepayment risk”.

At each payment date the mortgagor must decide whether to regularly make the next scheduled payment or to exercise the prepayment option. If he does so, he decides that is not convenient to keep paying the current amount of interest and reimburses the remaining outstanding notional, thus closing the contract without waiting until contractual maturity. The available possibilities are

- Curtailment (partial prepayment): repay only part of the outstanding, without closing the contract but benefit from reduced interests
- Conversion: closed the loan and replace it with a new one
- Full Prepayment: fully repay the outstanding notional, thus closing the contract before maturity.

To understand the main reasons of this behaviour, we can have a look at what happens empirically. First, we should distinguish between residential and not residential mortgages. Residential mortgages show a relationship between prepayment rates and loan age which is usually S-shaped: prepayment is generally low shortly after origination and sharply rises as the mortgages’ life goes on, reaching a maximum level, to slowly decline until arrive at a steady-state level near maturity. Not residential mortgages instead, tend to last much longer and are repaid later. Second, prepayment for residential mortgages is negatively correlated with house price appreciation, while it is positively correlated to the divorce rate. Focusing the analysis only on residential mortgages, and looking at the relevant prepayment variables, they can be classified as follows[6] :

- Financial market conditions (the stochastic path of market interest rates)
- Loan characteristics (loan amount, loan-to-value ratio and note rate)
- Housing market conditions

6 Unobserved heterogeneity in Mortgage Termination Risk. Clapp, Deng and An. 2013

- Personal characteristics (salary, age, etc.)

1.1.1 Refinancing

The main cause of prepayment is refinancing, i.e. exploit the upgraded commercial conditions to obtain better financing conditions.

Some refinancing incentives are the variation in the mark-up applied to the reference interest rate, mainly driven by mortgage competition, so that it may be profitable for the borrower to refinance. Also changing in the borrower's credit standing can influence his prepayment decisions because a rating upgrade makes the borrower think to deserve a lower rate and therefore he is more likely to prepay (anyway this the last case is more correct for companies then for residential mortgagors, since it is very difficult for an householder to be aware of its rating).

Mortgage specific characteristics, like disbursed amount, loan duration, Loan to Value ratio can also impact on the borrower's decision to refinance. It is proved[7] that the higher the current Loan to Value, the lower the probability to prepay. This happens because when the value of collateral drops below the loan balance, the borrower will find hard to access to capital markets or he needs to add additional cash to refinance, therefore he is discouraged of prepaying. In the same way, the higher is the disbursed amount the less likely is for the borrower to have the economic means to reimburse the entire outstanding debt before maturity. Loan duration has a different interpretation: it is shown that the prepayment path is S shaped, so that mortgages with very high or very low duration are less likely to be prepaid than those in between.

However, the most important refinancing incentive is the stochastic path of market interest rates, which is also the most accessible information for borrowers.

Indeed, if the mortgagor were rational, he would exercise or not the prepayment option according to the value of market interest rates, following refinancing reasons. Refinancing is attractive when the currently paid mortgage rate is higher than the mortgage rate available on the market. Literature showed[8] that the right to refinance the

7 Bennett, Peach and Peristiani (2001)

8 Hendershott and Van Order (1987) and Kau and Keenan (1995)

mortgage provides the borrower an American call option on the mortgage debt with a strike price equal to the unpaid mortgage balance: at any time during the life of the mortgage, the borrower can decide to return the outstanding principal, so that no further claim of any kind will be pursued by the lender against him.

Following the option pricing theory, a rational mortgagor would exercise the refinancing option when the interest rate at which the mortgage may be refinanced is lower than the contractual rate paid by the borrower. However the empirical evidence shows that borrowers do not exercise the option as optimally as do owners of other financial options[9], even if there is an inverse relation between the contractual mortgage rate paid and the effective duration of the mortgage. Many mortgagors fail to exercise the option to prepay when the prevailing mortgage rate is below the contract rate, whereas the option is frequently exercised even if the market rate is above the contract rate. This is true especially for retail householders, where the generally low level of financial culture prevents mortgagors from operating in an efficient market. Moreover, the presence of significant prepayment rates also on floating rate mortgages can't be explained by the optimal exercise of the financial option.

Indeed, borrowers show diverging abilities to time optimal exercise as well as other non-rational heterogeneities in borrower behaviour. One consequence of this heterogeneity is the so called "burnout effect", which describes what happens in a pool of mortgages under refinancing incentive: the first refinancing incentive generates a wave of prepayments from the most financial active borrowers (called "fast refinancers") while the other part (that does not prepay the first time) is unlikely to prepay even when a new refinance incentive occurs[10].

Summing up, it is true that financial variables are helpful but refinance incentive alone is not able to explain why prepayment occurs (or not occurs), because people are often irrational in their financial decisions. Therefore, the choices of mortgagor must be driven also by non-financial factors, which are called "behavioural factors".

1.1.2 Household mobility

9 Deng, Quigley and Van Order (2000)

10 Gonchanov (2002)

Regarding residential mortgages, the choice to sell a property can go beyond financial reasons. Moving is the best example. Move decision is the consequence of the so called “housing dissatisfaction” which in turn is caused by “changes in the needs of a household, changes in the social and physical amenities offered by a particular location, or a change in the standards use to evaluate this factors”[11].

The main elements driving the household mobility are customer specific behavioral factors and can be divided into household characteristics and exogenous circumstances.

Household characteristics are physical characteristics like age or the marital status: the age is negatively correlated to household mobility because transfers for work and change in family size or composition are less frequent the older is the mortgagor, thus there are less prepayment events for moving reason. Indeed, being married, having children and currently having a job significantly deterred household mobility. On the other hand household mobility increased with salary because the improved family salary makes the mortgagor aim for a better home (bigger or in a more prestigious area).

The exogenous circumstances can be divided into economic factors and the so called “trigger events”[12] .

Economic factors[13] that influence move decisions are socioeconomic indicators like local unemployment rate or local house price. Intuitively, local unemployment rate is positively correlated to prepayment for moving reasons so the geographical location of the property can give information about the probability of emigration for working transfer. The house market path, instead, should be taken into account because in periods of house price appreciation, “home sales and mortgage originations may increase as the expected return on investment rises”[14]. Therefore, when the house market goes up, the number of prepayment events due to relocation increase, while the opposite happens during periods of price depreciation or price uncertainty as home sales and mortgage originations tend to decrease.

However, after controlling for economic and customer personal variables, many other unforeseeable events can happen and lead the mortgagor to prepay. Those behavioral

11 Speare (1974)

12 Deng, Quigley and Van Order (1996)

13 Pavlov (2001)

14 Agency (2014)

factors are called “trigger events” and their key determinants can be shortly listed[15] as follows:

- Unemployment or Curtailment of salary
- Illness or death of mortgagor
- Excessive obligation
- Marital difficulties or divorce
- Illness or death in family
- Extreme hardship
- Business failure
- Property problems
- Inability to sell or rent properties
- Employment transfer or military service

1.1.3 Other events

Prepayment does not always happen for refinancing or household mobility reasons. For example, prepayment can be caused by debt restructuring, when the borrower early terminates the old contract because he opens a new mortgage with the same financial institution, but changing notional or some other financial conditions. In this case the risk is less, because the bank keeps to margin.

One special case of prepayment is default. The borrower in default will not pay the contractually scheduled instalments, exercising the “default option”. This is an event whose frequency has spike in the last years after the financial crisis. Banks use their internal profiling models to discriminate the credit deserving customers. The assignment of the proper rating or probability to default should enable financial institutions to charge the customers for their default risk. However many Italian banks are facing problems because they should improve the credit risk process

15 Modeling of Mortgage Prepayments and Defaults. Lakhbir Hayre. Citigroup. (2006)

The relevance of prepayment risk for Italian banks has grown a lot during the last decade. The low interest rates level enhanced the refinancing incentive and encouraged mortgagors, and especially householders with fixed-rate residential mortgage, to review their contractual interest rates. This was facilitated by a new law on mortgages: before 2007 the prepayment option was exercised by paying a penalty proportional to the residual debt of the mortgage. So designed, this penalty was quite high and served as deterrent for early termination decisions. Art. 7 of Law 40/2007 (called "Decreto Bersani") eliminates, for mortgage agreements entered into after 2 February 2007, the possibility of the bank to charge a penalty in case the borrower requests for early full or partial termination. Moreover, the law facilitated the possibility for the mortgagor to change bank, following more profitable financial conditions. Given the huge amount of mortgages in the banking books affected by the law, the scenery for banks changed radically.

The new law and the reduced market interest rates led to higher competition on mark-up margins between financial institutes, while rapidly increasing the volume of prepaid mortgages. This behaviour affected both fixed and floating rate mortgages, even if is the fixed rate case that arises the biggest risk, as it will be explained later. Financial institutes are forced to face prepayment risk, because its consequences have impact on balance sheets that can no more be neglected.

Prepayment is a big source of risk for banks, not only because it causes an interruption or a reduction of the expected cash flows, but also for its implications regarding some important banking risks such as Liquidity risk, Interest rate risk, Securitisation risk, Internal Transfer Rate risk. Therefore it is very crucial for the lender to have reliable estimates of the future prepayment rate and understand the main factors driving it. A correct model may provide the bank a better knowledge of its customers' financial needs, and this can be the base for improving the commercial relation with the client. Moreover, a prepayment model will help banks to predict the amount of inflows and outflows, which in turn will have big consequences on banks' balance sheets. The aim of the following paragraphs is to illustrate these aspects more in detail.

1.2 Liquidity risk

“Liquidity is the ability of a bank to fund increases in assets and meet obligations as they come due, without incurring unacceptable losses. The fundamental role of banks in facilitating the maturity transformation of short-term deposits into long-term loans makes banks inherently vulnerable to liquidity risk, the risk that demands for repayment outstrip the capacity to raise new liabilities or liquefy assets.”[16]

Basel III standards[17] require banks to measure, manage and monitor liquidity risk in relation to both a short term time horizon (“Liquidity risk” tout court or “Operational Liquidity”) and a medium-long term time horizon (“Funding Risk” or “Structural Liquidity”).

The short term liquidity is measured through the Liquidity Coverage Ratio, defined as the ratio between liquidity reserves and net liquidity outflows in a stressed period of 30 days. The structural liquidity is measured using the Net Stable Funding Ratio, defined in a 1 year time horizon as

$$\frac{\text{Available amount of stable funding (liabilities)}}{\text{Required amount of stable funding (assets)}} > 100\%$$

Paraphrasing, the NSFR requires that long term assets are mainly financed through stable available liabilities.

The Basel Committee defines specifically the assets that compose the denominator and assigns to each one a specific ponderation factor in relation to the following logic: more liquid assets receive lower ponderation because they require less funding while less liquid assets need an higher level of stable finding and so receive an higher ponderation.

For their usual characteristic of medium-long time horizon investment, mortgages and in particular Residential Mortgages belong to the stable funding and so the amount of the residual debt conveys at the computation of NSFR denominator. In detail, “Unencumbered residential mortgages, with a residual maturity of 1 year or more and a risk weight of less or equal to 35%[18]” enter in the computation of the Required Amount of Stable Funding with a ponderation percentage of 65%.

16 Principles for Sound Liquidity Risk Management and Supervision. Basel Committee on Banking Supervision (2008)

17 Basel III: International Framework for liquidity risk measurement, standards and monitoring (2010)

18 Basel III – The Net Stable Funding Ratio. Consultation paper (2014)

Residential mortgages do not receive a ponderation factor or 100% because the Committee recognizes that mortgages have some peculiar aspects that can influence their liquidity and the amount of funding that can be considered stably required. One of this is the prepayment effect, which is recognized by the regulators as having a considerable impact on the representation of the structural liquidity profile of banks: prepayment reduces the loan residual debt, so decreasing the required amount of stable funding, and increasing corresponding NSFR value.

Other considerable aspects considered in the 65% ponderation are the capability of mortgages to create funding by themselves, and this often happens when banks pool mortgages and receive liquidity through the issue of Covered Bonds or by the Securitization Process or fund themselves using them as collateral through the ABACO system (Italian system to pledge credit claims to ECB pooling).

So banks takes into account the prepayment effect while computing the NSFR indicators sent to Regulators. However, it is crucial for them to consider prepayment even while computing their internal monitoring indicators. Gap Ratios, for example, are monitoring indicators used to evaluate the equilibrium of the maturity ladder, where inflows and outflows over the different time horizon are compared. Each maturity bucket has its Gap Ratio.

$$Gap\ Ratio\ x\ Y = \frac{Liabilities\ with\ maturity\ over\ x\ years}{assets\ with\ maturity\ over\ x\ years}$$

In general banks used this indicator with 1Y-5Y maturity bucket range. $Gap\ Ratio\ x\ Y$ different from 100% indicates a funding unbalance for that maturity *x* and this means:

- a need to raise new funding or to refinancing current assets (<100%);
- an overfunding situation to be solved (>100%).

A problem can arise if a bank computes these indicators considering only the contractual cash flows, because in that case it will be unable to catch the cash flow changes due to behavioural dynamics such as prepayment. This means that in monitoring its funding needs the bank would trust on biased indicators, and this can lead management to wrong funding decisions. More in particular, a wrong definition of its liquidity profile exposes the bank to overfunding risk, that is to say, the risk to issue on higher maturities than needed, with higher cost of funding: funding the entire notional amount of a mortgage until its contractual maturity costs more than funding part of the notional amount (in case of partial prepayment) or the entire notional until a shorter maturity.

Liquidity problems arise both on fixed rate and floating rate mortgages, independently on the rate typology. Therefore it is very important for banks to evaluate and include prepayment risk in their indicators, to avoid overfunding problems and unjustified liquidity costs. The result will be more realistic liquidity management and a funding cost reduction that will have great impact on their balance sheets.

1.3 Interest rate risk

“Interest rate risk is the exposure of a bank's financial condition to adverse movements in interest rates. Accepting this risk is a normal part of banking and can be an important source of profitability and shareholder value. However, excessive interest rate risk can pose a significant threat to a bank's earnings and capital base. Changes in interest rates affect a bank's earnings by changing its net interest income and the level of other interest sensitive income and operating expenses. Changes in interest rates also affect the underlying value of the bank's assets, liabilities, and off-balance-sheet instruments because the present value of future cash flows (and in some cases, the cash flows themselves) changes when interest rates change. Accordingly, an effective risk management process that maintains interest rate risk within prudent levels is essential to the safety and soundness of banks."[19]

Usually banks centralize the management of interest rate risk belonging to its business units to benefit from netting (one access to market). Then the Asset and Liability management department of the bank has the mission to apply hedging strategies to manage the risk at a centralized level. The simplest hedging instruments are Interest Rate Swaps, Options (cap and floor) or Swaptions (normal or bermudan).

For example, suppose a bank with the only risk position represented by a fixed rate mortgage: the bank receives from the customer a fixed flow of payment, therefore it is exposed to the risk of missing earnings in case interest rate increase.

So if the bank wants to hedge the fixed rate position against the risk of increasing interest rates it would empower the ALM department to enter into an opposite Interest

19 Principles for the Management and Supervision of Interest Rate Risk. Basel Committee on Banking Supervision (2004)

Rate Swap where it pays the fix rate and receive the floating rate, or it can decide to buy a cap option where the bank receives money if rates are above the fixed level.

Alternatively, bank could prefer to decide in the future if it is convenient to enter or not into an interest rate swap according to changes in market conditions. In this case the ALM department could buy a Swaption that gives to the bank the right to enter into an interest rate swap at one (Normal) or more (Bermudan) predefined schedule of potential exercise dates.

Centralizing interest rate risk means that the ALM department has to evaluate an hedging strategy for the overall interest bearing portfolio of assets and liabilities (banking book). Given the typical relevance of mortgages in the portfolio of commercial banks, the prepayment phenomena can deeply affect interest rate risk.

First of all partial or total prepayment impacts directly on the bank's balance sheet, decreasing the net interest income by reducing (or eliminating) the flows of interests earned on mortgages.

Moreover, high levels of notional prepayment can modify the overall interest rate exposure for a certain time horizon. With the exposure changed it may be the case to modify also the correspondent hedging strategy, and in many cases this means closing the hedging transaction that now have become useless, thus facing closing costs.

In addition, prepayment introduces a new source of risk: an incorrect valuation of the amount prepaid and, by consequence, of the future cash flows exposes banks to the risk of over-hedging, that is to say the risk of hedge the position entering into an interest rate swap with an excessive notional with respect to the real one, having unjustified costs. In case of fixed rate mortgages that means exposing the bank to the risk of paying higher fixed rates on the derivatives than the one collected on the newly stipulated loans, with significantly impacts on the balance sheets.

On the other hand, if the bank could evaluate correctly the prepayment dynamics it would benefit from hedging cost reduction, because it didn't need more to pay for hedging the overall notional till contractual maturity but it just need to hedge for a shorter time horizon (in case of total prepayment) or a lower notional (in case of partial prepayment).

Prepayment on floating rate mortgages arises interest rate risk too, even if in a slighter way. In that case the bank just loses the commercial spread earned on the transaction so has the opportunity cost of missing earnings derived from reinvesting the commercial spread. The impact on balance sheets is less. But we should not forget that floating rate mortgages can be “hedged” too by floating-to-fixed instrument. Therefore the floating rate mortgage case arises the same problems in terms of over-hedging of the fixed rate mortgage case.

1.2 Securitization risk

Securitization is the process through which illiquid assets are transformed into financial securities and sold in the market. The typical example are mortgages loans, but this process can hold for many other kind of credit obligations. Mortgages loans are impossible to sell directly on the market because each one has its unique characteristic and should be administrated separately. In the securitisation process the loans are pooled and repackaged into interest-bearing securities. The advantage for the holder of the underlying debt, the bank, is that it can get immediate liquidity from the market to refinance a debt that is born as illiquid. Moreover, the credit risk (and interest rate risk) of the notes are transfer to the market, while the bank can benefit from the removal of the assets from the balance sheet. Banks commonly do this issuing Covered Bonds or Asset Backed Securities (ABS).

The mortgage prepayment conditions this process: prepayment modifies the estimated duration of the mortgages and this determines a mispricing risk of the securities.

Usually Covered Bonds are priced following this steps: first of all the mortgage duration is estimated, assuming a certain interest rates level and prepayment speed, then the bond is priced setting a yield equal to the treasury ZC yield with the evaluated duration plus a spread. Prepayment accelerates the duration decline, reducing the price of the mortgage portfolios underlying the Covered Bond and the overall bond price.

The adoption of a prepayment model, constantly tested and updated would allow banks to better estimate the effective fair value of the reference portfolio. This would mean more stable balance sheets and reduction of the funding cost through the market issuing of Covered Bonds.

1.3 Internal transfer rate risk

Every bank uses a system of internal transfer rates in their daily activity. This system is made by a series of fictitious transactions, priced at the internal interest rate (called Internal Transfer Rate - ITR), between the business units and the central Asset Liability Management department (ALM).

The main purpose of the ITR is to transfer the interest rate risk and liquidity risk (sometimes also currency risk) from the business unit where they are generated to the ALM department, which has the mission to evaluate and hedge (manage in general) these risks at a centralized level. ITR remunerates ALM for assuming the liquidity, interest and currency risk and leaves credit risk and the other risks inside the business unit, which is remunerated by the commercial mark up (or mark down for liabilities). In this sense, the ITR is commonly used to evaluate the profitability of each business unit.

Banks use this system to centralize risks, and defines the cost of funding for the business unit. This cost is reflected inside the pricing policies definition thus, despite being a figurative system, the consequences are very concrete.

The ITR value is determined in relation to the effective interest rate that can be traded by ALM department on the market. Market interest rates changes with maturities, so usually different levels of internal interest rates are set for different maturities.

Exemplifying, in a common mortgage operation, the figurative ITR system would work as follows:

- The customer asks the business unit for a mortgage with a notional of 100.000
- The business unit asks the ALM to fund that notional
- The ALM gives funding receiving the ITR rate
- The business unit gives the mortgage to the customer and receives cash flows with an interest rate given by the ITR rate plus the commercial mark-up.
- The ALM can decide to bear the risk or use the ITR rate to raise funds on the market. The ITR rate is usually greater or equal to the financing market interest rate

In conclusion, if the overall system works well, each business unit is remunerated and the cost of all this structure is reversed to the customer in its mortgage rate.

In this context the exercise of the prepayment option embedded in the mortgage contract can generate an internal transfer rate risk, because the ITR value computed without taking into account the possibility of prepayments may be incorrect. Indeed, the established value of the internal transfer rate includes the usual amortizing process but it doesn't take into account the possibility that the funded notional could drop in the future. This means that the resulting ITR is higher than it should be, because ALM assumes to need funding for the original notional till the contractual maturity, when market rates are higher.

The main consequence of applying a higher ITR level is that the final customer pays too much, since the ITR conveys at determining the final rate.

On the other hand, a correct evaluation of the customer prepayment risk would enable the bank to better calibrate the funding needs of that position, with a reduction of its funding costs. The lower internal cost for that position would allow the business unit to apply a discount to the customer final rate while preserving its mark-up margins. This will lead to many commercial implications, in terms of pricing competitiveness, trust and reputation reinforce and so on. The next paragraph describes them more in detail.

1.4 Commercial implications

Managing all the risks illustrated in the above paragraphs is a costly activity for the bank. Usually the remuneration of all those risks is reflected into the pricing policies applied to the customer in the mortgage contract.

Without the adoption of a proper prepayment model the arising costs to be remunerated could be higher for the bank and resulted less competitive on the market. The use of a correctly backtested model would generate commercial benefits for the bank in many ways:

- get a better profiling of its customer, allowing a better personalization of the product offer

- define different pricing policies for different groups of customers in relation with the expected duration of their loans, derived from the model using the customer specific information
- higher level of pricing competitiveness on the market, because the bank is able to reduce the final price applied to the customer while preserving the commercial spreads in a profitable way as a consequence of the reduced funding and risk management costs
- achieve new clients, and reinforce the trust of the old ones, as a consequence of the pricing competitiveness

Chapter 2. Dealing with prepayment risk

Chapter 1 illustrated why managing prepayment risk has become so important for Italian financial institutes, but very similar considerations can be made for all European countries. That's why European financial institutes started to develop their own prepayment forecasting models. This chapter will show the main models used to deal with prepayment risk by the most relevant European institutes and the recent standard given by regulators to guide the implementation of future prepayment models by 2018.

The following paragraph illustrates the current situation.

2.1 The 2017 Stress Test Results [20]

On 28 February 2017 the European Central Bank announced a stress test exercise on the Sensitivity Analysis of Interest Rates Risk in the Banking Book, which interested the 111 significant banks under the ECB's direct supervision. The exercised was focused on the analysis of how an interest rate shock would affect banks through changes in the economic value of the assets and liabilities that are not related to the banks' trading activities (banking book). Moreover "the exercise also aims to analyse how banks' models of customer behaviour impact their interest risk measurement, as such behaviour may change in response to changes in interest rates." The results were published on 9 October 2017.

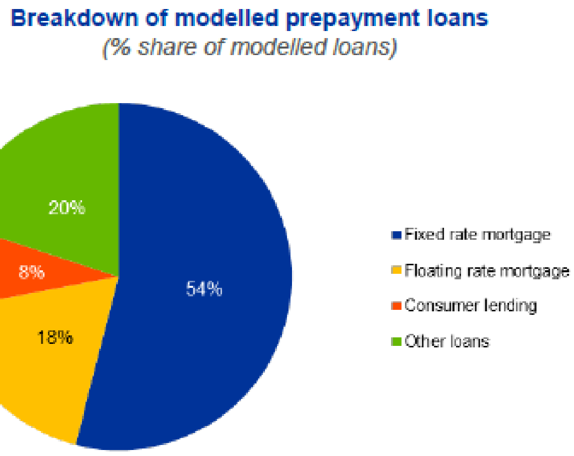
The exercise provided a useful summary of the situation regarding the perceived importance of loan prepayments by the European institutions. First of all the exercise found evidence of the already mentioned refinancing incentive guiding the choice to prepay: "borrowers pay back or renegotiate some of their loans ahead of schedule, especially when market rates decrease below loan rates and in jurisdictions where penalty fees are low or absent".

Secondly, the exercise reports the situation regarding the volume of loans involved in the application of prepayment models: on average, the total amount of modelled loans is

20 Sensitivity Analysis on IRRBB – Stress test 2017 Final Results

3.5 trillion of euro[21], the 54% of which regards fixed rate mortgages. This reflects the major significance of prepayment risk on fixed rate mortgages with respect to the floating rate ones (which represents only the 18% of the modelled loans).

Graph2: Sensitivity Analysis on IRRBB – Stress test 2017 Final Results



To deal with the loan early termination risk banks heavily rely on models of customer behaviour, which is used by 55% of the institutions (61 banks over 111). More in detail, 46 institutes over 111 (40% of the sample) have developed a model for the specific case of fixed rate mortgages. This is also the type of mortgages for which banks have developed the most sophisticated modelling approaches.

The majority of the prepayment models for fixed interest rate mortgages use the interest rate level as explanatory variable for predicting the amount of expected prepayments.

The application of those models shorten the duration of some loans, especially house mortgages that are long-dated and fixed rate.

In particular, for fixed rate mortgages the application of prepayment models meant that the average duration of the sample portfolio of loans reduces at 5.6 years from the previous 7.6, meaning "the weighted average remaining time until repricing". Not only, the final results provide an estimate of the Conditional Prepayment Rate, that is to say the share of loans expected to be paid back within one year applying prepayment models (and assuming to maintain the current interest rate environment), which for the most relevant case of fixed rate mortgages is 7.5%.

21 The report specifies how the amount is generated:"Each institution was asked to report the five most relevant loan prepayment models for each of the four categories above. Information is related to EUR positions only"

Moreover, the usage of prepayments models deeply mitigates the impact on banking account under stress conditions: the average Economic Value of Equity impact in a +200 bps shock scenario (that is to say, how much would the net present value of the banking book change in response to IR changes) would be -2.7% of the Common Equity Tier 1, applying loan prepayments models, instead of -11.1% CET1.

2.2 Best Practices

For a deeper examination of the possible solutions it is worth to have a look at how prepayment risk is modelled by some significant European financial institutions:

UnicreditGroup:

The Unicredit Group well specifies on its balance sheets how they deal with prepayment risk. The estimates of the prepayment behaviour are based on historical prepayment data as well as trend analysis. The results of the prepayment model are integrated in the interest rate management strategy. Different models and different hedging strategies are identified for the different countries, in relation to the importance of the phenomenon and to the main factors driving prepayment, which may change from one country to another. As written in Pillar III: ”In Italy the prepayment expected profile is implicitly taken into account by treasury while hedging for commercial assets interest risk. The prepayment risk for the German mortgage portfolio is driven by the level of the interest rates and by the behaviour of the customers independent of the level of the interest rates. The interest rate sensitive prepayments are rather small at the current level of the interest rates and are hedged by swaptions. The not interest rate sensitive prepayments are hedged via swaps according to the Interest Rate Risk strategy of the bank. The prepayment risk in the Austria and CEE countries loan portfolio is deemed residual therefore no prepayment hedging strategy is applied.” The internal Risk Management function is responsible for validating and implementing the model, “to map the liquidity profile of balance sheet items”

Intesa San Paolo Group

Intesa San Paolo Group specifies that they compute shift sensitivity of the fair value of the banking book, calculated at individual cash flow level for each financial instrument, based on different instantaneous rate shocks.
In measurements, capital items are represented based on their contractual profile, but for those instruments that can have a different risk profiles it is used a behavioural representation to calculate the risk measures. In particular “for mortgages, statistical techniques are used to determine the probability of prepayment, in order to reduce the Group's exposure to interest rate risk (overhedging) and to liquidity risk (overfunding)”.
This model is more sophisticated than the previous one since it makes also dynamic margin simulation analyses to combine shifts in yield curves with changes in customer behaviour in different market scenarios.

UBI Banca

The UBI Banca approach is more traditional. They measure the exposure to interest rate risk by using gap analysis and sensitivity analysis models on banking book instruments.
However, in performing the sensitivity analysis they include an (not better specified) estimate of the “impacts resulting from the early repayment of mortgages and long-term loans, regardless of whether early repayment options are contained in the contracts”.

Deutsche Bank

The Group manages the interest rate risk exposure of its loans using a replicating portfolio approach to determine the average repricing maturity. They include also the prepayment behaviour of its customers. The model uses parameters chosen following historical observations, statistical analyses and expert assessments.

Commerz Bank

The bank doesn’t seem to use a mathematical approach to model prepayment risk, but they define the impact of early repayments and on investor behaviour using assumptions.

BNP Paribas and Montepaschi Group

The Group models prepayment integrating constant prepayment rates in their cash flow model.

Société Générale S.A.

Société Générale includes prepayment in determining the maturity of its loans. To model prepayment they use models based on customers' historic behaviour patterns.

Banco Bilbao Vizcaya Argentaria, S.A.

Banco Bilbao adopts a more complex solution to model prepayment risk. They use a statistical scenario-simulating model to elaborate the impact of prepayment on balance sheets both in terms of deviations in net interest income and in terms of the impact on economic value. The model includes changes in market interest rates but also other more personal variables such as inheritance, divorce, change of residence that may create incentives for the Bank's customers to cancel loans or deposits early, thus modifying the future behaviour of the balances on the balance sheet with respect to forecasts, in accordance with the contractual calendar of maturities. Historical information relating to prepayments, and to interest rates, are used to estimate the relation with variables and customer behaviour. The hypotheses are reviewed at least on an annual basis.

2.3 Basel Committee on Banking Supervision Standards

Prepayment has always been recognized as a source of risk by regulators, although a complete formalization of its discipline has been done only recently. The Basel Committee on Banking Supervision firstly issued in 2004 the "Principles for the Management and Supervision of Interest Rate Risk" where the prepayment topic was mentioned as a source of interest rate risk together with the other embedded optionalities in the formula "loans which give borrowers the right to prepay balances"[22]

22 Principles for the Management and Supervision of Interest Rate Risk. Basel Committee on Banking Supervision (2004)

However, in the last decade the focus on prepayment risk has grown and in the revised version of the “Standards on Interest Rate Risk in the Banking Book (IRRBB) ”, issued on April 2016 the Committee included some specific considerations about how to deal with prepayment risk in the revised principles for banks.

Principle 5 states: “In measuring IRRBB, key behavioural and modelling assumptions should be fully understood, conceptually sound and documented. Such assumptions should be rigorously tested and aligned with the bank’s business strategies.” A new attention is given to behavioural aspects that can influence the fair value of a product.

In the subsequent explanation of the principle the Standards also make a list of the common products with behavioural optionalities, including prepayment risk in its more serious declination, that is to say, with regard to fixed rate loans. “Fixed rate loans subject to prepayment risk: Banks should understand the nature of prepayment risk for their portfolios and make reasonable and prudent estimates of the expected prepayments. The assumptions underlying the estimates and where prepayment penalties or other contractual features affect the embedded optionality effect should be documented. There are several factors that are important determinants of the bank’s estimate of the effect of each interest rate shock and stress scenario on the average prepayment speed. Specifically, a bank must assess the expected average prepayment speed under each scenario“.

In the revised principles the Committee goes deep in detail in explaining how banks should measure and manage prepayment risk. Standards are technical documents so they also deal with technical considerations regarding the model assumption: “Modelling assumptions should be conceptually sound and reasonable, and consistent with historical experience.”

Moreover, they define specifically the factors that should be considered while modelling prepayment risk: ”Loan size, loan-to-value (LTV) ratio, borrower characteristics, contractual interest rates, seasoning, geographical location, original and remaining maturity, and other historical factors. Other macroeconomic variables such as stock indices, unemployment rates, GDP, inflation and housing price indices should be considered in modelling prepayment behaviour”.

With regard to prepayment of floating rate loans, the Committee just says "… banks should consider the materiality of the impact of behavioural optionalities within floating rate loans. For instance, the behaviour of prepayments arising from embedded caps and floors could impact the banks' economic value of equity."

Finally the standards stress that the model and its assumption should be regularly backtested, and give some suggestions about how to keep the prepayment model updated, making an example: "As market conditions, competitive environments and strategies change over time, the bank should review significant measurement assumptions at least annually and more frequently during rapidly changing market conditions. For example, if the competitive market has changed such that consumers now have lower transaction costs available to them for refinancing their residential mortgages, prepayments may become more sensitive to smaller reductions in interest rates."

The modelling prepayment approach of the Basel Committee was confirmed in July 2018 by the EBA's published "Guidelines on the management of interest rate risk arising from non-trading book activities"[23]. Banks are expected to apply these guidelines from June 2019, so there is still time to see how the standards are interpreted by banks and to measure their effectiveness. However some initial consideration can be done exploiting the 2017 stress test results as we did in the first paragraph.

2.4 Survival Analysis: literary approach

Prepayment is not a recent issue, even if is importance has been recognized only recently. Before the last decade the most popular framework to deal with prepayment risk was the option theoretic model, which modelled the prepayment option as a financial option embedded in the contract, replicating a mortgage subject to prepayment with a callable bonds issued by the mortgagor or, very similarly, with a lookback put option.

The problem with this approach was that it assumed that borrowers behaved purely rational and did not allow for the use of explanatory variables. Moreover, as explained

[23] Available on the web site:
https://www.eba.europa.eu/documents/10180/2282655/Guidelines+on+the+management+of+interest+rate+risk+arising+from+non-trading+activities+%28EBA-GL-2018-02%29.pdf

in the first chapter, history has shown that mortgagor, especially retail mortgagor, does not exercise the option optimally, but can take prepayment decisions without any rational explanation, therefore trying to predict their behaviour using rational models is useless.

In the last years, also encouraged by regulators, the best practices on the topic have moved to the so called "behavioural models", i.e. statistic and econometric models, which try to explain the customer choice using a set of covariates. The advantage of these models is that they allow a more flexible approach in the determination of the covariates and are built to take into account the behavioural factors driving the prepayment choice.

Among the wide range of statistic and econometric solutions adopted, a special attention should be given to the models belonging to the "Survival Analysis", a branch of statistics developed mainly in the latter half of the 20th century, which aims at modelling the lifetime (or survival time) of an individual, that is to say the time elapsing between a starting point to an event. The starting point is often called "failure" and can happen at most once for any individual, the length of time is called "failure time" and the data that observe this construction are called "survival time data".

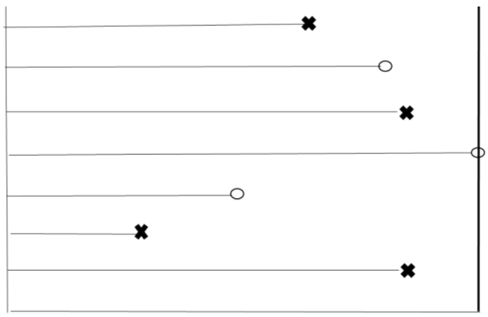
The application of Survival Analysis started in medicine but the methodology has already been extended to many other fields such as public health, social science, engineering and economics, where it is commonly known as "duration analysis".

The choice of Survival Analysis for modelling the time elapsing between the origination and the prepayment of a mortgage is due to the ability of this methodology in dealing with peculiar aspects of this data, such as censoring and non-normality, which complicate a lot the analysis through traditional models like multiple linear regression, logistic regression or ANOVA.

Censoring is defined as "the possibility that some individuals may not be observed for the full time to failure"[24]. It can be easily understood in the medical field, where usually a sample is observed from the starting time when the therapy begins, till an ending time when each individual can be still alive or dead. However it can happen that an individual in the sample die for a different reason than the disease for which he is under therapy or that the individual is still alive at the moment when the study ends so that we

24 Cox D.R., Oakes D., "Analysis of Survival Data", Chapman and Hall, 1984

are unable to say if he will die or not in the future. In this case the data are right truncated because we know the starting point but not the ending time. Other censoring possibilities are “left censoring”, when we don’t know the exact time when the individual started the therapy, or “interval censored”, when the data are simultaneously left and right censored. Censoring arise a problem of missing data that may bias or make not reliable the traditional models results.



Censoring may exist in any type of survival data. In our case, mortgage data can be seen as survival data for that a mortgage is observed from origination to its last payment. In particular we will define as failure the prepayment event, and we might have right censored observations, because a mortgage can die before its contractual time for other reasons, such as default, and above all because we are unable to say if mortgages that are still paying when our study ends will experience a prepayment event in the future.
Usually this kind of observations are not left censored because the time when the mortgage begins is written in the contract and stored in the banking databases.
Obviously in mortgage data the time in the study does not coincide with a fixed calendar time but the origination time for each contract will be considered as starting time.

The aim of the Survival Analysis is to give a shape to the Survivor Function of the random variable T, the failure time, that is to say find the probability that the failure will happen after a certain time t. In analytical terms, if T has a cumulative distribution function $F_T = P\{T > t\}$, the survivor function of T is

$$S_T(t) = P\{T \geq t\} = 1 - F_T$$

And the density function is $f_T(t) = -S'(t)$

To understand the shape of the survivor function it is fundamental to know the hazard function, which represents the instantaneous probability to fail exactly at time t, given that the observation is still alive before t

$$h_T(t) = \lim_{\Delta t \to 0+} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t} = \frac{f_T(t)}{S_T(t)} = \frac{-S'_T(t)}{S_T(t)}$$

For example if the hazard function is constant over time this means that the probability to fall in every instant time interval is fixed. From the hazard function we can know the hazard rate of each individual and make comparison, understanding who has the greatest risk of failure at time t. The most common shapes of hazard functions are monotone, ∪-shaped or ∩-shaped[25].

Notice that, opposite to the survival function which describes the probability of not having an event at time t, the hazard function describes the probability of the event occurring at time t. However there is a strong connection between the aforementioned functions: knowing the shape of $S_T(t)$ we can derive the correspondent $h_T(t)$ and, vice versa, the survival function can be written in terms of the integral of the hazard function

$$S_T(t) = \exp\left(-\int_0^t h_T(u)du\right)$$

The hazard function is commonly represented in terms of its integral over time, the cumulative hazard function $H_T(t)$, which tells the accumulation of the risk over time

$$H_T(t) = \int_0^t h_T(u)du = -\,log[S_T(t)]$$

Survival Analysis is mainly divided into three approaches: the Non-Parametric analysis, the Proportional hazard approach (which includes the semiparametric and the fully parametric analysis) and the Accelerated Failure Time models. These methods have different properties and interpretations, therefore some considerations must be made in order to choose the model that best fits survival data.

25 Meeker and Escobar (1998), Bagdonavicius and Nikulin (2006), Martinussen and Scheike (2006)

2.4.1 Non-parametric models

Non-parametric models require no assumptions about the distribution of survival time, they are built just on the basis of the observed survival times, both uncensored and censored. They are very useful to understand the shape of the survival function and during the explorative analysis when we need to see the marginal contribution of each variable. The most common non-parametric estimate of the survival function utilize the Kaplan-Meier method, which in its simplest form is

$$S_T(t) = \prod_{j=1}^{k} \frac{n_j - d_j}{n_j}$$

where k represents the total number of time points at which one or more event occurred, j represents a time point, and n_j represents the number of people surviving at time t(j) and d_j represents the number of events that occurred at time t(j)[26].

The value of $S_T(t)$ is constant between the time at which one event occurs and the subsequent time, therefore the estimated function is a step function that changes value only at the time when a new event occurs.

The Kaplan-Meier survival curves can show differences between the survival curves of two or more groups, so that it is possible to graphically discriminate if a variable has a marginal contribution on the survival function because, when this is true, the Kaplan-Meier survival curves generated by its multiple determinations will have very different shapes. However, to be sure that this difference is statistically significant we need to apply another important non-parametric analysis, the Log-Rank test, which is a hypothesis test that discriminates whether two (or more) Kaplan-Meier survival curves follow the same distribution[27].

The Log-Rank statistic test is

$$Q = \sum_{j=1}^{n} \frac{\left(O_j - E_j\right)^2}{E_j} \sim {\chi^2}_{n-1}$$

where O_j and E_j are the observed and expected number of events under the null hypothesis, n is the number of survival functions to compare. Of course, a p-value less

26 Dobson and Barnett, 2008, p. 193

27 Mantel, Evaluation of survival data and two new rank order statistics arising in its consideration., in Cancer Chemotherapy Reports, vol. 50, nº 3, 1966, pp. 163–70

than 0.05 indicates than the null hypothesis of equivalence in the two (or more) survival curves is rejected, so we can reasonably suppose that the variable has some kind of influence to the survival probability.

However, both the Kaplan-Meier estimate and the Log-Rank statistic test are examples of univariate analysis because they describe they describe the survival probability in relation only to the variable under investigation but they don't consider the impact of the other factors. Moreover, log-rank test can only evaluate differences between survival curves but it is not able to quantify how relevant is one variable to the overall survival probability.

Therefore, in a situation where we suppose that more than one variable can affect the survival probability non-parametric analysis is useful to describe data, but can't be used to make inference. For this purpose it is necessary to apply regression methods to the analysis of survival data. We have already mentioned that the standard multiple linear regression or logistic regression are not suitable to survival data, for their non-normality and censoring peculiarities, but survival analysis provide other models to link the survival function with the explanatory variables. Those models are mainly divided into two classes: proportional hazard models and accelerated failure time models.

2.4.2 Proportional Hazard models (PH models)

The class of Proportional Hazard models is mainly divided into two categories: semiparametric and fully parametric models.

Semi-parametric models are an intermediate solution between non-parametric and parametric models. In the semi-parametric models no assumption is made on the structural form of the survival and the hazard functions. The only functional form that must be specified regards the influence of the explanatory variable on the survival function.

The most famous model of this class is the Cox regression model[28], known also as Cox proportional hazard model, which describes the link between the failure event, represented in terms of the hazard function and the explanatory variables, assuming that

28 Cox 1972

the hazard function is proportional to the baseline hazard function $h_0(t)$, with the proportionality constant given by the n predictors on the hazard and their coefficients

$$h(t) = h_0(t)\exp(b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)$$

The baseline hazard function $h_0(t)$ is the value of the hazard function when all the explanatory variables are 0. The correspondent survival function is

$$S(t) = S_0(t)^{\exp(b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}$$

That is of course proportional to the baseline survival (i.e. the survival probability without considering the effects of the explanatory variables)

The advantage of using Cox model is that is simple to use because the baseline hazard function is estimated non-parametrically so there is no need to know the distributional form of the survival times. Nevertheless, it is still possible to obtain reliable estimates of the regression coefficients and therefore of the survival and hazard function.

The Cox model can be though as a multiple linear regression of the logarithm of the hazard function on the explanatory variables, with the baseline hazard as intercept. The baseline changes with time and it is multiplied by the covariates, which act as a proportional factor. This is the reason why this model refers to the class of proportional hazard model: the hazard function of a failure event is proportional to the hazard function of the others. Given two individuals with different covariates x and x', the quantities

$$HR = \frac{h_0(t)\exp(b_i x)}{h_0(t)\exp(b_i x')}$$

are called "hazard ratios" and are time-independent: Proportionally implies that the hazard ratio of two individuals is constant over time. Of course, a hazard ratio greater than one means a greater hazard and so a shorter survival time with respect to the other individual.

When it is possible to recognize the functional form of the survival times another possible solution is to use the Fully Parametric Proportional Hazard models. This class of models is very similar to the Cox regression but, unlike it, they assume a functional form for the hazard function. The coefficients of the Cox model are estimated through partial likelihood while the ones in the parametric proportional hazard model are estimated through maximum likelihood, but the interpretation of hazard ratios is the same and the proportionality assumption is still required.

The advantage of using fully parametric models is their increased efficiency in the coefficient estimates with respect to Cox model. Moreover, the predictions over mean and median of survival times and hazard ratios are immediate to find and to understand. Some common distribution used are Exponential, Weibull and Gomperts, depending on the corresponding functional form of the survival times and their main comparison must be made in terms of the derived hazard function: the Exponential distribution has a constant hazard, the Weibull has a monotonic hazard. Other distributions like the Log-Logistic, that admits a hazard that is increasing and then decreasing, are not possible under the proportional hazard framework since the functional form is kept only by the baseline hazard while changes in the hazard function.

Fully parametric PH models are not used very frequently, because parametric AFT model are as simple to apply and easier to interpret. In can also happen that the two approach give the same model, as it happens when the survival times follow a Weibull distribution (but the parameter interpretation changes).

However, the application of both Cox PH model and Fully Parametric PH model is not possible when the hazard function is not proportional, therefore the proportionality assumption must be checked graphically or with goodness of fit tests such as the Schoenfeld residual test. In case the proportionally assumption fails other solutions should be used, as the accelerated failure time model explained in the next session.

2.4.3 Accelerated Failure Time (AFT) models

The AFT models are a parametric class that proposes an alternative for analysing survival data with respect to the aforementioned proportional hazard model. Under this class of models it is possible to measure the direct effect of the covariates directly on the survival times, while in the proportional hazard model the effect is measured on the hazard and so it requires a transformation to understand the same effect on the survival function. For this reason the results of the AFT models are much easier to interpret with respect to the PH models.

The survival function under the AFT model is assumed to be as follows:.

$$S(t) = S_0(t\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n))$$

Where S_0(t) is the baseline survival function and $x_1, x_2, \ldots, x_n$ are the covariates.

Looking at the formula we can understand where the name accelerated failure time comes from: the survival function obtained from the x_i can be viewed as the baseline survival function accelerated by the factor $\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n)$. The explanatory variables have a constant effect on the survival function given by the quantity $\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n)$ which is called "acceleration factor": the covariates increment or reduce the survival time depending on the acceleration factor to be less then or more than 1.

The corresponding hazard function is

$$h(t) = h_0(t\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n))\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n)$$

Let T be the random variable representing the failure time, it has the following relation with the failure time in the baseline scenario (where all the covariates are zero): $= T_0/\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n)$. This lead to the usual expression of the AFT model in log-linear form with respect to time:

$$\log(T) = \mu - (b_1x_1 + b_2x_2 + \cdots + b_nx_n) + \sigma\varepsilon$$

Where T is the failure time, μ is the intercept term representing the logarithm of the baseline failure time $E(\log T_0)$, ε is a random variable of zero mean and independent on x_i, σ>0 is the scale parameter. Under this model the effect of covariates on the survival time is assumed to act additively on the log time scale.

The exponentials of the covariate coefficients, i.e. the quantities $exp(b_i)$, are called "time ratios" and are used to compare different levels of the covariate x_i. Their interpretation is quite simple: when the time ratio is above 1 it means that the effect of the covariate is prolonging the survival time with respect to the baseline survival, while the survival time is reduced if the time ratio is below 1.This simplicity makes the AFT model very easy to understand.

The AFT is a parametric model since the form of the distribution of baseline failure time should be specified, to estimate parameters and make inference. The most used distributions in AFT models are Exponential, Weibull, Log-logistic, Log-Normal, Gamma and take their name from the corresponding distribution of $log(T_0)$. The parameters are estimated through the maximum likelihood method.

The survival distribution of T can be expressed also in terms of the survival distribution of ε:

$$S(t) = S_\varepsilon\left(\frac{\log t - \mu - \beta X}{\sigma}\right)$$

so that we have different AFT models depending on the distribution of ε. The distribution of ε that give the above mentioned failure time distributions are respectively Extreme value with 1 parameter, Extreme value with 2 parameters, Logistic, Normal, Log-Gamma.

The application of AFT models requires the verification of two assumptions: the AFT assumption and the parametric assumption. The following paragraph will show how to check the goodness of fit.

2.4.4 Check the goodness of fit

Literature proposes a wide variety of instruments to check how much the survival regression results fits the data. Every step of the analysis can be monitored using statistical tools, from the choice of the survival class of model to be used (AFT or PH) and the type or model (parametric or non-parametric or semiparametric model), until the goodness of fit check of the final model, passing through the variables selection.
However, checking the goodness of fit of a survival regression is quite tricky, because there is not a unique tool which is efficient for every survival class of model, but every instrument should be adapt to the chosen class and the parametric model. For this reason the literature on the topic may be difficult to understand. This paragraph tries to gather the main approaches:

- Checking if the parametric distribution fits the data:
    - Graphical check: the plot of the logarithm of the cumulative hazard function in each group versus the logarithm of time must show straight lines. This assures that the parametric model we assumed is correct.

- Statistical check: use the likelihood ratio test to understand which parametric model has the best fit to data. This test can be used also to find the best selection of covariates, through a “stepwise” (or “backword”) selection, i.e. computing log-likelihood ratios every when adding (or removing) a covariate to the model, to understand if the variable is relevant for the analysis using the Log-likelihood Ratio Test. The applied formula is

$$LRT = -2log\left(\frac{L_S(\hat{\theta})}{L_g(\hat{\theta})}\right)$$

  where $L_S(\hat{\theta})$ is the log-likelihood of the null model and $L_g(\hat{\theta})$ is the log-likelihood of the alternative model.

  The test statistic will be chi-squared distributed, with degrees of freedom given by the difference in dimensionality of the two models.
- Hazard function check: the best parametric distribution should be also verified through the shape of the hazard function that is assumed to be followed by survival times. As mentioned before, when data are exponential their hazard is constant over time, when data are Weibull distributed their hazard monotonic so always increasing or decreasing, when data follow a log-logistic distribution the hazard is increasing and then decreasing or the opposite.

- Checking if the AFT model is appropriate:
  - Graphical check: the survival proportion in one group at any time t must be equal to the second group proportion at time t multiplied for the acceleration factor. The graphical tool to check this assumption is the Q-Q plot of survival percentiles: must give a straight line passing for the origin
  - Statistical check: find the best fitting AFT model using Akaike Information Criterion (AIC) defined as $AIC = -2l + 2(k + c)$ where l is the log-likelihood, k is the number of covariates and c is the number of model-specific parameters. The lower the AIC result, the better is the model fit to data.

- Residual analysis: the most powerful tool to check the goodness of fit of the model is the residual analysis. For the peculiarities of the Survival Analysis models, the usual regression residuals don’t work well and it is necessary to find other solutions such as:

- Cox-Snell Residuals plot: Cox-Snell residuals are defined as[29]

  $$r_{Ci} = \widehat{H}_0(T_i)\exp(b_1 x_{1i} + b_2 x_{2i} + \cdots + b_n x_{ni}) = \widehat{H}_i(T_i) = -\log(\widehat{S_\iota}(t_i))$$

  where $\hat{S}_i(T_i)$ is the estimate of the survival function, $t_i$is the observed survival time for the i^{th} observation.

  The Cox-Snell residuals exploit the property that, despite the distribution of T, H(T) has an exponential distribution with unit mean. Therefore if the model is appropriate, the plot of $r_{Ci} = -\log(\widehat{S_\iota}(t_i))$ versus r_{Ci} will give a straight line with unit slope and zero intercept. Moreover, Cox-Snell residuals will never be negative or symmetrically distributed around zero.

  In AFT models Cox-Snell residuals can also be computed using the property that $S(t) = S_\varepsilon\left(\frac{\log t - \mu - \beta X}{\sigma}\right) = S_\varepsilon(r_s)$ where r_s is the standardized residual. Therefore the Cox-Snell residuals depend on the parametric distribution of T (from which the distribution of ε is derived).

  For example if T follows a log-logistic distribution, $S_\varepsilon = (1 + e^{\varepsilon})^{-1}$ so that the Cox-Snell residuals of the AFT model are

  $$r_{Ci} = log[1 + \exp(r_s)]$$

  And the AFT log-logistic model is appropriate if the plot of $\log(-\log(S(r_{Ci}))$ against $\log(r_{Ci})$ give a straight line with unit slope and zero intercept.

- Deviance residuals plot: Deviance Residuals are defined as

  $$\widehat{d_\iota} = sign(r_{si})\sqrt{2(\widetilde{l_\iota} - l_i)}$$

  Where r_{si} is the standardized residual, $\widetilde{l_\iota}$ is the actual log-likelihood of individual i and l_i is the maximum possible log-likelihood of that individual. Deviance residuals can be viewed as a micro likelihood ratio test for individual i , but are normally distributed. If the model fits the data, deviance residuals should be symmetrically distributed around zero. However, in case of censored data, the normal distribution is less accurate. Deviance residuals are used to analyse the outliers: a positive deviance residuals means that the failure event happened before the time predicted by the model, while a negative residual means that the event happened later.

29 Cox and Snell 1986

Chapter 3. Modelling Prepayment Risk

Chapter two showed the main approaches in modelling prepayment risk adopted by the European financial institutes and illustrated the different modelling possibilities proposed by literature.

The aim of this chapter is to implement the theory explained in chapter 2 on Montepaschi Group data, developing a model able to predict at a certain confidence level the prepayment time of each subject depending on a set of explanatory variables. The model results could help the institute in improving the management of its risk and reaching a more competitive position on the market.

3.1 Data description

Chapter 1 illustrated the possible drivers that may guide the mortgagors in its choice of paying the next scheduled mortgage instalment or interrupt the contract and reimburse the residual debt ahead of time. Very briefly those drivers were identified in:

- Mortgage characteristics: Loan to Value, Original and Residual Life, Rate Type, Guarantee, ect… (this class was enriched with the credit rating of the lending bank, represented by its CDS Rate)
- Household characteristics: Age, gender, marital status, family composition, internal credit rating class, salary change and extraordinary events
- Socioeconomic factors: Unemployment rate, House Price Index, GDP
- Market interest rate changes

The following analysis aims at checking if those variables are actual drivers when applied on real data. In doing so it has been selected the free statistical software R.

The model is built on a sample of mortgages data provided by Montepaschi Group. The sample is composed by monthly data on around 750.000 mortgages observed during a life period of nine years, starting from June 2009 until May 2018.

The information available for each mortgage are the following:

- CODE NUMBER: internal identification number of the mortgage
- STATUS: life status indication. Further details will be given later
- CUT_OFF: monthly observation date (reference date), starting from June 2009 and ending in May 2018
- CTP_NDG: bank identification number of the mortgagor, or of the first mortgage co-signer in case of mortgages jointly held.
- CTP_SAE: indicates the economic activity sector of the mortgagor, or of the first mortgage co-signer in case of mortgages jointly held, following the classification indicated by Bank of Italy (circ.140[30])
- CTP_SEG: internal client commercial segment, divided into Retail or Wholesale
- CTP_AGEM: mortgagor age class. From 1 to 12 covering all ages. The first class contains mortgagors that are less than 20 years old when signing the contract. The other classes cover 5 years old each, until 70 years old. The last class indicates mortgagors older than 71 year. When there are mortgages jointly held it is used conventionally the age of the first co-signer.
- CTP_GEND: gender of the mortgagor or of the first mortgage co-signer in case of mortgages jointly held. This variable has a third value indicating when the mortgagor is a legal subject.
- CTP_ECAA: group of economic activity of the mortgagor or of the first mortgage co-signer in case of mortgages jointly held.
- CTP_NATI: mortgagor region of address
- CDG_AGE: tells if the loan has some facilities or not
- CDG_RSTR: indicates the kind of loan, if it regards a real property mortgage or not
- CDG_LEAS: indicates residential, guaranteed or non-guaranteed mortgages
- CTP_CLARAT: rating class of the mortgagor according to the internal credit rating model. It is divided into 19 rating class including the Default grade.
- CDG_LTV: Loan to Value Class at the start date. Loan to Value is an indicator given by the ratio between the value of the Loan and the value of the property for which the loan is asked. Banks usually don’t lend money when the LTV is over 80%. The sample contains 15 LTV classes from 0 to 100%

[30]Bank of Italy “ ISTRUZIONI RELATIVE ALLA CLASSIFICAZIONE DELLA CLIENTELA “ update 2014

- DUR_RES: Residual Duration, computed assuming French amortization, so that the residual duration of the mortgage is the half part of the time distance between reference date and maturity date
- RES_LIFE: Residual Life, computed as time distance between maturity date and reference date
- DUR_ORIG: Original life time of the mortgage
- FIN_DTVA: Starting date of the mortgage (first value date)
- FIN_DTSC: Maturity date of the mortgage
- FIN_TIPTAS: type of interest rate applied on mortgages (Fixed/Floater/Mix)
- FIN_DEBRES: residual debt
- FIN_RIMB: type of amortization schedule
- FIN_AMERO: original outstanding
- RCI_TASSO: mortgage rate
- OPR_TIPO: dummy field indicating if the mortgage has an early termination penalty or not
- AEV_AMM: amount prepaid in a totally prepayment transaction
- ERA_CAP: amount prepaid in case of partially prepayment

The variable STATUS summarises the contractual situation of the mortgage at each reference date. Three typologies of status are possible:

1. Totally prepaid mortgages, when the total residual amount has been repaid in one solution before the contractual maturity.
2. Partially prepaid mortgages, when the mortgagor repays part of the residual outstanding during the life of the contract so that the future interest payments are reduced. This event can happen many times during the life of the contract and it can also be the case than many partially prepayments cause a total prepayment.
3. Alive or naturally ended mortgages, when no prepayments have happened during the life of the contract.

To obtain a model compliant with the European Central Bank standards, the original information were integrated adding for each observation date some macroeconomic and financial indicators, such as:

- Gross Domestic Product: annual Italian Real GDP rate[31]
- Unemployment Rate: annual Italian Unemployment rate[32]
- House price index: annual Italian House Price index rate[33]

Finally, the dataset was enhanced with the credit rating of the lending bank, represented by its CDS rate (annual CDS rate variation[34]) with the aim of including the risk perceived by the household when keeping his relationship with the bank. Empirical evidence showed that when a customer closes his accounts with the actual bank and transfers them to another bank, usually he moves all the other banking products, including mortgages. So it is easy to expect an increment of the CDS rate to reduce the survival time of the mortgage[35].

3.2 Refinancing Incentive

Looking at the recalled prepayment drivers, it is clear that mortgage characteristics are totally available in the dataset, while household characteristics are partially available since it is impossible for law reasons or expensive to collect and timely update personal information such as family composition or salary change. Basic socioeconomic factors are included in the dataset, even if their number can be extended.

What is not yet part of the data is the market interest rate change, which in Chapter 1 was described as the one of the main drivers of the *Refinancing Incentive*, therefore a new variable must be constructed. Of course, this variable is not important for floating rate mortgages, in which the payment rate changes accordingly to market interest rate changes. For fixed rate mortgages, instead, the movements of market interest rates are crucial: intuitively, when interest rates go down, the mortgagor has an incentive to

31 Italy Real GDP Year on Year Seasonally Adjusted and Working Day Adjusted (EUGNITYY code): "Gross domestic product (GDP) measures the final market value of all goods and services produced within a country. The GDP by expenditure approach measures total final expenditures (at purchasers' prices), including exports less imports. This concept is adjusted for inflation." Source: Bloomberg

32 Italy Unemployment Rate SA (ITEMUNES Index): "The unemployment rate tracks the number of unemployed persons as a percentage of the labor force (the total number of employed plus unemployed)". Source: Bloomberg.

33 Italy Residential House Price index, quarterly average yearly computed on basis 2015. The final variation rate is computed on basis 2009=100. Source: Istat

34 MPS Bank CDS 5y rate (CBMP1E5) computed on basis 2009=100. Source: Bloomberg

35 Notice that the mortgagor is charged for the lender's credit rating, because it is included in the bank's cost of funding, which is part of the final rate applied to mortgage. Therefore it is important to include the CDS rate in the analysis because if it is found that the bank's credit deterioration reduces the survival time of the mortgage, this will mean that the bank has to fund the amount for less time, thus saving money, and this in the end will mitigate the price applied to the customer.

prepay his mortgage. This effect is bigger the higher is the spread between the contractual interest rate he pays and the interest rate he will pay on a new mortgage (considering the same amount of money and same residual maturity).

To construct the Refinancing Incentive variable it was assumed that the borrower forms his decision only on the basis of the market interest rates, so the different commercial mark-up levels that may be applied by competitors are not considered. This assumption is justified by the fact that it is easier for people to access to market rates values (which are available on newspaper, television or internet) than visiting competitors to simulate the effective mortgage rate.

Moreover, it was assumed that the mortgagor wants to refinance exactly the same mortgage, so that the new debt will be equal to the current residual debt, without extending the maturity or incrementing the amount.

Considering these assumptions, what really influences the prepayment decision is just the spread between the original market interest rate, computed on the original duration of the mortgage, and the current level of market interest rates on the current residual duration.

An adjustment must be made for mortgages born before February 2007, for which Art. 7 of Law 40/2007 is not applicable. Those mortgagors still have to pay an extra amount in case of early termination of the contract, therefore the value of the penalty must be included in the computation of their refinancing incentive. In particular, the penalty acts by increasing the current market interest rates, so that a refinancing incentive exists if and only if the current market interest rate level is so low to be below the difference between the original market interest rate and the penalty.

Usually the penalty is proportional to residual debt, but notice that also the two compared market interest rates should be multiplied for the residual debt. Therefore the formula simplifies and assuming that the mortgagor refinances exactly the same residual debt means that the penalty can be considered a fixed amount to be divided by the residual duration.

In mathematical terms:

$$Refinancing\ Incentive = i_0 - \left(i_t + \frac{penalty * f}{d_t}\right)$$

Where:

- i_0 is the original level of market interest rates on the original duration (percentage level),
- i_t is the current level of market interest rates on the current duration (percentage level),
- $penalty$ is the contractual percentage to be applied on the residual debt in case of early termination for mortgages born before 2007, and zero otherwise
- d_t is the mortgage duration at time t
- f is the payment frequency

Notice that in defining the market interest rates level it was used the current and residual duration instead of the real time distance. This is done according to the way banks price mortgages, assuming a coupon paying mortgage (not bullet) with French amortization. In this case the effective cash flows have a duration that is nearly the half of the real time distance, therefore banks pick from the market curve the interest rate relative to the duration time instead of the effective time distance when defining the price of a mortgage.

3.3 Performing Survival Analysis

The analysis is performed on data applying the survival regression theory explained in Chapter 2.

For the purpose of building a survival regression model, it is necessary to identify a *Survival Object* that consists in a dataset containing for each observation two variables: the reference time and the linked survival status. The reference time is defined as the number of months between the origination of the mortgage contract and the observation month (reference date).

The approach followed in this analysis to construct the Survival Object was to extract observation time and survival status from a subset of the initial sample, created on the basis of the different status, using the following criteria:

1. Totally prepaid mortgages: only the information related to the prepayment month were included in the subset. The status assigned to totally prepaid mortgages is “dead”. In survival analysis these are defined uncensored observations since the object event of the analysis, i.e. prepayment, is really observed. The information related to the event observation enter into the sample with full weight 1.
2. Partially prepaid mortgages: the information at each prepayment date are part of the subset. Partially prepayments are assigned status “dead” and are considered uncensored observations since the event is actually observed, but the connected information enter in the sample with a weight given by the ratio between the amount prepaid and the residual debt.
3. Alive or naturally expired mortgages: only the information related to the most recent date are included in the subset. The two cases are merged since in both of them the event is never observed, that’s why the assigned status is “live” in both cases. Nevertheless, in the survival object there is a distinction between the two events: mortgages that are still alive when the study ends are identified as “right censored” because it is impossible to know if they will face a prepayment event after the end of the study. This lack of knowledge results in missing data when computing regression analysis and makes unreliable other regression estimates besides the survival regression.

The analysis will be performed following these steps:

1. Define the problem
2. Find the best model solution
3. Reclassify the covariates
4. Find the max-likelihood model
5. Present the final model
6. Interpret the model results
7. Check the goodness of fit

3.1.1 Define the problem

Usually the first thing to do when looking at data is descriptive analysis. The original dataset was composed by nearly 1 million mortgages, but some filters had been applied before performing the analysis:

- Exclusion of bullet mortgages, because bullet mortgages are paid at once at the end of the period and prepayment analysis can be not sound
- Exclusion of defaulted mortgages, because the cash flows of mortgages on default follow other reasons than the behavioural ones, therefore prepayment analysis makes no sense
- Exclusion of expiring mortgages: because the prepayment of a mortgage that has only one payment left makes no sense and gives no risk
- Exclusion of divided mortgages, to avoid biases due to data repetition

The sample obtained applying those filters is of nearly 750.000 mortgages.

Looking at the main mortgage characteristics, the composition of the sample is the following:

Fixed Rate	Floating Rate
30%	70%

Real Property	No Real Property
81%	19%

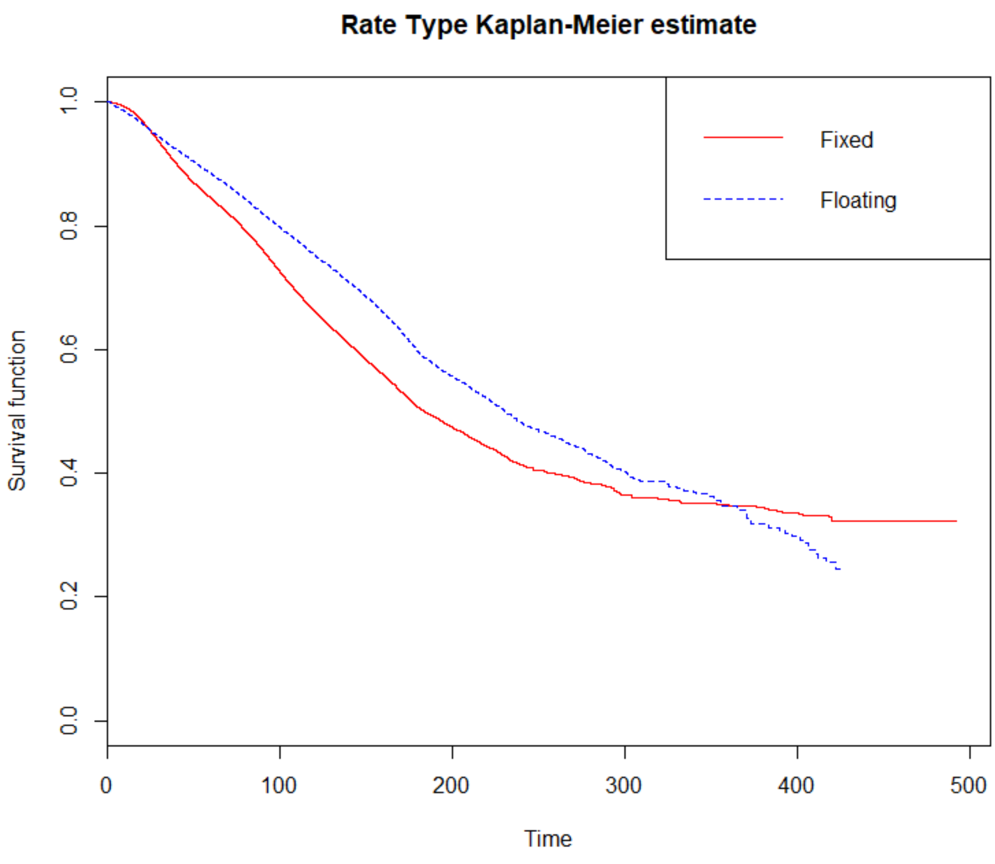
Retail	Wholesale
85%	15%

Residential	Non Residential
72%	28%

Now, it is clear that the composition of the mortgages portfolio is unbalanced. It presents mainly Floating Rate mortgages, coherently to the overall Italian situation. However, the first chapter explained that the prepayment phenomena arises a problem much more serious for fixed interest rates mortgages than for floating rates one, so it would be the case to see if there is a difference in the prepayment behaviour of fixed and floating rate mortgagor, to develop a more useful prepayment model.

Fortunately, the Survival Analysis provides a very powerful tool to understand the shape of the (baseline) survival function, of the hazard rate function and to verify if a variable has a significant contribution to the survival function profile: the Kaplan-Meier non-parametric methodology (described in Chapter 2). This methodology will be used to understand the differences between the two survival curves of fixed rate and floating rate mortgages.

Applying the Kaplan-Meier non parametric analysis the shape of the survival function in case of fixed rate and floating rate mortgages is the following:

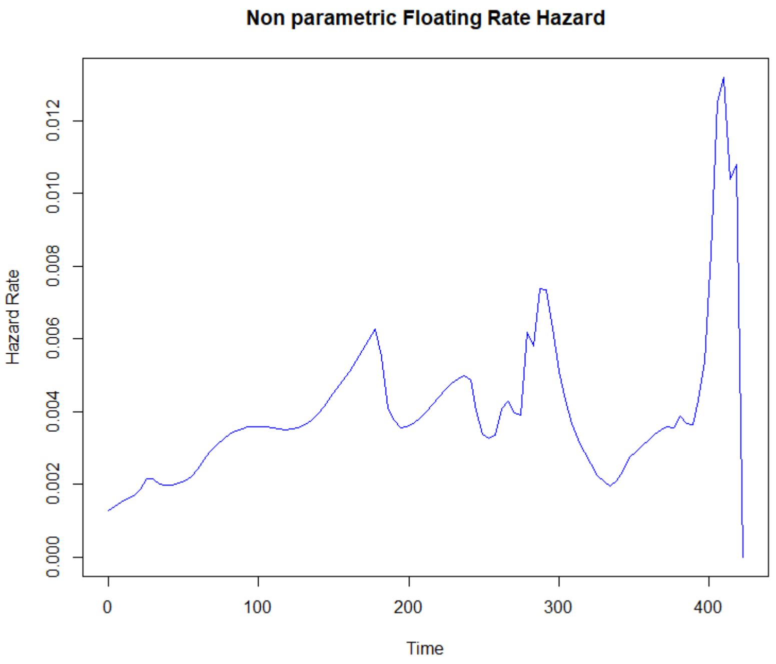


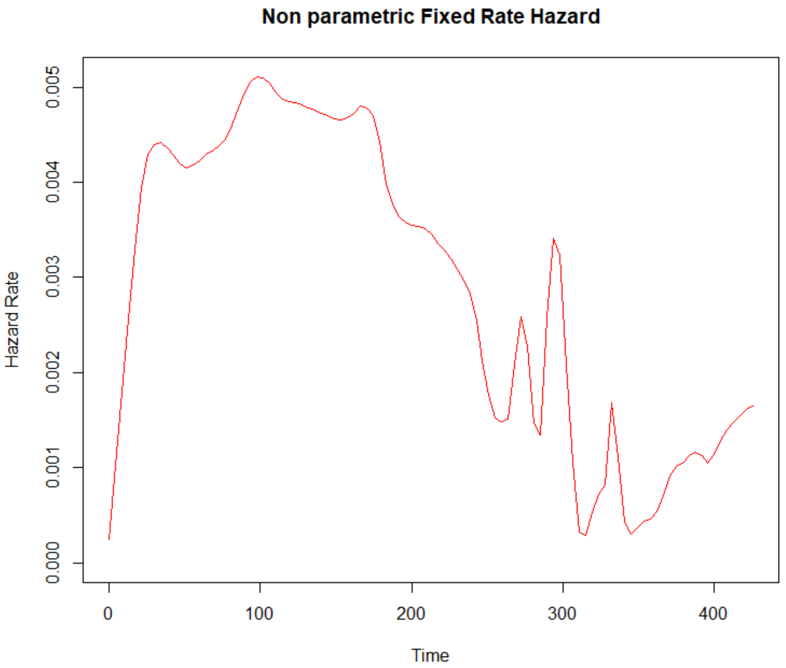
The two survival curves seem to be slightly different, and this difference can be checked statistically using the non-parametric Log-Rank test:

```
Chisq= 1656  on 1 degrees of freedom, p= 0
```

which gives a very high chi-squared value and, consequently, confirms that there is difference between the two curves at a high significance level.

Moreover, what determines the shape of the survival function is the shape of the relative Hazard Function, which represents the instantaneous probability to fail. The two hazard functions can be plotted using non-parametric methodologies:





The non-parametric hazard rate function confirms the difference between the prepayment behaviour of fixed and floating rate mortgagor. Therefore, to find a model that best fits the data it is necessary to split the analysis and focus on the riskiest case for banks, the fixed rate one, which is composed by around 207.000 mortgages[36]. The following analysis will be made only on this case. A short overview of the floating rate model will be given in the appendix.

Now, to better delineate the problem, it should be considered the updated composition of the other variables, in the fixed rate scenario:

Real Property	No Real Property
81%	19%

Retail	Wholesale
90%	10%

Residential	Non Residential
73%	27%

It is clear that the fixed rate portfolio is concentrated on Real Property mortgages made for Residential purposes by Retail customer and therefore only these ones will be part of the analysis.

The final sample will be composed by around 133.000 observations, because it is possible to compute the Refinancing Incentive only for mortgages originated after 2000, for lack of market data[37]. Therefore the max time value is 225 months.

Notice that with the information available in the dataset it is not possible to tell whether the mortgage prepaid is transferred to another bank, but this phenomena should be partly caught by the refinancing incentive and the credit spread: when interest rates go down, the incentive to move the mortgage to another bank increases; in the same way when the bank credit worthiness get worse, people are more easy to close all the products with the bank.

36 Only the pure fixed rate mortgages are considered, for the sake of simplicity. Modular mortgages are omitted (10% of the sample) because they have switching rate possibilities that may bias the overall prepayment behaviour.

37 Mortgages originated before the year 2000 are 12% of the total. The final number is obtain removing them and excluding observations for which some relevant variable has null values.

Another phenomenon that is not possible to evidence in the data is debt restructuring, i.e. when people close the mortgage to open a new one with different characteristics in the same bank. Both old and new mortgage are considered in the sample.

Summing up, the considered mortgages have the following characteristics:

- No bullet
- No defaulted
- No expiring
- No divisions
- Fixed rate
- Residential
- Retail
- Real property
- Originated after year 2000

3.2.2 Find the best model solution

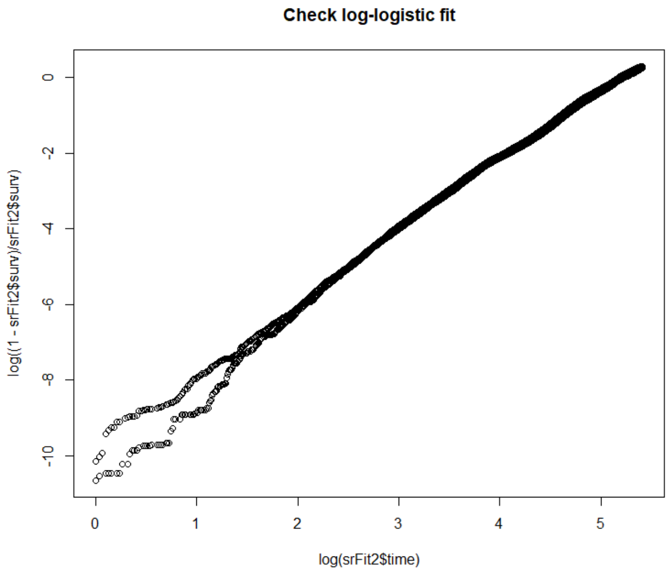
After defining the sample, the next step is to find the best model solution to fit the data. Chapter 2 illustrated two main classes of model that can be used when dealing with survival data: proportional hazard model and accelerated failure time model.

Proportional hazard models are mostly used in their semiparametric declination, which is used when it is not possible to assume any functional form for the shape of the survival function and the hazard function. This is not the case, because the Kaplan-Meier estimation showed that the hazard function is firstly increasing and then declining. Literature tells that the hazard function of an exponential distribution has a constant shape, while it is monotonic in case of a Weibull and the hazard is increasing and then decreasing or the opposite in case of Log-Logistic model, so it might be the case to use a log-logistic model. Later on, it will be checked statistically if this assumption is true.

Moreover, in case of log-logistic distribution the use of fully proportional hazard model is not correct because the log-logistic hazard shape is maintained only in the baseline scenario, while in the other ones the log-logistic fit is lost.

Therefore, if it is found that the best fitting parametric model is the log-logistic one, the only possible survival model to use is the accelerated lifetime model.

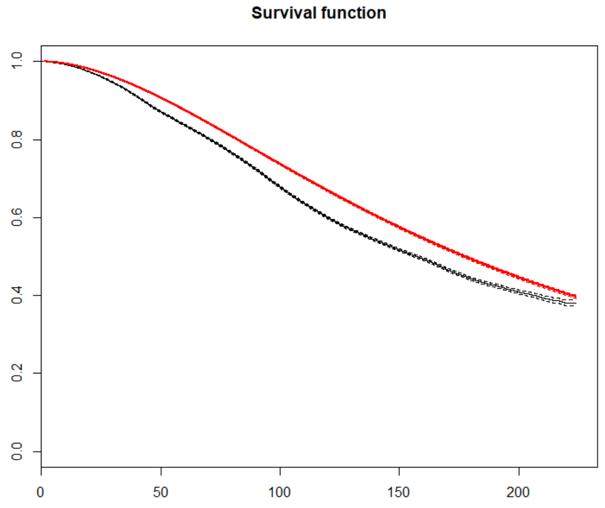
Recalling Chapter 2, there are many ways to verify if the chosen parametric distribution fits the data. One first way is the graphical check: the plot of the logarithm of the cumulative hazard function in each group versus the logarithm of time must show straight lines. In the data one possibility is to divide data following the Loan to Value criteria, assuming for the sake of simplicity only two classes, below 70% and above 70%. The log-logistic distribution passed this first test, as visible in the following graph, where the logarithm of the cumulative hazard function in each class is plot versus the logarithm of time, showing straight lines.

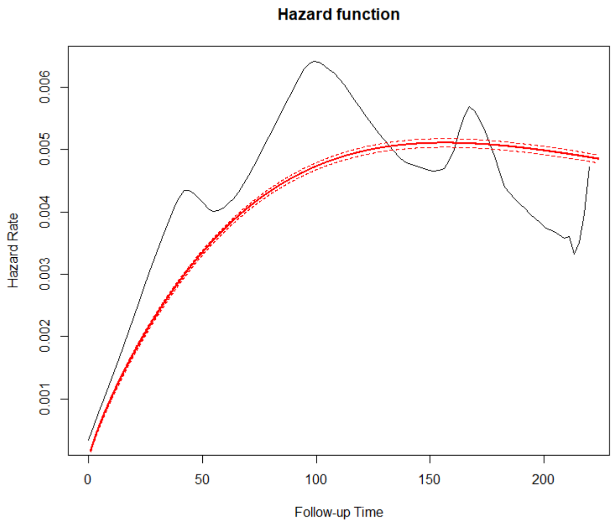


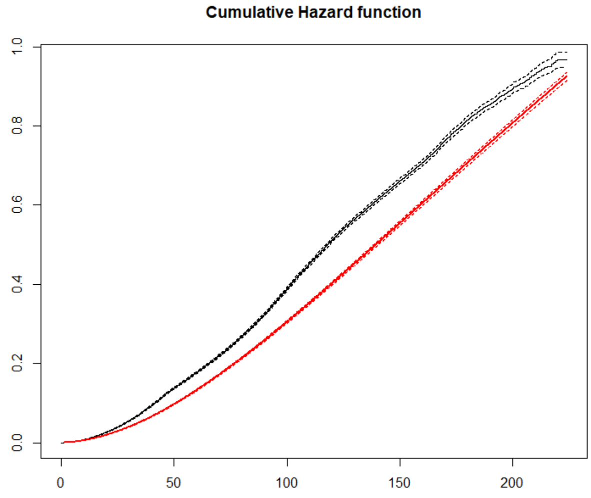
The statistical check uses the Akaike Information Criterion (AIC), which is based on the comparison between the log-likelihood of the different model. The lower the AIC result, the better is the model fit to data. The table below compares the results of the AIC test for the most common models:

Parametric Model	**AIC**	**Log-Likelihood**
Log-Logistic	474,910	- 237,453
Weibull	475,542	- 237,769
Exponential	484,304	- 242,151
Gaussian	487,440	- 243,718
Logistic	492,013	- 246,005
Lognormal	475,530	- 237,763

The AIC test confirms statistically the log-logistic assumption: the log-logistic model is the only one that gives the lowest AIC and the highest log-likelihood. Notice that all the models are performed in the "full scenario", where the effect of the covariates is not considered yet. It is now possible to plot the estimated full log-logistic survival function, hazard function and cumulative hazard function against the equivalent non parametric functions to check the fit graphically:

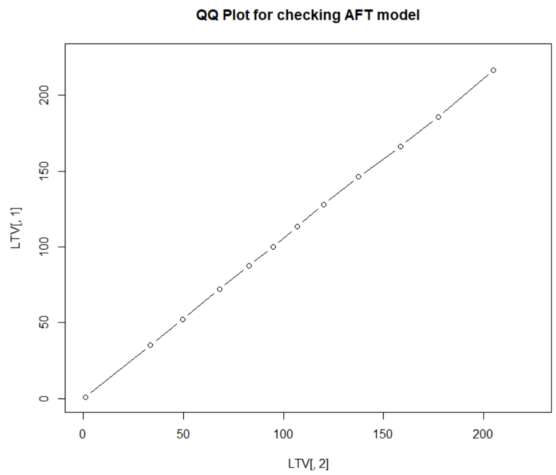






Since the baseline failure time was verified to follow a log-logistic distribution, the analysis should proceed applying the Accelerated Failure Time model, which is the only model that preserves the log-logistic distribution of failure time in all the possible scenarios[38].

However, there is also a graphical plot to verify the AFT assumption: the quantile-quantile (Q-Q) plot. "A plot of percentiles of the K-M estimated survival function from one group against another should give an approximate straight line through the origin if the accelerated failure time model is appropriate. The slope of this line will be an estimate of the acceleration factor"[39]. The Q-Q plot of the survival functions of the two Loan to Value classes confirms the AFT assumption:



38 Cox D.R., Oakes D., "Analysis of Survival Data", Chapman and Hall, 1984

39 "Comparison of Proportional Hazards and Accelerated Failure Time Models". MsC Thesis by Jiezhi Qi 2009

3.2.3 Reclassification of variables

The Accelerated Failure Time model is mostly used for its simplicity of interpretation compared with other models like proportional hazard models. It starts from identifying a baseline scenario and then adds one by one all the covariates. Each covariate acts modifying the baseline scenario, that is to say, extending or reducing the survival time accordingly to the coefficient estimation. The effect on the survival curve is reached through the hazard function, whose new shape is proportional to the baseline one (with proportionality given by the constant acceleration factor).

In particular, a positive coefficient means that the presence of the associated covariate decelerates the baseline failure time, while a negative coefficient tells that the covariate accelerates the verification of the event.

The choice of covariates should be made rationally, including the ones that make sense, and applying a stepwise selection, adding one variable at a time until the model reaches the highest log-likelihood.

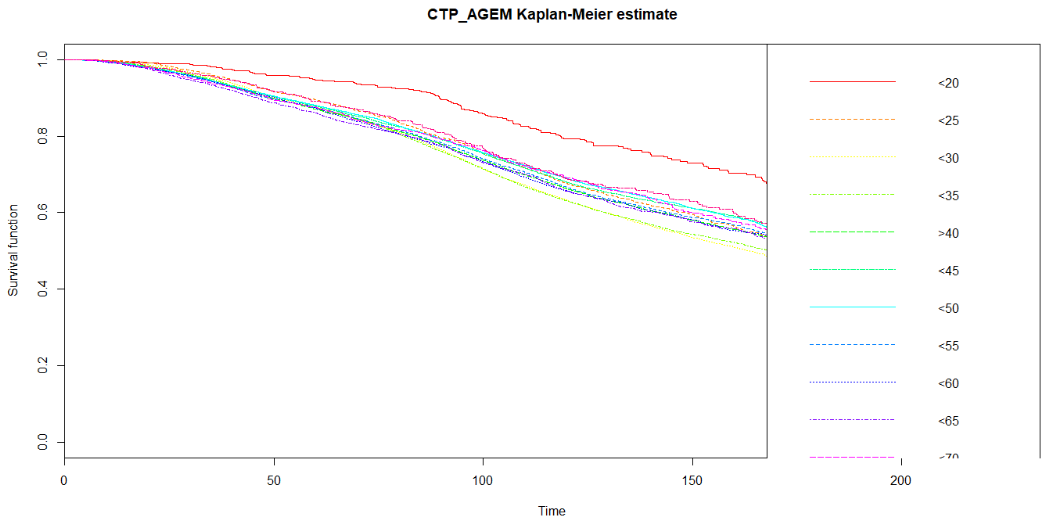
Again, the first thing to do is the Kaplan-Meier analysis of the covariates, that provides an easy tool to see graphically if there is difference between the survival curves generated by the different variable values: if so, it is worth to add the covariate to the model and check the resulting log-likelihood. In the proceeding pages all the variables will be scanned.

CTP_SAE: it is strictly linked to the CTP_SEG variable, therefore in the final sample is deeply concentrated on the retail segment and doesn't give any extra information

CTP_ECAA: unfortunately this information is useless because is missing mostly of the time

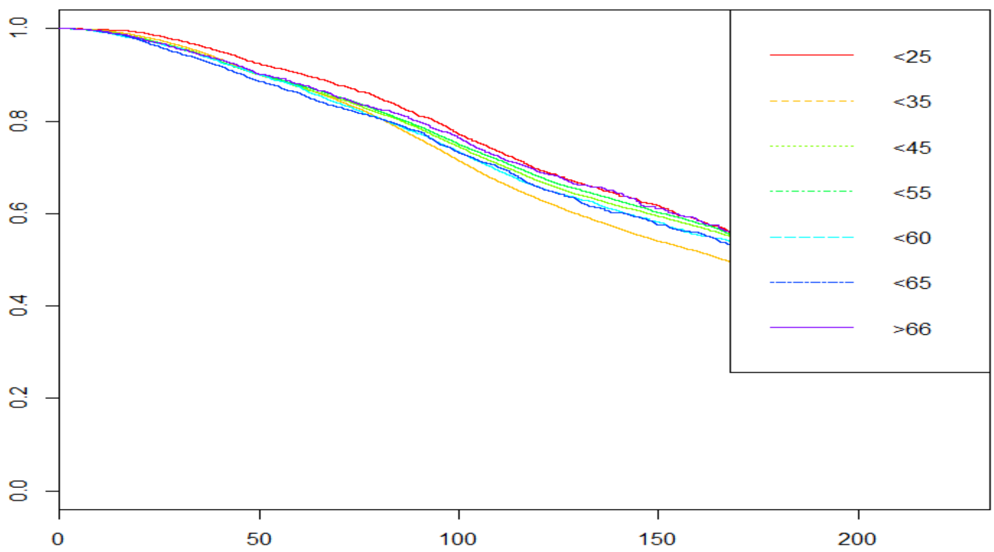
FIN_RIMB: useless, in the final sample almost all the mortgages have French amortizing plan

CTP_AGEM: the original variable has too many classes for a clear interpretation. To aggregate the classes it is good to look both at the significance level of their regression coefficients, and at their non-parametric baseline path and put together the similar ones:



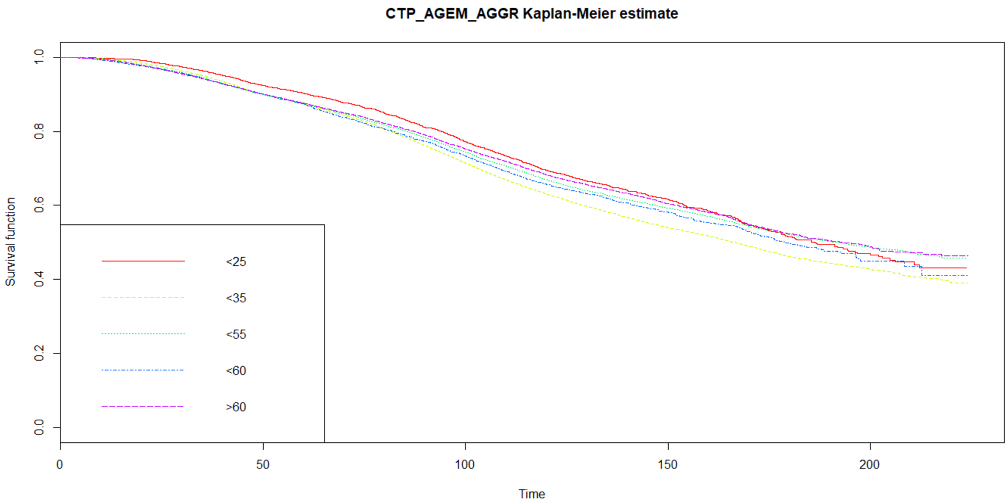

```
Call:
survreg(formula = mySurv_fix ~ CTP_AGEM, data = working_fix,
    weights = WEIGHTS, dist = "loglogistic")
                  Value Std. Error        z        p
(Intercept)      5.2757    0.04401  119.877 0.00e+00
CTP_AGEM<=20     0.3016    0.06713    4.493 7.01e-06
CTP_AGEM(21:25] -0.0522    0.04791   -1.089 2.76e-01
CTP_AGEM(26:30] -0.1559    0.04497   -3.466 5.29e-04
CTP_AGEM(31:35] -0.1536    0.04465   -3.440 5.82e-04
CTP_AGEM(36:40] -0.0922    0.04476   -2.061 3.93e-02
CTP_AGEM(41:45] -0.0402    0.04503   -0.892 3.72e-01
CTP_AGEM(46:50] -0.0364    0.04543   -0.802 4.23e-01
CTP_AGEM(51:55] -0.0872    0.04593   -1.898 5.77e-02
CTP_AGEM(56:60] -0.1098    0.04672   -2.350 1.88e-02
CTP_AGEM(61:65] -0.1222    0.04859   -2.516 1.19e-02
CTP_AGEM(66:70] -0.0526    0.05430   -0.968 3.33e-01
CTP_AGEM>=71     0.0000    0.00000      NaN      NaN
Log(scale)      -0.5880    0.00441 -133.427 0.00e+00

Scale= 0.555

Log logistic distribution
Loglik(model)= -236770.8   Loglik(intercept only)= -236888.3
        Chisq= 235.02 on 12 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 132285
```

The significance levels of classes 21-25, 26-30, 41-45, 46-50, >=71 is low, so they can be aggregated:



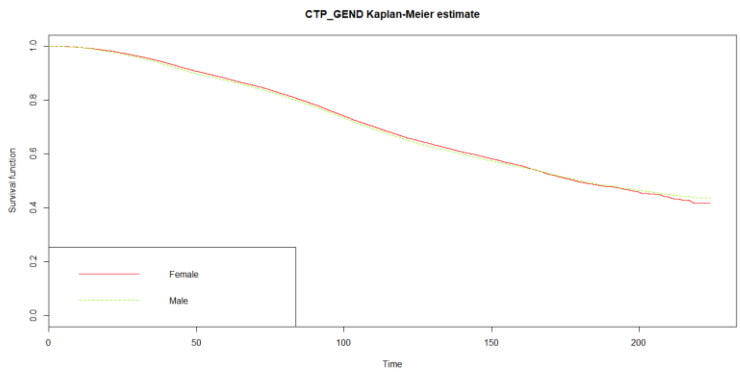

```
Call:
survreg(formula = mySurv_fix ~ CTP_AGEM_AGGR, data = working_fix,
    weights = WEIGHTS, dist = "loglogistic")
                    Value Std. Error       z        p
(Intercept)        5.2734    0.01798  293.32 0.00e+00
CTP_AGEM_AGGR<=35 -0.1521    0.01885   -8.07 7.04e-16
CTP_AGEM_AGGR<=45 -0.0674    0.01896   -3.55 3.78e-04
CTP_AGEM_AGGR<=55 -0.0554    0.01986   -2.79 5.29e-03
CTP_AGEM_AGGR<=60 -0.1074    0.02388   -4.50 6.89e-06
CTP_AGEM_AGGR<=65 -0.1198    0.02735   -4.38 1.18e-05
CTP_AGEM_AGGR>=66 -0.0318    0.03139   -1.01 3.11e-01
Log(scale)        -0.5877    0.00441 -133.35 0.00e+00

Scale= 0.556

Log logistic distribution
Loglik(model)= -236806.5   Loglik(intercept only)= -236888.3
        Chisq= 163.73 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 132285
```

Finally, classes <=45 and <=55 show the same path so they can be put togheter. The same is true for classes <=60 and <=65. Therefore the final classification of variable CTP_AGEM will be the formed by 5 classes: <=25 years (baseline), 26-35 years, 36-55 years, 56-60 years, over 60 years.



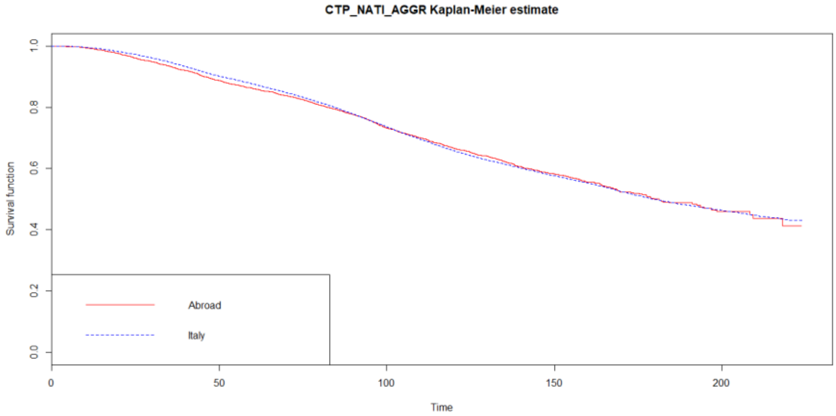

```
Call:
survreg(formula = mySurv_fix ~ CTP_AGEM_AGGR, data = working_fix,
    weights = WEIGHTS, dist = "loglogistic")
                    Value Std. Error       z        p
(Intercept)        5.2735    0.01798  293.29 0.00e+00
CTP_AGEM_AGGR<=35 -0.1521    0.01885   -8.07 7.06e-16
CTP_AGEM_AGGR<=55 -0.0706    0.01876   -3.77 1.66e-04
CTP_AGEM_AGGR<=60 -0.1074    0.02388   -4.50 6.91e-06
CTP_AGEM_AGGR>=60 -0.0508    0.02016   -2.52 1.17e-02
Log(scale)        -0.5875    0.00441 -133.32 0.00e+00

Scale= 0.556

Log logistic distribution
Loglik(model)= -236809.8   Loglik(intercept only)= -236888.3
        Chisq= 157.02 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 132285
```

CTP_GEND: the variable gender is not significant in describing the prepayment of fixed rate mortgages. This is proved by the similarity of the survival curves, which is more statistically certified by the significance of the log-rank test



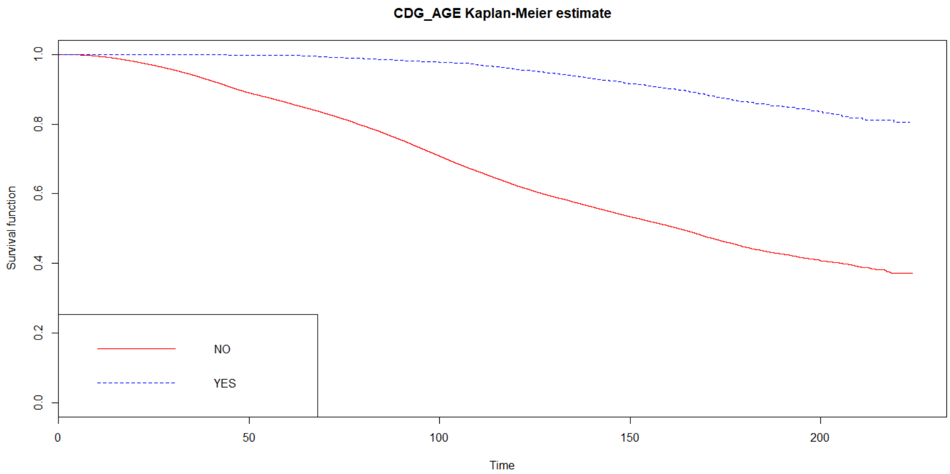

```
Call:
survdiff(formula = mySurv_fix ~ CTP_GEND)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
CTP_GEND=F 42924    14821    14695     1.076      1.58
CTP_GEND=M 89361    31070    31196     0.507      1.58

 Chisq= 1.6  on 1 degrees of freedom, p= 0.208
```

CTP_NATI: only 10% of the sample belongs to mortgagors with foreign residence, and the survival shape of the two classes looks the same. Nevertheless the log-rank test suggests to include the covariate in the model

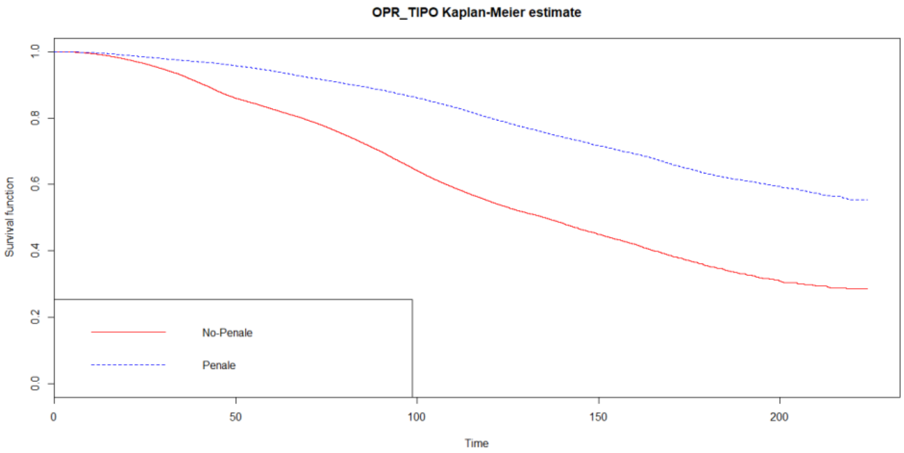



```
Call:
survdiff(formula = mySurv_fix ~ CTP_NATI_AGGR)

                          N Observed Expected (O-E)^2/E (O-E)^2/V
CTP_NATI_AGGR=Estero  10303     2698     2253     87.72      92.3
CTP_NATI_AGGR=Italia 121982    43193    43638      4.53      92.3

 Chisq= 92.3  on 1 degrees of freedom, p= 0
```

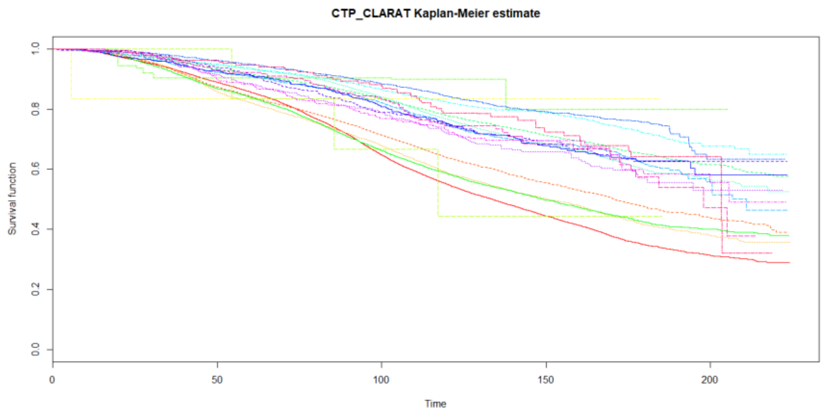
CDG_AGE: should be included, having or not facilitation seems to have high influence on the phenomenon

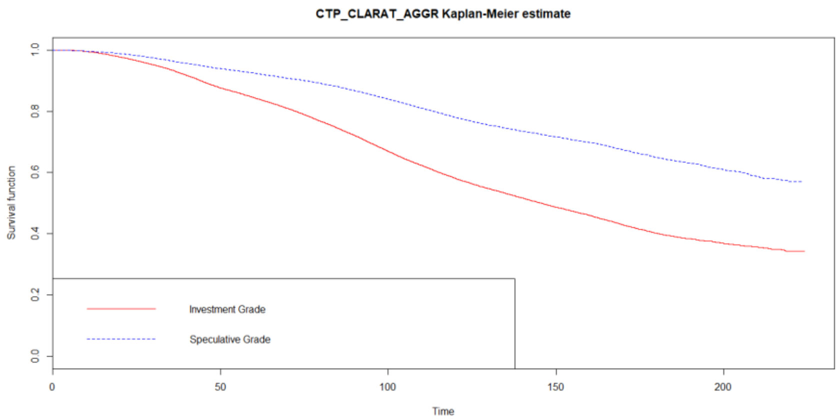


OPR_TIPO: should be included, having or not penalties seems to have high influence on prepayment

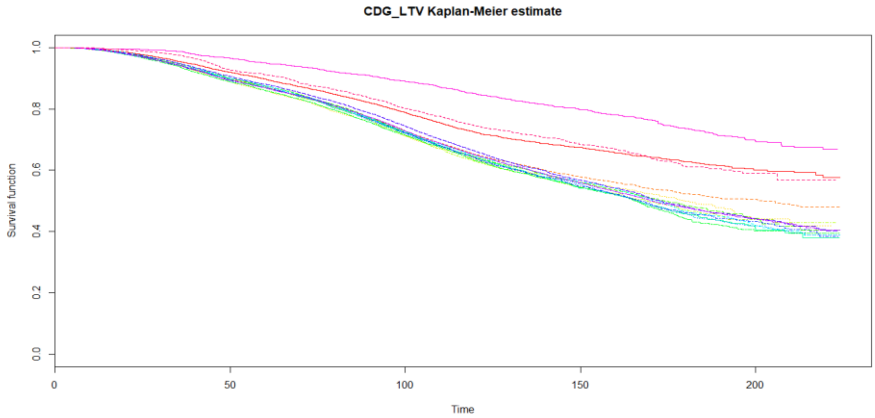


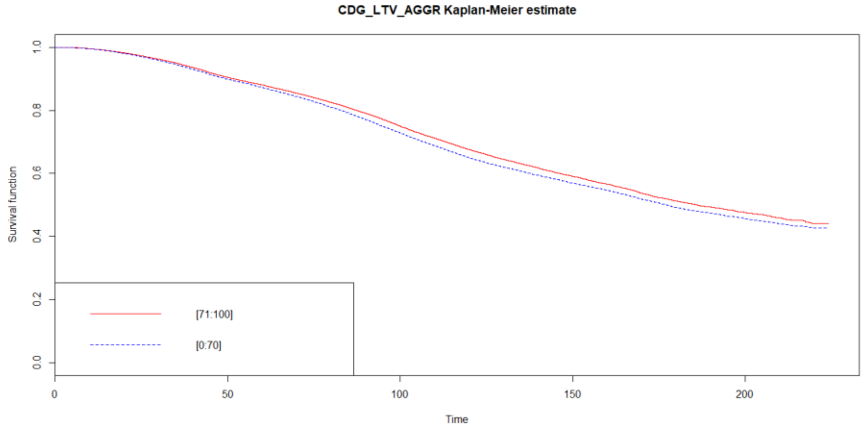
CTP_CLARAT: in aggregating this variable it was followed a rational criteria. The original 18 rating classes besides default were aggregated into 2 classes, according to their survival path, for an easier interpretation: higher rating classes, from AAA until B1 (scale grade Moody's) merged into "Investment grade" class, while lower rating composed the "Speculative Grade" class





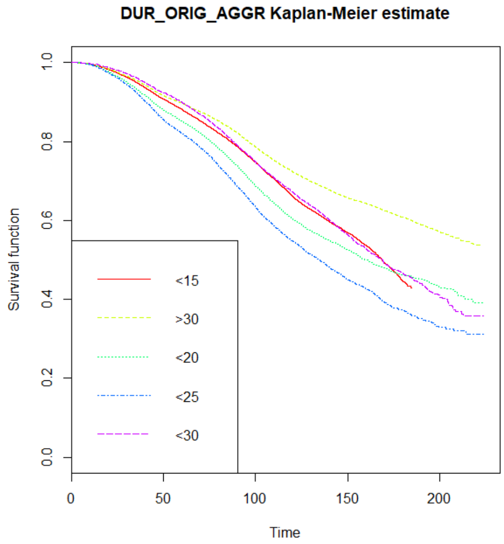
CDG_LTV: The sample contains 15 LTV classes from 0 to 100%, which where aggregated in two classes according to their survival path. The final classification of the variable is LTV<70% and LTV>70%. The log-rank test tells that the two survival curves are significantly different






```
Call:
survdiff(formula = mySurv_fix ~ CDG_LTV_AGGR)

                           N Observed Expected (O-E)^2/E (O-E)^2/V
CDG_LTV_AGGR=[71:100] 46615    15477    16042      19.9      30.7
CDG_LTV_AGGR=[0:70]   85670    30414    29849      10.7      30.7

 Chisq= 30.7  on 1 degrees of freedom, p= 3.09e-08
>
```

DUR_ORIG: the original life time of the mortgage has high influence on prepayment decision, as demonstrated by the very different survival curves. The aggregation of the variable is made reflecting the main mortgage life, with the aim of reach more interpretable results, so the new values are <=15 years, <=20 years, <=25 years, <=30 years and over 30 years.



Macroeconomic data like GDP, Unemployment rate, House Price Index and also financial variables such as CDS rate and the constructed Refinancing Incentive were not classified but treated as numeric variables. Their significance will be checked together with the other covariates in the final model. However, some consideration must be made. In particular, those variables are time dependent and this can be a problem that can bias the final results if not properly managed. A variable is defined as time dependent if it takes for the same subject two different values depending to different periods of time. The correct solution would be to repeat the observations each time the variable changes, to be able to follow all the time path. However, this requires very powerful computational means because the dataset extended from 132 thousands to more than 3 million of data. Therefore the model results are exposed to this source of bias.

3.3.4 Find the maximum likelihood model

Until now it was found that the log-logistic AFT model seems to be the wright parametric solution to model fixed rate retail residential (on real property) mortgages in the full scenario. Then it was found the subset of variables which look like possible explaining variables for the mortgage prepayment time.

Next step is to add one by one the possible covariates to the baseline scenario and verify which combination gives the highest log-likelihood. Moreover, the hypothesis of equality of the two models is checked using log-likelihood ratio test, which follows a Chi-Squared distributions. The tables below show the composition of the two models, their likelihood and the chi-squared value of the equality test every time a new covariate is added

Step	Model1	Model2
1	CTP_AGEM_AGGR	+CDG_LTV_AGGR
2	CTP_AGEM_AGGR+CDG_LTV_AGGR	+CTP_CLARAT_AGGR
3	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR	+CTP_NATI_AGGR
4	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR	+OPR_TIPO
5	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO	+DUR_ORIG_AGGR
6	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO+DUR_ORIG_AGGR	+GDP
7	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO+DUR_ORIG_AGGR+GDP	+Unempl
8	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO+DUR_ORIG_AGGR+GDP+Unempl	+HPI
9	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO+DUR_ORIG_AGGR+GDP+Unempl+HPI	+Refinancing_Incentive
10	CTP_AGEM_AGGR+CDG_LTV_AGGR+CTP_CLARAT_AGGR+OPR_TIPO+DUR_ORIG_AGGR+GDP+Unempl+HPI+Refinancing_Incentive	+CDS

Step	LL1	LL2	CHI2	pvalue	H0: Model1=Model2
1	- 236,811	- 236,776	69	0.000000	reject
2	- 236,776	- 234,561	4,431	-	reject
3	- 234,561	- 234,560	2	0.130942	can't reject
4	- 234,561	- 231,902	5,319	-	reject
5	- 231,902	- 231,482	840	0.000000	reject
6	- 231,482	- 221,873	19,217	-	reject
7	- 221,873	- 220,746	2,255	-	reject
8	- 220,746	- 211,899	17,694	-	Reject
9	- 211,899	- 201,213	21,373	-	Reject
10	- 201,213	- 189,880	22,666	-	Reject

The log-likelihood ratio test suggests that the variable CTP_NATI is not statistically significant for explaining the prepayment event. Therefore it should not be considered in the final model. All the other variables, instead, are statistically significant and should be added to the final model as drivers of prepayment.

3.3.5 The final model

The final model is illustrated in the table below. Notice that all the coefficients are significant at 95% confidence level and also the overall p-value is significant.

The log-logistic AFT survival and hazard mathematical formulas are explained in the appendix.

```
Call:
survreg(formula = mySurv_fix ~ Refinancing_Incentive + as.factor(CTP_AGEM_AGGR) +
    as.factor(DUR_ORIG_AGGR) + as.factor(CDG_LTV_AGGR) + as.factor(CTP_CLARAT_AGGR) +
    as.factor(OPR_TIPO) + GDP + Unempl + HPI + MPS, data = working_fix,
    weights = WEIGHTS, dist = "loglogistic")
                                             Value Std. Error       z        p
(Intercept)                                5.81340   4.12e-02  140.95 0.00e+00
Refinancing_Incentive                     -0.29576   1.54e-03 -192.43 0.00e+00
as.factor(CTP_AGEM_AGGR)26-35             -0.02447   1.02e-02   -2.40 1.63e-02
as.factor(CTP_AGEM_AGGR)36-55             -0.04146   1.01e-02   -4.12 3.84e-05
as.factor(CTP_AGEM_AGGR)56-60             -0.09379   1.29e-02   -7.24 4.39e-13
as.factor(CTP_AGEM_AGGR)>60               -0.10147   1.31e-02   -7.72 1.15e-14
as.factor(DUR_ORIG_AGGR)[15Y:20Y]         -0.08648   5.26e-03  -16.44 9.64e-61
as.factor(DUR_ORIG_AGGR)[20Y:25Y]         -0.08762   6.41e-03  -13.68 1.39e-42
as.factor(DUR_ORIG_AGGR)[25Y:30Y]         -0.07056   6.27e-03  -11.25 2.30e-29
as.factor(DUR_ORIG_AGGR)>30Y               0.06943   7.14e-03    9.72 2.46e-22
as.factor(CDG_LTV_AGGR)[0:70]             -0.04083   4.28e-03   -9.54 1.44e-21
as.factor(CTP_CLARAT_AGGR)Speculative Grade  0.06429   4.35e-03   14.77 2.14e-49
as.factor(OPR_TIPO)Penale                  0.89295   4.89e-03  182.48 0.00e+00
GDP                                       -0.37735   3.64e-03 -103.78 0.00e+00
Unempl                                    -0.22610   4.42e-03  -51.21 0.00e+00
HPI                                       -0.11280   1.03e-03 -109.21 0.00e+00
MPS                                       -0.00361   2.63e-05 -137.47 0.00e+00
Log(scale)                                -1.40466   4.37e-03 -321.29 0.00e+00

Scale= 0.245

Log logistic distribution
Loglik(model)= -189879.6   Loglik(intercept only)= -236888.3
        Chisq= 94017.45 on 16 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 132285
```

In conclusion, the fitted survival function on the i^{th}individual is

$$\hat{S}_i(t) = \left\{1 + t^{\frac{1}{0.245}} e^{(\hat{\eta}_i)}\right\}^{-1}$$

where $\hat{\eta}_i$ is the acceleration factor given by $\frac{-\mu-\hat{b}_i x_i}{\hat{\sigma}}$. In detail:

μ is the intercept of the model

$\hat{\sigma}$ is the reciprocal of the scale factor

$\widehat{b_i}$ is the coefficient vector of individual i

x_i is the covariate vector of individual i

$$\hat{\eta}_i = \frac{1}{0.245}\{-5.81 + 0.29x_1 + 0.02x_2 + 0.04x_3 + 0.09x_4 + 0.1x_5 + 0.09x_6 + 0.09x_7 + 0.07\,x_8 - 0.07\,x_9 + 0.04\,x_{10} - 0.06\,x_{11} - 0.89x_{12} + 0.38\,x_{13} + 0.23\,x_{14} + 0.11x_{15} + 0.003x_{16}\}$$

where:

$x_1 = RI$

$x_2 = CTPAGEM_{26-25}$

$x_3 = CTPAGEM_{36-55}$

$x_4 = CTPAGEM_{56-60}$

$x_5 = CTPAGEM_{>60}$

$x_6 = DURORIGAGGR_{15-20}$

$x_7 = DURORIGAGGR_{20-25}$

$x_8 = DURORIGAGGR_{25-30}$

$x_9 = DURORIGAGGR_{>30}$

$x_{10} = CDGLTVAGGR_{<70}$

$x_{11} = CTPCLARATAGGR_{Speculative\ Grade}$

$x_{12} = OPRTIPO_{Penalty}$

$x_{13} = GDP$

$x_{14} = UNEMPL$

$x_{15} = HPI$

$x_{16} = CDS$

And the corresponding hazard function on the i^{th} individual is

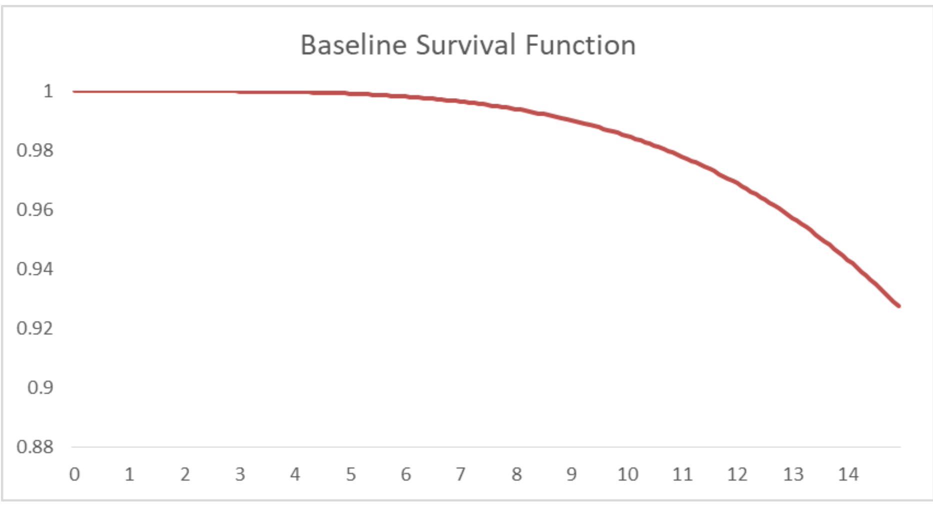
$$\hat{h}_i(t) = \frac{1}{0.245t}\left\{1 + t^{-\frac{1}{0.245}}e^{(-\hat{\eta}_i)}\right\}^{-1}$$

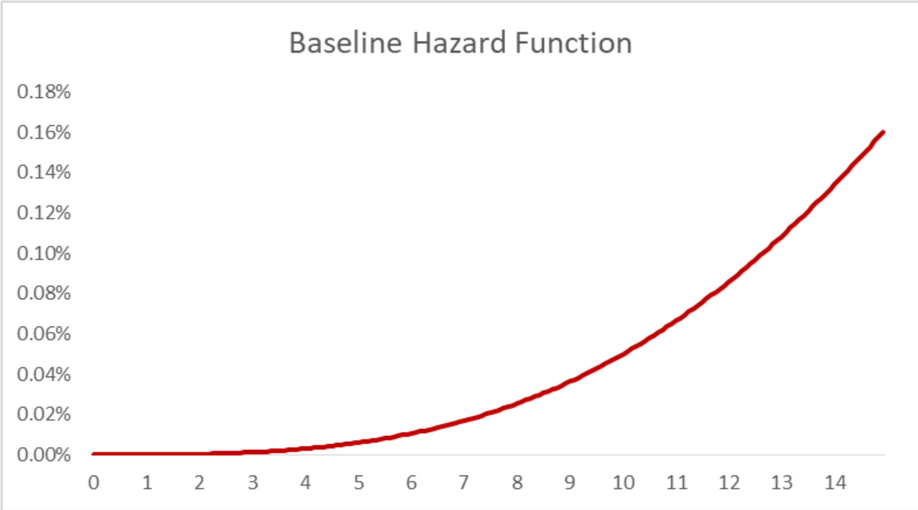
3.3.6 Interpreting the model results

The intercept of the model tells what happens in the baseline scenario, i.e. in the first possible scenario. In the final model the baseline scenario was determined by the first determination of each covariate. Some other choices are possible, like for example assuming as baseline the most frequent scenario in the dataset. The chosen baseline scenario is the following:

- Mortgagor is less than 25 years old
- Mortgage original life time is less than 15 years
- Mortgage Loan To Value is over 70%
- Mortgagor has a credit rating level classified as Investment Grade (higher level)
- The mortgage has no prepayment penalties
- Market interest rates, GDP, Unemployment rate, House Price Index and the credit rating of the lending bank are assumed constant

The reciprocal of the scale parameter of the model determines the S shape of the survival and hazard function (see appendix). In the model this value is 4.07, so that the shape of the baseline survival and hazard function is the following:

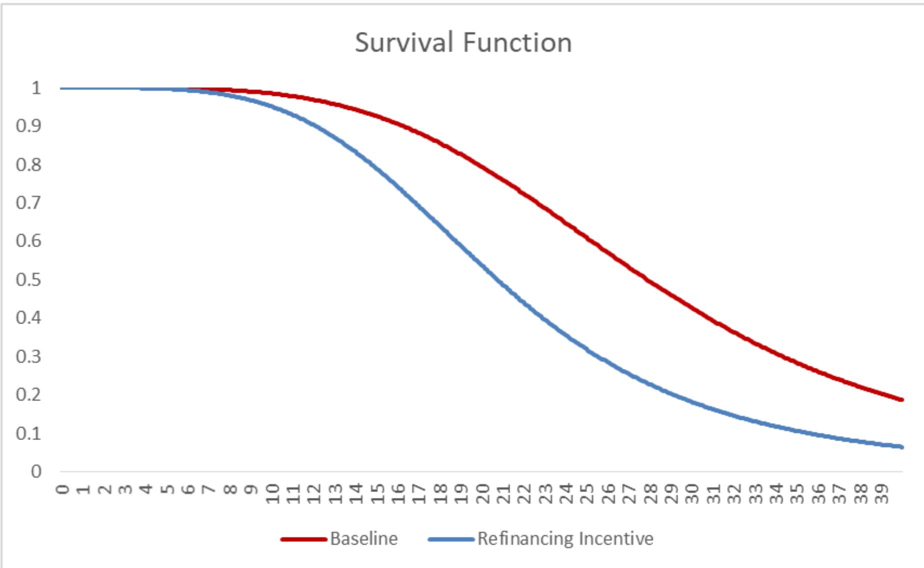




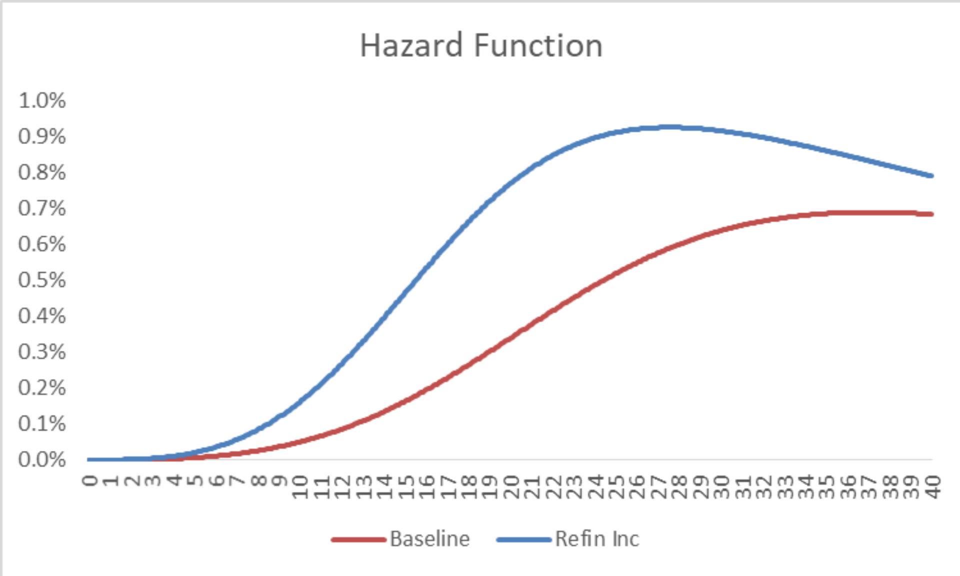
Survival function should be interpret in this way: considering the baseline scenario, the probability to prepay of a 15 years mortgage that has a LTV over 70%, when the householder is less than 25 years old and has a high rating, assuming constant market and macroeconomic variables is 1% in 5 years and 3% in 10 years, while the probability to prepay in the last year is 6%.

Of course, this is just a theoretical exercise because assuming constant macroeconomic and financial variables makes no sense. It is necessary to change the covariates value used in the baseline scenario to reach more significant results.

What happens to the survival probability when covariates change status from their first value depends on the sign of their coefficients. The effect of adding a covariate with negative coefficient, for example the Refinancing Incentive, is to accelerate the speed of the survival function, thus anticipating the failure time. Graphically this means:



The effect is reached because the acceleration factor emphasizes the S shape of the hazard function, meaning that the instantaneous failure risk is higher.



Notice that this is coherent with what we expected. The refinancing incentive was defined as the spread between the original market interest rates when the mortgage was originated, computed on the original mortgage duration and the current interest rate level on the actual mortgage duration. When the interest rates go down, mortgagor receives an incentive to prepay that is higher when the spread level increases. The computation of this effect is made looking at the time ratio

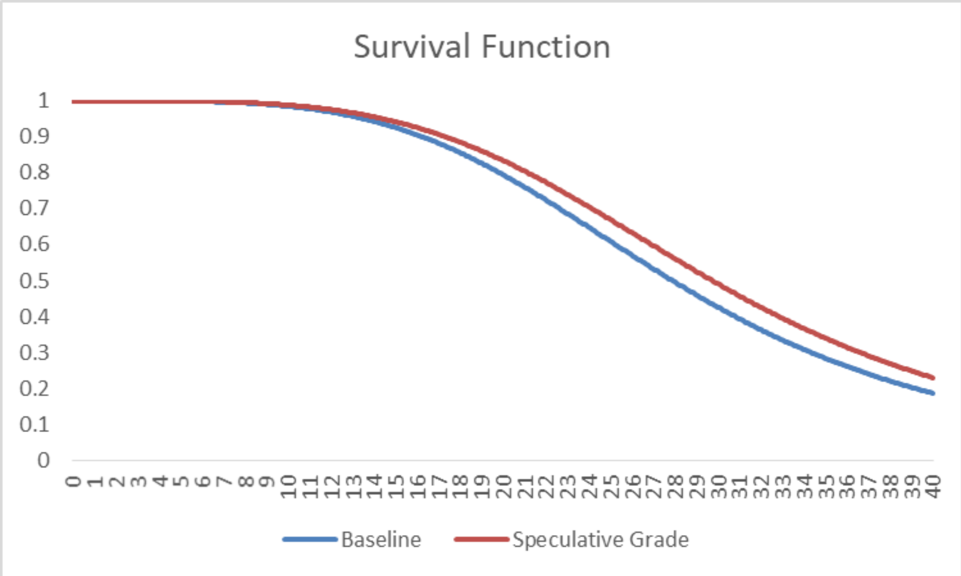
$$exp(b_i) = \exp(-0.29) = 0.74$$

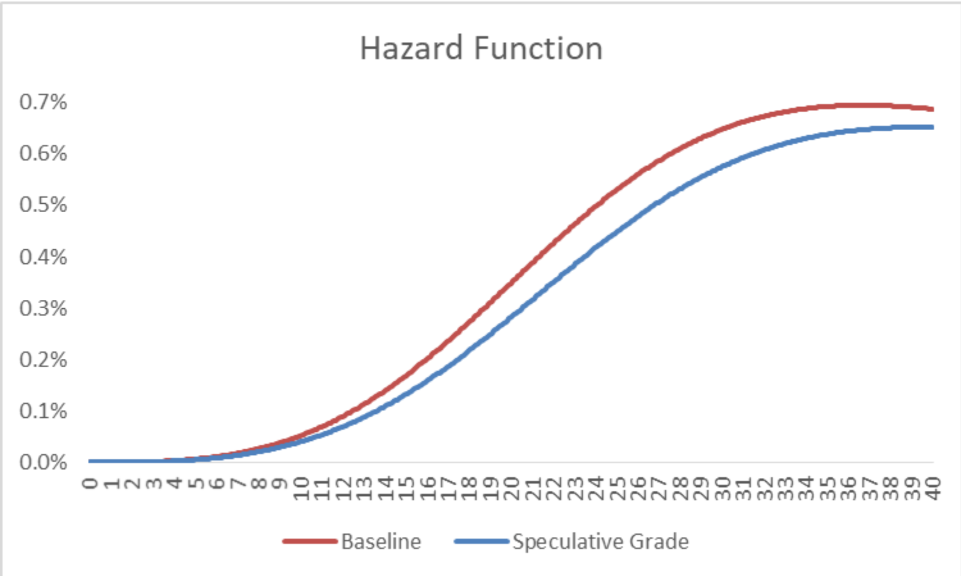
This means that the baseline survival time decreases from 100% to 74% or, more easily, that the Refinancing Incentive accelerates the probability of default by 26%. This is the effect assuming a Refinancing Incentive value of 1. The effect is proportional to the covariate value, so that if the Refinancing Incentive is 2 the acceleration factor is $\exp(-0.29 * 2) = 0.55$ which means that the survival probability is reduced by 45%.

A similar interpretation can be given to positive coefficients. Looking for example at the CTP_CLARAT variable, the model tells that if the mortgagor has rating value Speculative Grade, the associated probability of early termination decelerates. This effect can be quantified as

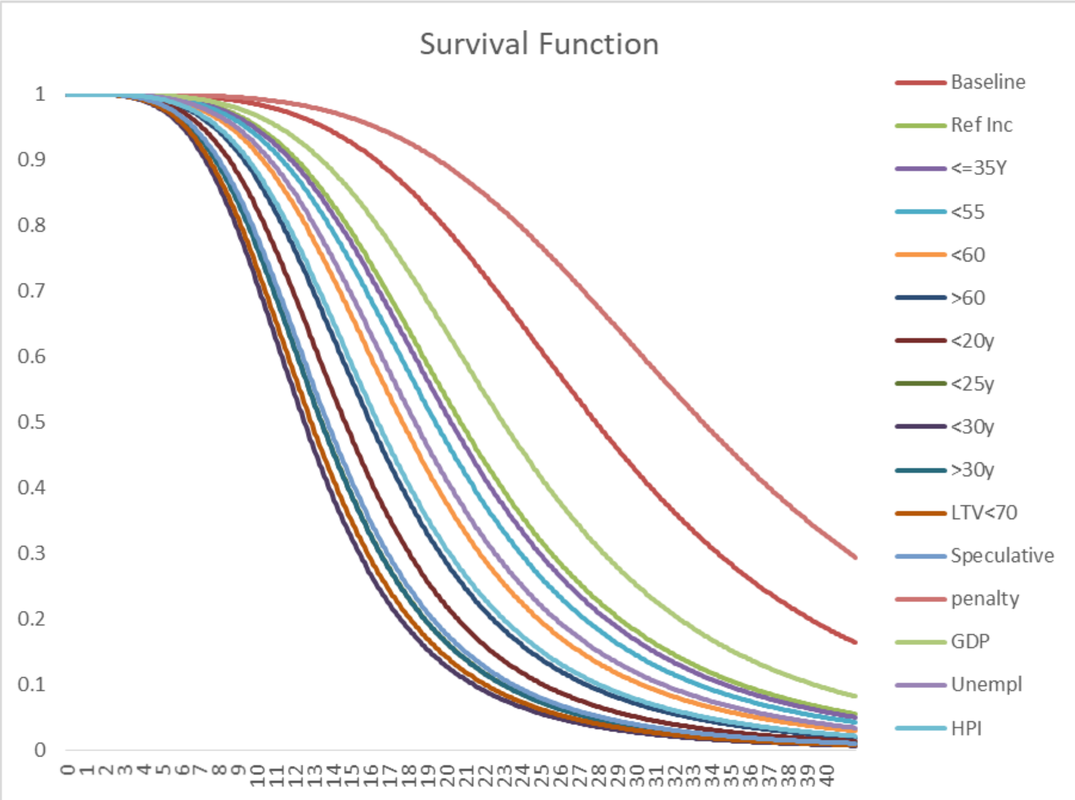
$$exp(b_i) = \exp(0.06) = 1.13$$

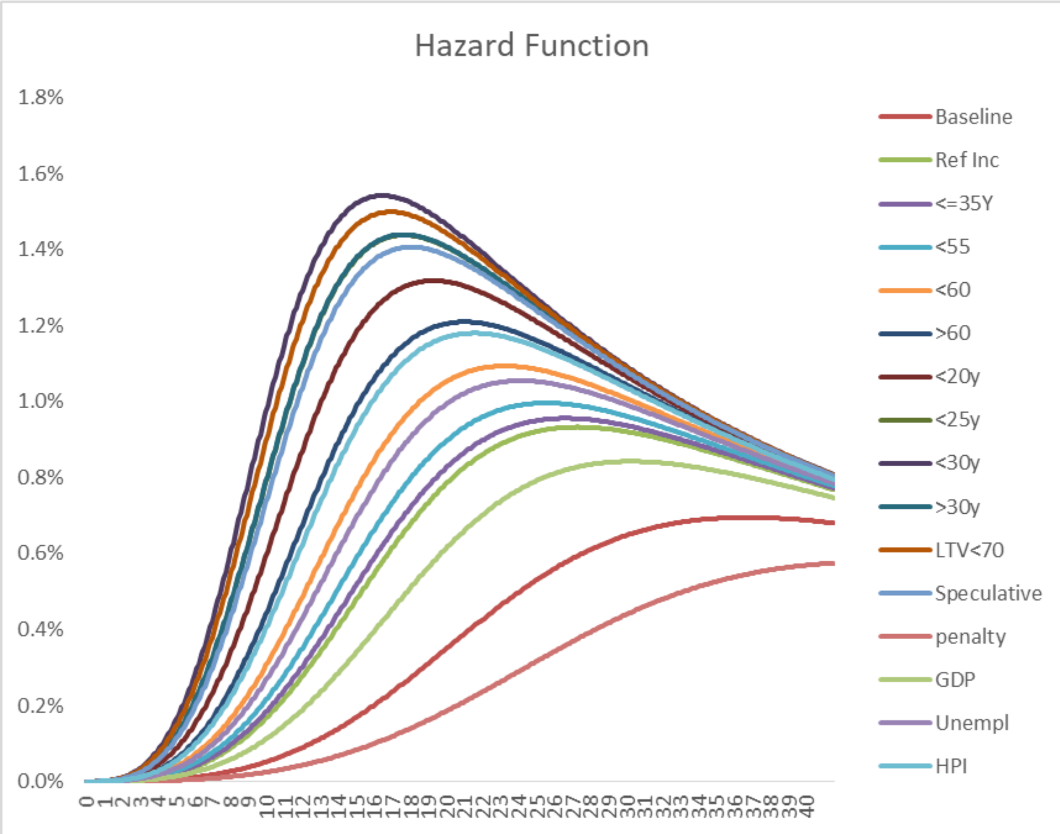
which means that the survival probability increases by 13%. This is obtain through a linearization of the hazard ratio, which moves close to the x-axis as can be seen graphically





What happens when adding up all the variables is that the effect of each new covariate effect sums up to the previous ones. Of course, the qualitative variables are dummy variables that assume the value 1 in the presence of the specific characteristics, zero otherwise. Since the effect is additional, the covariate order inside the model should be taken into consideration when trying to interpret the results because the effect of each explanatory variable is not separated by the preceding one, but should be combined with them. The new explanatory variable can mitigate or increase the acceleration effect of the preceding covariates depending on the sign of its coefficient. Consequently a range of hazard and survival functions are obtained depending on the combination of covariates.





It is now possible to interpret the model results analysing the coefficient signs and magnitude:

- Refinancing Incentive: negative coefficient. Baseline scenario assumes constant market interest rates level, but this is very unlikely to happen. As financial theory says, the higher the spread between the current and previous level of market interest rates, the greater is the incentive to prepay to refinance the mortgage at lower rates
- CTP_AGEM_AGGR: in the baseline scenario it was assumed a less than 25 years mortgagor. Usually his salary is very low and the need to move in a new house as marital status changes or family increase is still low, comparing to later ages, therefore his probability to prepay is lower. That's why all the other age class add prepayment risk to the baseline scenario. The magnitude of prepayment risk seems to increase with age, since 60 years old householders have the highest negative coefficient
- DUR_ORIG_AGGR: the baseline scenario assumes a 15 years mortgage. Extending the mortgage maturity seems to increase the prepayment effect, since all the classes add a negative coefficient to the model. Class 20-25 years seems

to prepay the most, looking at the magnitude of its coefficient. As in chapter 1, prepayment path is S shaped, so that mortgages with very high or very low duration are less likely to be prepaid than those in between. This is confirmed by the positive sign of the class over 30 years

- CDG_LTV_AGGR: negative coefficient. The baseline scenario assumes a Loan to Value ratio higher than 70%. The coefficient sign is coherent with the reason explained in chapter 1: the higher is the disbursed amount the less likely is for the borrower to have the economic means to reimburse the entire notional before maturity and vice versa
- CTP_CLARAT_AGGR: acts positively on prepayment, decelerating the time speed, comparing to the baseline Investment Grade scenario. This is expected result: low credit rating people have less possibility to raise funds to prepay or to get a new mortgage from other banks. Therefore prepayment probability of Speculative Grade is less than Investment Grade
- OPR_TIPO: positive coefficient, with high magnitude. The baseline scenario assumes no penalty. A very high positive coefficient means that, as historically verified, having to pay a prepayment penalty fee is a deterrent for the prepayment choice
- GDP rate: negative coefficient, with high magnitude. Baseline scenario assumes constant GDP. When GDP increases, people receive an incentive to prepay. This can be explained by the fact that when the economy raises, it is likely for mortgagors to get higher wages and a better economic situation may encourage family enlargement and household mobility.
- Unemployment rate: negative coefficient. This is coherent with what explained in chapter 1, i.e. the Unemployment rate is positively correlated to prepayment for moving reasons.
- House Price Index rate: negative coefficient. The higher is the index, the lower is the probability for the mortgage to reach its natural maturity. As described in chapter 1, when the house price appreciates, the number of prepayment events due to relocation increases because the householders can get an higher return from selling their houses or may expect an higher return from buying other ones. For example when HPI raises by 1% the new survival probability of the

mortgage is the 89% of the previous one, which means that the prepayment probability increases by 11%

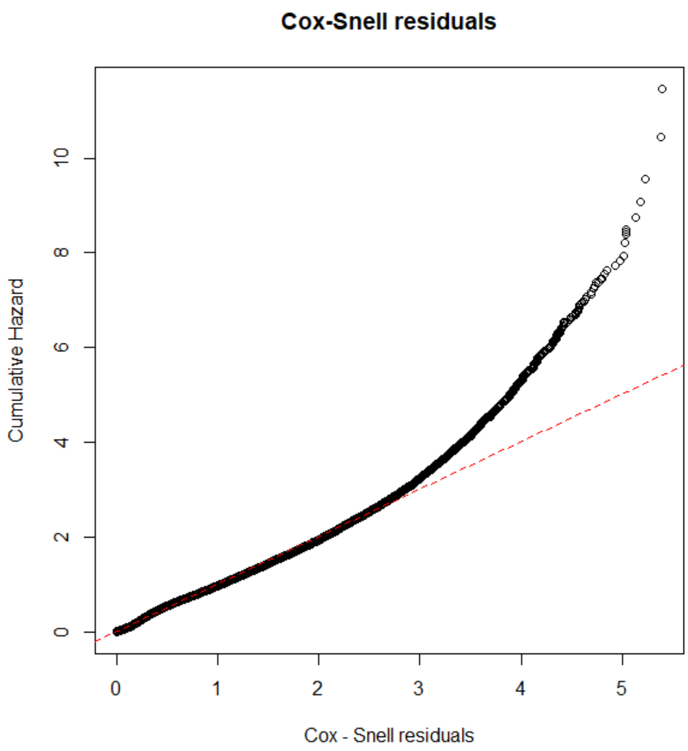
- CDS rate: negative coefficient. As mentioned before, when the credit rating of the bank get worse, customers are more likely to leave the bank and move all the commercial products to another bank.

3.3.7 Goodness of fit tests

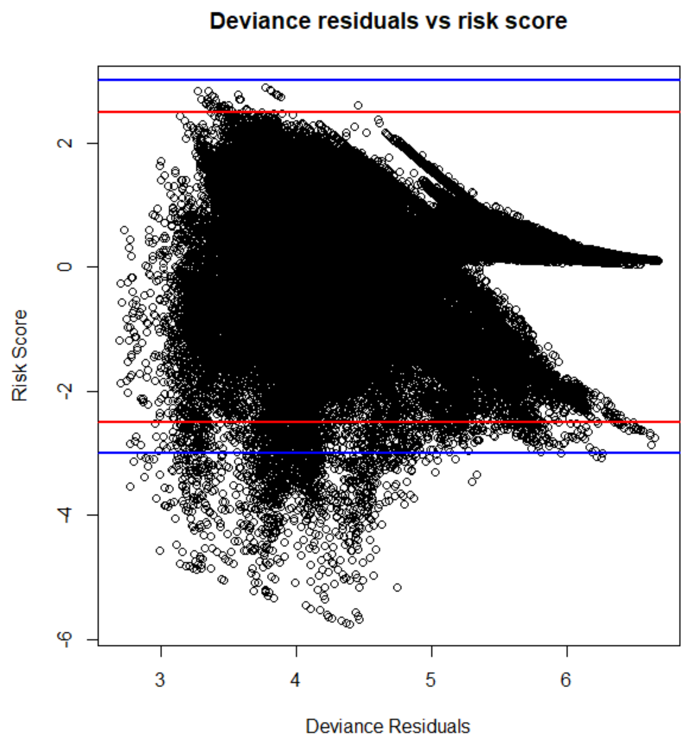
The last thing to do after finding the final model is checking the goodness of fit of the log-logistic model. This can be done with AIC test or with the residuals check. It was compared another time the AIC and Log-Likelihood value of the log-logistic model against the other possible parametric models, applying in each model the same covariates. Results are showed in the table below. The log-logistic model is confirmed to be the best: it preserves the lowest AIC and the highest log-likelihood even when moving away from the baseline scenario.

Parametric Model	AIC	Log-Likelihood
Log-Logistic	379,795	- 189,880
Weibull	382,523	- 191,243
Exponential	424,162	- 212,064
Gaussian	383,704	- 191,834
Logistic	384,695	- 192,330
Lognormal	383,280	- 191,622

However, the most powerful tool to check the goodness of model fitting is the residual analysis. As illustrated in Chapter 2, residuals in the survival analysis can be checked through the Cox-Snell residuals. If the model is appropriate, the plot of the cumulative hazard of the Cox-Snell residuals over the residuals values will give a straight line with unit slope and zero intercept.



Cox-Snell residuals show that the model fits well the data at the beginning, but the fit on the extreme data is worse. Therefore, it might be useful to check the outliers using the deviance residuals.



```
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-5.746663  0.006488  0.226417  0.078187  0.518322  2.887688
> var(resi.d)
[1] 0.7274965
```

Deviance residuals seems to be mostly distributed around 0, with no path. Departure from normality is due to the high level of censored observations in the sample. However, the majority of residuals are between the range (-3, 3). This means that the model does not have many outliers: 1.142 observations over more than 132.000 mortgages, which means the 1% of the sample.

In conclusion, the found model seems to be reasonable. Of course, the residuals show that the survival model is not able to capture all the elements that may cause prepayment. In many cases the decision to prepay is not due to financial or personal characteristics but to external events, or extraordinary events, like divorce or changes in the number of family members (new births, deaths, illnesses). Moreover, monitoring the mortgagors working sphere would be important, because salary increases, job movements or layoffs can encourage household mobility. All those effects are very difficult to be known by the bank and therefore can’t be caught by the survival model.

An important aspect to be mentioned regards the economic and financial variables. Those variable have a time-dependent nature because may assume different values when time changes. However, it was chosen not to consider this characteristic in the model. This choice was made for computational reasons, because to include a time-dependent variable it is necessary to know its value for every mortgage in every reference date, which means modifying the original dataset passing from 132.000 to more than 3 million of data. However not considering the time-dependent nature of those variables may bias the model results.

Chapter 4. Managing Prepayment Risk

In the previous chapter it was developed a survival model to predict the prepayment probability of a mortgage depending on the mortgage characteristics, the householder characteristics and the macroeconomic and financial context.

Far away from being a statistical exercise, the model results have several important consequences on the overall business that may convince financial institutes that it is worthwhile to implement it on their data. In this chapter they will be explained using a case study.

4.1 Case study

To discover what happens when using a survival model for modelling mortgage prepayment probability, the following fixed rate mortgages can be considered (assuming stable macroeconomic scenario and same starting date):

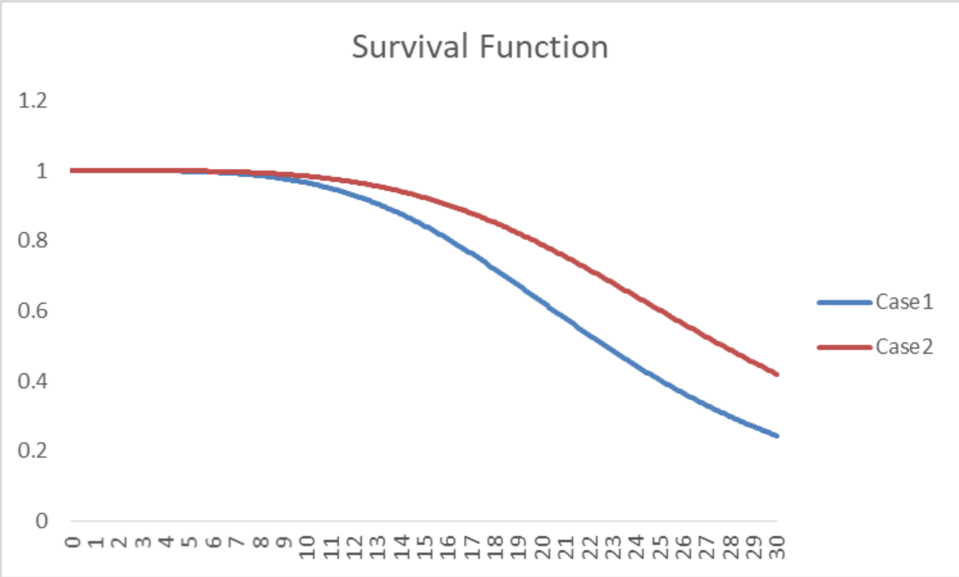
Mortgage 1:

- Mortgagor is between 56 and 60 years old
- mortgage maturity 30Y
- LTV less than 70%
- No penalties
- Investment Grade

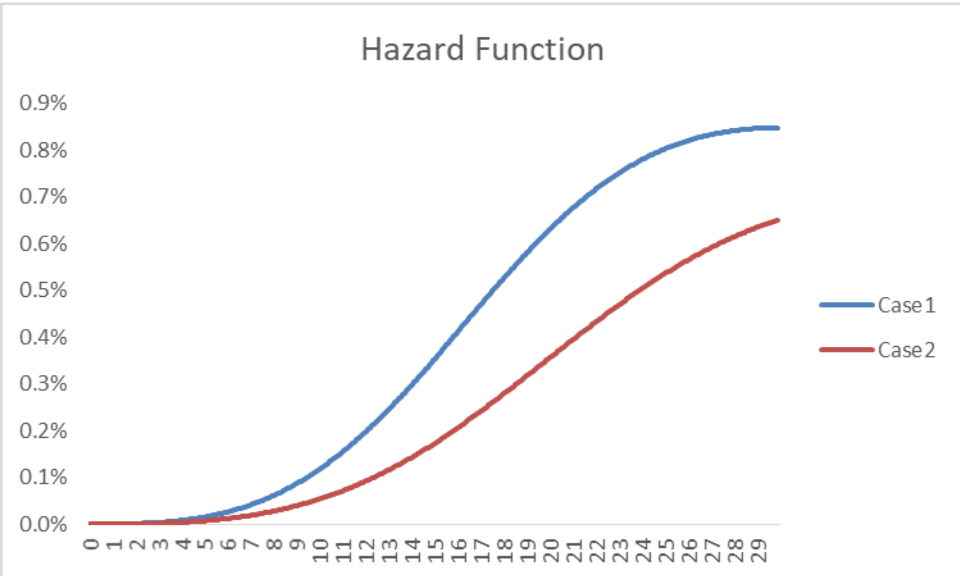
Mortgage 2:

- Mortgagor is less than 25 years old
- mortgage maturity 30Y
- LTV is over 70%
- No penalties
- Speculative Grade

The two Survival Functions are

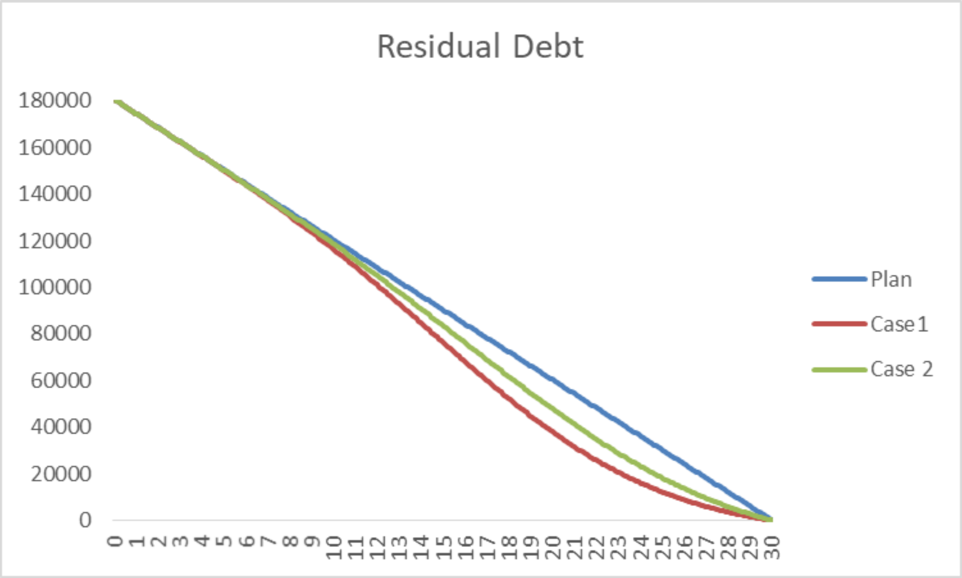


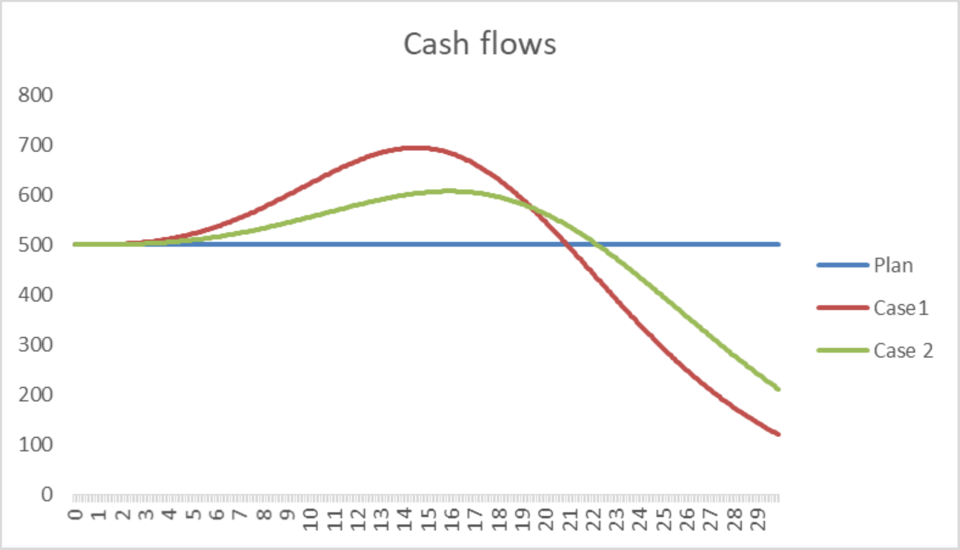
Mortgage 2 seems to last more than Mortgage 1, for example after 15 years the survival probability of Mortgage 2 is 92% while the same probability of mortgage 1 is 83%. The same results is confirmed by the higher hazard rate of Mortgage 1: the instantaneous probability to prepay when the mortgage reaches 15 years is 0.4% in case 1 while 0.2% in case 2.



It is also possible to compare their cash flow scheduling. Of course the cash flow path will be influenced by prepayment probability. In particular in case the mortgage doesn't prepay, it is expected to receive the coupon at each scheduled payment date, while if the mortgage prepays, it is expected to receive the residual debt. Therefore, the combination of the two probabilities gives the expected cash flow at each payment time: fixed coupon times the probability of not prepaying, plus the residual debt times the probability of prepaying.

Comparing the resulting cash flows in both scenarios the following plots are obtained:





The plots demonstrate that the scheduled cash flows are modified by prepayment, with an effect that is greater the higher is the prepayment speed, so Model 1 pays the expected cash flows much before the contractual time. This means that the duration of the original mortgage reduces, because the cash flows are paid before. In the case study the original mortgage maturity was 30 years, which means around 15 years duration adopting the French amortization. Assuming prepayment, mortgage duration reduces to 13 years for Model 1 and to and 14 years for Model 2.

4.2 ALM and Risk management results

One first implication of the model regards the ALM and Risk Management. On the liquidity risk point of view, if the real position has a lower duration it should also be funded on a lower tenor in the funding curve, which generally means paying a lower funding rate.

Moreover a better knowledge of the duration buckets of its mortgage portfolio will result in more reliable gap ratios and therefore in a more accurate management of liquidity risk, reducing the risk of overfunding.

Similar considerations can be made for interest rate risk management: if the bank were able to identify ex ante the actual duration of its mortgage portfolio, which in case of prepayment is lower than the contractual one, it would hedge on a lower maturity, which usually is much less expensive.

From the point of view of securitization risk, being able to understand the actual duration of the reference portfolio means being able to price it correctly. Assuming for example a portfolio composed on just the two case study mortgages: the overall duration would be 13.5 years instead of 15y. Therefore the bond will be priced accordingly to the yield (plus spread) of a treasury ZC bond with 13.5 years of maturity instead of 15 years, which is usually lower.

4.3 Balance Sheets results

It is worth to mention some very important results regarding balance sheets improvements. These are mainly connected to the new accounting standards IFRS9 (International Financial Reporting Standards), which introduce some changes in the calculation of the expected credit loss. The new Impairment model splits the credits into

3 stages depending on the credit deterioration: stage 1 for credits that have the lowest credit risk, stage 2 for those credits experimenting a significant risk increment, stage 3 when the credit have experienced an impairment event.

The problem for financial institutions is that IFRS9 require for credit in stage 2 and in stage 3 to compute the Expected Loss over a lifetime horizon rather than the previous 12 months, which is now left only to stage 1 credit. This means applying Probability of Default and the Loss Given Default over the overall lifetime horizon, therefore the resulting Expected Loss would be much bigger than before, especially when the duration of the credit portfolio is very high.

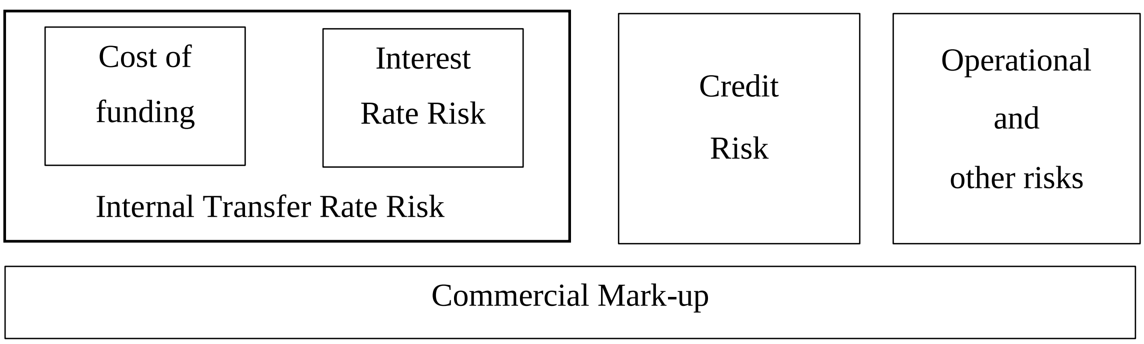
Mortgages gain importance in this scenario, for their natural very long time horizon. As a consequence, if on the one hand banks are reducing the maximum possible maturity of the mortgage, on the other hand it has become crucial to correct estimate the actual lifetime of the mortgage portfolio.

Prepayment can affect very deeply the lifetime of mortgages, therefore being able to capture the mortgagor behaviour would translate in higher competitive advantage with respect to competitors, in terms of less impairment, better balance sheets, and more money free for business purposes.

4.4 Commercial results

Another very important result coming from the application of the model could regard the commercial area. If the bank were able to know the real duration of the mortgage during the preliminary analysis, it would be able to significantly reduce the final price.

It is worth to recall the main building blocks of mortgage pricing:



The actual duration affects all the aspects:

- the bank needs to raise funds for a shorter maturity, reducing the funding spread,
- the interest rate risk is reduced because the effective interest rate to be applied on mortgage is not the one connected to the original duration but to the actual duration, which is usually lower (assuming a normal yield curve shape), therefore the hedging costs are reduced. Therefore the internal transfer rate to be applied on the transaction, which is made by internal interest rate and funding spread should be lower.
- credit risk lasts for a shorter period of time and, moreover, with respect to stage 2 and 3 mortgages, the impairment connected to the expected loss in the lifetime horizon is less,
- operational and other risks must be computed for a shorter period

Consequently, the bank may be able to reduce the price preserving the same commercial spread on the transaction. Moreover, the bank is able to define different pricing policies in relation to the different duration profile of the customers.

Both counterparts could benefit from this process: the customers could receive a more convenient price, while the bank could reach much more competitiveness on the price, which could be allowed by a reduced cost of risks of taking that position. The price competitiveness could convince new clients and reinforce the trust of the old ones.

Those considerations can be applied to the chosen case study, after simplifying the mortgage pricing process so to consider the remuneration of risks inside the mark-up level (market data as of June 2018[40]):

[40] Source: https://www.euribor.it/tassi-storici-eurirs

	Mortgage 1	Mortgage 2
Maturity	30y	30y
Original duration	15y	15y
Estimated duration	13y	14y
Original market interest rate	1.3%	1.3%
Mark-up applied	0.05%	0.05%
Original Price	1.35%	1.35%
New market interest rate	1.15%	1.22%
Mark-up buffer	0.15%	0.08%
New price	1.20%	1.27%

Before the application of the prepayment model the two mortgages would have had the same price of 1.35% given by the market interest rate computed on a duration of 15 years plus mark-up. After the application of the prepayment model the price is much lower and differentiated between the two mortgages: the market interest rate used in the pricing process is computed on the new duration which is lower and different between the two mortgages. The bank could apply the same mark up and give a lower price to the customer of 1.20% in the first case and 1.27% in the second case. Moreover, the financial institution would have a free mark-up buffer to use to increment the commercial spread, preserving the lower price until the increment covers all the buffer.

Conclusion

The prepayment problem is an old topic but it has gained importance in the last decade, especially in Italy, after the abrogation of the prepayment fee made by the Art. 7 of Law 40/2007 (called "Decreto Bersani"). In 2017 prepayment entered into the topics analysed by the ECB in the stress test exercise regarding Sensitivity Analysis on IRRBB and the BCBS suggests some important information to be considered when modelling prepayment risk. However, an agreed solution has not been found yet.

In this thesis the theoretical approach based on Survival Analysis was followed and a possible modelling solution tested on Montepaschi Group data. The aim was to compute the actual duration of the mortgage portfolio and illustrate all the benefits deriving from a correct knowledge of the expected cash flows. These benefits will cover many areas of the bank, especially commercial, ALM, risk management and accounting.

The found model tells that the expected duration of the mortgage portfolio is deeply affected by the mortgage characteristics, the macroeconomic and financial context and the mortgagor characteristics. The results reached on the sample seems to be reasonable.

However this thesis wants just to be a starting point in modelling prepayment risk. Many more tests (and backtests) and improvements must be made before being able to trust it. Suggested improvements of the model regard modifying the data to check the effect of time-dependent covariate, such as macroeconomic and financial variables, and to improve the dataset constantly updating the LTV class, which is now only the value at origination.

Appendix

A.1 Log-Logistic AFT model[41]

The log-logistic survival function is

$$S(t) = \frac{1}{1 + (t\rho)^{\alpha}}$$

where α =1/scale is the shape parameter and $\rho = -\exp(location\ parameter)$

The log-logistic hazard function is

$$h(t) = \frac{\alpha t^{\alpha-1}\rho^{\alpha}}{1 + (t\rho)^{\alpha}}$$

if α >1 the hazard has a single maximum and then decreases to zero; if α <1 the hazard is monotonically decreasing

the AFT log-logistic model in its linear version is:

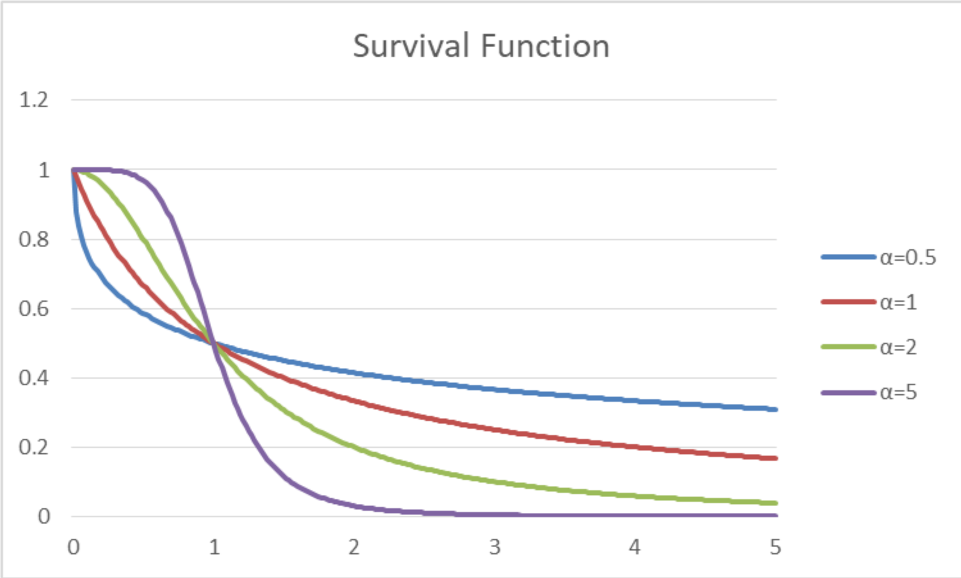
$$\log\left(\frac{S(t)}{1 - S(t)}\right) = -\alpha\beta x - \alpha log(t\rho)$$

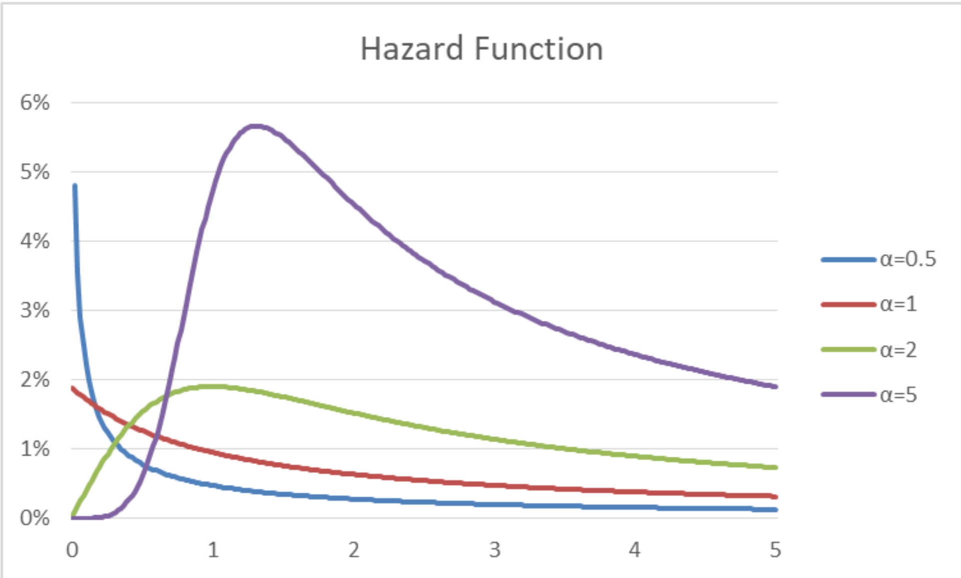
So that the accelerated form of the baseline survivor function is

$$S(t) = \frac{1}{1 + (t\rho e^{\beta x})^{\alpha}}$$

The following graphs show the possible shapes of the survival and hazard functions according to the different values of the scale parameter α

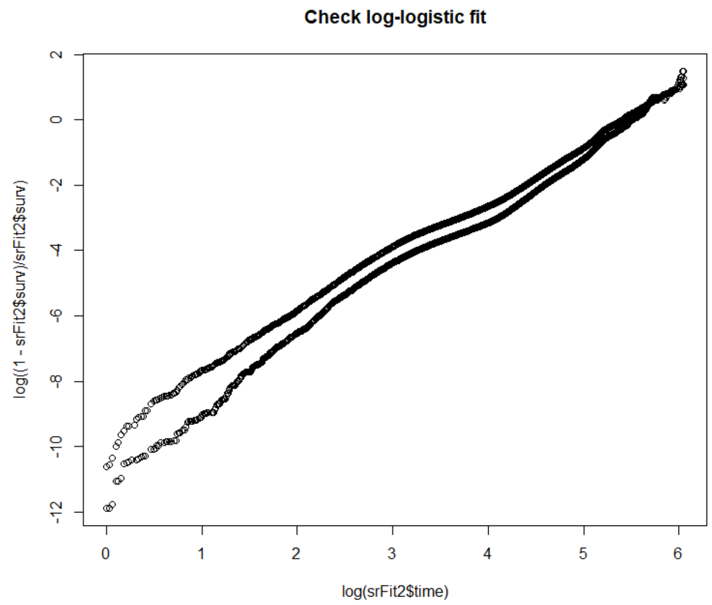
41 Cox D.R., Oakes D., "Analysis of Survival Data", Chapman and Hall, 1984

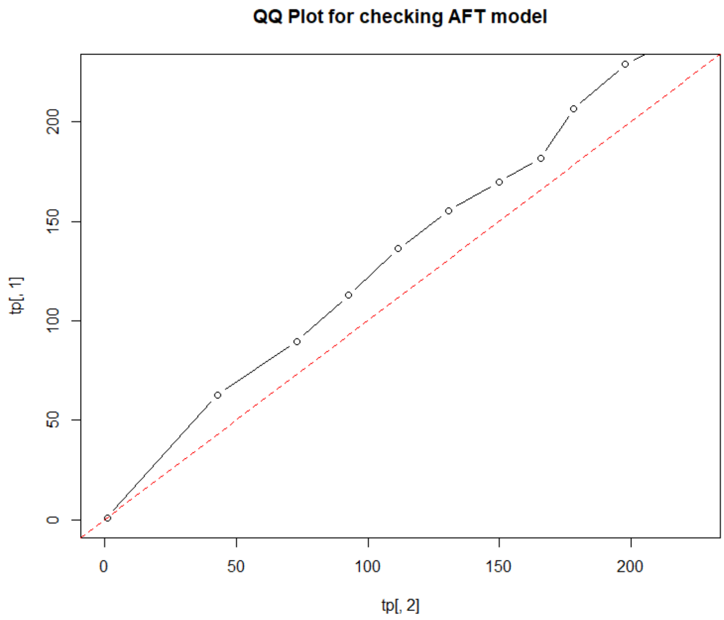




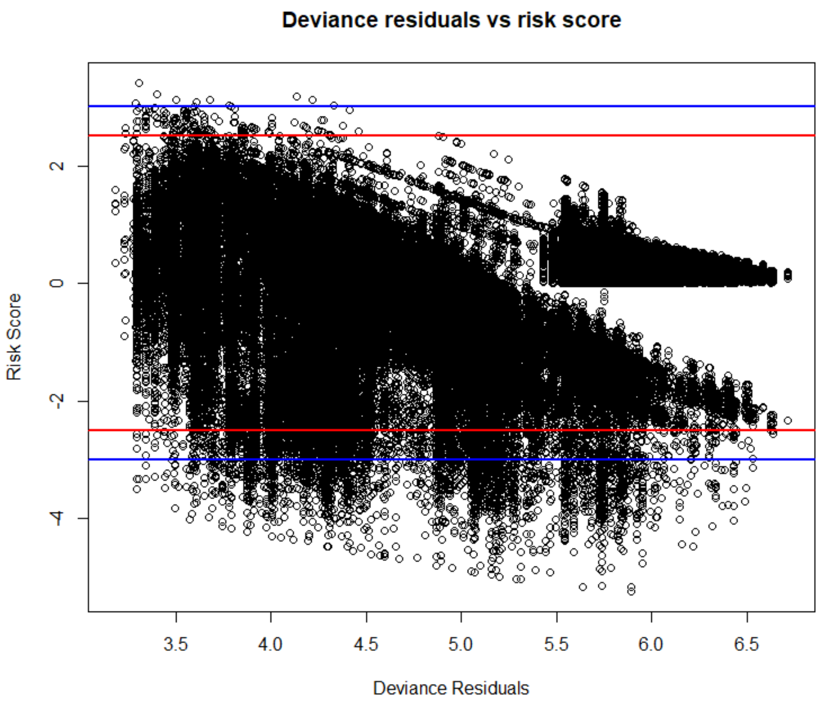
A.2 Floating Rate model

For the sake of completeness we report the results of the prepayment model in case of floating rate mortgages, using the same covariates (except the refinancing incentive, which makes no sense in the floating rate case)

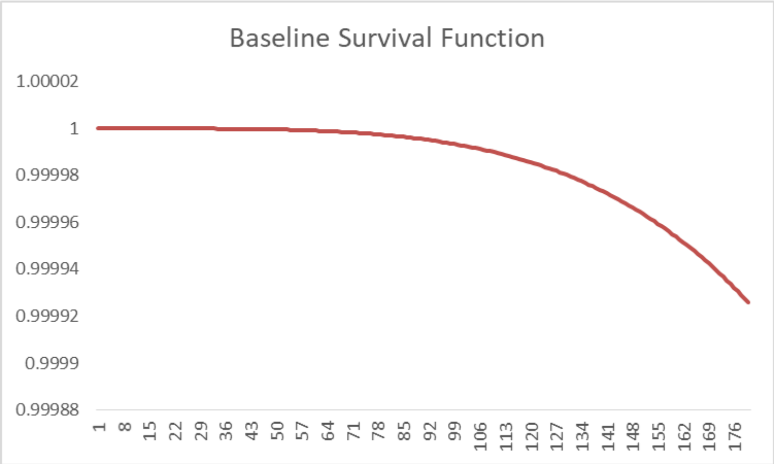

Check log-logistic fit
log((1 - srFit2$surv)/srFit2$surv)
log(srFit2$time)
2
0
-2
-4
-6
-8
-10
-12
0
1
2
3
4
5
6

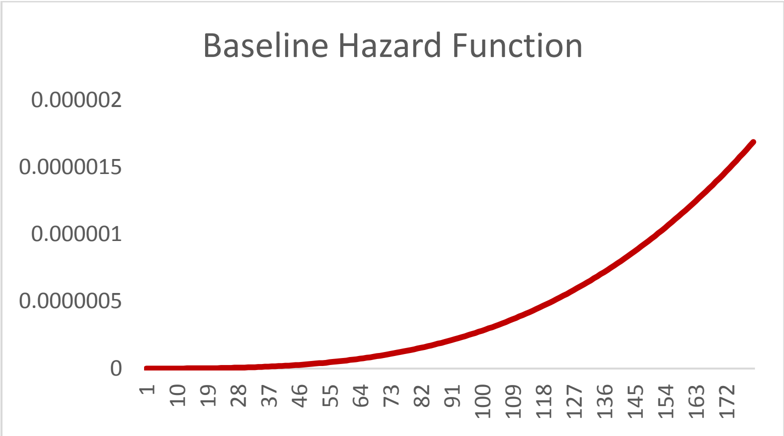

QQ Plot for checking AFT model
tp[, 1]
tp[, 2]
0
50
100
150
200

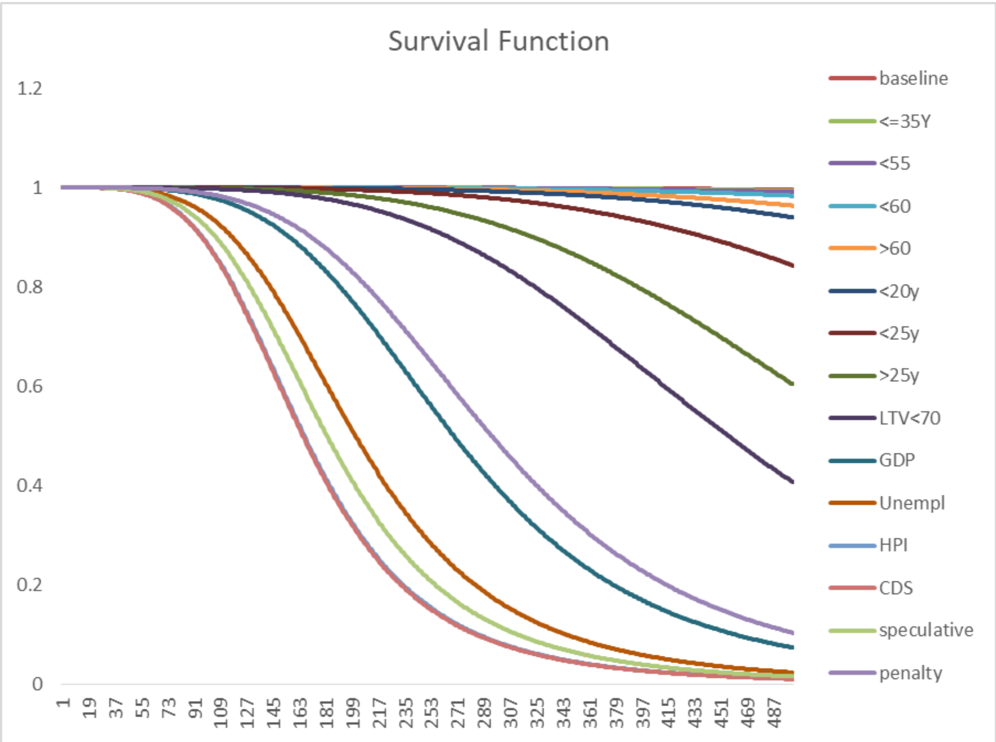
```
Call:
survreg(formula = mySurv_flt ~ as.factor(CTP_AGEM_AGGR) + as.factor(DUR_ORIG_AGGR) +
    as.factor(CDG_LTV_AGGR) + as.factor(CTP_CLARAT_AGGR) + as.factor(OPR_TIPO) +
    GDP + Unempl + HPI + MPS, data = working_flt, weights = WEIGHTS,
    dist = "loglogistic")
                                                 Value Std. Error       z        p
(Intercept)                                    7.53659   4.82e-02  156.30 0.00e+00
as.factor(CTP_AGEM_AGGR)26-35                 -0.07681   1.14e-02   -6.73 1.70e-11
as.factor(CTP_AGEM_AGGR)36-55                 -0.08913   1.13e-02   -7.87 3.49e-15
as.factor(CTP_AGEM_AGGR)56-60                 -0.14705   1.55e-02   -9.50 2.01e-21
as.factor(CTP_AGEM_AGGR)>60                   -0.20152   1.59e-02  -12.70 5.59e-37
as.factor(DUR_ORIG_AGGR)[15Y:20Y]             -0.12943   6.33e-03  -20.43 8.44e-93
as.factor(DUR_ORIG_AGGR)[20Y:25Y]             -0.26219   7.46e-03  -35.12 2.81e-270
as.factor(DUR_ORIG_AGGR)>25Y                  -0.31150   6.72e-03  -46.36 0.00e+00
as.factor(CDG_LTV_AGGR)[0:70]                 -0.19639   4.94e-03  -39.72 0.00e+00
as.factor(CTP_CLARAT_AGGR)Speculative Grade    0.09556   5.03e-03   18.99 1.88e-80
as.factor(OPR_TIPO)Penalty                     0.48147   5.49e-03   87.76 0.00e+00
GDP                                           -0.52934   4.63e-03 -114.34 0.00e+00
Unempl                                        -0.28853   5.12e-03  -56.31 0.00e+00
HPI                                           -0.19093   1.29e-03 -147.83 0.00e+00
MPS                                           -0.00574   3.04e-05 -188.74 0.00e+00
Log(scale)                                    -0.97647   3.59e-03 -271.71 0.00e+00

Scale= 0.377

Log logistic distribution
Loglik(model)= -332158.6   Loglik(intercept only)= -400013.9
        Chisq= 135710.7 on 14 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n=303014 (1323 observations deleted due to missingness)

>
```

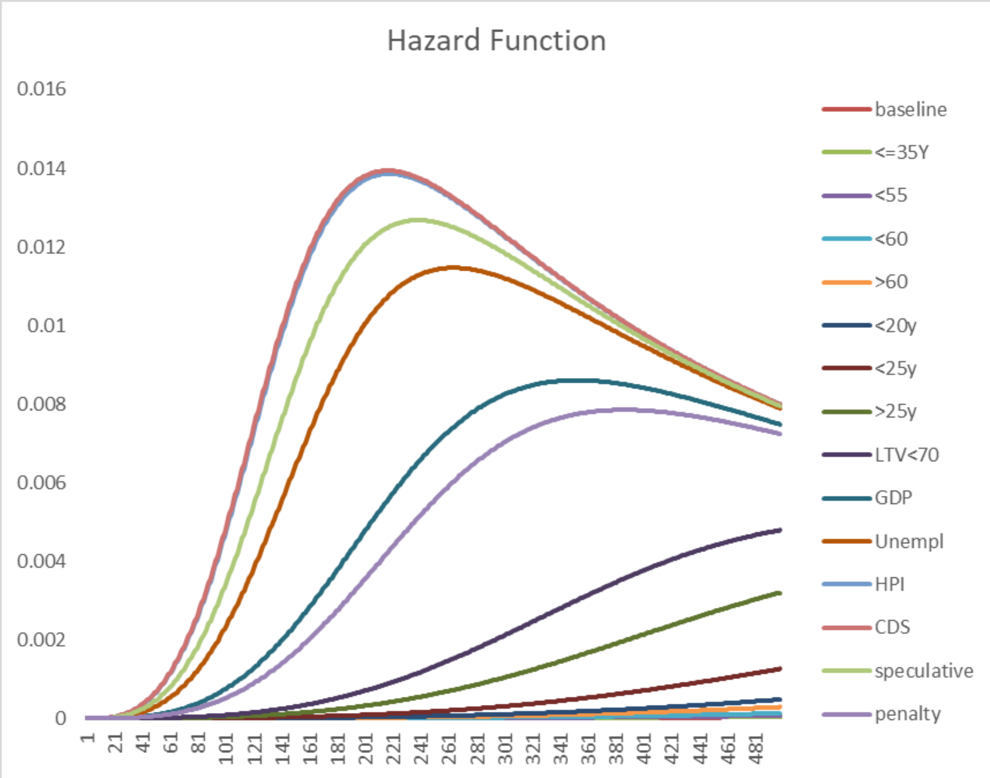


The log logistic fitting looks quite good, but the value of the intercept is very high, so the final model seems to have an higher survival probability comparing to the fixed rate case. Macroeconomic factors and the credit worthiness of the lending bank seems to matter the most:


Baseline Survival Function
1.00002
1
0.99998
0.99996
0.99994
0.99992
0.9999
0.99988
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176


Baseline Hazard Function
0.000002
0.0000015
0.000001
0.0000005
0
1 10 19 28 37 46 55 64 73 82 91 100 109 118 127 136 145 154 163 172


Survival Function
1.2
1
0.8
0.6
0.4
0.2
0
1 19 37 55 73 91 109 127 145 163 181 199 217 235 253 271 289 307 325 343 361 379 397 415 433 451 469 487
baseline
<=35Y
<55
<60
>60
<20y
<25y
>25y
LTV<70
GDP
Unempl
HPI
CDS
speculative
penalty



Hazard Function
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
1
21
41
61
81
101
121
141
161
181
201
221
241
261
281
301
321
341
361
381
401
421
441
461
481
baseline
<=35Y
<55
<60
>60
<20y
<25y
>25y
LTV<70
GDP
Unempl
HPI
CDS
speculative
penalty

Bibliography

Textbooks

Cox D.R., Oakes D. (1984). *Analysis of Survival Data*. Chapman and Hall

Richard Thaler and Cass Sunstein (2008). *Nudge. Improving decisions about health, wealth and happiness.*

Fondazione Rosselli (2015). *European Banking 3.0. Bank Industry and Supervision in the Behavioural Finance Revolution*. Bancaria Editrice

Speare A. (1974). Residential Satisfaction as an Intervening Variable in Residential Mobility. Demography

Mara Tableman and Jong Sung Kim (2004). *Survival Analysis Using S: Analysis of Time-to-Event Data.* Kindle Edition

Dobson A.J. and Barnett A.C. (2008). *An Introduction to Generalized Linear Models.* Chapman & Hall/CRC

Meeker and Escobar (1998). *Statistical Methods for Reliability Data*. Wiley series in probability and statistics

Moore D.F. (2016). *Applied Survival Analysis using R*. Springer

Resti A. Sironi A (2015). *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies.* Wiley Finance

Martinussen and Scheike (2006). *Dynamic Regression Models for Survival Data.* Springer

Papers

John H. Huston and Roger W. Spencer (2013). *Housing and behavioural factors.* Applied Economics Letters

Clapp, Deng and An. (2013). *Unobserved heterogeneity in Mortgage Termination Risk.* Real Estate Economics

Bennett, Peach and Peristiani (2001) *Structural Change in the Mortgage Market and the Propensity to Refinance*. EconPapers

Hendershott and Van Order (1987). Pricing mortgages: *An interpretation of the models and results*. NBER

Kau and Keenan (1995). *The valuation at origination of fixed-rate mortgages with default and prepayment.* Journal of Real Estate Finance and Economics

Deng, Quigley and Van Order (2000). *Mortgage termination, Heterogeneity and the exercise of mortgage options.* Econometrica

Shankar Khanal (2014). *Accelerated Failure Time Models: An application in the Survival of acute liver failure patients in India*. Research Gate

Consalvi and Scotto (2010). *Measuring Prepayment Risk: an application to Unicredit Family Financing*. Unicredit & Universities

J.P.A.M. Jacobs, R.H. Koning, E. Sterken. Dept. Economics, University of Groningen, The Netherlands. *Modelling Prepayment Risk*. Research Gate

Mulla Parveen Yusuf. (2015) *Behavioural Finance:a challenge to market efficiency*. CLEAR International Journal of Research in Commerce & Management

Pavlov (2001). *Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?*. The Journal of Real Estate Finance and Economics

Mantel (1966). *Evaluation of survival data and two new rank order statistics arising in its consideration*. Cancer Chemotherapy Reports

Bagdonavicius and Nikulin (2006). *On goodness‐of‐fit for homogeneity and proportional hazards*. BIG statistics

Pocock, Clayton, Altman (2002). *Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls*. The Lancet

Clark, Bradburn, Love, Altman (2003). *Survival Analysis Part 1: Basic concepts and first analyses*. British journal of cancer

Clark, Bradburn, Love, Altman (2003). *Survival analysis part II: multivariate data analysis--an introduction to concepts and methods.* British journal of cancer

Zare, Hosseini, Mahmoodi, Mohammad,Zeraati, Holakouie Naienia (2015). *Comparison between Accelerated Failure-time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients.* Iranian Journal of Public Health

Swindell W. (2008). *Accelerated failure time models provide a useful statistical framework for aging research.* PMC

Clark, Bradburn, Love, Altman (2003). *Survival analysis part III: multivariate data analysis—choosing a model and assessing its adequacy and fit.* British journal of cancer

Gonchanov (2002) *An intensity based approach for valuation of mortgage contracts subject to prepayment risk*. PhD thesis, University of Illinois, Department of Mathematics, Statistics and Computer Science

Jiezhi q. (2009) *Comparison of proportional hazards and accelerated failure time models.* Msc Thesis, University of Saskatchewan Department of Mathematics and Statistics. eCommons

Meis J. (2015). *Modelling prepayment risk in residential mortgages*. Msc Thesis, University of Rotterdam

Zheng Jing (2016) Survival Analysis *Modeling with Parametric Distributions of the Location-Scale Family*. Msc Thesis, McMaster University, Department of Mathematics and Statistics

Agency (2014). *Recent Trends in the Enterprises' Purchases of Mortgages from Smaller Lenders and Nonbank Mortgage Companies.* FHFA OIG

Saikia R, Barman M.P. (2017). *Comparing Accelerated Failure Time Models with Its Specific Distributions in the Analysis of Esophagus Cancer Patients Data.* International Journal of Computational and Applied Mathematics.

Regulatory publications

Bank of Italy. *Istruzioni relative alla classificazione della clientela* (2014) https://www.bancaditalia.it/statistiche/raccolta-dati/segnalazioni/normativa-segnalazioni/Circ_140_4_agg.pdf

EBA (2018). *Final Report. Guidelines on management of interest rate risk arising from non-trading book activities.* https://www.eba.europa.eu/documents/10180/2282655/Guidelines+on+the+management+of+interest+rate+risk+arising+from+non-trading+activities+%28EBA-GL-2018-02%29.pdf

BCBS (2008). *Principles for Sound Liquidity Risk Management and Supervision. Basel Committee on Banking Supervision*. https://www.bis.org/publ/bcbs144.htm

BCBS (2010). *Basel III: International Framework for liquidity risk measurement, standards and monitoring.* https://www.bis.org/publ/bcbs188.pdf

BSCS (2014). *Basel III – The Net Stable Funding Ratio. Consultation paper. https://www.bis.org/bcbs/publ/d295.htm*

ECB (2017). Sensitivity analysis of IRRBB – Stress test 2017. Final Results https://www.bankingsupervision.europa.eu/press/pr/date/2017/html/ssm.pr171009.en/ssm.pr171009_slides.en.pdf

European Bank Disclosure

Unicredit. *Unicredit Group Disclosure as at 31 December 2017*. https://www.unicreditgroup.eu/content/dam/unicreditgroup-eu/documents/en/investors/third-pillar-basel/2017/UNICREDIT-GROUP-DISCLOSURE.pdf

IntesaSanPaolo Group. *Basel 3 Pillar3. Disclosure as at 31 December 2017* https://www.group.intesasanpaolo.com/scriptIsir0/si09/contentData/view/20180509_PILLAR3_FY17_uk_def.pdf?id=CNT-05-0000000501A99&ct=application/pdf

Ubibanca. Pillar 3 disclosure https://www.ubibanca.it/contenuti/RigAlle/UBI_Informativa%20al%20pubblico_Pillar%203_al%2031.12.2017_pubblicazione_ing.pdf

Deutsche Bank. Risk Report 2017. https://annualreport.deutsche-bank.com/2017/ar/servicepages/downloads/files/dbfy2017_risk_report.pdf

Commerzbank. Disclosure report as at 31 Decembre 2017 https://www.commerzbank.com/media/en/aktionaere/service/archive/konzern/2018_2/Commerzbank_Offenlegungsbericht_2017.pdf

BNPP Paribas. Results as at 31 December 2017. https://invest.bnpparibas.com/en/results

Societe Generale. Risk Report 2018. Pillar 3 2017. https://www.societegenerale.com/sites/default/files/2018/societe_generale_-_pillar_3_2018_en_002.pdf

BBVA. Pillar 3 2017. https://shareholdersandinvestors.bbva.com/wp-content/uploads/2018/03/16032018PilarIIIen.pdf

Santander Consumer SA. Risk and Capital Managment 2017.

https://www.santanderconsumer.no/globalassets/om-oss/investor-relations/basel-ii/santander-consumer-bank-as-pillar-3-disclosure-2017.pdf

Web sites

Lakhbir Hayre (2006). *Modeling of Mortgage Prepayments and Defaults*. Columbia Edu. https://ieor.columbia.edu/files/seasdepts/industrial-engineering-operations-research/pdf-files/Hayre_Lakhbir.pdf

Breheny P. (2010). *Residuals and model diagnostics.* BIOS 7210. http://myweb.uiowa.edu/pbreheny/7210/f15/notes/11-10.pdf

Vergari D., Macellari M. *IFRS9 sfide e soluzioni: dalla metodologia alla messa a terra.* https://www.abieventi.it/Downloads/12977_M-Macellari-e-D-Veragri-CRIF-Credit-Solutions.pdf

Duke University Libraries: *Appendix Three: R Code for Analyses.* https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/8499/Appendix%20Three%20R%20Code.pdf?sequence=4

Dimitris Rizopoulos (2018). *EP03: Survival Analysis in R Companion.* http://www.drizopoulos.com/courses/emc/ep03_%20survival%20analysis%20in%20r%20companion#accelerated-failure-time-models

Qingxia (Cindy) Chen (2012). Bios 323: *Applied Survival Analysis*. http://biostat.mc.vanderbilt.edu

Dätwyler C, Stucki T (2011). *Parametric Survival Models*. ETH Zurich Seminars

CRAN.r-project. *Package Flexsurv*. (2017); *The survival Package* (2006) https://cran.rproject.org/web/packages/flexsurv/flexsurv.pdf

http://ftp.auckland.ac.nz/software/CRAN/doc/packages/survival.pdf

Parametric models in survival Analysis. Notes https://www.stat.nus.edu.sg/~stachenz/ST5212Notes8.pdf

Lindell, Berry (2017). Survival Analysis Tutorial. Rstudio-pubs https://rstudio-pubs-static.s3.amazonaws.com/300535_2a8382af47714d0aaa3f4cce9a7645a3.html