Measuring prepayment risk: an application to UniCredit Family Financing

Matteo Consalvi
Giovanni Scotto di Freca
Statement of Purpose
The Working Paper series of the UniCredit & Universities Foundation is designed to disseminate and to provide a platform for discussion of either work of the UniCredit Group economists and researchers or outside contributors (such as the UniCredit & Universities scholars and fellows) on topics which are of special interest to the UniCredit Group. To ensure the high quality of their content, the contributions are subjected to an international refereeing process conducted by the Scientific Committee members of the Foundation.

The opinions are strictly those of the authors and do in no way commit the Foundation and UniCredit Group.

Scientific Committee
Franco Bruni (Chairman), Silvia Giannini, Tullio Jappelli, Catherine Lubochinsky, Giovanna Nicodano, Reinhard H. Schmidt, Josef Zechner

Editorial Board
Annalisa Aleati
Giannantonio de Roni

The Working Papers are also available on our website (http://www.unicreditanduniversities.eu)
Measuring prepayment risk: an application to UniCredit Family Financing∗

Matteo Consalvi† matteo.consalvi@unicreditgroup.eu
Giovanni Scotto di Freca† giovanni.scottodifreca@unicreditgroup.eu

February 15, 2010

Abstract

Banking books contain numerous implicit options such as prepayment options on mortgages, early withdrawal options etc. As these options may be exercised in response to market changes, they induce significant liquidity risk and interest rate risk. We propose here a behavioral model for estimating and measuring prepayment risk which is both tractable and sufficient realistic.

1 Introduction

Mortgages granted to physical persons feature an implicit prepayment option in relation to maturity; such an option can usually be exercised by paying a penalty that is proportional to the residual debt still in place.

Art. 7 of Law 40/2007 eliminates, for mortgage agreements entered into after 2 February 2007, the possibility for the bank to charge a penalty in case of the borrower’s request for early or partial termination of a mortgage agreement.

1.1 Purposes of the analysis

The prepayment trend for loans granted to private individuals or companies exposes credit-granting Banks to several risks, which can be subdivided into three main categories:

• Interest Rate Risk: for fixed rate mortgages, typically hedged against the risk of changes in interest rates through Interest Rate Swaps in which the Bank pays the fixed rate and receives the floating rate, the
presence of significant prepayment rates exposes the bank to the risk of paying higher fixed rates on the derivatives than the one collected on the newly stipulated loans. For fixed rate mortgages, the prepayment also exposes banks to the risk of overhedging. On the other hand, for floating rate mortgages, the risk generated by prepayments is limited to the risk of reinvestment for commercial spread collected from customers.

- **Liquidity Risk**: the maturity profile of the loans subject to prepayment has a considerable impact on the representation of the liquidity profile of commercial banks: an incorrect evaluation of this profile exposes banks to the risk of overestimating their future liquidity requirements (overfunding) as well as to the risk of increased long-term liquidity costs. The liquidity risk generated by prepayment is independent from the type of the rate of the mortgages (fixed or floating).

- **Mispricing towards customers**: pricing policies for customers should take into account the costs deriving from managing the aforementioned risks and the Internal Transfer Rates system should consequently consider such costs by remunerating the ALM centre for managing the related risks.

- **Mispricing towards special-purpose vehicles as regards securitisation**: considering the risk of prepayment on the basis of internal evidence would help banks better estimate the reference fair value of the mortgage portfolios being securitised.

More specifically, the use of a suitably tested and updated prepayment model may enable banks to obtain substantial benefits in terms of: reducing their funding costs deriving from reallocating the medium/long term assets subject to prepayment on short/medium term buckets; more competitiveness in the pricing policy for customers deriving from reduced funding costs. Such advantages are immediately reflected in the Balance Sheet (debt structure in terms of maturity and products) and Income Statement of the Bank.

### 1.2 The analysis of the prepayment in the literature

The approach to the problem of measuring the prepayment can be associated to two main theoretical frameworks: the financial framework, which relies on the principle of absence of arbitrage (being analysed through the optional theoretical models) and the behavioural framework (being analysed through the econometric models). The financial models are suitable for use in the analysis of callable securities (or corporate loans). The behavioural models are better suited to explain the prepayment trend in the retail mortgage market, towards which commercial banks are predominantly exposed.
The difficulties of applying the financial models to the retail mortgage market relate to two sets of reasons: the average level of financial knowledge of the participants distances it from the efficient market theory and, at least for the Italian market, the predominance of floating rate agreements makes the trend in any case hard to justify in terms of absence of arbitrage. With regard to behavioural models, special consideration is given to the models developed within the survival analysis, in which the model we propose can be included. Therefore, this document specifically focuses on recalling the principles of this theory and describing its application to the prepayment trend.

The prepayment option granted to the borrower should be exercised with the intent of minimising the value of the mortgage: in particular, the latter should be refinanced when its value exceeds the option exercise price. Based on these considerations, the value of the mortgage could be calculated according to the option pricing theory, just like a callable bond. A study of the historical data regarding the exercise of the prepayment option reveals the presence of non optimal prepayment decisions: borrowers exercising the prepayment option when the refinancing interest rate is higher than the contractual rate, or a mortgage not being prepaid when the market rate is lower than the contractual rate and the exercise of the prepayment option would instead be ideal. The exercise of the prepayment option also on floating rate mortgages (for which the refinancing option has a value close to zero\(^1\)), combined with the inability of "option-based" models to explain the prices observed for MBS (Mortgages Backed Securities), has led to the development of prepayment models based on econometric models. The aim is to explain the relationship existing between the prepayment rates observed and explanatory variables that influence the prepayment decision. The models most frequently used in literature to estimate the probability of prepayment are part of the "survival analysis", whose purpose is to model the distribution of the "occurrence time" variable by defining the relationship that links it to some explanatory variables, usually called covariates. Within the survival analysis, the Accelerated Life Model is a category of models that applies well to the analysis of the prepayment risk, in which the explanatory variables determine that "time effectively 'runs faster' for some individuals compared to others" (Thomas Nye).

### 2 Survival Analysis

The survival analysis is a branch of statistics which studies the modelling of a special random variable: the time elapsing before the occurrence of a certain event; in particular, if \((T)\) is a random variable with a probability

---

\(^1\) The value of the option, ceteris paribus, may be conditioned by the ability to obtain lower spreads consequently to an improved credit standing.
distribution of $F(t) = Pr(T \leq t)$ defined on the interval $[0, \infty]$, the survival function of $T$, indicated with $S(t)$, is given by:

$$S(t) = PrT \leq t = \int_t^\infty f(u)du = 1 - F(t)$$ (1)

and is therefore defined as complementary to the distribution function of $T$. If $T$ expresses an occurrence time, the survival function represents the probability of an event occurring after a certain time $t$. The probability density function of $T$, $f(t)$, is defined as:

$$f(t) = \lim_{\Delta t \to 0} \frac{PR(t \leq T < t + \Delta t)}{\Delta t} = -\frac{dS(t)}{dt}$$ (2)

and represents the instantaneous probability of “death” or “failure” at time $t$.

An additional characterisation of the survival function is given by the hazard function, $h(t)$, which expresses the instantaneous risk of failure at time $t$, conditional to the non occurrence of the event until time $t$:

$$h(t) = \lim_{\Delta t \to 0} \frac{PR(t \leq T < t + \Delta t|T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{S'(t)}{S(t)}$$ (3)

Finally, the cumulated hazard function is given by:

$$H(t) = \int_0^t h(u)du = -lnS(t)$$ (4)

The survival models can be seen as ordinary regression models, where time is the response variable. However, the calculation of the likelihood function is compounded by problems relating to the possible incomplete observation of the failure time, the censoring. The censoring mechanism may be classified differently, depending on the type of information it provides; three types of censoring can be distinguished in particular: right censoring, left censoring and interval censoring.

An observation is right censored if the occurrence time of the event being observed is unknown\textsuperscript{2}, left censored if the initial time of exposure to the risk is unknown\textsuperscript{3}. Therefore, an observation can be both left censored and right censored: in this case it is usually called interval censored\textsuperscript{4}.

\textsuperscript{2}Observations are referred to as right censored every time the observation of the event terminates before its occurrence has been checked.

\textsuperscript{3}Left censored observations can be obtained in those cases in which the occurrence of the event of interest is only checked at time intervals, instead of continuously; if the observation takes place once a year, for example, and the event being analysed occurs during the year, it is only known that $t < l$, where $l$ is equal to one year.

\textsuperscript{4}Interval censored observations may also be obtained when the occurrence of the event of interest is not checked continuously: if the observation takes place once a year and the event does not occur during the first year, but occurs during the second year, it is then known that $r < t < l$, where $r$ equals one year and $l$ equals two years.
The likelihood function for survival models in the presence of censored observations is formulated by partitioning the sample of observations into four categories, discriminated by the fact that the observation is uncensored, left censored, right censored or interval censored. In particular, by hypothesising that the observations are independent, given the parameters, the likelihood function is defined as the product of the likelihood functions of each observation.

For an uncensored observation \( i \), the contribution to the likelihood function is given by \( f(t_i|\Theta) \). For a right censored observation the occurrence time of the event is known to be greater than a time \( r_i \); then, the likelihood for the right censored observation \( i \) is:

\[
Pr(T > r_i|\Theta) = S(r_i|\Theta)
\]

In a left censored observation, it is only known that the occurrence time \( t_i \) of the event is shorter than a time \( l_i \); therefore, the contribution to the likelihood is:

\[
Pr(T < l_i|\Theta) = F(l_i|\Theta) = 1 - S(l_i|\Theta)
\]

Finally, for an interval censored observation the occurrence time ranges between \( r_i \) and \( l_i \), therefore:

\[
Pr(r_i < T < l_i|\Theta) = S(r_i|\Theta) - S(l_i|\Theta)
\]

In the presence of censored observations, the likelihood is therefore given by:

\[
L(\Theta|t) = \prod_u f(t_i|\Theta) \prod_l F(l_i|\Theta) \prod_r S(r_i|\Theta) \prod_v [S(r_i|\Theta) - S(l_i|\Theta)] \quad (5)
\]

where products are developed on the set of uncensored, left censored, right censored and interval censored observations, respectively. The most frequent and significant problem relates to the presence of right censored observations, which may occur every time the observation period has a predefined duration; the termination of the analysis interval may result in the failure to observe the event, even though its occurrence is possible at a subsequent time.

Considering a sample of \( N \) observations of the failure time \( t_i \) and assuming the existence of a period \( r_i \) such that the observation of the failure time ends in \( r_i \), if \( T_i > r_i \), then the sample consists of the set of observations \( X_i = \min(T_i, r_i) \) and the dummy variable \( I_i = 1 \) if \( T_i \leq r_i \) (uncensored) and \( I_i = 0 \) if \( T_i > r_i \) (right censored). In terms of failure or censoring time observed, the likelihood function becomes

\[
L(\Theta|x_i) = \prod_u f(x_i|\Theta) \prod_r S(x_i|\Theta) \quad (6)
\]

using the relationship \( f(t) = h(t)S(t) \), the log-likelihood function is:
3 Explanatory variables and survival functions

The survival analysis typically has the purpose of analysing the relationship existing between the survival function and some explanatory variables, called covariates. To this end, it is often worthwhile to define the vector $z$ of explanatory variables in a way to develop the model in two parts: a model for the distribution of the failure time in correspondence with a zero value of the explanatory variables ($z = 0$) and the representation of the change induced on the distribution of the failure time by a value of $z$ other than zero. Within the survival analysis, the link between the survival function and the explanatory variables is usually expressed on the basis of two main models: accelerated life model and proportional hazard model. The following notation will be used in the following paragraphs:

1. $z$, a vector of explanatory variables,
2. $\psi(z)$, a positive function linking vector $z$ to the failure time distribution,
3. $S_0(t)$, the survival function of the random variable $T_0$ in correspondence with $z = 0$,
4. $h(t)$, the hazard function under the standard condition $z = 0$.

3.1 Accelerated Life Model

In the Accelerated Life model (AL) the survivor function of $T$, for $z \neq 0$, is given by:

$$S(t; z) = S_0[t\psi(z)]$$

with $\psi(0) = 1$; therefore, the density and hazard functions are:

$$f(t; z) = f_0[t\psi(z)]\psi(z)$$

and

$$h(t; z) = h_0[t\psi(z)]\psi(z)$$
The random variables \( T_0 \) and \( T \) are linked by the relationship:

\[
T = \frac{T_0}{\psi(z)}
\] (10)

The function \( \psi(z) \) is usually expressed in parametric form: \( \psi(z; \Theta) \); since \( \psi(z; \Theta) \geq 0 \) and \( \psi(0; \Theta) = 1 \), a natural candidate is

\[
\psi(z; \Theta) = e^{\Theta'z}
\] (11)

in this way, considering the logarithm of the failure time and with \( \mu_0 = E(\ln T_0) \), the (10) can be rewritten as

\[
\ln T = \mu_0 - \Theta'z + \epsilon
\] (12)

a linear regression model for \( \ln T \).

In Accelerated Life models the occurrence time is decreased or increased in relation to the value taken in correspondence with \( z = 0 \). Therefore, a parametric model for the survival analysis may be obtained by clarifying the distribution followed by the occurrence time and specifying a functional form for \( \psi(.) \).

### 3.1.1 Exponential distribution of occurrence times

Defining a constant hazard rate, \( h_0(t) = \lambda \), means hypothesising an exponential distribution for the variable \( T_0 \), with survival function \( S_0(t) = \exp(-\lambda t) \) and density function \( f(t) = \lambda \exp(-\lambda t) \). The constant hazard function reflects the lack-of-memory property of the distribution: for each \( t_i > 0 \), the conditional distribution of \( T - t_i \), given \( T > t_i \), coincides with the unconditional distribution of \( T \).

Considering the case of parameter exponential failure time distribution \( \lambda(z_i; \Theta) \) depending from the explanatory variables and a vector \( \Theta \) of unknown parameters, the log-likelihood function in the presence of right censored observations is given by:

\[
\sum_u \ln \lambda(z_i; \Theta) - \sum \lambda(z_i; \Theta)x_i
\] (13)

where \( x_i = \text{min}(t_i, r_i) \) and the sums are developed on the set of uncensored observations and on the entire sample of observations, respectively. Considering the following specification for function \( \lambda \):

\[
\lambda(z; \Theta) = \exp(\Theta'z)
\] (14)

becomes

\[
\sum_u \Theta'z_i - \sum \exp(\Theta'z_i)x_i
\] (15)
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

3.1.2 Weibull distribution of occurrence times

The random variable $T_0$ with Weibull distribution has the survival function:

$$S_0(t) = \exp[-(\lambda t)^\alpha] \quad (16)$$

Parameter $\alpha$ is called the shape parameter of distribution, while $\lambda$ is the scale parameter.

The hazard and density function are given by:

$$f_0(t) = \lambda \alpha (\lambda t)^{\alpha - 1} \exp[-(\lambda t)^\alpha] \quad (17)$$

respectively.

Parameter $\alpha$ plays an important role in characterising the hazard function; if $\alpha > 1$ in particular, the hazard function for the Weibull distribution is monotonically increasing; on the other hand, if $\alpha < 1$, $h(t)$ is monotonically decreasing; finally, for $\alpha = 1$ the Weibull distribution coincides with the exponential distribution of the occurrence times and the hazard function is constant; therefore, the Weibull distribution includes, as a special case, the exponential distribution.

Figure 1 reports the graph of the Weibull survival function for different values of parameter $\alpha$. Figure 2 and figure 3 show the density function and the hazard function of the Weibull distribution for the different values of $\alpha$, respectively.

Defining $\psi(z; \Theta) = \exp(\Theta' z)$, the survival function of $T$ for $z \neq 0$ in the AL model is given by:

$$S(t) = \exp[-(\lambda e^{\Theta' z})^\alpha] \quad (18)$$

The survival function of random variable $T$ defined by (18) is still a Weibull survival function.

Defining also in this case $x_i = \min(t_i, r_i)$ and using the relationship $f(t) = h(t)S(t)$, the log-likelihood in the presence of right-censored observations is:

$$\sum_u [(\alpha - 1) \ln x_i + \alpha \Theta' z_i + \ln \alpha + \alpha \ln \lambda] - \sum [e^{\alpha \Theta' z_i} \lambda \alpha x_i^\alpha] \quad (19)$$

3.1.3 Log-logistic distribution of occurrence times

Considering the log-logistic distribution, the survival function in correspondence with $z = 0$ is given by:

$$S_0(t) = \left[1 + (\lambda t)^\alpha\right]^{-1} \quad (20)$$

while the hazard and density functions are:

$$h_0(t) = \frac{\lambda \alpha (\lambda t)^{\alpha - 1}}{1 + (\lambda t)^\alpha} \quad (21)$$
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

Figure 1: Weibull survival function for different values of parameter $\alpha$

Figure 2: Weibull density function for different values of parameter $\alpha$
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

Figure 3: Weibull hazard function for different values of parameter $\alpha$

$$f_0(t) = \frac{\lambda \alpha (\lambda t)^{\alpha-1}}{[1 + (\lambda t)^\alpha]^2} \quad (22)$$

respectively.

Parameter $\alpha$ defines the distribution form; the hazard function defined by the log-logistic distribution for $\alpha > 1$, in particular, is first increasing and then decreasing, reaching a single maximum in correspondence with $t^* = (\alpha - 1)^{1/\alpha}/\lambda$; instead, for $\alpha \leq 1$ the hazard function is decreasing.

Figure 4 shows the log-logistic survival function for different values of $\alpha$; figure 5 and figure 6 show the density function and the hazard function for different values of parameter $\alpha$, respectively.

By defining the function $\psi(z; \Theta)$, which links failure times $T$ and $T_0$, as $\psi(z; \Theta) = \exp(\Theta' z)$, and introducing, in vector $z$ of explanatory variables, a component $z_0$ with constant value equal to one, in a way that $\lambda = \exp(\Theta_0)$, the survival function of $T$ in the AL model is given by:

$$S(t; z) = [1 + t^\alpha e^{\alpha \Theta' z}]^{-1} \quad (23)$$

Finally, the log-likelihood function for the log-logistic distribution in the presence of right censored observations is:

$$\sum_u [(\alpha - 1) \ln x_i + \ln \alpha + \alpha \Theta' z_i - \ln (1 + x_i^\alpha e^{\alpha \Theta' z})] \quad (24)$$

with $x_i = \min(t_i; r_i)$. 
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

Figure 4: Log-logistic survival function for different values of parameter $\alpha$

Figure 5: Log-logistic density function for different values of parameter $\alpha$
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

3.2 Proportional Hazard Model

The second type of models more widely used to describe the relationship linking explanatory variables to failure time distribution is usually specified through the definition of the hazard function.

In the proportional hazard model (PH) the hazard function for \( z \neq 0 \) is given by:

\[
h(t; z) = \psi(z)h_0(t)
\]  

(25)

with \( \psi(0) = 1 \).

In the PH model the hazard function for \( z \neq 0 \) is a constant multiple of the baseline hazard function, where the proportionality constant depends on the explanatory variables.

The survival, density and cumulated hazard functions are:

\[
S(t; z) = [S_0(t)]^{\psi(z)} = \exp[-H_0(t)\psi(z)]
\]

(26)

\[
f(t; z) = \psi(z)[S_0(t)]^{\psi(z)-1}f_0(t) = \psi(z)[S_0(t)]^{\psi(z)}h_0(t)
\]

(27)

\[
H(t; z) = \psi(z)H_0(t)
\]

(28)

respectively.

As for the AL model, a parametric proportional hazard model is obtained by specifying a functional form for \( \psi(z) \) and clarifying the distribution followed by the failure times, in this case through the baseline hazard function \( h_0(t) \).
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

3.2.1 Weibull distribution of occurrence times

A proportional hazard model based on the Weibull distribution is obtained by specifying the following form for the baseline hazard function:

\[ h_0(t) = \lambda \alpha (\lambda t)^{\alpha - 1} \]  

(29)

By specifying a log linear relationship between the explanatory variables and the hazard function of the type \( \psi(z; \Theta) = \exp(\Theta' z) \), a parametric PH model is given by:

\[ h(t) = \lambda \alpha (\lambda t)^{\alpha - 1} \exp(\Theta' z) \]  

(30)

the hazard function defined by (30) is still the Weibull hazard function, with scale parameter \( \lambda \exp(\Theta' z) \) and form parameter \( \alpha \).

The survival function is given by:

\[ S(t) = \exp\left[-(\lambda t)^{\alpha} \exp(\Theta' z)\right] \]  

(31)

and the log-likelihood function:

\[ l = \sum_u \left[ \alpha \ln \lambda + \ln \alpha + (\alpha - 1) \ln x_i + (\Theta' z_i) \right] - \sum \left[ \exp(\Theta' z_i)(\lambda x_i)^{\alpha} \right] \]  

(32)

The Weibull distribution (and consequently the exponential distribution) is the only distribution for which the AL model and the PH model are equivalent. Indeed, in the Accelerated Life Model the survival function of \( T \) for \( z \neq 0 \) is given by:

\[ S(t) = \exp\left[-(\lambda e^{(\Theta' z)} t)^{\alpha} \right] \]  

(33)

and coincides with (31) unless the coefficients are reparametrised.

In particular, by indicating the parameter vector in the Accelerated Life Model with \( \Theta^{AL} \) and the parameter vector in the Proportional Hazard Model with \( \Theta^{PH} \), the relationship between the coefficients of the two models is given by:

\[ \exp[\alpha(\Theta_1^{AL} z_{i1} + \Theta_2^{AL} z_{i2} + \ldots + \Theta_k^{AL} z_{ik})] = \exp[\Theta_1^{PH} z_{i1} + \Theta_2^{PH} z_{i2} + \ldots + \Theta_k^{PH} z_{ik}] \]

therefore:

\[ \alpha \Theta_j^{AL} = \Theta_j^{PH} \]  

(34)

with \( j = 1, \ldots, k \)
3 EXPLANATORY VARIABLES AND SURVIVAL FUNCTIONS

3.2.2 Log-logistic distribution of occurrence times

Considering a proportional hazard model based on the log-logistic distribution with baseline hazard function:

\[ h_0(t) = \frac{\lambda \alpha (\lambda t)^{\alpha-1}}{1+\lambda t^\alpha} \]  
(35)

and specifying a linear log relationship between the explanatory variables and the hazard function:

\[ \psi(z; \Theta) = \exp(\Theta' z) \]  
(36)

the hazard function for the random variable \( T \) is given by

\[ h(t; z) = \frac{\lambda \alpha (\lambda t)^{\alpha}}{1+\lambda t^\alpha} \exp(\Theta' z) \]  
(37)

The hazard function defined by (37) is not a log-logistic hazard function: a proportional hazard model specified in this way in reality does not define a log-logistic distribution for the occurrence times, since only the baseline hazard function has the correct form. Instead, the log-logistic distribution is defined correctly in the AL model case.

Despite the considerations made on the distribution form generated by the PH model with log-logistic baseline hazard function, Schwartz and Torous (1989) and He and Liu (1998) use the parametric model described by equations (35) and (36) in applying the survival analysis to the estimate of the prepayment risk.

The choice of the distribution form is justified by the will to incorporate the prior knowledge on the dependence of the prepayment from the age of the mortgage (so-called seasoning); indeed, as already highlighted, the log-logistic hazard function admits different relationships between the prepayment probability and the age of the mortgage: for \( \alpha > 1 \) the prepayment probability grows from zero to a maximum amount in correspondence with \( t^* = (\alpha - 1)^{1/\alpha} / \lambda \), then decrease back to zero. Such behaviour is consistent with the observation according to which the prepayment rate is typically low at the beginning of the agreement and increases during the life of the mortgage, then decreases as its maturity approaches.

The survival function for the model defined by equations (35) and (36) is:

\[ S(t; z) = [1 + (\lambda t)^\alpha]^{-\exp(\Theta' z)} \]  
(38)

thus the log-likelihood function, for the parameters included in vector \( \Theta \) as well as \( \alpha \) and \( \lambda \) of the baseline hazard-function, in the presence of censored observations is:
4 PREPAYMENT DETERMINANTS

\[ l = \sum_u [\ln \lambda + \ln \alpha + (\alpha - 1) \ln(\lambda x_i) - \ln[1 + (\lambda x_i)^\alpha] + (\Theta' z_i)] - \sum \exp(\Theta' z_i) \ln[1 + (\lambda x_i)^\alpha] \]  

(39)

3.2.3 Cox Proportional Hazard Model

As already underlined for the PH model, the impact of the explanatory variables is proportional on the hazard function: a parametric model can be specified by defining a functional form for \( \psi(z) \) and clarifying the distribution followed by the failure times under standard condition \( z = 0 \).

In the semiparametric Cox model (1972), the premise of proportionality between covariates and hazard function is maintained:

\[ h(t; z) = \psi(\Theta; z) h_0(t) \]  

(40)

which is typical of the PH model, but no assumption is made as regards the distribution followed by the failure times: the baseline hazard function \( h_0 \) is left unspecified. Furthermore, in the Cox Proportional Hazard Model (CPH), function \( \psi(z; \Theta) \) assumes the following form:

\[ \psi(z; \Theta) = \exp(\Theta' z) \]  

(41)

Through the CPH model, it is therefore possible to obtain estimates of the coefficients of vector \( \Theta \) without clarifying the distribution followed by the failure times. The estimate of the model parameters, which in this case does not include the parameters relating to the baseline hazard function, can be obtained through the partial likelihood method developed by Cox (1972).

The main difference with the parametric AL and PH models analysed in the paragraphs above lies in that the CPH model leads to an estimate of the relative risk among parties and not to an estimate of the absolute risk, which instead can be obtained through the parametric estimate of the survival function or the hazard function.

Therefore, the CPH model has no practical implications if the objective of the analysis is the forecast, but can be used in preventive analyses to select the covariates.

4 Prepayment determinants

The reasons underlying the decision of prepaying a mortgage may be summarised into four main categories:

- **Refinancing incentive**: as mentioned above, the prepayment option for fixed rate mortgages should be exercised on with the intent of minimising the debt value: the borrower should prepay the mortgage when its
value exceeds the option exercise price, i.e. when the refinancing rate is lower than the contractual rate, including the charges deriving from the prepayment. Due to the borrowers' heterogeneous behaviour as regards the collection of information, the speed and the correctness of the reaction to variations in market conditions, a non optimal behaviour was observed in exercising the prepayment option; in particular, given a constant refinancing incentive, not all of the borrowers in a pool will exercise the option identically: as the past prepayment opportunities increase (the refinancing option has been “in the money” several times), the pool will only include those who exercise the option late or do not exercise it at all, this causing a decrease in the prepayment rate (so-called pool burnout effect). The prepayment decision may also be due to variations in the mark-up applied to the reference interest rate; in this case the trend also concerns floating rate mortgages. Finally, the refinancing incentive for floating and fixed rate mortgages may be due to the borrower’s improved (worsened) credit standing: a variation in the rating of the counterparty, particularly in the case of companies, may cause an increase (decrease) in the prepayment rate observed.

- **Housing turnover**: for residential mortgages, the decision to sell a property may derive from different reasons, including transfers for work, divorce, real estate portfolio recomposition, emigration; these reasons may, in turn, be linked to different factors, such as the geographical location of the property (the prepayment rate may vary based on the geographical regions due to different emigration / growth rates), Loan to Value ratio (borrowers with a high ratio may find it hard to access the capital market and may consequently assume a different behaviour in terms of property sale decisions as compared to parties with a low ratio), movements in the real estate market, guided by variations in the price of properties, also depending on changes in the interest rate level; improved family income or seasonality of the transactions in the real estate market. As a result, the factors underlying the prepayment decision due to housing turnover may be further classified into “personal characteristics” (such as income, age, gender, location of the property, marital status, etc.), “mortgage characteristics” (disbursed amount, Loan to Value, interest rate applied) and “conditions of the real estate market”.

- **Default**: in this case the prepayment of a mortgage derives from the exercise of the "default option"; therefore, the prepayment deriving from this type of motivation is alternative to the prepayment deriving from exercising the refinancing option or from a housing turnover.

- **Debt restructuring**: in this case, the prepayment option is exercised when borrower enter into a new loan with the same institution, at
Partial prepayment is also a type of prepayment: the capital is paid in advance as compared to the original instalment plan, resulting in a reduction of the residual debt; therefore, the consequence of partial prepayment is the revision of the mortgage amortisation schedule.

4.1 Refinancing incentive

Among the reasons for the prepayment of fixed rate mortgages, special importance is given to the prepayment that derives from exercising the refinancing option. The explanatory variables most frequently used in the literature for the modelling of such a determinant have the main objective of representing the link existing between the prepayment probability, given by the survival function, and the refinancing incentives deriving from the moneyness of the prepayment option. Therefore, such variables constitute the proxy for the value of the refinancing option available to the borrower and, consequently, for the probability of exercising the same.

4.1.1 Risk-neutral probability of refinancing

The information on the refinancing incentive may be synthesised by one single variable: the risk neutral probability of excercise of option, which may be calculated within an absence of arbitrage opportunity framework. As mentioned above, a mortgage featuring a prepayment option is comparable to an amortizing callable bond, issued by the borrower and undersigned by the Bank: the prepayment option available to the borrower is American or Bermudan.

In order to calculate the risk neutral prepayment probability in closed form, the option may be broken down into a portfolio of N European options on amortizing instruments with exercise dates that correspond to the instalment payment dates $\tau_j$.

Each single maturity call $\tau_j$ will have a strike equal to:

$$K_{\tau_j} = \left[ Dr_{\tau_j}(1 + c) \right] + c^*$$  \hspace{1cm} (42)

where:

$K_{\tau_j}$ is the strike at time $\tau_j$
$Dr_{\tau_j}$ is the residual debt at time $\tau_j$
$c$ is the penalty proportional to the residual debt
$c^*$ represents the fixed costs implied by the prepayment of the mortgage

Supposing that:

- the conditional probability$^5$ of exercise of the $j$-th European option $\delta_j$ is independent from the probability of exercise of the subsequent ones,

$^5$Conditional to the non exercise of any of the options with antecedent exercise date.
the prepayment, if occurring, can only be in full, and thus the j-th option may be exercised only provided that none of the precedent options has been exercised,

- the spread\(^6\) applied to the risk free rate to determine the client rate is constant over time,

it is possible to apply the Black model in the risk free measure. Considering, for simplicity purposes, a bullet mortgage featuring a single exercise date \(\tau\), the value of a call option for this mortgage, according to the Black model, is:

\[
\text{Call}_{\tau_0,\tau} = p(\tau_0, \tau)[F_{\tau_0,\tau}N(d_1) - KN(d_2)]
\]

where:

- \(F_{\tau_0,\tau}\) is the forward value of the mortgage at option exercise date \(\tau\),
- \(p(\tau_0, \tau)\) is the risk free discount factor from the valuation date to \(\tau\),
- \(d_1 = \frac{\ln\left(\frac{F_{\tau_0,\tau}}{K}\right) + \frac{\sigma_{\tau_0,\tau}^2(\tau-\tau_0)}{2}}{\sigma_{\tau_0,\tau}\sqrt{\tau-\tau_0}}\)
- \(d_2 = d_1 - \sqrt{\tau-\tau_0}\sigma_{\tau_0,\tau}\)
- \(\sigma_{\tau_0,\tau}\) is the volatility of \(F_{\tau_0,\tau}\)

As a result, the risk neutral probability of exercise of option could be obtained with the Black model and is:

\[
\delta = N(d_2)
\]

In case of an amortizing mortgage, \(\delta_j\) represents the conditional probability of exercise of the j-th European option; given the aforesaid hypotheses, the unconditional probability of exercise of the j-th option can be obtained through the theorems on conditional probability.

The unconditional probability of exercising the option with maturity \(\tau_j\) synthesises the probability that the time elapsing between valuation date \(\tau_0\) and the exercise of the refinancing option is \(t_j\), with \(t_j = \tau_j - \tau_0\); in other words, it is the (risk neutral) probability that \(t_j\) is the failure time observed for party \(i\).

As recalled above, the information on the refinancing incentive may be completely synthesised by using the risk neutral probability of refinancing, as an explanatory variable within a behavioural model. However, the complexity of the calculations for its derivation may make its practical use difficult. For this reason, the explanatory variables more frequently used in the literature for the modelling of refinancing incentives, consist in proxies of the

---

\(^6\)Two trends might explain a spread reduction due to risk: the client’s varied riskiness or the pricing algorithms adopted by the Bank; clients are quite unlikely to be aware of these elements and, therefore, it would seem reasonable to assume that, when measuring the moneyness of their options, these would initially focus on market rates.
moneyness of the prepayment option given by differences in the interest rate level.

The next paragraph describes the explanatory variable for the modelling of refinancing incentives within the analysis model developed in this paper.

4.1.2 Coupon Incentive

The prepayment option should be exercised when the value of the mortgage exceeds the option exercise price, i.e. when the interest rate at which the mortgage may be refinanced is lower than the contractual rate paid by the borrower, including the prepayment charges. The mortgage refinancing rate may depend, in addition to the level of market interest rates, also on the spread applied by the bank to the risk free rate to determine the client rate. As mentioned above, a variation in the spread may, in turn, depend on two different trends: the borrower’s improved (worsened) credit standing or a variation in the mark-up applied to the reference interest rate (for example, caused by different pricing policies adopted by the bank). For mortgages granted to retail clients, in particular, the borrower would rarely be aware of its credit standing and would be unlikely to know about changes in the pricing algorithms adopted by the bank; therefore, by focusing the analysis on prepayments for residential mortgages, one may reasonably assume that borrowers determine the moneyness of their options mainly in consideration of market rates. In light of these considerations, a synthetic measure of the refinancing incentive for fixed rate residential mortgages has been introduced (and therefore adopted as an explanatory variable) in the framework of the analysis. It is called Coupon Incentive (CI) and is defined as:

\[ CI = y_{\tau_0} - (y_{\tau_j} + \frac{P}{RL_{\tau_j}}) \]  

(45)

where

\( \tau_0 \) is the mortgage disbursement rate.

\( y_{\tau_0} \) is an approximation of the Internal Transfer Rate of the transaction on the mortgage disbursement date \( \tau_0 \); in particular, it is given by the par rate with a tenor equal to the approximate duration of the mortgage upon the disbursement date.

\( \tau_j \) is the date of observation of the failure or censoring time. For a mortgage discharged within the observation sample (so-called uncensored), \( \tau_j \) is the prepayment date; on the contrary, in the case of a mortgage for which a prepayment event does not occur within the observation sample (so-called right censored), \( \tau_j \) is the date of termination of the observation period for mortgage \( j \). Therefore, \( \tau_j = \text{Min (Payment date, Observation end date)} \).

\( y_{\tau_j} \) is an approximation of the Internal Transfer Rate of the substitute transaction on the date of observation of the failure or censoring time \( \tau_j \); in par-
5 Analysis of prepayment risk

The objective of the analysis is to estimate the probability of prepayment, in order to adequately predict the significance of the phenomenon and, accordingly, enable the risks associated to the same to be properly managed. As mentioned above, the existence of the prepayment option in the hands of the borrower exposes banks to a number of risks, such as interest rate risk, liquidity risk and mispricing risk.

The correct valuation and representation of the prepayment phenomenon can, therefore, lead to benefits in terms of: lower risks of overhedging for fixed rate mortgages, a better valuation of short and long term liquidity requirements (thus reducing the risk of overfunding), more accurate pricing policies vis-à-vis the customer (permitted by the potential reduction of the cost of funding and by better valuation and management of prepayment-related risks).

5.1 Prepayment classification

A correct estimation of prepayment risk necessarily implies the proper identification and classification of the phenomenon under analysis. Section 4 illustrated four reasons underlying the decision of prepaying a mortgage. In addition to the full exercise of the prepayment option (total prepayment), the phenomenon may also take different forms: in particular, partial prepayment, mortgage restructuring and default of the counterparty.

Partial prepayment takes the form of payments made in advance with respect to those scheduled, with the objective of reducing the residual debt of the mortgage, without, however, discharging the same; the failed identification of this phenomenon could lead, if erroneously classified as total prepayment, to an overestimation of the prepayment probability, or, if not detected, to an underestimation of the same. In both situations, if the phenomenon is particularly significant, it could lead to a distortion in the valuation of prepayment probability, with consequent repercussions on the results of the
liquidity risk analysis and interest rate risk analysis. Correctly identifying cases of partial prepayment therefore implies checking the database in order to correctly quantify and classify the phenomenon: the analysis should be conducted by checking the intertemporal coherence between the amortisation schedules and the residual debts of mortgages, as well as the size of the sample of live mortgages with respect to those marked as prepayment.

Prepayment resulting from restructuring arises in cases in which the repayment, in this case total, of the mortgage is in reality only formal: the discharged mortgage is actually simultaneously replaced by a new mortgage drawn up with the same bank. In this case, any failure to identify the phenomenon would lead to an overestimation of the repayment probability: restructuring does not necessarily entail a cash-in above that envisaged by the original amortisation schedule. It does not appear to be possible to treat mortgage restructuring phenomena in the same way, as regards the analysis of interest rate risk and liquidity risk; the decision regarding the treatment of those observations will therefore depend on the relative significance of the two analyses. As regards interest rate risk, the restructured mortgage should be considered prepaid on the date of restructuring and the establishment of a new mortgage should be recorded simultaneously. As regards liquidity risk, on the other hand, the restructured mortgage should not be recorded as an observation of a prepayment and the new mortgage drawn up should continue to be monitored in the place of the previous one in the database.

Default is another type of mortgage prepayment. As already mentioned, default risk can be classified as a risk in competition with the prepayment risk. As this type of risk is already monitored and managed when credit and counterparty risks are assessed, only mortgages with performing status should form part of the database analysed for prepayment purposes. The passage of a mortgage to non-performing status can therefore be correctly marked, classifying the observation as right censored and therefore forcing its exclusion from the sample when it becomes non-performing (this is the solution adopted for analysis purposes).

5.2 Methodological approach

Section 2 introduced the most common models used to study the effects of the explanatory variables on the probability of survival, as well as the distribution of failure times.

Among the distributions analysed, the log-logistic distribution is the distribution that best fits the prepayment risk analysis. The hazard function for this distribution permits different relationships between the prepayment probability and the age of the loan: for \( \alpha > 1 \) prepayment probability rises from zero to a maximum corresponding to \( t^* = (\alpha - 1)^{1/\alpha} \) and then falls again. As demonstrated by Schwartz and Torous (1989) and He and Liu (1998) log-logistic distribution enables prior knowledge regarding the depen-
dence of prepayment probability on the seasoning of the loan to be incorpo-
rated in the model; the form of the log-logistic hazard function is, in fact,
consistent with the observation of a prepayment rate that is typically low
at the beginning of the contract, rises as the loan matures and then falls
again as the maturity date approaches. In paragraph 3.2 it was shown how
the choice of the hazard log-logistic function in a Proportional Hazard Model
does not, in reality, entail a log-logistic distribution for the occurrence times:
in fact, only the baseline hazard function possesses the correct form.
On the other hand, log-logistic distribution is correctly specified in an Accel-
erated Life Model; in this case, the survival function of the random variable
$T$ for $z \neq 0$, with $\psi(z; \Theta) = \exp(\Theta'z)$, is given by
\begin{equation}
S(t; z) = \frac{1}{1 + [t\lambda \exp(\Theta'z)]^\alpha}
\end{equation}

and is still a log-logistic survival function.
The term $\exp(\Theta'z)$ in (46) is known as the “acceleration factor”: as $\Theta'z$
increases (decreases), the acceleration factor increases (decreases) and the time
of the subject accelerates (slows down) with respect to the timescale of the
baseline survival function, decreasing (increasing) the subject’s probability
of survival for each time $t$. In the AL Model the covariates have a positive
impact on the logarithm of survival time $t$: the form of the survival function,
determined by the $\alpha$ parameter, is instead the same for all mortgages.
Equation (46) describes the modelling used for prepayment risk analysis in
this paper.
The following two paragraph illustrate how the database for model estima-
tion is constructed and the main types of explanatory variables used in the
model as well as their treatment for the purpose of the estimation.

5.2.1 Construction of the database

The estimation of the vector of parameters $\Phi' = (\Theta', \beta')$, where $\Theta$
indicates the parameters relative to the model’s explanatory variables and $\beta$
indicates the parameters of the baseline survival function, is made within the survival
analysis through the likelihood method; in particular, the estimation of $\Phi$ is
obtained by solving the following optimum problem:
\begin{equation}
\Phi = \arg \max l(\Phi; x, z)
\end{equation}

where $l(\Phi; x, z)$ is the log-likelihood function calculated in the above para-
graphs and given by equation (24) for an AL Model with a baseline survival
log-logistic function.
The maximisation of the log-likelihood function requires the availability of
a sample of occurrence times (mortgage prepayment times) with details of
individual observations: aggregating the sample of mortgages with similar
characteristics is therefore not compatible with the type of observations required by survival analysis. For econometric analysis, the sample should mostly comprise right censored observations, rather than information regarding the trend of the phenomenon analysed.

As already highlighted in paragraph 5.1, within the sample of prepaid loans, it must be possible to distinguish between prepayments and repayments on maturity; within the prepayment category, it must also be possible to differentiate between total prepayments, partial prepayments, passages to non-performing status and restructuring. These events are characterised differently, have a different information value and are treated differently within the analysis; in particular:

- **Total prepayments**: in this case, the occurrence time is given by the time between the date of disbursement of the mortgage and the date of prepayment of the same; the failure time $t_i$ observed must be, in this case, classified as uncensored and the contribution made by said observation to the likelihood function is given by $f(t_i)$.

- **Partial prepayments**: in this case, the occurrence time is given by the difference between the date of disbursement of the mortgage and the partial prepayment date. The failure time $t_i$ is classified as uncensored but the observation contributes to the estimate of the prepayment probability to a lesser extent than a total repayment event; the observation is, in fact, weighted on the basis of the significance of the partial prepayment event: the weight assigned is given by the relationship between the prepaid amount (principal) and the amount disbursed. A mortgage for which one or more events of partial prepayment are observed, will also present a further observation: within the sample of live mortgages at the end of the observation period, or within the sample of discharged loans. This observation is also assigned a weight, in this case given by the complement to one of the sum of the weights assigned to the partial prepayment events of that mortgage.

- **Defaults**: on several occasions, we have emphasised how default risk must be classified as a risk in competition with prepayment risk. The time $r_i$ observed is, in this case, classified as right censored and is given by the time between the date of disbursement and the date on which the mortgage becomes non-performing. The contribution made by a prepayment, marked as default, to the likelihood function is therefore given by $S(r_i)$.

- **Restructuring**: prepayments resulting from renegotiations or restructuring of mortgages will not be recorded as actual prepayment observations; in this case, the prepayment does not entail a cash-in before
that envisaged by the contractual plan and, to avoid an overestimation of the phenomenon, the observation is not entered into the database used for estimation purposes.

- **Repayments on contractual maturity:** in this case, repayment occurs following an event different from prepayment events, which entails the exclusion of the mortgage from the sample. The occurrence time $r_i$ is therefore given by the original maturity of the loan, namely the difference between the contractual maturity date and the loan disbursement date; time $r_i$ is classified as right censored and the contribution made by the observation to the likelihood function is given by $S(r_i)$.

Mortgages, that are still live at the end of the observation period, contribute to the estimation as right censored observations: the only information available as regards the failure time (prepayment time) $t_i$ is that $t_i > r_i$, defining $r_i$ as the period of observation of the i-th loan. In this case, the censoring time is given by the time between the date of disbursement of the loan and the end of the observation period (date of the analysis); the contribution made to the likelihood function is given by $S(r_i)$.

The sample of failure times, calculated as described above, and the indicator variable $I_i = 1$ if $T \leq r_i$ (uncensored) and $I_i = 0$ if $T > r_i$ (right censored) represent the response variable for the estimation of the parameters of the survival function.

### 5.2.2 Characteristics of the covariates

The explanatory variables used in the analysis can be classified into different types, depending on the modes assumed by the covariates:

- **Dichotomic qualitative variables:** assume only two values, which can be encoded through the introduction of a dummy variable; an example of this type of variable within prepayment analysis is the sex of the borrower, which assumes “male” or “female” mode: for the purposes of the model estimation, a binary variable is introduced, which assumes the value 1 in correspondence to one mode of the explanatory variable (for example “male”) and 0 otherwise. The parameters estimated for the dichotomic qualitative variable have an additive effect on the argument of the function $\Phi(.)$ defined by (11). In this analysis, the dichotomic qualitative variables present in the model are the sex and the nationality of the borrower, which assumes “Italian” or “Foreign” modes.

- **Polytomic qualitative variables:** the variable can assume $p$ mode, with $p > 2$; an example of this type of variable within prepayment analysis is the geographic location of the property, which could assume modes
“North”, “Centre” and “South”; in this case, for the purposes of the model estimation, the modes must be encoded through the introduction of dichotomic variables. In the example of geographic location, two dummy variables have to be introduced: the first assumes the value 1 in correspondence to “Centre” mode and zero otherwise, the second assumes the value 1 in correspondence to “South” mode and zero otherwise. “North” mode is identified by other combinations. In this analysis, the following polytomic qualitative variables are present: area in which the property is located, Loan to Value (LTV), borrower’s business activity, Sector of Economic Activity (SAE), age of borrower at time of disbursement and original maturity of the loan. The age of the borrower at time of disbursement, the LTV and the original maturity are ordinal quantitative variables which form part of the analysis as qualitative variables following their prior subdivision into ordinate classes. The business activity of the borrower and the SAE assume a high number of modes in the original database; in this analysis, in order to guarantee the robustness of the estimates, a subset of the modes was considered, depending on the significant percentage of the observations. The remaining modes were grouped into a residual class.

- **Quantitative variables**: these variables assume quantitative values; an example of this type of variable within the prepayment analysis is the Coupon Incentive.

A further point of note in the construction of the database, and in particular of the explanatory variables, regards the treatment of missing values. If the percentage of missing values did not compromise the robustness of the estimates (or the possibility of using the information in the application of the model to predict the prepayment rate), the information was enriched. In particular, the database was enriched by assigning the missing information a value randomly extracted from the subset without missing values of the initial population. This method of integrating missing information also takes into account whether the observation to be enriched belongs to the censored or uncensored set of observations.

### 6 The results

The analyses was performed on Italian residential mortgages booked in Unicredit Family Financing. The sample was divided into two subgroups, differentiated by the financial characteristics of the mortgage:

- Fixed Rate Mortgages,
- Floating Rate Mortgages,
therefore, separate prepayment models were estimated for each of the two subsets.

For each explanatory variable in the model, a preliminary analysis was conducted to verify the significance of the variable, or, if polytomic, of the modes assumed, through the comparison between nested models and through hypothesis testing on parameter values. The variables and the modes selected in this way become explanatory variables in the aggregate model. At this stage, for variables with ordinal modes (such as, for example, the age of the borrower at time of disbursement and the LTV), the process is structured in such a way as to only permit the aggregation of contiguous models; the aim of the latter is to permit a reduction in the number of modes that can be used also for the subsequent application of the model, without compromising the interpretability of the results in reports.

The next stage of model building regards the selection of the significant variables and modes to construct the final model, through the comparison of nested models via backward elimination.

The comparison of the models and the selection of the variables are made through the likelihood ratio test, LR, with a significance level of 1%.

The process used to select the variables and the results obtained are illustrated in the paragraphs below.

6.1 The sample

The sample of mortgages available for Unicredit Family Financing includes the set of repayments made, for the different reasons already mentioned, between 31/12/2005 and 30/09/2009, as well as the set of live mortgages on the end date of the observation period (date of the analysis). The subset of floating rate mortgages with typical principal repayment plans represents, for UniCredit Family Financing, the largest sample, encompassing 76.52% of the observations of live and discharged mortgages. Fixed rate mortgages with typical plans represent 23.41% of the sample, while only 0.07% falls into the category of mortgages with atypical principal repayment plans. The following paragraphs illustrate the results of the analyses conducted on two subsets for UniCredit Family Financing.

6.2 Floating rate mortgages

As already mentioned, for each explanatory variable in the model, a preliminary analysis was conducted in order to verify its significance through the comparison of nested models and hypothesis testing on the parameter values. The objective of the variables selection process, for polytomic variables, is the selection of the significant modes assumed by the variables; for the other variables, the objective is their exclusion, if not significant, in order to construct the aggregate model.
Next paragraphs show an example regarding the variable selection and the process for the model building.

### 6.2.1 Variable selection: an example

In this paragraph, the age at disbursement variable is taking into consideration as an example of variable selection. The figure 7 shows the modes assumed by the age at disbursement variable and the relative percentage frequencies of the sample. The variables selection process for this variable is shown in figure 8, while the results of the tests are shown in figure 9.

The process can be summarised as follows:

- **Step A**: shows the estimate of the model with only the explanatory variable age at disbursement, open on the 10 above-indicated modes. As can be seen in figure 8, for age brackets [41, 45] and [46, 50] the null hypothesis of equality to zero of the parameter is accepted for a significance level of 1% (P-Value of the test t > 1%). Given the non-contiguity of the modes compared to the reference mode, represented by age bracket [0, 20], the aggregability of the brackets on the intercept is not tested, but an equality test is performed on the value of the parameter between the two brackets.

- **Step B**: shows the estimated model under the null hypothesis of equality of coefficients of brackets [41, 45] and [46, 50]. The outcome of the LR test to verify the null hypothesis, against the alternative hypothesis represented by the initial model, is shown in figure 9: the test is passed at 1% confidence level (P-Value LR Test > 1%); passing this test leads to the aggregation of modes [41, 45] and [46, 50] into a single bracket [41, 50], as shown in Figure 7.

- **Step C**: shows the estimated model under the null hypothesis of equality of coefficients of brackets [61, 65] and 'Over 65', together with the aggregation of modes [41, 45] and [46, 50] into bracket [41, 50]. The
outcome of the LR test to verify the null hypothesis, against the alternative hypothesis represented by the initial model, is shown in 9: the test is passed at a 1% confidence level; passing this test leads to the aggregation of modes [61, 65] and ‘Over 65’ into a single bracket ‘Over 60’, as shown in figure 8.

- **Step D**: the further constraint of equality of the coefficients of modes [56, 60] and ‘Over 60’ is added to the model estimated in the previous step. Again in this case, the test is passed, leading to the aggregation of the modes into a single bracket ‘Over 55’.

- **Step E**: a further hypothesis of equality of coefficients of modes [21, 25] and [26, 30] is added to the constraint imposed in Step D. The p-value of the likelihood ratio test leads, again in this case, to accepting the null hypothesis at a significance level of 1%. Modes [21, 25] and [26, 30] are therefore merged into a single bracket [21, 30].

- **Step F**: a further hypothesis of equality of the parameters between brackets [21, 30] and [31, 35] is tested; in this case the test is not passed at a level of significance of 1%, the modes are not aggregated and will enter the model building stage as separate brackets.
6 THE RESULTS

Figure 9: Results of modes selection process for the age at disbursement variable

<table>
<thead>
<tr>
<th>Age at disbursement</th>
<th>Frequency %</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,20]</td>
<td>0.73%</td>
</tr>
<tr>
<td>[21,30]</td>
<td>22.31%</td>
</tr>
<tr>
<td>[31,35]</td>
<td>21.40%</td>
</tr>
<tr>
<td>[36,40]</td>
<td>19.36%</td>
</tr>
<tr>
<td>[41,50]</td>
<td>22.95%</td>
</tr>
<tr>
<td>[51,55]</td>
<td>6.40%</td>
</tr>
<tr>
<td>Over 55</td>
<td>7.89%</td>
</tr>
</tbody>
</table>

Figure 10: Age at disbursement variable

The variables selection process for the age covariate concludes at Step E, bringing the interval of the variable to the modes shown in figure 10.

6.2.2 Model Building

Figure 11: Initial aggregate model at the model building stage, Floating Rate Mortgages

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0430</td>
</tr>
<tr>
<td>Age Bracket [21,30]</td>
<td>0.3570</td>
</tr>
<tr>
<td>Age Bracket [31,35]</td>
<td>0.3180</td>
</tr>
<tr>
<td>Age Bracket [36,40]</td>
<td>0.2890</td>
</tr>
<tr>
<td>Age Bracket [41,50]</td>
<td>0.3270</td>
</tr>
<tr>
<td>Age Bracket [51,55]</td>
<td>0.2060</td>
</tr>
<tr>
<td>Maturity Class [11,15]</td>
<td>0.1090</td>
</tr>
<tr>
<td>Maturity Class [16,20]</td>
<td>0.2160</td>
</tr>
<tr>
<td>Maturity Class [21,25]</td>
<td>0.3010</td>
</tr>
<tr>
<td>Maturity Class Over 25</td>
<td>0.5440</td>
</tr>
<tr>
<td>Foreign Nationality</td>
<td>0.1900</td>
</tr>
<tr>
<td>Business Activity Class 2</td>
<td>0.1410</td>
</tr>
<tr>
<td>Business Activity Class 3</td>
<td>0.1220</td>
</tr>
<tr>
<td>Business Activity Class 7</td>
<td>0.1340</td>
</tr>
<tr>
<td>Business Activity Class B</td>
<td>0.1090</td>
</tr>
<tr>
<td>SAE 2 Class</td>
<td>0.1670</td>
</tr>
<tr>
<td>Log(scale)</td>
<td>0.7090</td>
</tr>
</tbody>
</table>

Once the variables selection stage has been concluded, the aim of the next stage of model building is to select the significant variables and modes to
be used to construct the final model. This process entails the comparison of nested models via backward elimination; the selection criteria are, again in this case, the likelihood ratio test with a significance level of 1%.

![Figure 12: Results of tests of the model building process, Floating Rate Mortgages](image)

Figure 12 shows the result of the estimation of the initial aggregate model, namely the initial model used for the selection of the variables and the modes that will constitute the model for floating rate mortgages with typical plans of UniCredit Family Financing. This model contains the variables and the modes as selected at the preliminary stage of variables selection described above; the model building process is similar to that used at previous stage and is described in detail for the Age of the borrower at disbursement variable.

![Figure 13: Final model, Floating Rate Mortgages](image)

The steps of the process and the results of the tests conducted are summarised in figure 12, while figure 13 shows the final model estimated for UniCredit Family Financing. As can be seen from figure 12, the tests conducted on the Age at disbursement variable lead to a further aggregation, with respect to that already performed at the preliminary variables selection stage, of modes [51, 55] and 'Over 55' and lead to the acceptance of the hypothesis of equality between these modes and the age bracket [31, 35] (steps 1 to 3). At step 4, the outcome of the test leads, on the other hand,
to the rejection of the hypothesis of equality of coefficients of modes [36, 40] and [41, 50], which will therefore be separate modes in the final model. Lastly, Steps 5 and 6 lead to the acceptance of the hypothesis of equality of coefficients of business activity classes 2, 3 and 7. The final model estimated for UniCredit Family Financing, floating rate mortgages with typical plans subset, is shown in Table 13.

6.3 Fixed rate mortgages

<table>
<thead>
<tr>
<th>INITIAL AGGREGATE MODEL</th>
<th>Parameter Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.9881</td>
<td>0.0000</td>
</tr>
<tr>
<td>CIncentive</td>
<td>-0.0165</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age Bracket [21, 25]</td>
<td>-0.3337</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age Bracket [26, 30]</td>
<td>-0.4413</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age Bracket [31, 35]</td>
<td>-0.3761</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age Bracket [36, 65]</td>
<td>-0.2646</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age Bracket Oltra 65</td>
<td>-0.2146</td>
<td>0.0066</td>
</tr>
<tr>
<td>Maturity Class [11,15]</td>
<td>-0.0640</td>
<td>0.0000</td>
</tr>
<tr>
<td>Maturity Class [16,20]</td>
<td>0.5094</td>
<td>0.0000</td>
</tr>
<tr>
<td>Maturity Class [21,25]</td>
<td>0.4333</td>
<td>0.0000</td>
</tr>
<tr>
<td>Maturity Class Oltra 25</td>
<td>0.6411</td>
<td>0.0000</td>
</tr>
<tr>
<td>Foreign Nationality</td>
<td>-0.3662</td>
<td>0.0000</td>
</tr>
<tr>
<td>Business Activity Class 2</td>
<td>-0.1066</td>
<td>0.0000</td>
</tr>
<tr>
<td>Business Activity Class 6</td>
<td>-0.0712</td>
<td>0.0000</td>
</tr>
<tr>
<td>Business Activity Class 8</td>
<td>0.2241</td>
<td>0.0000</td>
</tr>
<tr>
<td>Log(sigma)</td>
<td>-0.0083</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Figure 14: Initial aggregate model for the model building stage, Fixed Rate Mortgages

<table>
<thead>
<tr>
<th>Step</th>
<th>Equality test on parameters for Age brackets [36, 65] and Oltra 65</th>
<th>(Intercept)</th>
<th>Log likelihood</th>
<th>Number of Parameters</th>
<th>P-value</th>
<th>Test Result (Confidence 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Model</td>
<td>-57835.00</td>
<td>70</td>
<td>Passed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Constrasted Model 1</td>
<td>-57834.97</td>
<td>69</td>
<td>0.860927</td>
<td>Passed</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Equality test on parameters for Age brackets [21, 25] and [31, 35]</td>
<td>Constrasted Model 2</td>
<td>-57835.99</td>
<td>70</td>
<td>0.101207</td>
<td>Passed</td>
</tr>
<tr>
<td>4</td>
<td>Equality test on parameters for Business Activity classes 2 and 6</td>
<td>Constrasted Model 3</td>
<td>-57836.83</td>
<td>72</td>
<td>0.024405</td>
<td>Passed</td>
</tr>
</tbody>
</table>

Figure 15: Results of tests of the model building process, Fixed Rate Mortgages

The explanatory variables analysed for the fixed rate mortgages subset correspond to those used for the prepayment model for floating rate mortgages described above, with the addition of the quantitative variable Coupon Incentive. The stages of the analysis repeat the variables selection and model building processes described in the previous paragraph for the model developed for floating rate mortgages.

Figure 14 shows the initial aggregate model, used in the model building process for the fixed rate mortgages of UniCredit Family Financing. This model contains the variables and the modes, selected at the preliminary stage of variables selection.

The steps of the process and the results of the tests conducted are summarised in figure 15.
7 CONCLUSIONS

Figure 16: Final model, Fixed Rate mortgages

- Step 1: the null hypothesis of equality of the coefficients between the Age at disbursement brackets [36, 65] and 'Over 65' is tested against the alternative hypothesis represented by the initial aggregate model shown in 15. The null hypothesis is accepted at a significance level of 1% and the modes are further aggregated into the 'Over 35' bracket.

- Step 2: a further restriction of equality of coefficients of non contiguous modes [21, 25] e [31, 35] is added to the constraint imposed by step 1. Again in this case, the P-Value of the LR test leads to the acceptance of the null hypothesis; in this case, as the modes are not adjacent to an ordinal variable, the classes cannot be aggregated into a single mode, so will constitute separate classes with a common coefficient. The estimated model in this step ('Constrained Model 2' in figure 15), represents the alternative hypothesis for the test conducted in the next step.

- Step 3: the null hypothesis of equality of coefficients between Business Activity classes 2 and 6 is tested; the test was passed at a 1% confidence level.

The final estimated model for fixed rate mortgages with typical plans subset, is shown in figure 16.

7 Conclusions

This paper has given a synthesis about risks connected to prepayment and reviewed different approaches within the survival analysis. Among them, we used an accelerated life model with a log-logistic distribution to measure the prepayment risk for UniCredit Italian customers. This type of model enables prior knowledge regarding the relationship between prepayment probability
and the age of the loan. Moreover, the form of the chosen distribution is consistent with the empirical evidences of prepayment rate: low at the beginning of the contract, increasing during the life of the loan, then decreasing when maturity date approaches.

The results show the important role economic variables play, along with individual-specific characteristics, in determining customers prepayment behaviours. In particular, the coupon incentive is highly significant in the estimation of fixed rate mortgages: it measures the refinancing incentive when the interest rate at which the mortgage may be refinanced is lower than the contractual one. For both floating rate mortgages and fixed rate mortgages, the outcome of the analysis demonstrates the highly consistency of the variables original maturity and age of borrower, in explaining the prepayment phenomenon.

The usage of a prepayment model, correctly tested and frequently updated, could allow to have benefits in terms of:

- reducing cost of funding deriving from reallocation the medium/long term assets subject to prepayment on short/medium term buckets,
- more competitiveness in the pricing policy for customer deriving from lower funding costs.

Such advantages have an immediate effect in the Balance Sheet and in the Income Statement of the Bank.

The future extension of the model to other type of products (i.e. personal loans) and customers (i.e. corporate and small business) will help to better understand the prepayment phenomenon and will improve benefits for the bank.

References


REFERENCES


