American Journal of Applied Sciences 9 (9): 1337-1346, 2012
ISSN 1546-9239

A Discrete-Time Hazard Model for Loans: Some Evidence from Italian Banking System

Giambona Francesca
Department of Statistical and Mathematical Sciences,
University of Palermo, Viale delle Scienze, Palermo, Italy

Abstract: Problem statement: The probability of default, PD, is a crucial problem for banks. In the last years international accords (Basel, Basel 2 and Basel 3) have incentived banks to adopt objectives systems to evaluating and monitoring risk of default in order to predict PD for new loans based on borrower's characteristics. The aim of this study is to introduce a discrete survival model to study the risk of default and to propose the empirical evidence by the Italian banking system. **Approach:** Survival analysis is used if we are interested in whether and when an event occurs. In this context the event occurrence represents a borrower's transition from one state, loan in bonis that is not in default, to another state, the default. In this study through a survival model (in particular a discrete-time hazard model) it is possible verify when the probability of default is the highest considering, for each group of loans, a set of explanatory variables as risk factors of PD. **Results:** The empirical application obtained through a discrete time hazard model have provided clear evidence that time when the default occurs is an important element to predict the probability of default in time. Regarding Italian data the hazard model shows that explanatory variables (i.e., territorial area, productive economic sector, size of loan and generation of belonging) have effects both on if and on when loan bankrupts. **Conclusion:** The hazard model estimated for a population of loans involve different probability of default considering conjointly the explanatory variables and the time when the default occurs. Considering jointly the time and the risk factors a probability of default has been modelled for two main groups of loans: "Good borrowers" for which the risk of default is the lowest and "bad borrowers" for which this risk is the highest.

Key words: Credit scoring, hazard models, probability of default

INTRODUCTION

During the second half of the Nineties, banks have developed credit risk models to measure the potential loss, with a predetermined confidence level, that a portfolio of credit exposures could suffer within a specified time horizon, generally one year (BIS, 2004; 2011).

It is very important for banks to predict the probability of default for a homogeneous group of loans: the probability of default may be affected by some borrower's characteristics and losses on any single loan will not cause a bank to become insolvent. .

Borrower's characteristics (individual and social/economic conditions) have effects on default as well as the macro-economic and business cycle. Lenders in rich countries score potential borrowers based on a comprehensive credit history.

Banks should be able to attribute a default score for each potential borrower. This score is better if this is a reliable synthesis of the borrower's characteristics that influence the capacity of reimbursement.

In the last ten years banks are been encouraged to introduce standardized methodologies on monitoring and assessing the risk of default.

Credit scoring is a suitable objective model to evaluate the risk of default. This is a multivariate statistical model that examines the different borrower's characteristics attributing a different weight to explanatory variables on risk of default reaching a Probability of Default (PD) for each loan.

The purpose of credit scoring is to identify the characteristics that effect the insolvency of loan and to quantify the expected loss. In this study it is introduced a model to quantify PD proposing a credit scoring model that, also, introduce the time when the default occurs. The purpose is obtained by using discrete-time hazard model, that is a tool well known in social and, recently, in economic sciences also.

Thus, a discrete time hazard model for a population of loans assess the evaluation of PD considering, conjointly, the effects of explanatory variables, or risk factors and the time when a default occurs. This is useful, for banks, to predict a suitable PD.

The study uses a discrete time hazard model (in particular a non proportional hazard model) to evaluate the PD for a population of loans granted by Italian banks in a certain period. Some cohorts of loans have been selected and for these the characteristics of borrowers have been taken jointly to the time when the default occurr.

The following application shows the usefulness of this approach in the phases of evaluating and monitoring PD involving the time variable in credit scoring model.

MATERIALS AND METHODS

During the past twenty years marked progress has been made to measure credit risk. Most approaches involve the estimation of three parameters: the probability of default on individual loans or pools of transactions (PD), the estimation of the Losses-Given-Default (LGD) and the correlation between defaults (Crouhy *et al*., 2000; Duffie and Singleton, 2003).

The most common model to measure PD in credit risk measurement methodology is credit scoring analysis. A credit scoring model is a formula that puts weight on different characteristics of a borrower, lender and loan.

Credit scoring models are commonly structured along the lines of Altman (1968) Z-score model using historical loan and borrower data to identify which borrower characteristics are able to distinguish between defaulted and non defaulted loans. Based on the estimated of credit scoring, a credit score can be calculated for each new loan where a higher score indicates better expected performance of the borrower and thus a lower PD.

There are five methodological forms of multivariate credit scoring models: (1) the linear probability model, (2) the logit model, (3) the probit model and (4) the multiple discriminant analysis model, (5) decision trees.

The logistic regression technique overcomes this problem by directly estimating this probability and has therefore been the methodology of choice for retail credits. This technique assumes the existence of a continuous variable Z_j which is defined as the probability that a loan *j* defaults and can be modelled as a linear function of a set of variables x_j which describe the loan (Eq. 1):

$$Z_j = W'x = w_1 x_{j1} + w_2 x_{j2} + ... + w_k x_{jk} \quad (1)$$

where, w_k is the coefficient of the k_{th} variable and x_{jk} is the value of variable k for applicant j. Z_j is known as Z-Score of the j_{th} applicant. Z_j is ex-ante unobservable and default can only be defined ex-post as a 0-1 dummy (RFI, 1989; Altman *et al*., 1977; Lewis, 1990; Malik and Thomas, 2010; Tong *et al*., 2012; Thomas, 2000; Viganò, 1993).

The probability of default, π, is derived by using an iterative maximum likelihood estimation method as in a logistic regression model (Eq. 2) (Fahrmeir and Tutz, 1994; Hosmer and Lemeshow, 2000):

$$\pi_j = \frac{1}{1+\exp(-(W'x))} \quad (2)$$

Here, larger values of π reflect a higher PD.

Credit scoring models are relatively inexpensive to implement and do not suffer from the subjectivity and inconsistency of expert systems (used in the past).

But credit scoring does not consider the time when the default occurs. An approach in this sense is the mortality rate introduced by RFI (1989) used in different applications (Altman *et al*., 2001; 2004; Altman and Saunders, 1998; Altman and Suggitt, 2000; Dermine and Carvalho, 2006).

In this study, according to this approach, a survival model has been specified to measure the PD.

Researchers use the survival analysis in a variety of contexts that share a common characteristic: interest centers on describing whether or when event occurs. Time can be measured in years, months, days or seconds; the choice depends on the data

In this context the event occurrence represents a borrower's transition from one state, loan in bonis that is not in default, to another state, the default.

In survival analysis it is necessary to identify: (i) the target event, the occurrence event that represents an individual's transition from one state of interest to another; (ii) an initial starting point when no under study has yet experienced the target event (beginning time) and (iii) an appropriate metric for time in which an event occurrence is recorded.

The fundamental tool for summarizing the sample distribution of event occurrence is the life table that tracks the event histories of a population from the beginning of time (when no one has yet experienced the target event) to the end period considered. When the individual experiences the target event (or is censored) in one time period, he drops out of the risk set in all future time period. The life table provides information about the hazard rate, the survival function and the cumulative hazard rate.

The hazard rate, $h(t_j)$, is calculated as ratio between the number of occurrence event in a some period and the population at risk during the same period (the

population at risk is composed by the individual that are not yet experienced the target event).

The survivor function, $S(t_i)$, provides another way of describing the distribution of event occurrence over time. Unlike the hazard function, which assesses a unique risk associated with each time period, the survivor function cumulates these period-by-period risks of event occurrence together to assess the probability that a randomly selected individual will survive. It is defined as the probability that an individual will survive past some past time period (Klein and Moeschberger, 1997).

Thus if the interest is upon the risk factors that influence the probability that the target event occurs, it is necessary to specify a statistical model to control the effects of explanatory variable. In this case, at a given time point *t*, the hazard is the probability of experiencing the event of interest at time t conditional on being still at risk and on the value of the covariates (Eq. 3):

$$h_i(t \mid x_{it}) = P(T_i = t \mid T_i \geq t, x_{it}) \quad (3)$$

where, the vector x_{it} includes all the covariates of subject *i* at time *t*. The covariates can be time-invariant or time-varying. Time-varying covariates are extremely useful in building a proper model for the hazard, but they are rarely available in practice.

Since the hazard function is bounded between 0 and 1, a linear model for the hazard itself is not suitable, but one can apply a linear model to an appropriate transformation of the hazard (Eq. 4):

$$g(h_i(t \mid x_{it})) = \alpha_t + x_{it}'\beta \quad (4)$$

where, the transformation g, called link function, maps the (0,1) interval onto the real line.

On the right-hand side, β is the vector of regression coefficients and $(a_1,a_2,\ldots a_3)$ are time-specific intercepts representing the baseline hazard, i.e., the hazard for the hypothetical subject with all the covariates set to zero. The number of time-specific intercepts is P, the maximum number of time points (intervals) in the data.

Therefore, using the time indicators (α_i) as well as explanatory variables (x_i), each intercept parameters represents the value of logit hazard (the log odds of event occurrence) in that particular time period for individuals in the baseline group; each slope parameter assesses the effect of a one unit difference in that predictor on event occurrence, statistically controlling for the effects of all other predictors in the model.

When the link g (.) is the logit function $\log\left(\frac{x}{1-x}\right), x \in (0,1)$ the corresponding model is called logit or proportional odds (Eq. 5):

$$\log\left(\frac{h_i(t \mid x_{it})}{1-h_i(t \mid x_{it})}\right) = \alpha_t + x_{it}'\beta \quad (5)$$

Or, in terms of the hazard function (Eq. 6):

$$h_i(t \mid x_{it}) = \frac{1}{1+\exp(-\alpha_t - x_{it}'\beta)} \quad (6)$$

The interpretation of the regression coefficients requires some care, since β_k is the change in the logit of the hazard following a unit increase in the k-th covariate (Collett, 2003a; 2003b).

A key feature of survival analysis is the study of the dynamics of the covariates' effects. In fact, a time-invariant covariate may have a time-varying effect.

As in logistic regression it is rare to interpret the parameters estimated. More commonly, the odds ratio is defined as is the odds of an event occurring in one group to the odds of it occurring in another group, or to a sample-based estimate of that ratio (Eq. 7). An odds ratio of 1 indicates that the condition or event under study is equal in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group. And an odds ratio less than 1 indicates that the condition or event is less likely in the first group. The odds ratio must be greater than or equal to zero. When the odds of the first group approaches zero, the odds ratio approaches zero. When the odds of the second group approaches zero, the odds ratio approaches positive infinity:

$$OR = \exp(\beta) \quad (7)$$

A discrete time proportional hazard model shape is the same for all explanatory variables and the distance between each logit hazard functions is identical in every time period; the effect of explanatory variables on the log odds of event occurrence is hypothesized to be constant over time.

This is a restrictive assumption (the proportional assumption is violated in many social and economic phenomenon) that is possible to relax it by including interactions with time (time-dependent or duration-dependent effects).

When the effect of covariates is not proportional in time, it is necessary to adapt a non proportional hazard model. To represent adequately a time varying effect, there are different kinds of time interaction models (Singer and Willett, 2003; 1993). A parsimonious

model, considering the change of these effects, is to adapt a discrete hazard model where βi assesses the effect of Xi in time period c (for example in the first period) and γi describes how this effect linearly increases (if γ_i is positive) or decreases (if γ_i is negative) across time periods (Eq. 8):

$$\log\left[\frac{h_i(t \mid x_{it})}{1-h_i(t \mid x_{it})}\right] = \log it\left[h_i(t \mid x_{it})\right] = \alpha_t + \beta_i x_{it} + \gamma_i x_{it}.(t-c) \quad (8)$$

The interpretation of time indicators is identical to a proportional odds model.

By comparing deviance statistics for this model from the main effects model in (1), it is possible to test the null hypothesis that the effect of explanatory variables does not differ linearly over time.

Estimation can be carried out using standard software for binary response models. In fact, the likelihood of a discrete-time survival model on the original dataset is the same as the likelihood of a binary response model on the person-period dataset.

To obtain the person-period dataset, each original record i is replicated as many times as the observed time t_i and the new response variable is the indicator of the event of interest (For example, the record of a subject experiencing the event of interest at time 5, it is replicated 5 times and the values of the new response variable are (0,0,0,0,1). Also for a subject censored at time 5, the record is replicated 5 times, but the values of the new response variable are (0,0,0,0,0)). Finally, it is possible to calculate the cumulative hazard function that assess, at each point time, the total amount of accumulated risk that an individual has faced from the beginning of time until the present period (Eq. 9):

$$H(t) = \sum_{j=1}^{T} h(t_j) \quad (9)$$

Some authors (Cox and Oakes, 1984) prefer to define the cumulative hazard for discrete time as (Eq. 10):

$$H(t) = \sum_{j=1}^{T} In\left[1-h(t_j)\right] \quad (10)$$

It is useful to note that cumulative hazard is not a probability but is a rate.

RESULTS

The database used in this study is provided by Italian Central Bank.

It consists of 1.302.186 borrowers followed for the first 10 years after the grant of the loan.

For these borrowers, some characteristics as possible explanatory variables or risk factors for the default have beenc selected.

The explanatory variables include:

- Territorial area (T): Northern regions, Central regions, Southern regions
- Productive economic sector (S): producer families and firms
- Size of the loan (A): less than € 250.000, more than € 250.000
- Generation of loan or cohort (C): loans granted between 1985 and 1995

The last explanatory variables have been grouped in two main categories:

- 1985-1993: all loans granted before 1993
- 1994-1995all loans granted between 1994 and 1995

This categorization has been made because in 1993 in Italy has been introduced the law named "Testo Unico bancario" (Decreto Legislativo 1/9/1993, n. 385) to regulate the Italian banking system. Thus, in this study, the population of loans has been divided into two main categories to analyse the risk profile of loans before and after this regulation.

Table 1 reports the size of loan defaulted "dead" and in bonis "survivors" differentiated for the categories of explanatory variables considered.

The life table (Table 2) shows the elimination of loans in the period considered.

The life table shows how the loans in time (first ten years) dead.

The empirical hazard rate has a decreasing evolution and it shows that at the end of the period about 10 per cent of borrowers are defaulted.

Considering the classification of loans by territorial area the values of survivor function show a different evolution.

In Fig. 1 is shown the survivor function by area. Immediately we note that for Northern regions survivor function has higher values in comparison with the other regions. This is a first result that point out differences in PD on Italian Banking System.

To specify a discrete- time hazard model, loans have been tracked for 10 years in order to study the survival function in the first ten years from their origination.

Remember that a loan is censored when, in the period of study, it is not in default or it goes out of the study to verify an event different than default. Principally a loan is censored when: (1) it is in bonis, so it survives, (2) it has been repaid.

A first measurement used to describe the process of elimination of a generation of loans is the empirical hazard rate, obtained by relating the loan that in a certain period is defaulted to loans survived, to define, if and when the default occurs (Fig. 2).

The hazard rate shows a decreasing evolution of the default, but it has been calculated considering the population of loans as homogenous (the borrowers were not distinguished on the basis of the explanatory variables).

Now, considering that the population is heterogeneous (observed heterogeneity) each borrower will have different values on observed predictors. After defining a reference category (baseline), it is used a non proportional hazard model to estimate the influence of risk factors and time indicators variables on PD jointly.

Time indicator variables, or intercept parameters (α_t), represent the value of hazard in a particular time period for individual in the baseline group.

Table 1: Distribution of loans survived and defaulted, loans granted between 1985-1995

	Population of loans		
Regions	Survivors	Dead	Total
North	701,018	51,197	752,215
Centre	240,621	31,398	272,019
South	237,459	40,493	277,952
Total	1,179,098	123,088	1,302,186
Size of loan			
< 125.000 euro	1,032,495	95,417	1,127,912
> 125.000 euro	146,603	27,671	174,274
Total	1,179,098	123,088	1,302,186
Institutional sector			
Producer families	774,584	67,696	842,280
Firms	404,514	55,392	459,906
Total	1,179,098	123,088	1,302,186
Generation of loans			
1985-1993	831,600	99,041	930,641
1994-1995	347,498	24,047	371,545
Total	1,179,098	123,088	1,302,186

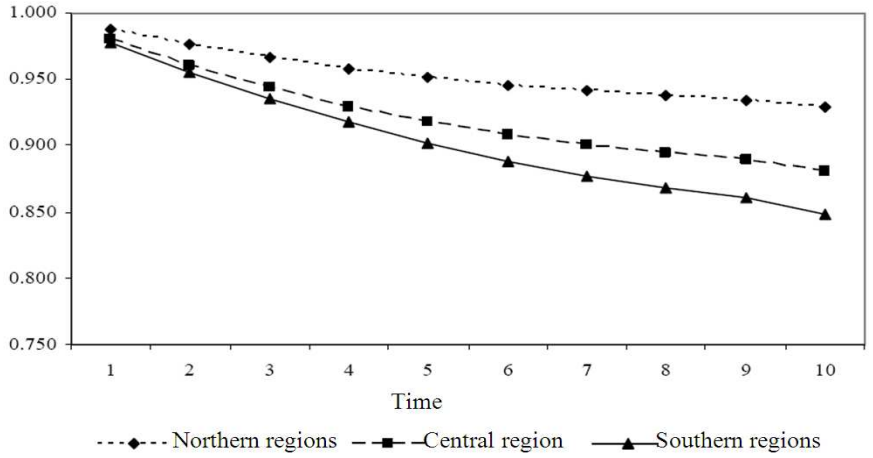

Fig. 1: Survivor function by area, loans granted between 1985-1995

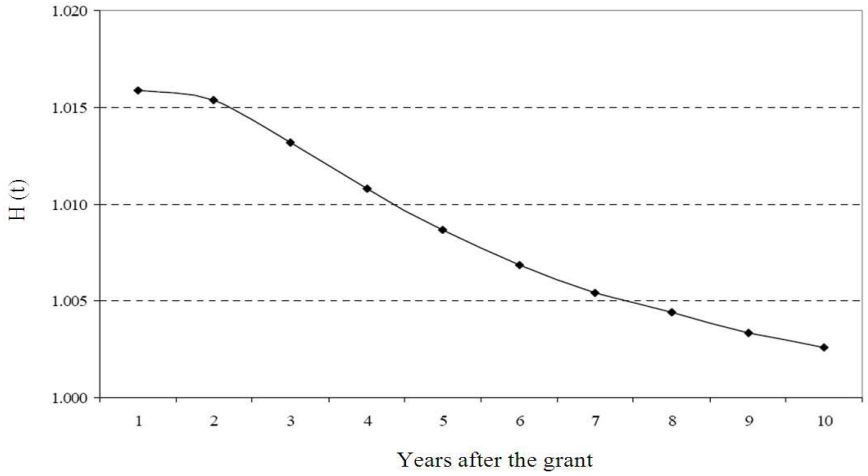

Fig. 2: Hazard function, loans granted between 1985-1995

Table 2: Life table, loans granted between 1985-1995

Time interval		Population at risk	Hazard rate	Cumulative Hazard	Survival function
0	1	1,302,186	0.0160	0.0159	0.9841
1	2	1,281,523	0.0159	0.0313	0.9687
2	3	1,261,363	0.0139	0.0447	0.9553
3	4	1,243,929	0.0124	0.0565	0.9435
4	5	1,228,639	0.0105	0.0664	0.9336
5	6	1,215,763	0.0087	0.0745	0.9255
6	7	1,205,180	0.0072	0.0811	0.9189
7	8	1,196,538	0.0061	0.0867	0.9133
8	9	1,189,304	0.0048	0.0911	0.9089
9	10	1,183,607	0.0076	0.0980	0.9020

The slope parameter (β_i) assess the effect of a unitary difference in a particular explanatory variable on event occurrence, statistically controlling for the effect of all other in the model.

The original dataset of 1.302.186 observations, since some explanatory variables have identical values, has been transformed and the observations have been groped into 3.374 main profiles.

To estimate a time discrete hazard model has been necessary to transform the dataset in a person-period matrix. Subsequently this dataset composed by 3.374 profiles has been transformed into a person-period dataset of 19.865 observations.

A logit regression has been used to regress the event indicator (the default) on the time indicators and on the selected explanatory variables in the person-period dataset.

Explanatory variables introduced in the model are dichotomous (productive sector, size of the loan and generation of the loan) and polytomous (territorial area), in order to study the effects of these in the probability of default (which determine the drop out of the cohort).

The loans have been shared in homogeneous groups. A score is attributable not only in relationship to the "risk factors" that cause higher value of PD, but also considering the years in which the loan could enter in default. This model is able to attribute, to a new loan, a diversified score for each year (survival score) that is the predictors variables of default and the year in which the default occurs.

The baseline has been selected with reference to the loan with the lowest PD value; in such way it is possible to define two bound profiles in terms of PD values, the lowest and the highest, inside which are included all possible combinations of risk factors in the selected period (in this case ten years).

The baseline is identified with a producer family, in Northern regions, for a size of the loan < 125.000 euro that has request a loan between 1994 and 1995.

Besides, the weight matrix, W, has been constructed to attribute to every period a weight equal to the number of default loan n_i, to the generic period j.

Table 3: Discrete time survival model, probability of default (baseline omitted1)

		Coeff.	St. dev.
Time indicators variables (αt)	T1	-4,6033	0,0124
	T2	-4,7514	0,0111
	T3	-5,0257	0,0107
	T4	-5,2911	0,0109
	T5	-5,6026	0,0120
	T6	-5,9441	0,0139
	T7	-6,2977	0,0163
	T8	-6,6314	0,0191
	T9	-7,0316	0,0224
	T10	-7,4319	0,0260
Explanatory variables (β_i)	Central regions	0,5191	0,0114
	Southern regions	0,6501	0,0107
	Size of the Loan > 125.000 euro	0,4389	0,0119
	Cohort 1985-1992	0,1263	0,0110
	Firms	0,0627	0,0098
Interactions(γi)	Central Regions * (T-1)0,003	0,0028	
	South Regions * (T-1) 0,049	0,0026	
	Size of the Loan > 125 * (T-1)	0,0439	0,0028
	Cohort 1985-1992 * (T-1)	0,1199	0,0030
	Firms *(T-1)	0,0649	0,0024

Table 4: Discrete time survival model, probability of default (*baseline* omitted1) measures of fit

Measures of fit			
Time-discrete proportional hazard model vs. Time discrete non proportional hazard model		Time-discrete non proportional hazard model	
Deviance		Log-Lik Intercept Only:	
- Time-discrete proportional hazatd model	1324126		-689303,967
Parameters	15		
		Log-Lik Full Model:	
- Time-discrete non proportional hazatd model	1320523		-666544,352
Parameters	20		
		LR Test	
X^2	c2 3601		45519,23
p-value	0,0000	Prob > LR:	0,0000

Table 5: Odds Ratios (OR)

Risk factors	Odds ratios
Central (/ Northern)	1, 6806 (+ 68 %)
Southern (/ Nordhern)	1, 9157 (+ 92 %)
Size of the Loan > 125.000 euro (/ < 125.000 euro)	1, 5511 (+ 55 %)
Cohort 1985-1992 (/Cohort 1993-1995)	1, 1346 (+ 13 %)
Firms (/Productive Households)	1, 0647 (+ 6 %)

First of all, it has been specified a proportional odds model, but the effects of explicative variables is not the same in all periods. So a non-proportional hazard model has been specified with a term of interaction.

In this study it is adopted the model (8) with c=1: β_i assesses the effect of the explanatory variables considered in the first time period and γi describes how this effect linearly increases (if γ_i is positive) or decreases (if γ_i is negative) across the follow time periods.

The non proportional hazard model specified is:

$$\log\left[\frac{h_i(t \mid T_{it}, S_{it}, A_{it}C_{it})}{1-h_i(t \mid G_{it}, S_{it}, A_{it}C_{it})}\right]$$
$$= \log it\left[h_i(t \mid T_{it}, S_{it}, A_{it}C_{it})\right] = \alpha_1 + ... + \alpha_{10}$$
$$++\beta_1 T_{it} + \beta_2 S_{it} + \beta_3 A_{it} + \beta_4 C_{it} +$$
$$+\gamma_1 T_{it}.(t-c) + \gamma_2 S_{it}.(t-c) + \gamma_3 A_{it}.$$
$$(t-c) + \gamma_4 C_{it}.(t-c)$$

where, T is territorial area, S is productive economic sector size of the loan A is size of loan and C is the generation of belonging of loan.
The results are shown in Table 3 with some measure of fit (Table 4).

Time indicator variables show that for the baseline the probability of default decreases over time and the explanatory variables are risk factors for default: a loan granted in Central/Southern regions, or to a firm, or in the year between 1993-1995, or for a size >125.000 euro, increases the probability of default.

The measures for fit confirm that the deviance for the non proportional odds hazard model is lower than the proportional odds model; thus the choice for the first model. Finally, the LR test shows that the explanatory variables in the model are risk factors for PD. Table 5 shows for each risk factor the corresponding OR.

DISCUSSION

Observing the results the loan cohort is an explanatory variable statistically significant: loans granted between 1985 and 1992 have an hazard of default higher than those granted between 1993 and 1995. The estimate confirms that the actions taken by Italian banks in order to decrease the default have produced some expected results.

Differences are found also with respect to the institutional sector of the borrower; the hazard is higher for firms than for a producer families (OR = 1, 06).

This latest estimate allows one consideration about the Basel Accord. Someone has underlined the difficulty of applying a scheme of its kind to Italy, persisting in its many small businesses, dependent on bank loans, that would have difficulties to pay higher costs than implementing a credit scoring model (Zadra, 2002). Indeed the estimates from the model 9 show that small firms (producers families) have a lower probability of default than other firms; this evidence resizes the preoccupations about the impact of the new Basel Accord on small businesses.

Territorial coefficients show higher values of hazard for borrowers in Central or Southern regions, especially for the latest. The OR are respectively 1.69 and 1.92; the PD for a loan in Southern regions is about 2 times more defaulted than a borrower in Northern regions, while for a borrower in Central regions this probability is 1.7 times higher.

The results confirm the dualistic Italian credit market: the defaults are more evident in the Southern than in Northern regions, where economic conditions promote the credit market.

The economic situation has been felt in the Southern regions where the financial system, which is strongly focused on banks, manages the entire savings of households (Cannari and Panetta, 2006).

The economy of the Southern regions showed negative differentials in terms of economic (and social) aspects than the Northern-Central regions, even with reference to the structure of the banking system.

This is confirmed by some empirical evidences: a lower GDP per capita; a higher degree of economic dependence by Northern regions or foreign countries; a less robust system of firms; a greater level of poverty among families; inadequate infrastructures (Mattesini and Messori, 2004).

The briefly mentioned aspects, that underline the fragility of Southern regions economy, have a feedback in credit market where the defaults are still too high as referred by the literature (Cusimano and Vassallo, 2007; Cusimano, 2006; Cannari and Panetta, 2006; Mattesini and Messori, 2004).

The explanatory variable related to the size of the loan shows that bigger loans have a higher PD (for loans more 125.000 euro, odds ratio = 1.55); this evidence conforms that higher PD are correlated to higher loan size.

The signs of logistic regression also confirm that a non-proportional odds model is more appropriate than a proportional odds model. The importance of the risk factors is not the same in the period considered (the first ten years), but the effects of these increase over time.

By the estimated coefficients, two differing groups of loans are easily defined. In the first group named "bad borrowers", all borrowers for which the PD is slower are included, while in the second group, named "good borrowers", all borrowers for which the PD is higher are included. The profiles are identified by the sign of the coefficients. Thus, "bad borrowers" have these characteristics: A loan granted between 1985 and 1992, to a producer family, in Central/Southern regions and for a size more than 125.000 euro. The "good borrowers" profile is the baseline.

By using the (6) expression it is derived the hazard function. The mentioned profiles are traced in Fig. 3. The profile "good borrowers" is associated with the estimated hazard h(G); "bad borrowers" have an estimated hazard h(B).

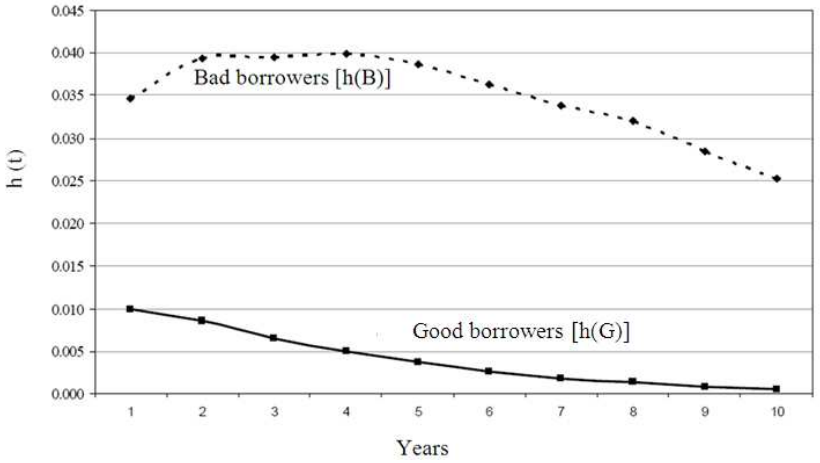

Fig. 3: "Good borrowers" and "bad borrowers", hazard in the first 10 years

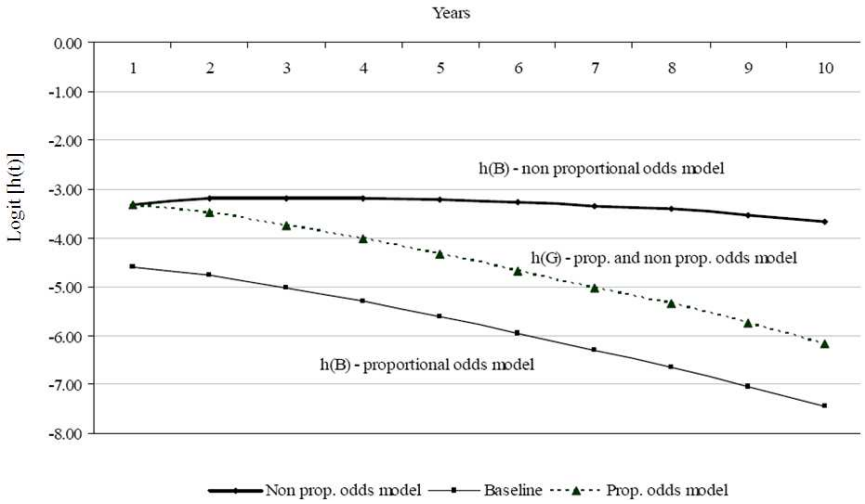

Fig. 4: Logit for "good borrowers" and "bad borrowers", comparison of models

Consequently, all other combinations of risk factors (i.e. explanatory variables) will have an hazard function between these two oppposite profiles. Figure 3 shows that the opposite profiles don't converge over time.

The different trend of the hazard after the first four years is also very interesting: hazard increase for "bad borrowers"; hazard decreases after the first year.
This evidence suggests differentiated bank policies related to borrowers' characteristics (already achieved in credit scoring models) and also by year of "loan life".

The results of the model and the graphic representation involve to attribute different score for different values of risk factor and different years.

For "good borrowers" it is desirable to assign a decreasing score already in its first year; for "bad borrowers" it is possible to attribute a decreasing score only after the fifth year.

Table 6 shows odds ratios for risk factors related to the "bad profile".

PD increases over time for all variables considered, thus the importance of risk factors is not the same in the period considered.

Considering, for example, its final years, a borrower in the Southern regions has about 3 times more than a borrower in the Northern regions; a borrower with a loan size more than 125.000 euro has a PD about 2.3 times higher than a borrower with less than 125.000 euro; for a company, the PD is about 2 times higher than for a producer; finally, loans granted between 1985 and 1992 have a PD 3 times higher than a credit disbursed between 1993 and 1995. Figure 4 displays that a non proportional hazard model is better; the effects of covariates increase in time.

Table 6: OR of explanatory variables in the first 10 years

Year after the beginning time	Territorial area (Southern Regions)	Loan size (>250 mila euro)	Sector (Firms)	Cohort (1985-1992)
1	1.916	1.551	1.065	1.135
2	2.012	1.621	1.136	1.279
3	2.113	1.693	1.212	1.442
4	2.219	1.769	1.294	1.626
5	2.330	1.849	1.380	1.833
6	2.447	1.932	1.473	2.066
7	2.570	2.018	1.572	2.330
8	2.699	2.109	1.677	2.626
9	2.835	2.203	1.790	2.961
10	2.977	2.302	1.910	3.338

Table 7: H (t) and H(t) in the first 10 years

	h(t)		h(t)	
Years	"bad borrowers" borrowers"	"good borrowers"	"bad borrowers"	"bad
1	0.04532	0.02004	0.04532	0.02004
2	0.05127	0.01728	0.09659	0.03732
3	0.05143	0.01313	0.14802	0.05045
4	0.05204	0.01007	0.20006	0.06052
5	0.05039	0.00738	0.25045	0.06790
6	0.04743	0.00524	0.29788	0.07314
7	0.04411	0.00368	0.34199	0.07682
8	0.04181	0.00264	0.38380	0.07946
9	0.03717	0.00177	0.42097	0.08123
10	0.03302	0.00118	0.45399	0.08241

This evidence confirms that the choice of a non proportional odds model is undoubtedly better than it necessary to introduce an additional term in the model which allows quantifying the direction of the variation (increasing/decreasing) and the intensity.

In Table 7 are showed h (t) and H(t) for each year relatively to the opposite profiles as introduced below. H(t) provide PD year by year. It has higher values for "bad borrowers": At the end of the period considered almost half of these become default (45%9; while for "good borrowers" only about the 10% is default.

CONCLUSION

The probability of default, PD, is a crucial problem for banks. In the last twenty years international accords, as Basel and the following Basel 2 and 3, have incentived banks to adopt objectives systems of evaluating and monitoring risk of default in order to predict PD for new loans based on borrower's characteristics. Literature confirms that credit scoring is the model utilised by banks.

In this study a revised version of credit scoring has been presented and a first application to Italian banking system has been reported. The time when the default occurs has been introduced in a credit scoring model by using a survival approach through a discrete time hazard model. It is used the dataset by Banca d'Italia and a non proportional odds model has been selected, in order to considering the variation explanatory variables effects in time, considered as risk factors for default. The discrete time non proportional hazard model has showed that PD is not constant over time and the explanatory variables considered (institutional sector, cohort of loan, territorial area and size of loan) are risk factors for default. Considering jointly the time and the risk factors a PD has been modelled for two main groups of loans: "good borrowers" for which the risk of default is the lowest and "bad borrowers" for which this risk is the highest. The last group of borrowers is identified with a loan granted between 1985 and 1992, to a producer family, in Central/Southern regions and for a size more than 125.000 euro. The "good borrowers" profile is the baseline. For "good borrowers" it is useful to assign a decreasing score already in the first year; for "bad borrowers" it is possible to attribute a decreasing score only after the fifth year. Results highlight that banks to improve the credit risk management should attribute a different score for categories of borrowers considering, jointly, the time.

REFERENCES

Altman, E., A. Resti and A. Sironi, 2004. Default recovery rates in credit risk modelling: A review of the literature and empirical evidence. Econ. Notes, 33: 183-208. DOI: 10.1111/j.0391-5026.2004.00129.x

Altman, E.I. and A. Saunders, 1998. Credit risk measurement: Developments over the last 20 years. J. Bank. Finan., 21: 1721-1742.

Altman, E.I. and H.J. Suggitt, 2000. Default rates in the syndicated bank loan market: A mortality analysis. J. Bank. Finan., 24: 229-253.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finan., 23: 589-609.

Altman, E.I., A. Resti and A. Sironi, 2001. Analyzing and explaining default recovery rates. The Pennsylvania State University.

Altman, E.I., R.G. Haldeman and P. Narayanan, 1977. ZETATM analysis A new model to identify bankruptcy risk of corporations. J. Bank. Finan., 1: 29-54.

BIS, 2004. International Convergence of Capital Measurement and Capital Standards: A Revised Framework. 1st Edn., Bank for International Settlements, ISBN-10: 9291316695, pp: 239.

BIS, 2011. Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements.

Cannari, L. and F. Panetta, 2006. Il Sistema Finanziario e il Mezzogiorno: Squilibri Strutturali e Divari Finanziari. 1st Edn., Cacucci, Bari, ISBN-10: 8884224942, pp: 398.

Collett, D., 2003a. Modelling Survival Data in Medical Research. 2nd Edn., Chapman and Hall, London, ISBN-10: 1584883251, pp: 391.

Collett, D., 2003b. Modelling Binary Data. 2nd Edn., Chapman and Hall, London, ISBN-10: 1584883243, pp: 387.

Cox, D.R. and D. Oakes, 1984. Analysis of Survival Data. 1st Edn., Chapman and Hall, London, ISBN-10: 041224490X, pp: 201.

Crouhy, M., D. Galai and R. Mark, 2000. A comparative analysis of current credit risk models. J. Bank. Finan., 24: 59-117.

Cusimano, G. and E. Vassallo, 2007. Caratterizzazioni territoriali nella distinzione dimensionale delle banche italiane. Minerva Bancaria.

Cusimano, G., 2006. Sul sistema bancario italiano trail 1999 ed il 2005. Minerva Bancaria.

Dermine, J. and C.N.D. Carvalho, 2006. Bank loan losses-given-default: A case study. J. Bank. Finan., 30: 1219-1243. DOI: 10.1016/j.jbankfin.2005.05.005

Duffie, D. and K.J. Singleton, 2003. Credit Risk: Pricing, Measurement and Management. 1st Edn., Princeton University Press, Princeton, ISBN-10: 0691090467, pp: 396.

Fahrmeir, L. and G. Tutz, 1994. Multivariate Statistical Modelling Based on Generalized Linear Models. 1st Edn., Springer-Verlag, New York, ISBN-10: 0387942335, pp: 425.

Hosmer, D.W. and S. Lemeshow, 2000. Applied Logistic Regression. 2nd Edn., John Wiley and Sons, New York, ISBN-10: 0471356328, pp: 373.

Klein, P.J. and M.L. Moeschberger 1997. Survival Analysis: Techniques for Censored and Truncated Data. 1st Edn., Springer, New York, ISBN-10: 0387948295, pp: 502.

Lewis, E.M., 1990. An Introduction to Credit Scoring. 1st Edn., Athena Press, San Rafael.

Malik, M. and L.C. Thomas, 2010. Modeling credit risk of portfolio of consumer loans. J. Operat. Res. Soc., 61: 411-420.

Mattesini, F. and M. Messori, 2004. L'evoluzione del Sistema Bancario Meridionale: Problemi Aperti e Possibili Soluzioni. 1st Edn., Il Mulino, Bologna, ISBN-10: 8815102809, pp: 265.

RFI, 1989. Default Risk, Mortality Rates and the Performance of Corporate Bonds. 1st Edn., Research Foundation of CFA Institute.

Singer, J.D. and J.B. Willett, 1993. It's about time: Using discrete-time survival analysis to study duration and the timing of events. J. Educ. Stat., 18: 155-195.

Singer, J.D. and J.B. Willett, 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. 1st Edn., Oxford University Press, New York, ISBN-10: 0195152964, pp: 644.

Thomas, L.C., 2000. A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. Int. J. Forecast., 16: 149-172.

Tong, E.N.C., C. Mues and L.C. Thomas, 2012. Mixture cure models in credit scoring: If and when borrowers default. Eur. J. Operat. Res., 218: 132-139. DOI: 10.1016/j.ejor.2011.10.007

Viganò, L., 1993. A credit scoring model for development banks: An african case study. Sav. Dev., 17: 441-482.

Zadra, G., 2002. Banche e piccole imprese: Per un nuovo rapporto dopo Basilea 2. Bancaria, 9: 2-7.