UNIVERSITÀ DEGLI STUDI
DI PERUGIA

ISSN 2385-2275

Working papers of the Department of Economics University of Perugia (IT)

Survival Models for Credit Risk Estimation in the context of SME

Alberto Burchi

Francesca Pierri

Working paper No. 18

December 2015

Survival Models for Credit Risk Estimation in the context of SME

Francesca Pierri
Department of Economics
University of Perugia
Via Pascoli, 06125 Perugia (PG) - Italy

Alberto Burchi
Department of Economics
University of Perugia
Via Pascoli, 06125 Perugia (PG) - Italy

December 2015

Abstract

The credit risk is the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms.

Using a large dataset of corporate balance sheets we develop a survival model to predict default.

Unlike previous works, we consider forecasts of probability of default for small corporate, we take into account regional macroeconomic conditions as well as national macroeconomic indicators and we use a large sample of balance sheet. We define our model after a series of steps designed to select only the significant variables; we examine the scale of the continuous covariates in the preliminary main effect model and we apply, when required, appropriate transformations to respect the linearity in the log hazard. Last but not least, we check the proportionality hazard assumption.

1 Introduction

The origins of modern banking can be traced to early Italian Renaissance. In the initial forms of banking business, banks receive surplus money from the people who are not using it and lend to those who need it for productive purpose. In more than six centuries, banks have changed their modus operandi but still must address the risks associated with this crucial role in the financial system. When these risks occur, they become effective losses. Consequently, the ability to distinguish good from bad borrowers is crucial for success in the credit industry. The recent financial crisis and regulatory innovations introduced by the Basel Committee highlight the central role of the credit risk assessment.

The credit risk is the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms (Basel, 2000). This risk has been the main character in banking, since the early days of the industry.

For conditions of optimal allocation of capital, the pricing of the loans has to reflect the risk. Credit scoring is a credit risk management technique that analyzes the borrowers risk. A good credit scoring model has to be selective: the more discriminative the scoring system, the better are the borrower ranked from high to low risk. Thus, the optimal allocation of capital is directly linked to appropriate credit risk models. The search for an approach for assessing credit risk has driven the interest of both academic and financial communities. The risk assessment takes the form of estimates of Probability of Default (PD) and Loss Given Default (LGD). Loss Given Default (LGD) is the loss incurred by a financial institution when a borrower defaults on a loan, corresponding to the fraction of exposure at default (EAD) unpaid after some period of time (Bellotti and Crook, 2012).

Logistic regression models are widely used in credit scoring context and they are generally known as static models. Usually, a retrospective data collection technique adopts for each company a set of explanatory variables observed several years before the event of interest (Pierri et al., 2015). However the characteristics of a business may change from year to year and its health is a function of its latest financial data, while the default probability that a static model assigns to a company does not vary with time.

Hazard models resolve the problem of static models by explicitly accounting for time. Classic examples of lifetime data can be found in medical, epidemiological or reliability studies (Stepanova and Thomas, 2002). However in the field of behavioral scoring survival analysis has already been adopted by several works (Banasik et al., 1999; Stepanova and Thomas, 2002; Andreeva, 2006; Bellotti and Crook, 2013).

In credit industry survival analysis may be considered a relatively new application that offers an advantage of predicting time to the event of interest and therefore lays the foundation of estimating the applicant's profitability (Banasik et al., 1999; Stepanova and Thomas, 2001)

Shumway (2001) states there are mainly three reason to prefer hazard models to static models for forecasting bankrupcy. When sampling periods are long, static models do not adjust for period at risk, that is they do not control when a company failed for bankruptcy. The second reason is that hazard models incorporate time varying covariates or explanatory variables that change with time and the third one is they may produce more efficient out-of-sample forecasts by utilizing much more data.

The purpose of this work is to propose an analytical approach aimed to implement the survival analysis in the context of SME. Using a large dataset of corporate balance sheets besides the Business Register dataset provided by the Chamber of Commerce of the Region of Umbria (Italy), we develop a survival model that exploits each firm's time series data by including annual observation as time varying covariates and incorporates some macroeconomic variables that are the same for all firms at a given point of time.

The novelties of our paper are that, unlike published work, we (1) consider forecasts of PD for small corporate, (2) we include both regional and national macroeconomic conditions in our model selection; (3) the analysis model is estimated after a series of steps designed to select only the significant variables and insert them in the best transformation to respect the linearity in the log hazard. Finally, (4) we test the usefulness of the Cox Proportional Hazard Model as a credit scoring system.

In Section 2, we present the literature review; in Section 3, we detail the applied methodology; in Section 4 we describe our data, the model estimation and the results; finally, in Section 5, we provide some conclusions.

2 Literature review

In the assessment of creditworthiness are both relevant the probability that the debtor fails to fulfill its obligations and when such adverse events may occur. Especially for medium and long term loans, the event of default can have a complexity that is difficult to be captured by models.

Survival models will predict not just the probability of whether an event will occur, and not limited to a pre-defined outcome period, but also the conditional probabilities of that event occurring over time. Although survival models can account for different types of events, via competing risks, they are based on the assumption that the risk of each event occurring is independent right up to when any event occurs, which might not necessarily hold true (Leow and Crook, 2014a).

The major strength of survival analysis is that it allows censored data to be incorporated into the model.

As stated by Stepanova and Thomas (2002), the idea of employing survival analysis for building credit-scoring models was first introduced by Narain (1992) and then developed further by Banasik et al. (1999). Narain (1992) applied the accelerated life exponential model to 24 months of loan data. The author showed that the survival analysis can add the time dimension to the traditional model of credit scoring. Also, the author noted that these methods can be applied to any area of credit operations in which there are predictor variables and the time of some event is of interest.

Banasik et al. (1999), comparing different approaches, found that survival-analysis methods are competitive with, and sometimes superior to, the traditional logistic-regression approach. The application of survival analysis models onto credit-related problems is welcomed for its ability to take into account factors that are inherent in the modelling of credit risk and the prediction of credit events, where regression methods are unable to. In particular, survival models are able to account for censoring. Second, they are able to incorporate time-dependent variables, allowing the inclusion of time-dependent account-specific covariates as well as time-dependent macroeconomic variables (Bellotti and Crook, 2009; Leow and Crook, 2014b).

Thomas et al. (2001) describe survival analysis as a means of building dynamic models, since this readily allows the inclusion of behavioral variables as time varying covariates. Bellotti and Crook (2009), following this path, use a Cox Proportional Hazards survival model to model the time to default for a large database of credit cards. This model has the advantage to naturally include macroeconomic time series data into the model as time varying covariates. They show that the inclusion of macroeconomic variables such as bank interest rates and earnings was significant and had the expected effect: that is, an increase in interest rates tends to raise risk of default whilst a rise in earnings tends to lower risk of default.

Moreover, the survival analysis can be combined with simulation, providing a reasonable platform for stress testing, as proposed by Rodriguez and Trucharte (2007) and Bellotti and Crook (2013, 2014).

At the last but not the least, survival models are able to generate probabilities of how likely an event is to occur over time, conditional on the event not having occurred before, and this provides a dynamic framework for the prediction of credit events. Because the likelihood of the credit event occurring over time can be estimated, the corresponding losses (McDonald et al., 2010) or profits (Ma et al., 2010) can also be predicted.

3 Methodology

Let the survival time T be the time the event object of study occurs. Moreover let T be a continuous random variable that follows a probability density function $f(t)$ and a cumulative density function $F(t)$. The survival function $S(t)$ which represents the probability that the event of interest has not happened by time t, can be defined as follows:

$$S(t) = Pr(T \geq t) = 1 - F(t) = \int_t^{\infty} f(u)\,\mathrm{d}u \tag{1}$$

The hazard function $h(t)$ can be measured as the conditional probability of experiencing the event at time t given survival at that time

$$\lim_{\Delta t \to 0} \frac{Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{2}$$

and can be interpreted as the instantaneous risk of event.

The most widely used hazard model is the semi-parametric proportional one that we can write as follows:

$$h(t|x_i) = exp\{x_i^{'}\beta\}h(t \mid 0) \tag{3}$$

where $h(t \mid 0)$ is the time dependent baseline hazard function that describes the shape of the hazard rate as a function of time; x_i are the covariates for each observation $i = 1, 2, \ldots, N$; $exp\{x_i^{'}\beta\}$ is the hazard ratio, denoted also as relative risk, that describes how the size of the hazard rate depends on covariates, which are assumed to be fixed over time.

In Cox (1972) the covariate vector affects the hazard rate proportionally as in equation (3) and any distributional assumption is required in the estimation procedure because the parametric part is canceled out. The parameters can be estimated maximizing the following "partial likelihood function" (Nam et al., 2008):

$$L\beta = \Pi_{i=1}^{n} \frac{exp\{x_i^{'}\beta\}}{\Sigma_{j\in R(t)} exp\{x_j^{'}\beta\}} \tag{4}$$

where the summation on the denominator is over all subject in the risk set at time t, denoted by $R(t)$. As shown by Lawless (2011) the maximum likelihood estimation of equation (4) is asymptotically normal distributed and its covariance matrix can be consistently estimated.

In recent years the Cox model (3) has been extended in order to include time varying covariates

$$h(t|x_{i,t}) = exp\{x'_{i,t}\beta\}h(t \mid 0) \tag{5}$$

where $h(t|x_{i,t})$ is the individual hazard rate of a subject i at time t and $x'_{i,t}$ is a covariate vector whose components are fixed or time varying. Shumway(2001) suggested to incorporate time varying covariates simply through the counting process formulation (Therneau and Grambsch, 2000): if we have data from n subjects, we consider the counting process $N_1, N_2, \ldots, N_n$ counting the number of occurrences of the event of interest for individual i in $[0, t]$. If by time t the event has been observed for subject i, then $N_{i,t} = 1$, otherwise $N_{i,t} = 0$.

4 Data, Analysis and Results

4.1 Data

The Chamber of Commerce of Perugia provided for this analysis two data sets: the Business Register of Umbria companies (last updated in February 2015) and the annual balance sheet data sets for capital companies for the period 2008-2013. The first data set supplies a complete picture of the legal situation of each company and therefore for each firm we know the creation, modification or cessation data, the legal form, the sector of economic activity and the belonging district. Starting from the financial statements for each firm and year we built indexes representative of the economic and financial situation (see Table 1 for details).

Moreover, following Bellotti and Crook (2014), we considered some of the suggested macroeconomic variables, that is the National Inflation Rate(Π), the Gross Domestic Product (GDP), the Debt-to-GDP ratio (DGR) and the Government Deficit-to-GDP Ratio (GDR), in addition, the Regional Unemployment Index(UNEMP). The main macroeconomic variables' source is the Italian National Institute of Statistics (ISTAT).

The event object of study is the business closure due to different causes such as insolvency (bankrupcy), voluntary or court winding-up ordered liquidation and or dissolution, closure not due to any action by creditors or the court. This choice was due to the fact that in Umbria Region there are a lot of Smalll Business Enterprises and they, because of the high costs of insolvency, tend simply to cease trading rather than applying for insolvency (Orton, 2013).

We arranged data in such a way we have multiple rows for each firm according to the number of annual balance data; for each row we have start and stop variables which denote the begin and end intervals (open on the left and closed on the right) and a status variable which takes value 0 if in that interval (year) the firm survives or is censored and 1 otherwise. Therefore each closed firm contributes only one failure observation and time varying covariates are incorporated simply by using each firm's annual data for its firm-year observation. On the basis of the Register data we determined the age at entry to the study for each firm considering January 2008 the start

Table 1: The potential covariates' selection

Abbreviation	**Variable**
CR	Current Ratio
QR	Quick Ratio
L	Leverage
IRR	Investment Rigidity Ratio
IER	Investment Elasticity Ratio
ER	Equity Ratio
DR	Debt Ratio
ROA	Return on Assets
ROE	Return on Equity
ROT	Return on Turnover
ROS	Return on Sales
TUR	Turnover
IRR decomposition	
TAR	Tangible Assets Ratio
IAR	Inangible Assets Ratio
FFAR	Financial Fixed Assets Ratio
OFAR	Other Fixed Assets Ratio
IER decomposition	
IIR	Inventories Impact Ratio
LR	Liquidity Ratio
STLR	Short Term Liquidity Ratio
LTLR	Long Term Liquidity Ratio
DR decomposition	
PDR	Permanent Debt Ratio
CDR	Current Debt Ratio

Table 2: Annual distribution of companies that have ceased their activity during the period January 2008-February 2015

Year	2008	2009	2010	2011	2012	2013	2014	2015
N	166	430	498	701	714	802	760	64

observational period; its stays the same from line to line, that is we have included the age as a time-constant covariate. According to balance sheets availability the end of our observational period was fixed at December 2013, therefore all the firms closed after this date where considered still active at this date.

4.2 Analysis

We have a total number of observation equal to 64442 for 15489 firms: of these 11550 conduct business in Perugia district and 3939 in Terni. At the end of observational period 21.39%(3313) were closed. On table 2 is reported the distribution of the firms closed between 2008 and February 2015.

The firms belong to six main sector activities: Agriculture (3.10%), Commerce (19.45%), Constructions (17.55%), Industry (17.46%), Tourism (11.69%) and Other Services (30.68%); only in the 0.07% of cases we met missing values and within each group the number of closed firms was between $20\% - 22\%$.

The prevailing juridic form of our firms is SRL (90.75%), then we have SocCoop (6.62%) and SpA (2.63%). These legal forms correspond respectively to: General Partnership, Cooperative Partnership and Joint-Stock

Company.

The mean age of the firms is 6.56 with an higher value for the active or censored (6.76) than for the closed (5.83). We observed the distribution of continuous variables and to ensure that statistical results were not influenced by outliers, we set all the observation higher than the ninety-ninth percentile of each variable to that value and all values lower than the first percentile were truncated in the same manner (Shumway, 2001).

Survival analyses was implemented first considering all the continuous variables in the data set (age and all built in indexes as in Table 3). The results of the Best Subset Method, which apply only for continuous variables, were compared with those coming from stepwise selection method, where we fixed as significance level for entering a variable into the model 0.25 and to remove 0.05. Considering the number of variables selected by stepwise method, the Best Subset Method gave the same results as reported on Table 4.

We checked the scale of the covariates applying the quartile design method (Hosmer et al., 2008). Variables ROA, PDR and ER showed a quadratic form, while ROT and FFAR where both turned into categorical variables. ROT was divided into three classes (ROT1, ROT2, ROT3) on the bases of the first ($<= 0.10$), second ($0.10 - |0.66$) and third and fourth quartile together (> 0.66). FFAR was shared into three (FFAR1, FFAR2, FFAR3) classes: the first class includes values belonging to the first and second quartile ($<=$ 0.0000006), the second and the third classes respectively correspond to the

Table 3: Descriptive statistics of all potential explanatory variables

Variable	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
CR	63119	3.274	11.405	0.000	97.752
QR	63119	1.879	5.349	0.000	43.585
L	64428	10.971	29.407	-75.437	193.153
IRR	64424	0.294	0.311	0.000	0.990
IER	64424	0.581	0.336	0.000	1.000
TAR	64424	0.215	0.285	0.000	0.977
IAR	64424	0.042	0.100	0.000	0.602
FFAR	64424	0.034	0.117	0.000	0.797
OFAR	16848	0.109	0.213	0.000	0.951
IIR	64424	0.160	0.264	0.000	0.971
LR	64424	0.421	0.314	0.000	1.000
STLR	64424	0.089	0.155	0.000	0.794
LTLR	64424	0.331	0.282	0.000	0.975
ER	64424	0.215	0.394	-1.917	0.994
DR	64424	0.758	0.408	0.000	2.856
PDR	64424	0.215	0.252	0.000	0.982
CDR	64424	0.540	0.392	0.000	2.408
ROA	64424	-0.032	0.198	-1.301	0.352
ROE	64428	0.030	1.124	-6.035	5.049
ROT	64424	0.913	0.989	-0.074	5.410
ROS	55568	1.116	0.788	-0.347	7.209
TUR	64424	0.892	0.995	0.000	5.487
DGR	64442	120.334	8.216	106.090	132.740
GDR	64442	-3.657	1.001	-5.450	-2.670
Π	64442	3.337	0.937	1.200	3.900
GDP	64442	-1.553	2.463	-6.300	1.000
UNEMPL	64442	7.335	1.994	4.207	10.421
AGE	64442	7.887	10.313	0.000	85.896

Table 4: Estimated coefficients, Standard Errors, Wald χ^2 statistic, p-value and Hazard Ratio for the Proportional Hazard Model containing variables significant at the 5% level after Stepwise selection with significance level for entering at 25%

Variable	**Coefficient Estimate**	**Standard Error**	**Wald χ^2**	**Pr $>\chi^2$**	**Hazard Ratio**
ROA	-1.552	0.059	697.516	<.0001	0.212
ROT	-0.364	0.022	286.570	<.0001	0.695
AGE	-0.022	0.002	97.382	<.0001	0.978
UNEMPL	-0.155	0.014	118.878	<.0001	0.856
LR	0.776	0.058	176.314	<.0001	2.174
FFAR	0.792	0.126	39.245	<.0001	2.207
ER	-0.316	0.037	73.362	<.0001	0.729
PDR	-0.593	0.072	67.539	<.0001	0.553
AIC	58022.439				
-2LogL	58006.439				

third $(0.0000006-|0.0080)$ and fourth quartile $(0.0080-|0.79705)$. On Table 5 we reported the coefficient estimates for the proportional hazard model including all transformed covariates.

As a second step we considered the sector activity, the district area and the juridic form as potential covariates in addition to the others. The sector activity showed an overall significance ($P > \chi^2 = 0.0007$) and therefore it was included into the model (Table 6). We checked for plausible interaction terms without significant results.

It should be emphasized that the process followed in the development of the model allows the identification of the most significant covariates and their appropriate transformation were applied to ensure the linearity in the

Table 5: Estimated coefficients, Standard Errors, Wald χ^2 statistic, p-value and Hazard Ratio for the Proportional Hazard Model with transformed covariates

Variable	**Coefficient Estimate**	**Standard Error**	**Wald χ^2**	**Pr $>\chi^2$**	**Hazard Ratio**
ROA	-3.089	0.150	426.386	$<.0001$	.
ROA^2	-1.699	0.123	190.295	$<.0001$	.
ER	-0.246	0.042	33.866	$<.0001$	.
ER^2	0.125	0.030	17.349	$<.0001$	.
ROT2	-0.510	0.048	113.912	$<.0001$	0.600
ROT3	-1.006	0.046	486.185	$<.0001$	0.366
FFAR2	-0.171	0.051	11.138	0.001	0.843
FFAR3	-0.034	0.045	0.590	0.443	0.966
LR	0.814	0.057	207.315	$<.0001$	2.257
PDR	-2.641	0.237	124.329	$<.0001$	.
PDR^2	2.489	0.266	87.561	$<.0001$	.
AGE	-0.016	0.002	52.793	$<.0001$	0.984
UNEMPL	-0.161	0.014	129.258	$<.0001$	0.851
AIC	57291.806				
-2LogL	57265.806				

Table 6: Estimated coefficients, Standard Errors, Wald χ^2 statistic, p-value and Hazard Ratio for the final Proportional Hazard Model

Variable	**Coefficient Estimate**	**Standard Error**	**Wald χ^2**	**Pr $>\chi^2$**	**Hazard Ratio**
ROA	-3.097	0.151	423.023	<.0001	.
ROA^2	-1.700	0.124	188.765	<.0001	.
ER	-0.235	0.043	30.605	<.0001	.
ER^2	0.124	0.030	16.937	<.0001	.
ROT2	-0.522	0.048	116.613	<.0001	0.594
ROT3	-1.029	0.047	477.858	<.0001	0.357
FFAR2	-0.177	0.051	11.852	0.001	0.838
FFAR3	-0.032	0.045	0.493	0.483	0.969
LR	0.810	0.057	202.983	<.0001	2.247
PDR	-2.664	0.238	125.157	<.0001	.
PDR^2	2.529	0.267	89.522	<.0001	.
AGE	-0.017	0.002	54.768	<.0001	0.983
UNEMPL	-0.163	0.014	131.261	<.0001	0.849
Commerce	0.299	0.119	6.375	0.012	1.349
Constructions	0.269	0.117	5.269	0.022	1.308
Industry	0.328	0.119	7.674	0.006	1.389
Other Services	0.138	0.114	1.457	0.228	1.148
Turism	0.251	0.122	4.230	0.040	1.285
AIC	57170.593				
-2LogL	57134.593				

log hazard. The continuous covariates FFAR and ROT were replaced with categorical ones after several comparisons between all possible adaptations.

This approach essentially lead us to avoid errors in measurement and in interpretation during creditworthiness assessment.

We verified the proportionality hazard assumption plotting for each variable the scaled Schoenfeld residuals against time: they appeared to be randomly distributed around the 0 without a discernible pattern.

The predicted survival probabilities over a six year period $(2008 - 2013)$ depend on some fixed and time varying covariates, therefore we could use the model to predict future survival probabilities if we would have covariates' data for the subsequent years. Unfortunately, at present we only know the legal status of each firm at February 2015. We decided to test the usefulness of Cox Proportional Hazard Model as a credit scoring system within a time period using the computed survival probabilities (Bellotti and Crook, 2009) . Given a cut-off threshold, if a firm has a survival probability greater than it is predicted as good, otherwise bad. The survival probabilities are therefore used as scores to predict the company closure. Taking into account only the firms still active at 2013 we built a misclassification table crossing the predicted scores with the legal status at 2015.

4.3 Results and Discussion

The ability of a company to meet its debt obligations is a result of two related phenomena: on the one hand, the competitiveness of products and services offered on the market, on the other hand, the sustainability of the financial structure (Roggi and Giannozzi, 2008).

The first characteristic, namely the competitiveness of the company, can be assessed through two aspects: the potential of the core business and the ability to generate positive returns on investment. The sustainability of the sources of financing essentially covers three aspects: the consistency of indebtedness, the ability to meet deadlines and the exposure to changes in the cost of debt.

These considerations led to the selection of indicators shown in Table 1 The objective of the statistical analysis is to identify which of these are more effective for the analysis of the health status of the debtor.

The observation of the results can highlight the reduction in size of the problem: the risk of default can be estimated through a limited number of variables.

The empirical analysis initially identifies eight significant variables (Table 4). It looks very interesting that the selection with the stepwise method produces a sample of variables which are included at least one from each family of indicators.

Indeed, in the resulting model are simultaneously present information about the profitability, firm age and capital structure: respectively ROA and ROT, Age and LR, FFAR, ER, PDR.

The survival model on Table 6 shows as if a firm has an FFAR value belonging to the second class, its rate of death is estimated to decrease by

16.7% and if its value lies on the third class it would decrease only by 3.4%. An increasing ROT allows a firm to sensibly decrease its probability of death, the same we can affirm for ROA, ER and PDR. An increasing local unemployment decreases by 1.49% the rate of death. Moreover older firms have a longer life expectancy and an increasing LR value decreases the probability to alive.

The signs of the coefficients are consistent with the existing literature. However, the negative relationship with the regional unemployment rate seems illogical. This result is affected by our period of analysis: the time horizon coincides with a period of deep crisis. Companies, before shut down as a result of the contraction of business, reduce staff. If the cost reduction is not enough to balance the accounts, the company can not avoid bankruptcy or closure. So, the regional unemployment rate anticipates the phenomenon of corporate crises which we detect only at final moment.

The Agricultural sector showed a significant difference when compared with Commerce ($p = 0.0116$), Constructions ($p = 0.0217$), Industry ($p = 0.0056$) or Tourism Sectors ($p = 0.0397$); the same result for Other Services sector when compared with Commerce (0.0023), Constructions (0.0112) or Industry (0.0004) sectors.

In order to assess the goodness of our survival model as a credit scoring system, we choose as cut-off a survival probability corresponding to the 23rd percentile (0.8156). Considering the legal status at February 2015 and excluding all the firms closed within December 2013, we buil a misclassification table. The hit rate was 79.73%, the sensitivity 82.90% and the specificity 45.18%. Note that sensitivity is much higher than specificity, but this is quite normal in credit scoring models (Baesens et al., 2005).

5 Conclusion

The results obtained in this study are useful both in the construction of rating models and as a guide to company management. The rigorous selection process of the most significant indicators has led to results that are consistent with previous studies. The approach has allowed the identification of covariates, the determination of their positive or negative relationship and the quantification of the effect on the probability of default. The availability of an evaluation that summarizes this impact enables operators to have a clear view on the business. On the one hand, the financial intermediaries have the ability to build a process of risk assessment. On the other hand, the businesses managers can have helpful indicators able to drive the decision-making process.

The adoption of survival analysis allows to take into account dynamic changes in the time of the covariates. Further investigation on the validity of the proposed model will be object of study as soon as financial information on companies will be available for the year later 2013.

The construction of statistical rating models must be preceded by a thorough analysis of the companies and a study on the functional form that best represents the influence of the covariates on the probability of default. The approach proposed in this work is a useful example of investigation.

References

Andreeva, G. (2006). European generic scoring models using survival analysis. *Journal of the Operational Research Society*, 57(10):1180 – 1187.

Baesens, B., Gestel, T. V., Stepanova, M., Poel, D. V. D., and Vanthienen, J. (2005). Neural network survival analysis for personal loan data. *Journal of the Operational Research Society*, 56(9):1089 – 1098.

Banasik, J., Crook, J. N., and Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, 50(12):1185 – 1190.

Basel (2000). Principles for the management of credit risk. *Basel Committee on Banking Supervision. BIS. September 2000.*

Bellotti, T. and Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60:1699 – 1707.

Bellotti, T. and Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting*, 28(1):171 – 182.

Bellotti, T. and Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4):563 – 574.

Bellotti, T. and Crook, J. (2014). Retail credit stress testing using a discrete hazard model with macroeconomic factors. *Journal of tthe Operational Research Society*, 65(3):340 – 350.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187 – 220.

Hosmer, D. W., Lemeshow, S., and May, S. (2008). *Applied survival analysis: regression modeling of time to event data*. Wiley-Interscience, New York, NY, USA, 2nd edition.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Leow, M. and Crook, J. (2014a). Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, 236(2):685 – 694.

Leow, M. and Crook, J. (2014b). The stability of survival model parameter estimates for predicting the probability of default: empirical evidence over the credit crisis. *European Journal of Operational Research*, pages 1 – 8.

Ma, P., Crook, J., and Ansell, J. (2010). Modelling take-up and profitability. *Journal of the Operational Research Society*, 61(3):430 – 442.

McDonald, R. A., Matuszyk, A., and Thomas, L. C. (2010). Application of survival analysis to cash flow modelling for mortgage products. *OR insight*, 23(1):1 – 14.

Nam, C. W., Kim, T. S., Park, N. J., and Lee, H. K. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, 27(6):493 – 506.

Narain, B. (1992). Survival analysis and the credit granting decision. In Thomas, L. C., Crook, J. N., and Edelman, D. B., editors, *Credit Scoring and Credit Control*. Clarendon Press, Oxford.

Orton, P. (2013). Measuring business failure in small and medium sized enterprises. Credit scoring & credit control. *XIII Conference, Edinburgh*.

Pierri, F., Stanghellini, E., and Bistoni, N. (2015). Forecasting probability of bankruptcy from unbalanced data. In *Current Topics on Risk Analysis: ICRA6 and RISK 2015 Conference*, pages 621 – 628.

Rodriguez, A. and Trucharte, C. (2007). Loss coverage and stress testing mortgage portfolios: a non-parametric approach. *Journal of Financial Stability*, 3(4):342 – 367.

Roggi, O. and Giannozzi, A. (2008). Predicting SME default risk. Does regional model make sense? In *8th Global Conference on Business and Economics*.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *The Journal of Business*, 74(1):101 – 124.

Stepanova, M. and Thomas, L. (2001). PHAB scores: proportional hazards analysis behavioural scores. *Journal of the Operational Research Society*, pages 1007 – 1016.

Stepanova, M. and Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2):277 – 289.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model*. Springer Science & Business Media.

Thomas, L. C., Ho, J., and Scherer, W. T. (2001). Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, 12(1):89 – 103.