European generic scoring models using survival analysis

Dr G. Andreeva, University of Edinburgh, UK

Credit scoring discriminates between 'good' and 'bad' credit risks to assist credit-grantors in making lending decisions. Such discrimination may not be a good indicator of profit, whilst survival analysis allows profit to be modelled. The paper explores the application of accelerated failure time and proportional hazard models to the data from the retail card (revolving credit) from three European counties. The predictive performance of three national models is tested for different definitions of default and then compared to that of a single generic model. It is found that survival analysis approach is suitable for building generic models and competitive with the current industry standard - logistic regression. Stratification is investigated as a way of extending proportional hazards models to tackle heterogeneous segments in the population.

Keywords: credit scoring; regression analysis; risk; banking

Introduction

Credit scoring is a technique mainly used in consumer credit to assist credit-grantors in making lending decisions. Its aim is to construct a classification rule that distinguishes between 'good' and 'bad' credit risks according to some specified definition. The rule is developed on a sample of the past applicants, whose performance is known. A number of modelling approaches is used from classical discriminant analysis to neural networks and genetic algorithms[1-4]. Most frequently the model is a weighted sum of the applicant's observed characteristics (age, marital status, etc.) that produces a score, which is a summary of the applicant's creditworthiness and reflects his or her ranking relative to other applicants. The classification into 'good' and 'bad' is achieved by comparing the score to a predetermined threshold or a cut-off level.

Traditionally a credit scoring model is constructed to fit a specific credit portfolio which normally consists of residents of one country (customised models). But the political desire for further integration of the European Union into a single internal market opens the possibility for the lenders to compete across national borders. Therefore the necessity arises to understand the performance of Pan-European credit scoring models as opposed to those based on separate countries. Does the predictive ability of the generic model compare favourably with national models? If it does then it will be possible to use the generic model, which is a considerably cheaper alternative to customised models, across a series of European countries. The attractiveness of generic models is further strengthened by the legal restrictions on the composition of credit scoring models. In certain countries the use of nationality can be regarded as illegal. For a more detailed discussion of legal restrictions on data used in credit scoring see Andreeva et al.[5]

The majority of published empirical tests demonstrate the superiority of customised models[6-9]. The current paper explores the possibility of applying a single generic model to credit score the applicants for a revolving store card from three EU countries (Belgium, the Netherlands, Germany), and demonstrates the competitiveness of generic models. Several generic models are developed using logistic regression and survival analysis, and their predictive accuracy is benchmarked against the performance of equivalent national models.

Whilst logistic regression is the standard approach in credit industry, survival analysis is a relatively new application that offers an advantage of predicting time to the event of interest and therefore, lays the foundation for estimating the applicant's profitability[10,11]. Applications of survival analysis in credit scoring began with Narain[12] showing that estimates of a lifetime of a loan obtained from the exponential model can significantly improve the credit-granting decisions. Several papers [10,11,13,14] explored different applications of proportional hazards models, including behavioural scoring. The dynamic exponential model was proposed[15] for new products, when there are no historical data to develop a model on.

These studies analysed fixed-term credit data and found the survival analysis approach competitive and in certain applications superior to the logistic regression. The current analysis extends the application of survival analysis to the area of revolving credit which has not been investigated before, and explores the family of accelerated failure time distributions that has not been addressed in detail by previous research. In this paper comparison is made between different parametric survival models, including the accelerated failure time ones, to non-parametric proportional hazards, in predicting the time to default. These models are also benchmarked against the logistic regression. In addition, the sensitivity of predictive ability of survival models to the presence of heterogeneous subpopulations (applicants from different countries) is investigated.

The paper is structured in the following way. The next section describes the data and presents an overview of the basic concepts and methods used. The subsequent section compares the national survival patterns and compares predictive accuracy of national and generic models under different modelling approaches. Alternative definitions of default are investigated for national models. The following section extends the generic proportional hazard model by means of stratification. The final section concludes and outlines some directions for further research.

Data description and methodology

The data for analysis were provided by a major international credit scoring consultancy and relate to the same retail card issue in 3 European countries: Belgium, Germany and the Netherlands (Table 1). The performance of the accounts was observed during 25 months from October 1998 until December 2000. The life of the account was measured from the month it was opened until the account became 'bad' or it was closed or until the end of observation. The account was considered to be 'bad' if payment was not made for two consecutive months. If the account did not miss two payments and was closed or survived beyond the observation period, it was considered to be censored. The list of characteristics collected in each country was different. However, it was possible to select 16 characteristics that were collected for all three countries (Table 2), and these characteristics were used for predicting default in national and generic models.

The traditional way of relating the vector of observed characteristics $\mathbf{x}$ to the probability of default is to fit the logistic regression model to estimate P, the probability of becoming 'bad' within the period of observation:

$P = \exp(\boldsymbol{\beta}'\mathbf{x})/(1- \exp(\boldsymbol{\beta}'\mathbf{x}))$.

This approach assumes that that accounts that do not experience default are 'good', non-defaulting accounts, whilst survival analysis treats such accounts in a more conservative way, as those that proved to be 'good' so far. Survival analysis allows to use characteristics $\mathbf{x}$ to estimate either time to default T or probability of surviving to a certain time:

$S(t) = S_0(\gamma(\mathbf{x})\, t)$,

or hazard function $h(t)$, which is the probability of the event occurring within the time interval $(t,t+\Delta t)$, given that the event did not occur before time t:

$h(t) = \gamma(\mathbf{x}) h_0(t\, \gamma(\mathbf{x}))$,

where S_0 and h_0 are baseline survival and hazard functions and $\gamma(\mathbf{x})=\exp(\boldsymbol{\beta}'\mathbf{x})$. The above given model - accelerated failure time (AFT) - assumes that the covariates $\mathbf{x}$ act multiplicatively on time t, thus influencing the speed that the account proceeds towards the event. In the proportional hazard (PH) model the covariates act multiplicatively on the baseline hazard rate, and so the relative ranking of risk presented by different accounts does not change with time:

$h(t)=h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x})$.

One can assume a particular distribution for a hazard function, the most commonly used in medicine[15] and reliability[16] are exponential, Weibull, log-normal, log-logistic and gamma, and they are considered in this paper. It should be noted that exponential and Weibull distributions can be both AFT and PH. Alternatively, one can use Cox proportional hazards model that does not specify the shape of the baseline hazard function[19]. Whilst the parametric AFT models provide more efficient estimation, PH has an advantage of robustness and flexibility. An important property that follows from an arbitrary nature of the baseline hazard $h_0(t)$ is that it can be allowed to vary between different groups (nations) in a population, a feature that can be useful for generic scoring models. A more detailed discussion of survival analysis methods is given elsewhere[16-24].

Cox non-parametric model and several parametric models (assuming the distributions described above) were fitted to three national datasets. Then three datasets were pooled together, and the generic models were built on the aggregated data. The predictive performance of the survival analysis models was benchmarked against the logistic regression (LR). All the datasets used in the analysis were randomly split into training (70%) and holdout (30%) samples. The model was developed on the training sample, and its predictive ability was measured on the corresponding hold-out sample. Generic models were tested on each national hold-out sample (to allow for comparisons with national models) and on the aggregated generic hold-out.

For Cox non-parametric PH model the estimate of $\exp(\boldsymbol{\beta}'\mathbf{x})$ was used as a score, for parametric models a score was given by the probability of 'surviving' in 25 months, which is similar to the way predictions are generated from logistic regression. However, the advantage of the survival analysis consists in the ability to produce predictions for several time periods from the same model, which logistic regression cannot do.

Comparison of the models was made by the Receiver Operating Characteristics (ROC) curve and percentage of incorrectly classified accounts. The area under ROC (AUROC) provides a measure of classification accuracy, which is not dependant on any threshold or acceptance rate[25]. It corresponds to the Wilcoxon or Mann-Whitney or *U* statistic, which estimates the probability that a ranking of a randomly selected bad account will be less than or equal to a ranking of a randomly selected good account[26]. The second measure is the error rate (ER), the sum of percentages for incorrectly classified goods and incorrectly classified bads. It is based on confusion matrix, which presents the counts of good and bad accounts correctly and incorrectly classified by the model. The use of the matrix requires the choice of a cut-off level. For the purpose of this analysis the cut-off was fixed at the level of the default rate in the hold-out sample, that is, such that the observed proportion of bads equalled the predicted proportion of bads in the hold-out sample.

Survival analysis compared to logistic regression

Before modelling the relationship between application characteristics and survival times, some exploration of survival patterns was done for each country by means of:

1) fitting Kaplan-Meier estimates of the survival distribution function (SDF) [18] ;
2) examining the hazard plot, with estimates of the hazard function obtained as described in Stepanova[14].

The SDF plots (Figure 1) overlap for Belgium and Germany, with SDF for the Netherlands decreasing faster and the difference becoming larger with time. The log-rank test and Wilcoxon test[19] showed that survival curves were significantly different for the three countries (Table 3).

The hazard plots (Figure 2) for all three countries increase rapidly in the first months of the account life and then decrease towards an asymptote. This supports a conventional wisdom - 'if they go bad, they go bad early'[9]. However, 'early' means 2 months for Belgium and Germany, and 5 months for the Netherlands. The height of peaks also differs between the countries, with Germany being the least risky in terms of early defaulters and the Netherlands being the most risky. After 9 months the confidence intervals overlap, but the Netherlands remains slightly higher and shows a slight increase at the end of the observation period. However, this may due to decreasing risk set.

The examination of plots suggested a number of possible approaches: log-normal, log-logistic and gamma are most suitable for modelling non-monotone hazards. But one should not discard exponential and Weibull distributions, they may offer a suitable approximation since the major part the national hazard functions are monotone and even constant. At the same time the country hazards looked roughly proportional, suggesting that PH models would be appropriate.

So exponential, Weibull, log-normal, log-logistic distributions were fitted along with the non-parametric Cox proportional hazards model. The model fit was measured by -LogLikelihood, with lower magnitude values indicating the better fit. This measure indicates how well the model describes the data but it can be used only to compare nested models. Models are nested if one is a special case of another, in our case the exponential, Weibull, log-normal models are nested within the generalised gamma[19]. This allows one to make tests of significance for the difference in model fit[22]. Judgements about models not nested in gamma (log-logistic, logistic and Cox PH) can be made based on their predictive ability, which was measured by the AUROC and error rate.

The results from different survival analysis models and logistic regression are presented in Table 4. The differences in model fit were highly significant (Table 5). In terms of model fit, the leader was the gamma distribution followed by the log-normal for all three countries. However, this did not translate into superior prediction results. In fact, gamma gave the worst prediction, which may be due to over fitting when better fit does not necessarily mean better prediction.

Since different distributions demonstrate very similar predictive accuracy, and given the desire for a more parsimonious and therefore robust model, the exponential distribution would be most suitable from parametric models. At the same time proportional hazards and logistic regression models give identical results. Since there is no or little difference in predictive accuracy, the decision as to which approach to use should be based on additional properties that a certain method can provide. If the exponential is chosen then it suggests that the default process is memoryless across the population for revolving credit and this explains the similarity in predictions obtained the proportional hazards and logistic models.

However, there may be periods within 25 months when certain models give superior prediction. This proposition would apply to AFT survival models since they allow for modelling the changing hazard rates between different groups over time, and therefore will produce different ranking of accounts in different time periods. To test this proposition the survival models were applied to two alternative definitions of ‘bad’. First, those that missed 2 payments within the first six months were considered to be ‘bad’, and the rest were treated as ‘good’. Second, those that defaulted within first twelve months were classified as ‘bad’, the remaining customers in the hold-out sample were considered good. The parameter estimates from AFT models developed earlier were used to generate predictions of surviving within 6 and 12 months, the cut-offs were chosen to match actual number of defaults within these time periods. For PH model the estimates obtained earlier where used, since they are not time sensitive.

The results in Table 6 suggest that log-logistic, log-normal and gamma models give slightly superior prediction for Belgium and the Netherlands, especially for ‘Default in 6 months’. This is in line with hazard plots (Figure 2) where early peaks in hazards are observed, so if one would expect some difference in results that would be during the first months. The hazard peak for Germany is least pronounced, and hence there is no marked difference between the predictive ability of survival models for this country. It should be noted though that even for Belgium and the Netherlands the differences are marginal and do not give enough grounds to conclude that log-logistic, log-normal or gamma should be preferred to more robust exponential and non-parametric Cox PH models. Perhaps, the superiority of AFT models may be more visible in different credit scoring applications (e.g. insurance) with more pronounced time structure.

Table 7 compares the performance of exponential, PH and LR generic models against the national LR model by means of AUROC and error rate (in brackets). The customised national models give slightly better prediction, but one can argue that from the business point of view the difference in performance is not dramatic and can be offset by the lower costs of using one model instead of three different ones. Out of three generic models, logistic regression is slightly superior, but again the difference is marginal. It should be noted that logistic regression is given an advantage by the fact that it is fitted specifically to a definition of being ‘bad’ within the fixed observation period, whilst survival analysis is not restricted by any arbitrary time horizon. In general, it is possible to say that all three approaches are fairly insensitive to the presence of different national segments.

Stratified proportional hazards model

As was noted before, the PH model can account for different subpopulations that may exist within the data, while producing a single set of parameter estimates that does not include a subpopulation indicator. This method (stratification) is commonly used for subpopulations that violate the proportionality assumption[22].

The following model was fitted to data :

$\log h(t)=\alpha_z(t) + \boldsymbol{\beta}'\mathbf{x}$,

where z is a subpopulation indicator. In this way the hazard function was allowed to vary between specified groups. This method presents a half way option between generic and customised models: to estimate the parameters the country indicator is required, but the subsequent scoring of new accounts can be done without making a distinction between the countries. It should be noted, though, that in order to use this property one needs to know the applicant's country of residence. Still it will be useful in situations when 'nationality' is not legally forbidden[5], and lenders can replace several national models with a generic one that accounts for different subpopulations due to stratification.

The stratified PH model has been fitted to the aggregated generic dataset (Table 7). AUROC shows no improvement on the generic hold-out sample, but when tested separately on the national samples, there is some increase in AUROC for Belgium and the Netherlands, although not a dramatic one. Error rate demonstrates some superiority of the stratified approach if compared to other models, including the logistic regression. For Belgium the stratified model shows the error rate even lower than the national LR model. In general, one can conclude that stratification brings some benefit but not a convincing one. Such modest improvements can be attributed to the fact that the hazards between the countries are roughly proportionate, so there is little scope for stratification to enhance the predictive performance.

Conclusions and further research

This paper addressed a crucial problem which is very likely to face European and eventually British banks in the near future, when households in countries of the European Monetary Union realize that the exchange rate risk they once faced has been removed, and start comparing the interest rates and loan terms across Europe rather than just within their own country. To maintain a competitive position and gain a competitive advantage banks must develop ways of assessing the creditworthiness of applicants from other countries.

The paper presents the first cross-country comparison of the application of survival analysis to predict when a borrower defaults and supports the previous findings that survival analysis is competitive with the logistic regression. The comparison of several approaches showed that there is little difference in classification accuracy between the parametric survival models, non-parametric Cox PH model and logistic regression. The current analysis demonstrated that in spite of significant differences between the model fit measures, the predictive accuracy is not affected. Such similarity in prediction can be attributed to relatively constant hazards (apart from the first months).

The analysis was carried out on a revolving type of credit, the area where the application of survival analysis was not investigated by previous published work. It was suggested that exponential distribution offered a suitable fit for the analysed dataset, therefore, implying that the data exhibits the memoryless property. This indicates that revolving credit may have a more random character than fixed term credit, a proposition that needs to be investigated further. Early defaulters also deserve further investigation, since they may represent fraudulent accounts.

So the choice of the model depends on some additional benefits any particular modelling approach can bring. The survival analysis offers a number of benefits that make it more attractive if compared to logistic regression. Time to default can be estimated, which provides a basis for profit scoring. Alternatively, predictions can be generated that give the probability of 'survival' in certain time period. It was shown that survival analysis is also suitable for building generic models.

As for the choice between different survival analysis models, it would depend again on the additional features that the lenders would consider important. The AFT models showed some superiority in predicting early defaulters, although the magnitude of improvement is dependent on how large is the number of such 'early defaulting' accounts as compared to the total number of defaults. The PH model offers a possibility to account for different segments in the population, especially for those that violate the proportionality assumption. This property makes it possible to produce models that are between generic and customised ones.

The emergence of the single European market in financial services makes generic scoring an important and timely application. And survival analysis generic models can serve a starting point of profit scoring in integrated Europe.

References

1 Hand DJ, Henley WE (1997). Statistical classification methods in consumer credit. *Journal of Royal Statistical Society, Series A* **160:** 523-541.

2 Rosenberg E, Gleit A (1994). Quantitative methods in credit management: a survey. *Operations Research* **42:** 589-613.

3 Thomas LC, Edelman DB, Crook JN (2002). *Credit scoring and its applications.* Society for Industrial and Applied Mathematics**:** Philadelphia, PA.

4 Thomas LC (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting* **16:** 149-172.

5 Andreeva G, Ansell J, Crook JN (2004). Impact of anti-discrimination laws on credit scoring. *Journal of Financial Services Marketing,* forthcoming in September 2004.

6 Overstreet GA, Bradley EL, Kemp RS (1992). The flat-maximum effect and generic linear scoring models: a test. *IMA Journal of Mathematics Applied in Business & Industry* **4:** 97-109.

7 Barron JM, Ellienhausen G, Staten ME (2000). Monitoring the household sector with aggregate credit bureau data. *Business Economics* **35:** 63-77.

8 Overstreet GA, Bradley ELJ (1996). Applicability of generic linear scoring models in the U.S. credit-union environment. *IMA Journal of Mathematics Applied in Business & Industry* **7:** 291-311.

9 Platts G, Howe I (1997). A single European scorecard. In: *Credit Scoring and Credit Control V, Conference Proceedings,* The University of Edinburgh Management School**:** Edinburgh.

10 Banasik J, Crook JN, Thomas LC (1999). Not if but when will borrowers default. *Journal of Operational Research Society* **50:** 1185-1190.

11 Stepanova M, Thomas LC (2001). PHAB scores: proportional hazards analysis behavioural scores. *Journal of Operational Research Society* **52:** 1007-1016.

12 Narain B (1992). Survival analysis and the credit granting decision. In: Thomas LC, Crook JN, Edelman DB (eds). *Credit Scoring and Credit Control,* Oxford University Press**:** Oxford, 109-122.

13 Stepanova M, Thomas LC (2002). Survival analysis methods for personal loan data. *Operations Research* **50:** 277-289.

14 Stepanova M . *Using survival analysis methods to build credit scoring models*. PhD thesis, 2001, University of Southampton.

15 Hand DJ, Kelly MG (2001). Lookahead scorecards for new fixed term credit. *Journal of Operational Research Society* **52:** 989-996.

16 Collett D (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall: London.

17 Ansell JI, Phillips MJ (1994). *Practical Methods for Reliability Data Analysis*. Clarendon Press**:** Oxford

18 Kaplan EL, Meier P (1958). Non-parametric estimation from incomplete observations. *Journal of American Statistical Association* **53:** 457-481.

19 Kalbfleisch JD, Prentice RL (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons: New York.

20 Cox DR (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, Series B* **74:** 187-220.

21 Cox DR, Oakes D (1984). *Analysis of Survival Data*. Chapman & Hall**:** London.

22 Allison PD (1995). *Survival analysis: using the SAS system*. SAS Publishing**:** Cary, NC, USA

23 Lawless JF (1981). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons: Chichester.

24 Parmar MKB, Machin D (1995). *Survival analysis: a practical approach*. John Wiley & Sons**:** Chichester.

25 Bamber D (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12:** 387-415.

26 Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143:** 29-36.

Table 1.

	Belgium		The Netherlands		Germany	
	Count	*Percent*	*Count*	*Percent*	*Count*	*Percent*
Bad	3090	11.75%	11213	13.63%	8909	11.75%
Good	23200	88.25%	71024	86.37%	66939	88.25%
Total	26290	100.00%	82237	100.00%	75848	100.00%

Table 2.

No	Characteristic	No	Characteristic
1	Home telephone	9	Employer's phone
2	Residential status	10	Card insurance
3	Marital status	11	Credit insurance
4	Occupation (Full-time, part-time, self-employed, etc.)	12	Number of dependants
5	Age	13	Spouse age
6	Time at address since 18 years old	14	Goods code
7	Time in employment	15	Goods price
8	Type of business (Manufacturing, banking, catering, etc.)	16	Payment date

Table 3.

Test	Chi-Square	Degrees of freedom	Pr > Chi-Square
Log-Rank	131.2216	2	0.0001
Wilcoxon	97.8208	2	0.0001

Table 4.

Model	Log L	AUROC	Error rate
	Belgium		
Exponential	-7842	0.7121	16.74%
Weibull	-7811	0.7119	16.78%
Loglogistic	-7780	0.7123	16.74%
Lognormal	-7708	0.7122	16.90%
Gamma	-7690	0.7116	16.90%
PH	-20093	0.7122	16.72%
Logistic	-5935	0.7129	16.92%
	The Netherlands		
Exponential	-24656	0.7802	16.50%
Weibull	-24185	0.7797	16.50%
Loglogistic	-23870	0.7800	16.50%
Lognormal	-23700	0.7801	16.54%
Gamma	-23693	0.7801	16.54%
PH	-79429	0.7802	16.50%
Logistic	-19229	0.7814	16.50%
	Germany		
Exponential	-22531	0.7408	15.78%
Weibull	-22380	0.7405	15.78%
Loglogistic	-22269	0.7406	15.78%
Lognormal	-22062	0.7406	15.86%
Gamma	-22016	0.7404	15.88%
PH	-65386	0.7412	15.76%
Logistic	-17389	0.7417	15.74%

Table 5.

	Chi-sq	df	p>0
	Belgium		
Exponential vs Weibull	44	1	2.76018E-11
Weibull vs Gamma	429	1	2.4511E-95
Lognormal vs Gamma	143	1	4.61618E-33
Exponential vs Gamma	473	2	1.5012E-103
	The Netherlands		
Exponential vs Weibull	941	1	1.4026E-206
Webull vs Gamma	985	1	3.5089E-216
Lognormal vs Gamma	13	1	0.000280096
Exponential vs Gamma	1926	2	0
	Germany		
Exponential vs Weibull	49	1	3.00302E-12
Webull vs Gamma	1177	1	6.7201E-258
Lognormal vs Gamma	335	1	7.43031E-75
Exponential vs Gamma	1225	2	7.7443E-267

Table 6.

Model	Default in 6 months		Default in 12 months	
	AUROC	**Error rate**	**AUROC**	**Error rate**
		Belgium		
Exponential	0.7225	8.30%	0.7225	13.20%
Weibull	0.7224	8.26%	0.7225	13.26%
Loglogistic	0.7232	8.30%	0.7229	13.26%
Lognormal	0.7235	8.22%	0.7223	13.40%
Gamma	0.7234	8.12%	0.7218	13.44%
PH	0.7226	8.32%	0.7224	13.26%
		The Netherlands		
Exponential	0.8241	7.36%	0.8036	12.24%
Weibull	0.8249	7.38%	0.8037	12.20%
Loglogistic	0.8255	7.24%	0.8046	12.22%
Lognormal	0.8262	7.26%	0.8052	12.24%
Gamma	0.8263	7.26%	0.8053	12.26%
PH	0.8242	7.26%	0.8036	12.22%
		Germany		
Exponential	0.7649	7.02%	0.7483	12.56%
Weibull	0.7639	7.02%	0.7482	12.54%
Loglogistic	0.7639	7.00%	0.7483	12.56%
Lognormal	0.7637	7.00%	0.7482	12.52%
Gamma	0.7635	6.98%	0.7480	12.50%
PH	0.7642	7.04%	0.7484	12.58%

Table 7.

Model	Hold-out sample			
	Belgium	**The Netherlands**	**Germany**	**Generic**
National LR	0.7129 (16.92%)	0.7814 (16.50%)	0.7417 (15.74%)	N/A
Generic LR	0.7027 (17.02%)	0.7790 (16.68%)	0.7316 (16.16%)	0.7465 (16.48%)
Generic EXP	0.7029 (17.06%)	0.7777 (16.68%)	0.7315(16.16%)	0.7464 (16.54%)
Generic PH	0.7031(17.08%)	0.7777 (16.68%)	0.7315 (16.16%)	0.7463 (16.54%)
Stratified PH	0.7048 (16.74%)	0.7786 (16.62%)	0.7314 (16.18%)	0.7438 (16.50%)

Figure 1.

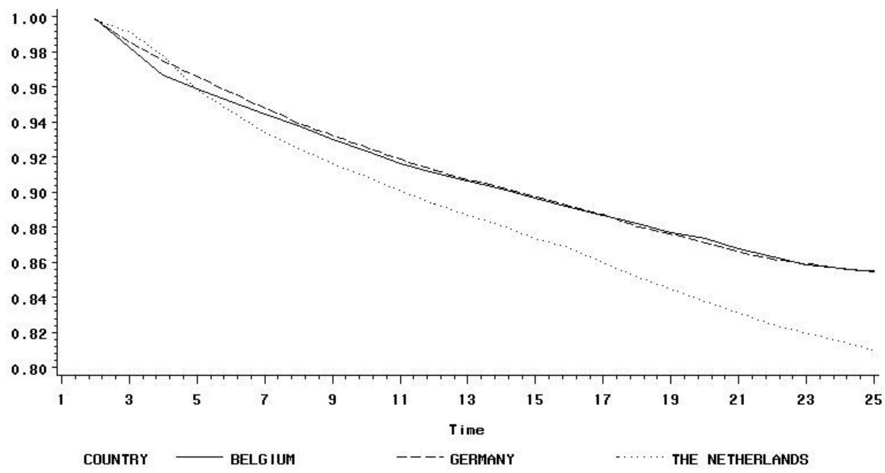
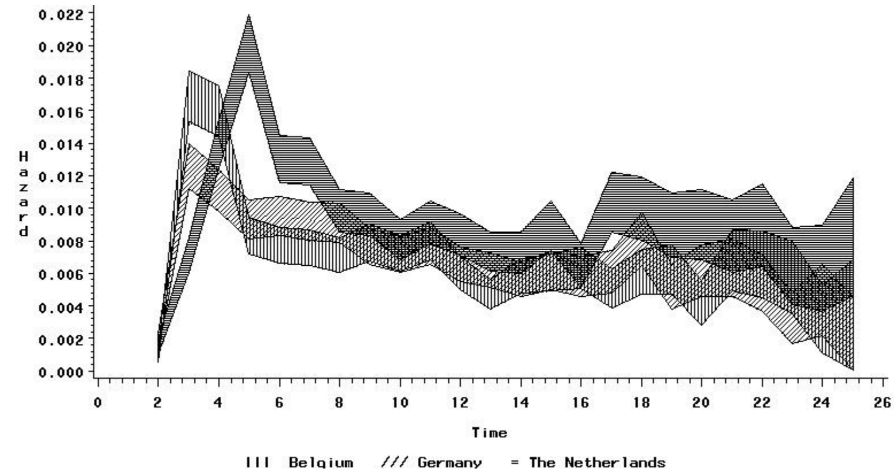

1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
1
3
5
7
9
11
13
15
17
19
21
23
25
Time
COUNTRY
BELGIUM
GERMANY
THE NETHERLANDS

Figure 2.



Hazard
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020
0.022
0 2 4 6 8 10 12 14 16 18 20 22 24 26
Time
||| Belgium /// Germany = The Netherlands

Captions for tables and figures

Table 1. Samples used in analysis

Table 2. List of characteristics used in analysis

Table 3. Test of equality of survival distribution function between countries

Table 4. Survival analysis and logistic regression models by country

Table 5. Model fit test of significance based on Log Likelihood

Table 6. Survival analysis models for different definitions of 'bad'

Table 7. Predictive performance of generic models

Figure 1. Survival Distribution Function by country

Figure 2. Hazard function with 95% confidence intervals by country