Received 14 May 2017
Revised 9 June 2017
Accepted 6 July 2017

Modeling credit risk in credit unions using survival analysis

M. Kabir Hassan, Jennifer Brodmann, Blake Rayfield and Makeen Huda
University of New Orleans, New Orleans, Louisiana, USA

Abstract

Purpose – The purpose of this paper is to investigate proprietary data from customers of a Southern Louisiana credit union. It analyzes the factors that contribute to an accelerated failure time (AFT) using information from customers' credit applications as well as information provided in their credit report.
Design/methodology/approach – This paper investigates the factors that affect credit risk using survival analysis by employing two primary models – the AFT model and the Cox proportional hazard (PH) model. While several studies employ the Cox PH model, few use the AFT model. However, this paper concludes that the AFT model has superior predictive qualities.
Findings – This paper finds that the factors specific to borrowers and local factors play an important role in the duration of a loan.
Practical implications – This paper offers an easily interpretable model for determining the duration of a potential borrower. The marketing department of credit unions can then use this information to predict when a customer will default, thus allowing the credit union to intervene in a timely manner to prevent defaults. Further, the credit union can use this information to seek out customers who are less likely to default.
Originality/value – This study is different from the previous research due to its focus on credit unions, which have distinct characteristics. Compared to similar lending institutions, the charter of the credit union does not allow management to sell off loans to other investors.

Keywords Survival analysis, Credit risk, Credit unions

Paper type Research paper

1. Introduction

Consumer loans provide a substantial amount of revenue for financial institutions and especially credit unions. The demand for consumer loans has grown tremendously since the 1980s, from 184 billion dollars to 1,305 billion dollars in May 2016, despite the 75 billion dollar-drop in consumer loans from January 2009 to February 2010 (FRED Economic data, Federal Reserve of St Louis, 2016). In May 2016, the Board of Governors of the Federal Reserve (2016) stated that "consumer credit increased at a seasonally adjusted annual rate of 6-1/4 percent." The demand for consumer loans is steadily growing and will continue to comprise a large proportion of financial institutions' and credit unions' balance sheets. It is important for financial institutions to target borrowers that will have a higher payment duration over the life of the loan. These types of loans can be problematic due to the difficulty of reliably assessing the risk associated with lending to this type of borrower. The existence of information asymmetry is a critical issue. Developing an accurate means of assessing each borrower's risk of default can provide a way to reduce this information asymmetry.

It is essential for credit scoring to accurately determine whether the lender should grant the borrower a loan based on information about the borrower as well as information about the current economic factors that may induce a borrower to default. A borrower may be fit to pay off the initial loan payments, but there may potentially be many events that

The authors would like to thank the editors and the anonymous referee for their enlightening comments and suggestions. The authors are grateful to the discussant and session participants from the Southwest Finance Association and Eastern Finance Association for their insightful feedback. The authors are also thankful to the Louisiana Credit Union for their support and providing the authors the access to the data that the authors have used in the model.

International Journal of Bank Marketing
Vol. 36 No. 3, 2018
pp. 482-495

0265-2323
DOI 10.1108/IJBM-05-2017-0091

occur over the duration of the loan which may inhibit a borrower from making payments. This is why modeling consumer credit risk is of vital interest to financial companies.

Credit unions have an especially heightened stake in credit risk analysis because of their unique characteristics, such as their inability to sell off loans to other investors after inception. Once issued, any loan must be held by the issuing credit union until repayment or default. Unlike other financial institutions such as commercial banks, credit unions are not-for-profit. As their mission statement is to serve the interests of their members, their earnings are disbursed to their members as higher interest rates, reduced fees, and decreased loan rates, rather than as dividend outlays to outside shareholders. The ownership structure of credit unions places great emphasis on democratic decision-making. Membership in a credit union is tantamount to ownership of that credit union. While each credit union is managed by the board of directors, each of its constituent members has an identical ownership stake. This means that each member has one vote, no matter how much money they have deposited in the credit union. The distinctive consumer-centric ethos of credit unions has been recognized by Congress, which exempted them from federal income taxes in the Credit Union Membership Access Act of 1998, proclaiming that "Credit unions, unlike many other participants in the financial services market, are exempt from Federal and most State taxes because credit unions are member-owned, democratically operated, not-for-profit organizations generally managed by volunteer boards of directors and because they have the specified mission of meeting the credit and savings needs of consumers, especially persons of modest means" (CUNA: The Credit Union Difference, n.d.). Credit unions' emphasis on serving their members has proven to be quite appealing to consumers. As of March 2016, there were almost 6,000 federally insured credit unions consisting of over 103 million members, with $799.5 billion in total loans (National Credit Union Administration, 2016).

Accurately assessing the credit risk has important implications to bank marketing. Soman and Cheema (2002) examine how consumers make decisions about using income to finance consumption. Shefrin and Thaler (1988) coined the term "consumption smoothing" to describe when consumers borrow money using credit to support their present lifestyle. The willingness to pay may surpass liquidity constraints when using credit. Assessing consumer credit risk can be part of determining the optimal consumer loan amount to encourage credit consumption. In addition, service features such as competitive interest rates can influence customer satisfaction and retention in retail banking (Levesque and McDougall, 1996). Marketers can use creditworthiness to target potential consumer loan borrowers who will have a higher duration of loan payments over the life of the loan. They can also use this information to offer more competitive rates for the loans.

Several empirical models have been developed to assess the consumer credit risk. Typically, these models involve a type of regression analysis or some form of probability model to determine the risk factors contributing to default. One model that is able to more naturally fit consumer credit risk questions is survival analysis. Survival analysis is a branch of statistics that deals with the analysis of lifetime data. It has been used previously for predicting the product life. Recently, researchers and banks have been becoming more aware of the usefulness of this model in assessing consumer credit risk.

"Credit scoring is a statistical method used to predict the probability that a loan applicant or existing borrower will default or become delinquent" (Mester, 1997). This is essential for financial and non-financial institutions to determine which borrowers to grant loans to. Credit scoring also contributes to the reduction of the information asymmetry involved in granting loans. This objective of credit scoring, also known as credit risk modeling, has recently shifted toward choosing the customers that will provide the highest profit. Because of this, loan offers must consider not only if a customer will default but also when they will default. The longer a customer pays without default, the greater the benefit to the firm. Therefore, the loan decision becomes a maximization problem where the distribution of a financial institution's assets is invested among the most profitable available opportunities.

We have chosen to use survival analysis methods for developing our credit scoring model. Survival analysis has a distinct advantage over other methods because it has characteristics that other models do not. One advantage of survival models is the ability to model time to default. Traditional methods focus on determining the factors that contribute to default. However, the interpretation of these results is limited in its ability to explain time to default. Thus, survival models can predict not only factors that contribute to default, but also the time when the default is more likely to occur. This characteristic is of heightened importance for our sample, because credit unions hold all loans until completion. This makes a portfolio maximization strategy more ideal for credit unions.

Another advantage of survival models is the ability to use incomplete information or "censored observations." Survival analysis applications typically involve censored data, data where an observation is incomplete for some of the credit records. This characteristic gives the researcher the ability to gather information about a sample or population even if some observations are incomplete. This is beneficial to this study, especially since credit applications may not always have complete information and in turn, observations can be incomplete.

This paper investigates the proprietary data on customers applying for consumer loans from a Southern Louisiana credit union. Our sample period has a time span of 35 years from 1981 to 2015. A proportion of the loans included in the sample were subject to charge-offs, an expense that both financial institutions and credit unions aim to minimize. The objective of this paper is to use survival analysis methods to create a credit scoring model that will predict time to default in such a manner as to minimize charge-offs. The information about time to default can then be used by marketers to tailor their financial products to the customers most likely to pay back the loan without default. We show the factors that contribute to an accelerated failure time (AFT) using information from customers' credit applications as well as information provided in their credit report. This study is exceptional because of the unique nature of credit unions. Compared to similar lending institutions (conventional banks), the charter of the credit union does not allow management to sell off loans to other investors. For any loan the credit union extends, it will hold this loan until repayment or default, which makes loan acceptance more important for credit unions.

2. Literature review

There have been several studies focusing on marketing for credit unions. Srivastava *et al.* (1984) propose a customer-oriented marketing approach to determine the market structures, using the financial/banking industry as an example. They examine the spending patterns of financial services clients by discussing the situations where customers are underserved by the current services offered. Lee (2002) analyzes financial service clients' preferences toward human interaction and self-service technology. Ralston and Wright (2003) study the measures credit unions need to take to achieve increased loan accessibility and reduced probabilities of loan default. They find this can be attained by identifying the high-risk borrowers, adjusting loan conditions, and tracking loan repayments after the loan is approved. Barboza and Roth (2009) use the order probit model to find the factors that impact overall customer satisfaction ratings. Byrne and McCarthy (2014) add to the perceived value and service quality literature relating to credit unions by analyzing the impacts of the technical and relational value propositions preferences of credit union members. Yee-kwong Chan (1997) states that transaction and financing convenience, such as higher credit limits, can encourage inactive credit card holders to increase their card usage. In the context of retail banks, Levesque and McDougall (1996) find that among the key determinants of customer satisfaction are service features such as competitive interest rates. Our study contributes to the bank marketing literature by providing a model that can be used for bank managers to assess borrower credit risk through survival analysis, hence allowing such banks to offer competitive interest rates.

Survival analysis techniques have been used in several studies to determine the probability of default. This can be seen in Cao *et al.* (2009), Bellotti and Crook, Madorno *et al.* (2013), Andreeva (2006), and Stepanova and Thomas (2002). Survival analysis can be used to uncover potential cross-selling and up-selling opportunities. Harrison and Ansell (2002) find that survival analysis can be used to predict who will be a financial institution client, what products that client may buy, and when the purchase is likely to be made. This is especially useful for determining what market segment to target, with what corresponding products, and in what time frame. Salazar *et al.* (2007) use survival analysis techniques to conclude that financial institutions should profile borrowers to develop business opportunities and design future retention strategies. Cao *et al.* (2009) find evidence supporting the effectiveness of survival analysis techniques by using the Cox proportional hazards (PH) model and the non-parametric conditional distribution estimation. Their results from these methods find "powerful discrimination between default and non-default credits," which shows the strength of these methods in determining the difference between credit issues that will or will not default.

Banasik *et al.* (1999) perform survival analysis on personal loan data from a foremost UK financial institution. Their data include 50,000 loans spanning the period of June 1994 to March 1997. Andreeva's (2006) data consist of a retail card issue in Belgium, the Netherlands, and Germany over a 25-month period from October 1998 to December 2000. Andreeva's implementation of survival analysis thus moves beyond the fixed-term credit studied in previous papers and into the arena of revolving credit. Andreeva uses various models, including AFT, PHs, and the more traditional logistic regression. Bellotti and Crook conduct a survival analysis on data from over 200,000 credit card accounts opened in the UK from 1997 to 2005. Tong *et al.* (2012) utilize a data set of 27,527 observations of consumer accounts from a single UK retail bank covering loan terms of 12, 24, and 36 months.

Several studies have shown findings that support the strength of survival analysis methods. Tong *et al.* (2012) provide research that connects the use of survival analysis in biostatistics to utilizing the same method in econometrics. They predict the probability of default on a UK loan portfolio using three methods: a mixture cure model, the Cox PHs method, and a standard logistic regression. The survival approaches were shown to provide more robust probability estimates in comparison to the standard logistic regression. Banasik *et al.* (1999) examine the probability of when a borrower will default as opposed to just the probability of if the borrower will default. They find that the ability of PH models to predict default in the first year of the loan term rivals that of logistic regression models. Furthermore, their findings suggest that PH models are better than logistic regression models in predicting whether a borrower will pay off their debt early within the first year. Bellotti and Crook find that survival analysis provides more predictive power than logistic regression due to its ability to incorporate time-variant macroeconomic variables such as retail banks' base interest rate, unemployment index, consumer confidence index, and house price index.

Madorno *et al.* (2013) use the Cox regression model "in order to write the default probability in terms of the conditional distribution function of the time to default." The findings from their analysis conclude that using survival analysis provides more robust results. Andreeva (2006) examines different "timescales of default" and finds strong support for the notion that the predictive power of survival analysis is comparable to that of logistic regression. Our paper further provides an empirical model that predicts time to default and contributes to the existing literature by creating a new data set from a Southern Louisiana-based credit union in the USA which illustrates the specific characteristics of credit union borrowers by pulling information about variables that influence the potential of a borrower to default.

The survival analysis techniques used in our paper are the AFT model and the Cox PH model. The main difference between the models is that the AFT is a parametric model while the Cox proportional model is non-parametric. Despite this, the results from these models tend to be quite similar. Credit scoring models should not just incorporate the factors associated with the potential borrower. Macroeconomic factors can have a great impact on whether a borrower will default on a loan. The global financial crisis has shown us that economic downturns affect both potentially "good" and "bad" borrowers. If a borrower experiences a layoff due to an economic downturn, this can greatly increase their chance of default. This is one of the reasons behind adding unemployment rate into our credit scoring model. Mishkin (2016) states that bank loan officers examine five main components when evaluating potential borrowers, namely "character, capacity (ability to pay), collateral conditions (in local and national economies), and capital (net worth)" (Mishkin, 2016). The empirical models used in our study factor in unemployment rates in order to incorporate collateral conditions into the model predictions. This extends to unemployment rate forecasts from the US Bureau of Labor Statistics, which will enable the model to be used in predicting the default probabilities of borrowers at any time in the future.

3. Data and methodology

This paper employs a unique proprietary data set. Furthermore, to our knowledge, no other paper available has researched credit scoring in the credit union setting. As stated in the introduction, this setting is unique because of the credit unions' inability to sell loans after they are created. This makes credit decisions more important for credit unions as they are exposed to credit risk for the duration of the loan. The loans are a mix of personal and auto loans. While we do not believe the mix of loans is an issue, we further investigate this issue later.

Using the customer base of a Southern Louisiana Credit Union, we gather credit information on a sample of customers. For each customer, we gather several factors that the literature has shown to be associated with credit scoring. The data come from individual credit reports, credit applications, and other sources used in the loan process.

Credit unions typically rank borrowers on a scale from A+ to E with A+ being considered as the safest borrower and E being considered as the most risky borrower. Loan officers have spoken about how this classification can be limiting because a "higher risk" borrower could be trying to put great effort into becoming a "lower risk" borrower. Such a borrower may be willing to do everything in their power to not default and make their payments on time in order to increase their credit score and no longer be considered a "higher risk" borrower. There are also "lower risk" borrowers who have higher incomes but have money tied up in risky investments or are spending an untenably high proportion of their income, which would over time increase their likelihood of defaulting on a loan. These situations have led to the coining of terminology such as "an E on its way up" and "an A on its way down" and provide reasons for looking at other means of assessing borrower risk than just the borrower's FICO credit score. Because of this, loan officers have relied on "soft" information in order to further determine which borrowers to provide loans to. We aim to provide a framework of a credit scoring model that incorporates additional factors beyond the credit score to assist loan officers in determining whom they grant loans to.

In order to gain insight beyond the blackbox "credit score," we gather detailed information from each customer's credit report. The variables we obtain from customers credit reports are: *PR*, total number of public records; *COL*, total number of collection accounts; accounts with a kind-of-business (*KOB*) code of "Y"; *NEG*, total number of accounts with a current manner of payment manner or payment (MOP) 2 or greater (30 days or more behind); *HISNEG*, historical negative: total number of accounts with an historical MOP 2 or greater and total number of times MOP 2 or greater ratings have historically occurred. Excludes current MOP; *TRD*, Total number of trades. *TRD* is the sum of *RVL*, *INST*, *MTG*, and *OPN*; *MTG*, total number of mortgage accounts (account type "M"); *OPN*, total number of open accounts (account type "O"); *INQ*, total number of inquiries; HighCredit, highest amount ever owed on an account; *CreditLim*, maximum credit amount approved by credit grantor; *Balance*, balance owed as of the date verified; *PastDue*, amount past due as of the date verified; and *MonthPay*, from the "TERMS" field on the account, subscriber reported monthly payment.

The credit application allows us to gather more personal information. Credit applications were pulled from the industry software Episys and the information needed to create the data variables was hand-collected and compiled. From the credit application, we gather age (in months), time to default (in months), time at current residence (in months), and months employed. From other sources we gather debt over income for both before and after the loan issuance. Table I shows the summary statistics for the data.

We have a total of 3,976 loans in our sample. Because of errors in the data or the collection process, some of the observations are missing. However, the number of missing variables is not more than ten percent of the total sample. To remove outliers, we winsorize the data, removing below the first and above the 99th percentile. This removes loans that default after one payment or never make any payments. We understand this is an import risk class for credit unions and we examine them later in the paper. After winsorizing, 49 percent of the sample consist of loans that have defaulted, with the average time to default being 30.72 months. This corresponds to approximately 2.5 years. The average age of the applicant is approximately 43 years. The maximum age in our sample is 93 and the minimum age is 20. The average loan amount is $12,251. Additional summary statistics can be found in Table I.

	Mean	Median	SD
Panel A: Consumer characteristics			
Time to default	30.72	22.00	53.67
Age (in months)	519.93	492.00	161.16
Credit score	619.63	647.00	182.11
Residence time	89.93	48.00	117.22
Months employed	85.31	48.00	98.69
Panel B: Credit characteristics			
DOI after	19.90	19.06	18.39
DOI before	17.21	15.43	17.08
PR	0.09	0.00	0.99
COL	0.89	0.00	2.43
NEG	0.61	0.00	1.76
TRD	9.36	1.00	12.88
RVL	4.35	0.00	7.27
INST	4.14	0.00	6.73
MTG	0.69	0.00	1.63
OPN	0.18	0.00	0.53
INQ	3.33	1.00	5.07
High credit	79,254.97	9,890.00	149,828.18
Credit limit	15,366.36	300.00	39,209.62
Balance	58,764.30	5,238.00	112,809.91
Past due	733.96	0.00	4,593.18
Month pay	854.19	295.00	1,452.76
Good/bad	0.49		

Notes: This table reports summary statistics for 3,976 consumer loans. The sample of loans is taken from a Southern Louisiana credit union. Panel A reports consumer traits as found on the initial loan application. Panel B reports credit characteristics as found on the credit report of the application

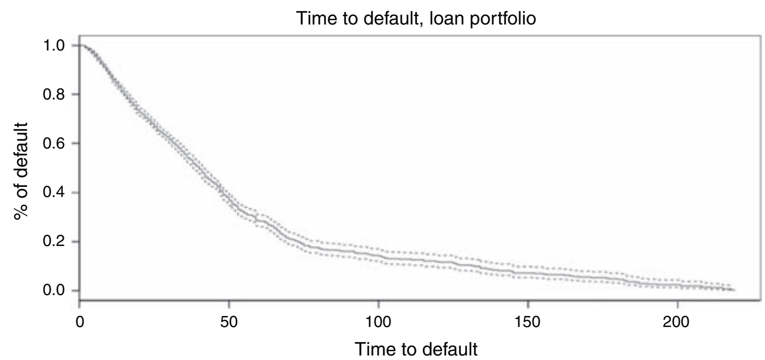
Table I.
Summary statistics

We test factors that may affect default time using the accelerated time model. While not as popular as the Cox PH model, because it does not require the data to fit to a distribution we believe the accelerated time model can validate our results.

Figure 1 shows the Kapen-Myer survival curve and (+) denotes a censored observation. We see that most loans that default do so within the first three years. Loans that are not in default are denoted as censored observations. In 40 months, more than 50 percent of the sample is no longer making payments. In 50 months, this percentage rises to more than 60 percent. Figure 2 shows the cumulative hazard function. Both of these figures provide a clearer picture of the sample of borrowers.

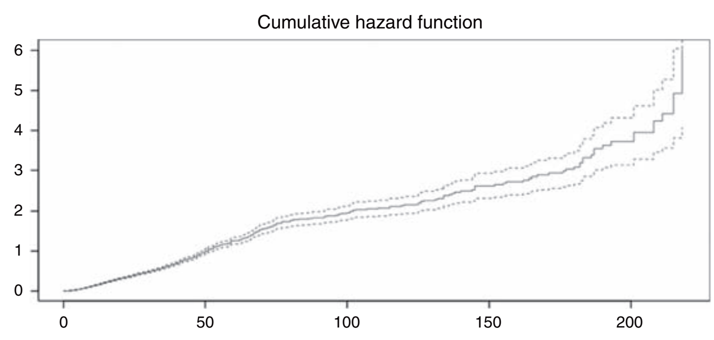
Accelerated time model

First, we test our data by creating a model using only residence time and the own or rent dummy variable, variables we gather only from the loan application. We use this as a baseline in order to judge the improvement in our model's explanatory power by using more detailed variables. Because the AFT model is a fully parametric model, we fit the survival



Notes: This figure shows the survival time of a sample of credit union consumer loans before winsorizing. The solid line depicts the observed time to default. The dashed lines present the 95% confidence interval proportion of surviving loans. The figure shows more than 50 percent of the sample defaults after 50 months

Figure 1.
Survival function



Notes: This figure shows the cumulative hazard of default for a sample of credit union consumer loans. The solid line depicts the observed cumulative hazard to default. The dashed lines present the 95% confidence interval proportion

Figure 2.
Cumulative hazard function

curve using two criteria, a log rank test and AIC. The results of both tests indicate log normal distribution is a good fit.

The hypothesis for Model (1) states that when taking into account a customer's credit history, residence history is not an important factor. The results are reported in Table II. Careful attention is taken to avoid correlation between variables. The leads us to debate the inclusion of *AGE* as a factor of loan duration. The inclusion of *AGE* may proxy for several credit report factors, such as, *INQ*, *MonthPay*, *TRD*, and *NEG*. For example, customers who are older may have a larger demand for goods and be more exposed to credit, thus having more opportunities for credit inquiries and higher monthly payments. Therefore, we decide to not include *AGE* in our final model. To investigate the importance of housing time and ownership status on survival time, we use *residence time* and *home ownership status* as covariates. In addition to housing status variables, we include months employed in every model. We expect this to have a positive effect on repayment. The results show *residence time* and *months employed* have a positive effect on survival time, which supports our hypothesis. *Residence time* is significant, which can be interpreted as an indication that a more stable housing and employment setting may be beneficial for repayment reliability.

The results of the credit application model can be found in regression (1) of Table III. Table III shows the other observable variables in the credit scoring process. We include the 12-month local unemployment rate and the variance over the same 12-month time period. We expect that the sign of the coefficients will be negative. This result would reflect that as unemployment rises, repayment suffers and during volatile employment periods, locally repayment also suffers. Model (1) of Table III shows that the inclusion of the two variables fits nicely. All selected variables from the credit report remain significant and both the

Model No.	(1)	(2)	(3)
Months employed	0.002***	0.001***	0.001***
	0.0003	0.0003	0.0004
Residence time	0.0004*	0.0001	0.0004
	0.0002	0.0002	0.0003
Ownership status	0.046	0.062*	0.004
	0.036	0.035	0.036
INQ			−0.032***
			0.005
Past due			−0.00001***
			0.00001
Monthly payment			0.0004***
			0.0001
TRD			0.037***
			0.006
NEG			−0.098***
			0.016
Credit score		0.001***	
		0.0001	
Constant	3.581***	2.770***	3.357***
	0.083	0.097	0.087
Observations	2,175	2,120	1,510
Log-likelihood	−4,107.69	−3,769.84	−2,608.37
χ^2	57.813*** (df = 3)	211.719*** (df = 4)	356.815*** (df = 8)

Notes: This table reports the application of the AFT model on consumer default. Model (1) investigates the impact of consumer characteristics on loan default time. Model (2) investigates the impact of consumer characteristics and credit score on loan default time. Lastly, Model (3) investigates the impact of consumer characteristics and credit scoring factors on loan default time. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table II.
AFT model

Model No.	(1)	(2)
Months employed	0.0004	0.001***
	0.0003	0.0003
Residence time	0.0002	0.0002
	0.0002	0.0002
Ownership status	0.022	0.063*
	0.031	0.034
INQ	−0.024***	
	0.004	
Past due	−0.00001***	
	0.00	
Monthly payment	0.0003***	
	0.0001	
TRD	0.035***	
	0.005	
NEG	−0.094***	
	0.013	
Credit score		0.001***
		0.0001
12 m unemp. observed	−0.205***	−0.083***
	0.029	0.03
12 m var. observed	−0.258***	−0.153*
	0.077	0.093
Constant	3.519***	2.878***
	-0.102	-0.118
Observations	1,444	2,055
Log-likelihood	−2,512.62	−3,744.21
χ^2	427.866*** (df = 10)	214.961*** (df = 6)

Notes: This table reports the application of the AFT model on consumer default. Model (1) investigates the impact of consumer characteristics, credit score factors, and 12 m observed unemployment on loan default time. Model (2) investigates the impact of consumer characteristics and 12 m observed unemployment on loan default time. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table III.
AFT model with employment

variance term and the unemployment term are negative. We see similar results in Model (2) when we replace the individual credit report variables with the numerical credit score. Model (1) leaves a lot to be desired. It shows that judging an applicant solely on the loan application is a poor method. Only the *residence time* variable is significant in Model (1). We believe that *residence time* is significant due to the possible correlation between *residence time* and *age*. In tests that are not reported, *age* is significant only in Model (1)'s specification, which therefore leads us to believe that it does not affect our results significantly.

In Model (2), we use the same variables from Model (1), however in this model we include the applicants' credit score. In this model, *credit score*, *months employed*, and *ownership status* are significant. The *credit score* variable is positive and highly significant. We expect the variable to be positive, showing that applicants who have a higher credit score are more reliable borrowers. While this may seem to contradict our previous hypothesis that *ownership status* does not affect repayment, we argue that *credit score* and *residence time* may be related, thus affecting our results. While taking into account that *ownership status* is only significant at a 10 percent level, we continue to propose that *employment history*, *credit score*, and *residence time* are important factors. Now that we have two models that are covered in the literature, we explore attributes from the data gathered from individual credit reports.

Similar to Model (1), Model (3) includes the variables gathered from the customers' credit reports. The results reflect a drop in the log-likelihood, with this decrease shown between Models (1) and (3). A log-likelihood ratio test between Models (1) and (2) as well as Models (2) and (3) confirm that Model (3) offers a greater predictive ability. Employment history remains significant showing that it is an important factor to consider even when taking into account credit history. The variable for credit inquires is negative and significant as expected, since a customer who has more inquires may default sooner. Both factors for poor payment history, namely *Past Due* and *NEG*, are negative and significant. *Monthly payment* is positive and significant. This variable may be acting as a proxy for a good credit score, i.e. more trustworthy borrowers can borrow more. All residence history variables are not significant. This may be evidence supporting the hypothesis that when credit history is taken into account, housing history does not offer additional useful information.

The factors we choose from credit reports are significant and there is an increase in the log-likelihood. Again, we find that residence situation does not affect repayment at any significant level. The *credit inquiry* variable is negative and significant, which indicates that more credit inquiries negatively affects survival time. *Days past due* and *negative credit* are also negative and significant. More interestingly, *month pay* and *trades* are positive. This leads us to an interesting interpretation, which is that approval and business from other trades is an indication of a good applicant. The results indicate it may be beneficial to consider individual credit score components as opposed to only considering the score.

Next, we consider the role of job outlook on a credit applicant's default time. Bellotti and Crook (2009) show that the macroeconomic factors can contribute to a survival model. Inspired by their methodology, we use several proxies for employment rate. We gather information about the credit union's region from the US Bureau of Labor Statistics. The first we use is a 12-month forward unemployment return. The return is calculated by $R_t = [U_{t+12} - U_t/U_{12}] \times 100$, were U is the unemployment rate and R_t is the return at time t. The results from the AFT model are presented in Table III. We also create a variable representing the 12-month variance of the unemployment rate for the local Southern Louisiana region. The primary advantage of using this methodology is banks can observe this return at the time of offering a loan.

The specification for this test is similar to Models (2) and (3) from Table III. The results confirm that the unemployment rate is a significant factor in credit default. The 12-month forward rate is negatively significant meaning a drop in the unemployment rate corresponds to a drop in the expected repayment reliability. Also, the variable representing the variance of unemployment is negatively significant. We interpret this to mean that in uncertain times there is high employment volatility and a decline in repayment reliability.

In Table IV, we replace the 12-month unemployment rate with unemployment projections from the BLS. The reasoning behind this stems from the analysis being able to utilize these variables during the actual credit decision. We expect the results to be similar, and the results confirm this. In Model (1), both unemployment projection terms are negative and significant. The results from Model (2) are similar. The results from this and the previous regressions are promising. To make these results useful in real time, historical unemployment projections can be used for the region. In unreported results, historical unemployment shows a similar pattern to the results of Table IV. From these results, we can be more confident that our method can be successfully applied in real time for making credit decisions.

Similar to the models reported in Table III, unemployment and unemployment variance are negative and significant. A drop in the unemployment rate corresponds to a drop in the expected repayment reliability, and in uncertain times there is high employment volatility and a decrease in repayment reliability.

Model No.	(1)	(2)
Months employed	0.0004	0.001***
	0.0003	0.0003
Residence time	0.0002	0.0002
	0.0002	0.0002
Ownership status	0.022	0.063*
	0.031	0.034
INQ	−0.024***	
	0.004	
Past due	−0.002***	
	0.00	
Monthly payment	0.0004***	
	0.0001	
TRD	0.025***	
	0.005	
NEG	−0.094***	
	0.013	
Credit score		0.001***
		0.0001
12 m unemp. projection	−0.105***	−0.103***
	0.069	0.13
12 m var. projection	−0.368***	−0.143*
	0.087	0.193
Constant	3.549***	2.872***
	0.102	0.118
Observations	1,444	2,055
Log-likelihood	−2,512.62	−3,744.21
χ^2	427.866*** (df = 10)	214.961*** (df = 6)

Notes: This table reports the application of the AFT model on consumer default. Model (1) investigates the impact of consumer characteristics, credit score factors, and 12 m projected unemployment on loan default time. Model (2) investigates the impact of consumer characteristics and 12 m projected unemployment on loan default time. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table IV.
AFT model with employment projection proxy

Cox PH model. In contrast to the AFT model, the Cox PH Model is a semi-parametric model of hazard rate. Where the AFT model determines the factors that affect repayment time, the PH model determines the factors that contribute to or decrease hazard (or default). This method is more popular because it is easy to deploy and does not require fitting the data to a distribution (for example the log normal distribution). However, data must fulfill the "PH" requirement.

To support the findings from the AFT model, we use the PH model with similar covariates. First, we must check the validity of the Cox PH assumption. We perform this by using the Grambsch and Therneau's (1994) test of weighted residuals. In unreported results, all covariates conform to the Cox PH assumption. For further verification, which is also not reported, we plot the residuals to examine if they are approximately normally distributed. In unreported tabulations, our results from the Cox model confirm the results of prior AFT models. For further verification of the results, we repeat the experiment using several of the most common tie methods, which are Efron (1977), Breslow (1974), and the exact method. Hertz-Picciotto and Rockhill (1997) state that the Efron method is a stronger method for handling ties. Despite this, the results of the unreported Cox Model regressions are strikingly similar. We therefore conclude that our findings are consistent and provide strong evidence of our hypotheses. Similar to the AFT model, we find that including factors from the customer credit report positively influences our model.

4. Marketing and managerial implications

The results from this study reveal a new means of assessing credit risk to consumer loans issued by credit unions by using the survival analysis. This can be used to help determine the optimal loan amount that should be issued to a borrower to achieve a longer survival time of payment on the loan. Emphasis on these factors can help tailor an effective marketing strategy to increase successful loans issued and decrease adverse selection. This model may be incorporated into an internal marketing strategy for credit union branches. Knowing how long a loan will survive will have implications on accounting and forecasting loan losses into the future. Integrating this knowledge into an internal marketing strategy can improve loan retention. When a credit union has information about when a customer is likely to default, it can be proactive and conduct a timely intervention before the customer defaults. Preventative actions such as financial counseling, following up with the customer (i.e. a friendly call to check their financial situation), and improved customer relations will have a more beneficial impact when they can be targeted at the customers who need them the most. Internal marketing concepts can bring effective results for credit union branches with limited resources (Tansuhaj *et al.*, 1991).

5. Conclusion

Our objective in this paper is to determine the relevant factors that can be incorporated into a reliable credit scoring model. Information gleaned by the use of this model can then be used by marketers to improve customer outreach and loan retention. We test the potential factors pulled from a borrower's credit report and credit application and also look at economic factors that may impact the probability and timing of a borrower to default on a loan. The results from our empirical tests suggest that certain factors such as months employed impact the probability of a borrower to default. Furthermore, we find several individual factors to be important in successfully predicting the timing of default. We can determine the relative impact of these factors found in a customer's credit report. Considering these factors individually is a significantly better predictor than considering credit score alone.

Approval and business from other trades can be used as an indication of a good applicant, as can macroeconomic factors such as unemployment rate. Taking into account individual factors from a consumer credit report can reveal a more accurate picture of a borrower's trustworthiness and reduce problems associated with the information asymmetry. The factors included in this study can be incorporated into an effective credit scoring model. Our findings, which provide a means for credit scoring that can be useful for credit unions as well as other financial institutions, are supported by both the AFT model and the Cox PH model.

6. Limitations and future research

Our study utilizes consumer loan data from a Louisiana-based credit union. Therefore, the results are limited. Conducting a survival analysis using multiple datasets from a larger sample of US credit unions will provide a more generalizable result. Survival analysis can also be utilized in studies examining the credit risk on loans issued by other types of financial institutions, peer-to-peer lenders (such as Lending Club, Prosper, and Sofi), and microfinance.

References

Andreeva, G. (2006), "European generic scoring models using survival analysis", *Journal of the Operational Research Society*, Vol. 57 No. 10, pp. 1180-1187.

Banasik, J., Crook, J. and Thomas, L. (1999), "Not if but when will borrowers default", *The Journal of the Operational Research Society*, Vol. 50 No. 12, pp. 1185-1190.

Barboza, G. and Roth, K. (2009), "Understanding customers' revealed satisfaction preferences: an order probit model for credit unions", *Journal of Financial Services Marketing*, Vol. 13 No. 4, pp. 330-344.

Bellotti, T. and Crook, J. (2009), "Credit scoring with macroeconomic variables using survival analysis", *Journal of the Operational Research Society*, Vol. 60 No. 12, pp. 1699-1707.

Board of Governors of the Federal Reserve (2016), " FRB: G.19 release – consumer credit. ", Board of Governors of the Federal Reserve, July 8, available at: www.federalreserve.gov/releases/g19/ (accessed July 12, 2016).

Breslow, N. (1974), "Covariance analysis of censored survival data", *Biometrics*, Vol. 30 No. 1, pp. 89-99.

Byrne, N. and McCarthy, O. (2014), "Value proposition preferences of credit union members and patronage activity", *International Journal of Bank Marketing*, Vol. 32 No. 6, pp. 567-589.

Cao, R., Vilar, J.M. and Devia, A. (2009), "Modeling consumer credit risk via survival analysis", *Statistics & Operations Research Transactions*, Vol. 33 No. 1, pp. 3-30.

CUNA: The Credit Union Difference (n.d.), available at: www.cuna.org/Thecredituniondifference/ (accessed July 15, 2016).

Efron, B. (1977), "The efficiency of Cox's likelihood function for censored data", *Journal of the American Statistical Association*, Vol. 72 No. 359, pp. 557-565.

FRED economic data, Federal Reserve of St Louis (2016), "Consumer loans at all commercial banks", July 1, available at: https://fred.stlouisfed.org/series/CONSUMER (accessed July 15, 2016).

Grambsch, P.M. and Therneau, T.M. (1994), "Proportional hazards tests and diagnostics based on weighted residuals", *Biometrika*, Vol. 81 No. 3, pp. 515-526.

Harrison, T. and Ansell, J. (2002), "Customer retention in the insurance industry: using survival analysis to predict cross-selling opportunities", *Journal of Financial Services Marketing*, Vol. 6 No. 3, pp. 229-239.

Hertz-Picciotto, I. and Rockhill, B. (1997), "Validity and efficiency of approximation methods for tied survival times in Cox regression", *Biometrics*, Vol. 53 No. 3, pp. 1151-1156.

Lee, J. (2002), "A key to marketing financial services: the right mix of products, services, channels and customers", *Journal of Services Marketing*, Vol. 16 No. 3, pp. 238-258.

Levesque, T. and McDougall, G.H. (1996), "Determinants of customer satisfaction in retail banking", *International Journal of Bank Marketing*, Vol. 14 No. 7, pp. 12-20.

Madorno, F., Mecatti, F. and Figini, S. (2013), "Survival models for credit risk estimation", in Carpita, M., Brentari, E. and Qannari, E.M. (Eds), *Advances in Latent Variables: Methods, Models and Applications*, Springer-Verlag, New York, NY.

Mester, L. (1997), "What's the point of credit scoring", *Business Review (Federal Reserve Bank of Philadelphia)*, Vol. 3, September/October, pp. 3-16.

Mishkin, F. (2016), *The Economics of Money, Banking, and Financial Markets*, 11th ed., Pearson, Boston, MA.

National Credit Union Administration (2016), "Industry at a glance", March 31, available at: www.ncua.gov/analysis/Pages/industry/at-a-glance.aspx (accessed July 15, 2016).

Ralston, D. and Wright, A. (2003), "Lending procedures and the viability-social objectives conflict in credit unions", *International Journal of Bank Marketing*, Vol. 21 Nos 6/7, pp. 304-311.

Salazar, M.T., Harrison, T. and Ansell, J. (2007), "An approach for the identification of cross-sell and up-sell opportunities using a financial services customer database", *Journal of Financial Services Marketing*, Vol. 12 No. 2, pp. 115-131.

Shefrin, H.M. and Thaler, R.H. (1988), "The behavioral life-cycle hypothesis", *Economic Inquiry*, Vol. 26 No. 4, pp. 609-643.

Soman, D. and Cheema, A. (2002), "The effect of credit on spending decisions: the role of the credit limit and credibility", *Marketing Science*, Vol. 21 No. 1, pp. 32-53.

Srivastava, R.K., Alpert, M.I. and Shocker, A.D. (1984), "A customer-oriented approach for determining market structures", *The Journal of Marketing*, Vol. 48 No. 2, pp. 32-45.

Stepanova, M. and Thomas, L. (2002), "Survival analysis methods for personal loan data", *Operations Research*, Vol. 50 No. 2, pp. 277-289.

Tansuhaj, P., Randall, D. and McCullough, J. (1991), "Applying the internal marketing concept within large organizations: as applied to a credit union", *Journal of Professional Services Marketing*, Vol. 6 No. 2, pp. 193-202.

Tong, E.N., Mues, C. and Thomas, L.C. (2012), "Mixture cure models in credit scoring: if and when borrowers default", *European Journal of Operational Research*, Vol. 218 No. 1, pp. 132-139.

Yee-kwong Chan, R. (1997), "Demographic and attitudinal differences between active and inactive credit cardholders – the case of Hong Kong", *International Journal of Bank Marketing*, Vol. 15 No. 4, pp. 117-125.

Further reading

Harrell, F. (2015), *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Springer-Verlag, New York, NY.

Corresponding author
M. Kabir Hassan can be contacted at: mhassan@uno.edu