“The role of credit card behavior in auto loan grant decision. An application of survival table”

AUTHORS	Clover M. Yeh Tsun-Siou Lee
ARTICLE INFO	Clover M. Yeh and Tsun-Siou Lee (2013). The role of credit card behavior in auto loan grant decision. An application of survival table. *Banks and Bank Systems*, *8*(1)
RELEASED ON	Tuesday, 30 April 2013
JOURNAL	"Banks and Bank Systems"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

Clover M. Yeh (Taiwan), Tsun-Siou Lee (Taiwan)

The role of credit card behavior in auto loan grant decision. An application of survival table

Abstract

Most of auto loan grant decisions are made on application data and these static data cannot capture the consuming behavior of applicants that contains more information about credit risk of auto loans. To improve the efficiency of auto loan grant decision model, not only auto loan application data but also credit card behavioral variables of applicants are included in this study because credit card is the most commonly held nonfinancial asset and most of auto loan applicants have credit cards. Based on above explanatory variables, a survival pre-warning model, proportional hazards model, is built in this study for auto loan grant decision since it tells not only if but also when will a loan default which may turn a loan of high default probability into a welcomed loan if the expected profit before default is higher than cost. This study also introduces a new credit-scoring system: survival table, similar to life table in insurance industry, provides probability of default or prepayment at every time point in the loan term which reduces the complication of auto loan grant decision. Evidence from Taiwan shows that both survival model and survival table are competitive with logistic model, the most widely-used credit model, in auto loan grant decision.

Keywords: survival analysis, auto loan, survival table, credit card, loan grant decision.
JEL Classification: G21.

Introduction

Accompanied with the change of consumer behavior, consumer credit loan market has grown rapidly so that the profit of consumer credit loan portfolios plays a decisive role in the revenue of financial institutions. Therefore, financial institutions over the world tried hard to expand their shares in consumer credit market in recent years. Unfortunately, the overexpansion of consumer credit market brought not only high return but also high risk. The credit card debt crises and subprime mortgage debacle occurred in Asia, the United States and Europe since 2003 brought financial institutions dramatic loss and led to the depression of consumer loan market. The economic impacts caused by the turmoil of consumer financial market are so critical that governments are now urged to take measures to combat the spread of financial collapse. Although credit risk is getting higher and higher, it is not smart for financial institutions to deflate consumer credit business because the global economic recession has already reduced their profits substantially. Facing the dilemma of increasing the profit or reducing the risk, an efficient grant decision process of consumer credit loans is indispensable for financial institutions. Credit debt crises have caused academic and industrial notices about consumer credit risk and relative researches bloomed. Prior consumer credit models, such as logistic regression, discriminant analysis and artificial neural networks, focus on the question whether consumer credit loans default by a given time in the future or not (Hayhoe et al., 1999; David, 2001; Limsombunchai et al., 2005). Based on the analysis results, a conservative financial institution may reject all consumer credit loans that have high probability to default in the loan term even though a loan defaults near the end of the loan term may bring a profit more than its cost before it defaults. Therefore, not only if but also when will a consumer credit loan default become a more important question (Thomas at el., 1999). To deal with the new concern, Narain (1992) first adopted survival analysis (Cox, 1972), a methodology usually used in medical science and biology, to build consumer credit models because it provides not only the expected default time but also the default probability of each time point in the future which is very helpful in consumer credit loan grant decision.

Among all consumer credit loans, automobile purchase is one of the most common loan purposes in consumer credit market so that improving the efficiency of auto loan grant decision is in haste for financial institutions to diminish the credit risk and enlarge the profit of consumer credit portfolios. Therefore, this study focuses on the credit risk of auto loans and uses survival analysis and survival table as tools for auto loan grant decision. What is the main limitation of current auto loan grant decision process? Most of auto loan grant decisions nowadays are made based on application data, including loan characteristics and demographic data (Eberly, 1994; Heitfield and Sabarwal, 2003; Agarwal et al., 2008). However, these static data cannot capture the consuming behavior of applicants that contains more information about credit risk of auto loans. To solve this problem, not only auto loan application data but also credit card behavioral variables of applicants are included in the auto loan grant decision model because credit card is the most commonly held nonfinancial asset and most of auto loan applicants have credit cards. Two more contributions about auto loan grant decision are built in this study. First, this study builds a survival model of auto loans based on application data and credit card behavioral variables. Second, a process of building survival table for auto loan grant decision is introduced in this study. A survival table, like a life table in insurance, provides the default probabilities at every time point in the future. The user-friendliness of survival table reduces the difficulty of auto loan grant decision because every employee can find out default risk of every auto loan applicant whether he can build a survival model himself or not. This study also contains an empirical study based on auto loan and credit card data of a major Taiwan financial institution to compare the prediction capabilities of survival model and survival table with logistic model, the most widely-used model in current consumer financial market.

The study is structured as follows. Section 1 outlines proportional hazards model, one of the most expansively-used survival analysis methods. It also describes the method of building survival table. Section 2 presents the empirical data and prediction variables of survival model. Section 3 contains the empirical study which includes the comparison of the survival model, survival table, and traditional logistic model. The final section concludes and outlines some directions for further research.

1. Proportional hazards model and survival table

Proportional hazards model is one of the most widely-used survival model that connects the explanatory variables to survival time. The survival time of a sample is defined as the period since the beginning of observation to the time of default event. A sample is called a complete data if this sample default event occurs in the period of research whose beginning and end of survival time can be observed. Otherwise, it is called a censored data or an uncompleted data. Suppose that the default event happens at time T, the probability for a sample to survive at time t before time T is represented by the survival function, and the default probability of unit time is measured by hazards function which means a sample defaults at the next moment in case that this sample is survival at time t. The relationship between survival function and hazards function is

$$S(t)=P(T\geq t)=\lim_{\Delta\tau_k\to 0}\prod_{k=0}^{r-1}[1-h(\tau_k)(\tau_{k+1}-\tau_k)] \tag{1}$$

where $0=\tau_0<\tau_1<...<\tau_{r-1}<\tau_r=t$

The proportional hazards function is defined as

$$h(t;X)=h_0(t)\exp(\beta X(t)), \tag{2}$$

where $h_0\,(t)$ represents the baseline hazards function at time t when $X(t) = 0$.

If an auto loan debtor i defaults at time t_i, than the information ratio of this sample compared with the whole risk set is

$$\frac{\exp\{\beta X_i(t_i)\}}{\sum_{l\in R\{t_i\}}\exp\{\beta X_l(t_i)\}}, \tag{3}$$

where $R\{t_i\}$ is the whole risk set at time t_i and $X_l(t_i)$ is the explanatory matrix of debtor l. And Cox models use maximum likelihood method to estimate the coefficients β. Assuming there are n auto loans default in the observation period and the default times are $t_1, t_2,..,t_n$ in turn, the log-likelihood function could be obtained by summing up the log value of risk information:

$$L(\beta)=\sum_{i=1}^{n}\left(\beta' X_i(t_i)-\log\left[\sum_{l\in R\{t_i\}}\exp\{\beta' X_l(t_i)\}\right]\right). \tag{4}$$

For an auto loan, there are two competing risks in survival models: default and prepayment. According to rule of thumb, an auto loan is defined as default if the time of overdue is more than 60 days, which is usually called as M2 status in consumer loan market. The survival function can be used to estimate the distribution of both T_d, the lifetime of the auto loan until default, and T_p, the lifetime of the auto loan until early repayment. Define T_m as the end of loan term, and the lifetime of an auto loan is

$$T=\min\{T_d,T_p,T_m\}.$$

The observation time period of this study is the first twelve months of loan term. That is, the survival time for a censored auto loan is from $P + 1$ to $P + 12$ where P is the initial time of auto loan. Otherwise, survival time is from $P + 1$ to T.
One of the most important contributions of this study is building survival table that contains default probabilities of all groups at every time point in the future, just like the widely-used life table in insurance industry. The first step of constructing a survival table is giving a credit score to each auto loan after building a proportional hazards model based on training samples. For each significant ordinal explanatory variable, give every auto loan a score from zero point to nine points based on the ordinal number. As for significant nominal variables, add different scores to auto loans according to their classification. By summing up all scores of significant variables, the credit score of every auto loan can be obtained and all auto loans can be sorted into groups with the credit scores. The credit score of each auto loan is connected to survival probability and the greater credit score is, the lower survival probability is. Therefore, auto loans in the first group have the highest survival probabilities and loans in the last group have the highest default probabilities.

Survival table is constructed as following steps. First, build a proportional hazards model with training data and get the value of baseline hazards function at every time point and coefficient matrix β. Second, give every auto loan a credit score based on above process and sort all auto loans into groups with their credit scores. Third, based on the coefficient matrix, compute the average value of βX for every group for training data as well as testing samples. Comparing the average βX curves of training set and testing set, it is found that there is a smooth shift which may be attributed to the macroeconomic change in different time horizon. Since the differences of these two curves are close to a constant, the average value of differences between training groups and testing groups, defined as increment C, can be a proxy that captures the effect of macroeconomic change. Add the average increment to the value of βX for every group in testing set, then values of modified $\beta X'$ can be obtained. Computing the hazard rate and survival probability of every group based on the values of baseline function and modified $\beta X'$, then survival table is completed. Survival table that contains prepayment risk information can be easily constructed by the same method, too. It should be noted that the increment is also engaged in the survival model because removing the effect of macroeconomic changes will help the model to fit better.

2. Loan data and prediction variables

The proportional hazards model of auto loans contains two kinds of prediction variables: auto loan application variables, including loan characteristics and demographic data, and credit card behavioral variables. The application characteristics include five variables: application date, application amount, loan amount, loan duration, and car used time. Demographic data include eight variables: gender, education level, marriage status, occupation, job title, work experience, annual income, and income certificates. It is noted that there are 7 kinds of income certificates, such as certificate of deposit and tax deductible receipt, provided by auto loan appli cants to support their income status. The behavioral variables of credit cards include the monthly performance data in the performance period, the last twelve months before observation period, i.e. P-1 to P-12. There are seven performance items: previous balance, sales amount, cash advanced amount, total amount payable, repayment amount, minimum repayment, and minimum repayment of last transaction.

In addition to monthly behavioral variables for the last twelve month, there are some extended variables. For each repayment item, there are two dimensions in concern: duration and statistic values. The duration defined in this study includes short-term data, mid-term data and long-term data, which means data of the last quarter (P-1 to P-3), the last six months (P-1 to P-6), and the last twelve months (P-1 to P-12), respectively. Moreover, to capture the deviant performance, two increments are included: the difference between long-term and mid-term and the difference between mid-term and short-term. The statistic values include mean, standard deviation, maximum, minimum, and summation. Finally, there are total 259 behavioral variables engaged in survival model. Because there are so many prediction variables in survival analysis which makes the model very complicated and the behavioral variables seem to be linear dependent that makes the model unreliable, the method of principal component is adopted in dealing with the behavioral variables to simplify the model and ensure the linear independence of variables. Finally, there are eighteen principal components, whose eigenvalues are greater than one, are included in the model and they are defined as factor 1, factor 2, etc."

Besides these principal components, four more credit card behavioral variables are included in the model: overdue, time of over due, block code, and time of block code. Overdue is a dummy variable that represents if the credit card is in the status of overdue. Block code represents the last record of overdue, reissue, over-consumption, and suspension, which is not surprising to capture the possibility of default. Finally, total thirty-five variables, including auto loan application information and credit card behavioral variables, are included in survival model.

Since the 1997 Asian financial crisis, financial institutions in Taiwan focused on consumer credit market because of the recession of corporate finance. In order to enlarge the profit, financial institutions increased the weight of consumer credit loans gradually and this changed consuming habit of individuals and increase their consumption cost in daily life. The rapid growth of consumer credit market finally led a serious double card debt crisis in the fourth quarter of 2005. Until the first quarter of 2006, the non-performing loans ratio of credit card and cash card peaked and hence inflicted severe damage on financial industry.

To compare the prediction power of candidate models, about nine thousand auto loan samples of a major Taiwan financial institution are engaged in the empirical study. To test the prediction capability of survival model in dramatic fluctuation, auto loan samples are divided into two sets, training set and testing set, according to their initial dates are before or after September 2005 because the double card debt crisis of Taiwan occurred in the last quarter of 2005 and consumer financial market expanded significantly before the crisis. That is, samples approved before September 2005 are included in training set to build the model. Since the observation period is the first twelve months in the loan term, the modeling structure is incomplete until September 2006 so that auto loan samples approved after October 2006 are included in the testing set. In order to make sure the completeness of credit card behavioral variables, auto loan samples are excluded if the applicants do not have complete credit card behavioral data before application date *P*. Finally, there are total 6954 auto loan samples included in the empirical study, 4766 of them are training samples and other 2249 samples are used to test the model. Table 1 shows the repayment behavior of auto loans in training set and testing set during the observation period.

Table 1. Repayment performance of auto loan samples

	Training set		Testing set	
Default	32	0.67%	29	1.29%
Not default	4734	99.33%	2220	98.71%
Prepaid	105	2.20%	44	1.96%
Not prepaid	4661	97.80%	2205	98.04%
Censored	4366	91.61%	2107	93.69%
Total	4766	100.00%	2249	100.00%

3. Results of empirical study

Most financial institutions nowadays take logistic regression into consideration when they make loan grant decisions because it is one of the simplest credit models. Therefore, reducing the complication of survival analysis by including the method of principal component is the first step to fit the need of financial industry. Furthermore, ensuring the simplification will not reduce the prediction power of survival model is the second step. This section will show the empirical results of survival analysis based on auto loan samples of a major Taiwan financial institution. Furthermore, a comparison of survival model and survival table with logistic regression will show that both of survival model and survival table are competitive with the major tool used in auto loan grant decision. To test the practicability of including credit card behavioral variables, two survival models and two logistic models are built in empirical study. First, survival model and logistic model are built based on application variables. Then based on not only application variables but also credit card behavioral variables, advanced survival model and advanced logistic model are built.

The results show that survival model contains four significant explanatory variables in default prediction: application date, gender, occupation, and work experience. And the logistic model contains all significant explanatory variables except occupation. On the other hand, the advanced survival model contains not only all significant explanatory variables of survival model, but also three significant principal components of credit card behavioral variables: factor 3, 4, and 7. Similarly, advanced logistic model contains the three factors besides all significant explanatory variables of logistic model. As for prepayment prediction, both of survival model and logistic model contain three significant explanatory variables: gender, loan amount, and loan duration. And both of advanced survival model and advanced logistic model contain above three significant explanatory variables and one significant principal component: factor 4.

After constructing proportional hazards models for default and prepayment, the process of building survival table mentioned in section 1 is applied to set up two survival tables for default and prepayment. Based on the results of advanced proportional hazards model for default prediction, most auto loan samples get credit scores in the interval between twenty points and one hundred and twenty points. Therefore, this study categorizes these loan samples into twenty-two groups. The first group includes auto loans which credit scores are lower than twenty points and the last group includes loans that get credit scores higher than one hundred and twenty points. Other auto loans are divided into twenty groups with their credit scores. Figure 1 shows the number of loans in every group. It also shows and the proportion of auto loan samples not default in the observation period and the expected probability of every group. It is shown that realized survival rates are close to the expected survival probabilities and both of them reduce with groups. That is, the higher credit score one auto loan gets, the higher probability this loan will default. Similar results were obtained for prepayment prediction. It should be noted that the descending survival probabilities from the first group to the last group show the validity of grouping rule and hence ensure the practicability of survival table.

To simplify the illustration, this study only shows part of survival table for default prediction in Table 2 to describe the structure of this credit scoring system. Survival table enables financial institutions know the change of an auto loan easily even for a staff not familiar with modeling approaches of survival model. To find the future survival probabilities of an auto loan, the staff only needs to calculate its credit score then he can find its survival probability for every time point in the future. For example, if an auto loan has a credit score of forty-three points, then it belongs to the sixth group. If the staff wants to find the expected default rate of the loan at the two and half year after approved, he can easily find the expected survival probability in the survival table so that he will know that the expected default probability of the loan after two and half year is 0.17%.

Table 2. Survival table for default prediction (part)

Group / Time	25	26	27	28	29	30
1	0.9998	0.9997	0.9996	0.9995	0.9995	0.9993
2	0.9995	0.9994	0.9992	0.9991	0.9989	0.9987
3	0.9995	0.9993	0.9992	0.9990	0.9988	0.9986
4	0.9994	0.9993	0.9991	0.9989	0.9987	0.9985
5	0.9994	0.9992	0.9990	0.9989	0.9987	0.9984
6	0.9994	0.9992	0.9990	0.9988	0.9986	0.9983
Group / Time	25	26	27	28	29	30
19	0.9982	0.9976	0.9971	0.9966	0.9960	0.9952
20	0.9979	0.9973	0.9967	0.9961	0.9954	0.9945
21	0.9976	0.9969	0.9962	0.9955	0.9947	0.9937
22	0.9971	0.9962	0.9954	0.9946	0.9936	0.9923

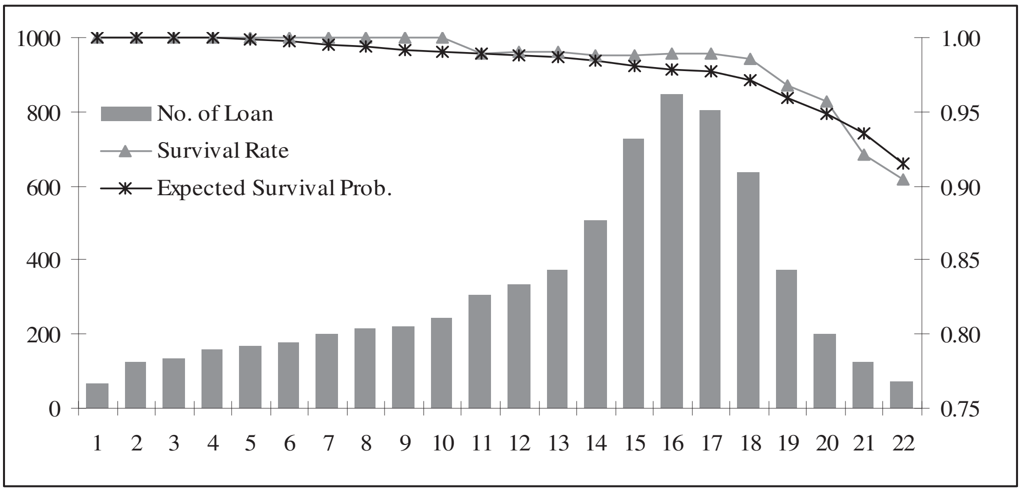

Fig. 1. Realized survival rate and expected survival probability (default prediction)

Table 3 and Table 4 show the comparisons of using survival models, survival tables and logistic models as tools of auto loan grant decision about default and prepayment, respectively. It is shown that the prediction capabilities of these three methods are rised by including the credit card behavioral variables because the type I errors of all advanced methods are much lower than these of their corresponding original models. Moreover, the participation of credit card behavioral variables reduces the type II errors of survival model and survival table at the same time. Thus, it is concluded that the efficiency of auto loan grant decision models is improved by including credit card behavioral variables among explanatory variables.

The results also show that survival method is competitive with the logistic regression approach since the type I errors of advanced survival models are lower than these of their corresponding logistic models in default and prepayment prediction. Although advanced survival models have higher type II errors than their corresponding logistic models, the lower type I errors make advanced survival model a qualified model for auto loan grant decision. Moreover, the characteristic of survival model that tells the default probability of every time in the loan term ensures the practicability of survival model because it helps financial institution to monitor the expected cash flow of every auto loan in the future. It is concluded that survival model, with the simplification of principle component method, is competitive with logistic model in auto loan grant decision. Similar results are found about using survival table as a tool of auto loan grant decision. Although advanced survival tables have higher type II errors than the corresponding logistic models, the lower type I errors make them qualified for auto loan grant decision method. Furthermore, the characteristic of user-friendliness reduces the difficulty of auto loan grant decision and ensures the practicability of survival table. It should be noted that, like other credit models, re-building a survival table once or twice a year will promise the accuracy of survival table.

Table 3. Prediction capabilities of survival model, logistic model, and survival table for default prediction

	Survival model		Logistic model		Survival table	
	Original	Advanced	Original	Advanced	Original	Advanced
Predict good and result in good	2147	2188	2200	2196	2139	2183
Predict good but result in bad	13	10	26	12	18	10
Predict bad but result in good	73	32	20	24	81	37
Predict bad and result in bad	16	19	3	17	11	19

Table 4. Prediction capabilities of survival model, logistic model, and survival table for prepayment prediction

	Survival model		Logistic model		Survival table	
	Original	Advanced	Original	Advanced	Original	Advanced
Predict good and result in good	2122	2158	2181	2164	2115	2149
Predict good but result in bad	24	14	31	16	28	15
Predict bad but result in good	83	47	24	41	90	56
Predict bad and result in bad	20	30	13	28	16	29

Conclusion and further research

Three contributions about auto loan grant decision are made by this study. In addition to application data, credit card behavioral variables are included in auto loan grant decision models because credit card is the most commonly held nonfinancial asset and most of auto loan applicants have credit cards. Second, this study constructs a pre-warning survival model, proportional hazards model, for auto loan grant decision. This study also establishes survival table as a tool for auto loan grant decision. An empirical study with auto loan samples of a major Taiwan financial institution shows that the efficiency of auto loan grant decision models is improved by including credit card behavioral variables among explanatory variables. It also shows that both of survival model and survival table are competitive with logistic model, the most widely-used model in current financial world. The lower type I errors guarantee the lower possibility of mistaking a loan with high probability of default or prepayment for a good loan. Moreover, the characteristic of survival model that tells the default probability of every time in the loan term ensures the practicability. Like life table in insurance industry, the user-friendliness of survival table makes it a doorkeeper in auto loan grant decision because every staff of financial institutions can easily find the default and prepayment probabilities of an auto loan at every time point in the future by checking survival tables instead of building a complicated model himself.

The empirical data of this study is the internal data of a major Taiwan financial institution which contain only application data of auto loans and behavioral data of credit card in the institution. To consider the completeness of data and the speed of data updating, if further researches can operate survival model and survival table in coordination with the internal data and nationwide data, such as data of Joint Credit Information Center, may raise the accuracy of models. On the other hand, these two survival methods can be widely used in other similar consumer loans such as mortgage loans in further researches with different explanatory variables setting.

References

1. Agarwal, S., Ambrose, W. and Chomsisengphet, S. (2008). Determinants of Automobile Loan Default and Prepayment, *Economics Perspectives*, Federal Reserve Bank of Chicago, pp. 17-28.
2. Cox, D.R. (1972). Regression Models and Life-tables, *Journal of the Royal Statistical Society*, Series B, Vol. 34, No. 2, pp. 187-220.
3. David, J.H. (2001). Modeling Consumer Credit Risk, *Journal of Management Mathematics,* Vol. 12, No. 2, pp. 139-155.
4. Eberly, J.C. (1994). Adjustment of Consumers' Durables Stocks: Evidence from Automobile Purchases, *Journal of Political Economy*, Vol. 102, No. 3, pp. 403-436.
5. Hayhoe, C.R., Leach, L. and Turner, P.R. (1999). Discriminating the Number of Credit Cards Held by College Students Using Credit and Money Attitudes, *Journal of Economic Psychology*, Vol. 20, No. 6, pp. 643-656.
6. Heitfield, E. and Sabarwal, T. (2003). What Drives Default and Prepayment on Subprime Auto Loans? *Journal of Real Estate Finance and Economics*, Vol. 29, No. 4, pp. 457-477.
7. Limsombunchai, V., Gan, C. and Lee, M. (2005). An Analysis of Credit Scoring for Agricultural Loans in Thailand, *American Journal of Applied Sciences*, Vol. 2, No. 8, pp. 1198-1205.
8. Narain, B. (1992). Survival Analysis and Credit Granting Decision, in: Thomas, L.C., Crook, J.N., Edelman, D.B. (eds.), *Credit Scoring and Credit Control*, Oxford University Press, Oxford, pp. 109-122.
9. Thomas, L.C., Banasik, J., and Crook, J.N. (1999). Not if but When will Borrowers Default, *Journal of the Operational Research Society*, Vol. 50, No. 12, pp. 1185-1190.