Duration of consumer loans and bank lending policy: dormancy versus default risk*

Kenneth Carling[†] Tor Jacobson[‡] Kasper Roszbach[§]

Working Paper Series in Economics and Finance No. 280
November 1998

Abstract

A bank that lends money to a household faces two types of risk. Most commonly mentioned is the risk of a default. Hardly ever referred to is the risk of an early redemption of the loan - leading to dormancy. We model consumer loans' transition from an active to a dormant state and estimate a semi-parametric duration model with a data set consisting of 4,733 individuals who were granted credit by a Swedish lending institution between 1993 and 1995. We analyze the factors that determine the time to maturity on a loan and investigate the model´s ability to match the maturities observed in the data. The model is used to evaluate loan applicants by their expected durations and - profits, and to derive the distribution of conditional expected durations and - profits for the loan portfolio. This enables us to draw some conclusions about the efficiency of bank lending policy.

JEL Classification: C14, C41, C53, D61, D81, G21.

Keywords: Bank lending policy, duration analysis, semi-parametric methods, dormancy, cost-benefit analysis.

*The authors thank Lars Ljunqvist for his helpful comments and Björn Karlsson and Yngve Karlsson at Upplysningscentralen AB for providing and discussing the data. Carling's research is funded by the Sasakawa Young Leaders' Foundation. Roszbach gratefully acknowledges financial support by the Jan Wallanders and Tom Hedelius foundation for research in the social sciences.

[†] Office of Labor Market Policy Evaluation (IFAU), Box 513, SE 751 20 Uppsala, Sweden; *Email:* kenneth.carling@ifau.uu.se.

[‡] Research Department, Sveriges Riksbank, SE 103 37 Stockholm, Sweden; *Email:* tor.jacobson@riksbank.se.

[§] Department of Economics, Stockholm School of Economics, and Research Department, De Nederlandsche Bank, Box 98, 1000 AB Amsterdam, The Netherlands; *Email:* nekr@hhs.se.

1. Introduction

Anyone applying for a loan at a bank can count on being subjected to one of the many procedures that exist to assess creditworthiness. Some credit institutions continue to analyze applicants' personal characteristics in subjective, non-formalized ways while others use so called credit scoring models. Their varying degrees of sophistication in spite, credit scoring models share a common objective: to separate loan applicants that are expected to pay back their debts from those who are likely to fall into arrears or go bankrupt. Altman, Avery, Eisenbeis and Sinkey [1] offers a good review of this literature.[1]

Unfortunately, credit scoring models leave much room for subjective factors in the loan approval process. By ranking customers according to predicted default probabilities, banks can minimize the expected default or misclassification rate subject to some exogenous acceptance rule. However, this is a far cry from solving the ultimate problem of profit or utility maximization. Scoring models make it possible for banks to predict bankruptcy, but require them to resort to ad-hoc methods to forecast profitability. Boyes, Hoffman and Low [8] and Jacobson and Roszbach [20] provide empirical evidence that confirms the authenticity and importance of this dilemma. In two independent studies of the bank credit scoring problem, they find that loan granting policies are consistent with neither default risk - nor loss rate minimization.

The quantitative importance of consumer credit for financial institutions in Sweden may be illustrated by the fact that it makes up 25 percent of total lending, excluding residential loans, to the public.[2] Even from the household perspective, consumer credit has come to play an increasingly important role as a means to maintain consumption at a level consistent with permanent income. Total lending, excluding residential loans, by banks and finance companies to Swedish households amounted to SEK 207 bn., or SEK 22.698 per capita, by the end of 1996. This is the equivalent of 12 percent of Swedish GDP or 22.7 percent of total private consumption. These numbers demonstrate that the process by which a credit institution decides to grant or refuse a loan can greatly affect households' ability to smooth consumption and thereby their welfare. Lending policy is thus of interest not merely because of its implications for bank profitability, but also

[1] The most commonly used methods are discriminant analysis and logistic regression. Recent studies have even applied k-nearest-neighborhood [17] and count data models [16], classification trees and neural networks [3].

[2] If one includes residential loans in total lending, this figure drops to 11 percent.

because of its implications for households' welfare. Nevertheless, credit rationing remains a common phenomenon in financial markets.

The expected return on a loan depends on several factors such as the interest rate, the amortization scheme, fees paid by the customer, fixed and variable costs incurred by the lender, the probability of a default, and the expected loss on the principal in case of a default. Except for fluctuations in the rate of interest, that we abstract from in this paper, a bank thus faces two types of risk when lending money to a household. Most commonly mentioned is the risk of a default. Hardly ever referred to is the risk of an early redemption of the loan, leading to dormancy. Since the provision and administration of credit involves mostly fixed costs while revenues are variable and spread out over time, loans with short maturities can result in losses for financial institutions, despite the fact that they never default. Moreover, because defaults generally occur after a period of regular installments and interest payments, the discounted loss on the average bad loan exceeds the profit on the average good loan by only little. In our dataset, the discounted loss on an averagely sized bad loan with the average duration to default is -36.7 percent compared to a return of 8.1 percent on a good loan. Combining these figures with the fact that defaulted loans generally make up only one percent of the portfolio for this type of loan, it appears a matter of course that dormancy, rather than default risk, is the economically important phenomenon to study.

The status of dormancy is a condition that every credit account sooner or later will arive at. It is therefore natural to study this phenomenon using survival, or duration, models. The classical application of duration models in economics is the analysis of unemployment survival. Kiefer [21] provides an accessible introduction to this class of models. Carling, Edin, Harkman and Holmlund [10] is a recent contribution in the area of labor market economics. Other examples of economic applications of duration analysis are Diebold and Rudebusch [15], who study duration dependence of expansions and contractions in the business cycle, and Baek and Bandopadhyaya [4], who analyze the factors that affect the length of commercial bank debt rescheduling negotiations. Bandopadhyaya [5] and Lane et al. [23] study bank failure.

In this paper, we model consumer loans' transition from an active to a dormant status and analyze the factors that determine the time to maturity. We also derive the distribution of conditional expected durations of loans in order to examine bank lending policy. In Section 2, we present the data set and its sources. The data consists of 4,733 individuals who were granted a consumer loan by a Swedish lending institution between September 1993 and August 1995. When monitored in October 1996, the loans had either been amortized in full or continued to pay regular installments and interest. Except for the duration of each loan, the data set contains detailed information on a range of demographic, financial and credit history variables for each borrower. This information has been extracted from a number of registers by using the personal number that every resident of Sweden has. In Section 3, we start by analyzing the duration of loans until failure. Default risk has traditionally been modelled by discrete choice models, letting the outcome variable be failure or not. To create a uniform framework for analyzing bad and good loans and to contrast the implications of the dormancy models, we also investigate the determinants of the time to failure. Next, we model the transition from activity to dormancy and relate this transition to the characteristics of the loan applicant. At the end of the section we address a number of conceptual problems that arise in modelling consumer credits by econometric duration models, as well as practical problems in revealing the functional relationship between the transition and the covariates. Finally, using the predicted time to dormancy we calculate the expected return on each loan and compare it with the revenue on a benchmark (break-even) loan. Section 4 concludes the paper with a discussion of the results.

2. Data

The data set consists of all 4,733 individuals who were granted a loan at a major Swedish lending institution between September 1993 and August 1995. When monitored on October 6, 1996, the loans were either still paying regular installments and interest or had been amortized completely. All loans were granted in stores where potential customers applied for instant credit to finance the purchase of a consumer good. They range in value between 3,000 and 30,000 Swedish kronor (approximately US$ 375 - 3750). The lending institution's policy was that no loans exceeding 30,000 kronor were supplied.

The evaluation of each application took place in the following way. First, the store phoned to the lending institution to get an approval or a rejection. The lending institution then analyzed the applicant with the help of a database with personal characteristics and credit history variables to which it has on-line access. The database is maintained by Upplysningscentralen AB, the leading Swedish credit bureau which is jointly owned by all Swedish banks and lending institutions. When approval was given, the store's salesman filled out a loan contract and submitted it to the lending institution. The loans are revolving and administered by the lending institution as any other credit facility. They are provided in the form of a credit card that can only be used in a specific store. The loan is free of interest during the month of application plus the following 2 months. After the interest free period, a payment of at least 5 percent of the outstanding debt is required each month. However, since the loan is revolving, there is no predetermined maturity of the loan. Earnings on the loan come from four sources: a payment by the store that is related to total amount of loans granted to its customers during some time-period, a monthly invoice fee of 17 kronor, and interest on the balance outstanding on the card.

We classify a loan as 'bad' once it has been forwarded to a debt-collecting agency. One can think of several different definitions. One alternative is, for example, 'all customers who have received one, two or three reminders because of delayed payment'. However, unlike 'being forwarded to debt-collecting agency", one, two or three reminders were all transient states in the register of the financial instititution. Once customers returned to the agreed-upon repayment scheme, the number of reminders was reset to zero. Such a property is rather undesirable if one needs to determine unambiguously which loans have defaulted and which have not.

Dormancy is defined as the state into which loans move when they are terminated or reach a debt balance below SEK 200 ($\approx$US$ 25). This definition may not appear completely appropriate for loans that are revolving. Due to the way in which the loans are administered, customers hardly ever reach a zero balance in the bank's computer system. A customer that makes his final payment on a loan, but sends away his money transfer a few days too late is likely to end up with a non-zero balance on his account. If the supposedly repaid loan is neglected and not formally terminated, accrued interest and automatically added late payment fees can quickly add up to 100-200 kronor. By classifying loans with a balance under SEK 200 as dormant, we thus make sure to assign the loans that are not active anymore but were never formally terminated to their correct group. Moreover, because it is exceptional for the customers in this sample to utilize the credit facility more than once, there is only minimal risk of incorrectly classifying active loans as dormant. This can be explained by the continuous supply of loans with an interest rate free period up front. In particular, because the loans in our sample do not involve any fixed cost for the customer, terminating the contract after amortization of the loan is obviously a cost minimizing strategy. Note that people cannot take a new interest free loan elsewhere after three months to amortize the old loan that is about to become interest bearing. To be eligible for a loan with an interest free period one must purchase a good at a store that cooperates with a bank.

Because the default rate on this type of loans is close to 1 percent (53 out of 4,786 observations), the analysis of the time to default requires a larger sample. For this purpose, we obtained a sample containing 1,103 defaulted loans with somewhat different payment conditions than the 'good' loans. These loans have up to 12 months of interest exemption, but pay a contract fee of approximately 250 kronor. Intuitively, the most natural thing to do would be to analyze dormancy with loans from the same sample. Unfortunately, because dormancy generally comes about after completion of the interest free period, a long lasting interest exemption makes it difficult to separate the economically induced duration process for good loans from the contractually determined minimum survival. The effect of interest exemptions on the behavior of people that renege on their committments is negligable, however. The time to default *can* therefore be studied with this larger sample. Using a different sample for the estimation of a model of the time-to-default is not ideal. The only purpose of the model for the time to default is, however, to contrast its properties with those of the process that governs dormancy, not to forecast. The main focus of our study is on modelling dormancy.

For this study, the lending institution provided us with a data file with the personal number of each applicant, the date on which the application was submitted, the size of the loan that was granted, the status of each loan (good, bad or active) on October 9, 1996, and the date on which defaulted and dormant loans ended their 'active' status. Upplysningscentralen provided the information that was available on each applicant at the time of application and which the financial institution accessed for *its* evaluation. By exploiting the unique personal number that each resident of Sweden has, the credit bureau was able to merge these two data sets. Before handing over the combined data for analysis, the personal numbers were removed. The database included publicly available, governmentally supplied information such as sex, citizenship, marital status, postal code, taxable income, taxable wealth, house ownership, and variables reported by Swedish banks like the total number of inquiries made about an individual, the number of unsecured loans and the total amount of unsecured loans. Table 1 contains definitions of all variables that are used in the analysis described in Sections 3. Table 2 provides some descriptive statistics.

A number of the variables in the dataset have not been used in the final estimation of the model described in Section 3. Among these are the number of

Table 1: Definition of variables.

Variable	Definition
PERSONAL	
AGE	age of applicant (in years)
MALE	dummy, takes value 1 if applicant is male
MARRIED	dummy, takes value 1 if appl. is married but not a separated woman
SEPARATE	dummy, takes value 1 if appl. is a widow(er) or a separated woman
NORDIC	dummy, takes value 1 if appl. is Nordic but not Swedish
FOREIGN	dummy, takes value 1 if appl. is non-Nordic
FINANCIAL	
HOUSE	dummy, takes value 1 if appl. owns (part of) a house
HOUSEVAL	value of all real estate ($\times$10k SEK), weighted by appl. owned share (log.)
ENTREPR	dummy, equals 1 if appl. has taxable income from a business
INCOME	annual taxable income from wages ($\times$100 SEK, logarithmized)
DUMMY180	dummy, takes value 1 if appl. has INCOME under 180k SEK
INCOM180	DUMMY180 * INCOME
DIFINC	INCOME_t - INCOME_{t-1} ($\times$100 SEK, not logarithmized)
CAPINC	dummy, takes value 1 if appl. has taxable income from capital
CREDIT	
NRQUEST	ln (1 + number of requests for information on the appl. that the credit agency received during the last 36 months)
LIMIT	collateral-free credit facilities already outstanding ($\times$100 SEK, log.)
ZEROLIM	dummy, equals 1 if appl. has no collateral-free loans outstanding
NRLOANS	ln (1 + number of collateral-free loans registered)
LIMUTIL	percentage of limit that is actually being utilized
BALINC	dummy {INCOME > 0} * ln(100 * BALANCE / INCOME)
BALINCSQ	BALINC^2
BALANCE	total collateral free credit facilities actually utilized
COAPPLIC	dummy, takes value 1 if appl. has a guarantor
LOANSIZE	amount of credit granted ($\times$ 1000 SEK)
LIMBAL	LIMIT-BALANCE
DUMMY20	dummy, takes value 1 if LIMBAL > 20k SEK
LIMBAL20	LIMBAL * DUM20

Table 2: Descriptive statistics for granted loans.

Variable	Mean	Min	Q1	Median	Q3	Max
PERSONAL						
AGE	44.00	20	35	44	52	81
MALE	.70	0				1
MARRIED	.57	0				1
DIVORCE	.03	0				1
NORDIC	.02	0				1
FOREIGN	.01	0				1
FINANCIAL						
HOUSE	.45	0				1
HOUSEVAL	3.47	0	0	5.24	6.01	11.21
ENTREPR	.04	0				1
INCOME	7.42	0	7.35	7.56	7.77	9.78
DUMMY180	.41	0				1
INCOM180	2.85	0	0	0	7.22	7.50
DIFINC	.62	-44.61	-.31	.53	1.81	68.16
CAPINC	.18	0				1
CREDIT						
NRQUEST	1.51	0	1.10	1.61	1.95	2.64
LIMIT	10.00	0	9.39	10.13	10.87	13.56
ZEROLIM	.01	0				1
NRLOANS	1.32	0	1.10	1.39	1.61	2.71
LIMUTIL	32.12	0	0	16	66	115
BALINC	1.38	0	0	.69	2.71	9.15
BALINCSQ	4.27	0	0	.48	7.34	83.70
BALANCE	6.08	0	0	8.21	10.28	13.55
COAPPLIC	.08	0				1
LOANSIZE	6.53	3.00	4.00	6.00	8.00	30.00
LIMBAL	9.28	0	8.92	9.55	10.05	12.21
DUMMY20	.27	0				1
LIMBAL20	2.84	0	0	0	10.05	12.21

months since the most recent change in marital status, number of months since immigration, number of houses a person (partially) owns, several measures of income, taxable wealth, a large number of entries on the two most recently submitted income-tax return forms, like total taxes due, back tax etc, and a number of transformations of these variables. Most of these were disregarded because they lack a relation with the dependent variable. Examples are back tax and real estate value. Section 3 contains a more detailed account of the methods we used in this selection process. Other variables were disregarded because they displayed extremely high correlation with covariates that measured approximately the same thing but had greater explanatory power. The numerous income measures in the dataset were eliminated in this way. Finally, taxable wealth could not be used as an explanatory variable. Wealth up to SEK 900,000 is tax-exempted, making the group of people with *taxable* wealth extremely small. Instead we used *taxable* income from capital - which is taxed from the first krona - to create a dummy explanatory variable.

3. Empirical analysis

Ideally, evaluating loan applicants or studying efficiency in the provision of bank loans would entail modelling the revenue on each loan as a function of a set of personal characteristics and macro-economic indicators. However, since few banks store complete time series of interest payments and amortizations on loans, the information presently available and useful for such a study is limited to the current balance and status (good or bad) of each loan. Therefore, we will instead model the survival time of each loan. With some simplifying assumptions imposed on the amortization scheme, and the rates of interest and discounting, one is then in principle able to calculate the return on each loan.

Because we have both censored and uncensored observations in our sample data, estimating a model of dormancy by means of OLS will lead to inconsistency in the parameter estimates. Duration models can take the censoring of data into account and will produce consistent and efficient parameter estimates. To make the exposition in the rest of this section more accessible for the reader, we start by briefly introducing some basic concepts in duration modelling.[3]

The probability distribution function of duration, denoted by $F(t \mid x) = Pr(T < t \mid x)$ and the corresponding density function $f(t \mid x) = dF(t \mid x)/dt$ are

[3]Kiefer [21] and Lancaster [22] discuss duration models and the analysis of transition data in greater detail.

two equivalent ways of describing the distribution of survival times conditional on a vector of characteristics x. Another equivalent concept, but with greater intuitive attraction in the context of duration analysis is the survivor function

$$\begin{aligned} S(t \mid x) &\equiv 1 - F(t \mid x) \\ &= Pr(T \geq t \mid x) \end{aligned}$$

Duration data are easiest interpreted, though, by means of the hazard function

$$\lambda(t \mid x) \equiv f(t \mid x) \ / \ S(t \mid x).$$

Roughly, the hazard rate is the probability of the immediate occurence of an event at a specific point in time, conditional on this event not having taken place earlier.[4] From the definition of $\lambda(t \mid x)$, it is easy to derive that $\lambda(t \mid x) = -d \ \ln S(t \mid x)/dt$ and

$$S(t \mid x) = exp[-\Lambda(t \mid x)]$$

where $\Lambda(t \mid x) = \int_0^t \lambda(v \mid x)\, dv$. The hazard function specification can been seen as a way to decompose an unconditional probability into a series of conditional probabilities. The hazard function also offers a practical interpretation of the meaning of duration dependence. When $d \ \lambda(t \mid x)/dt \ > \ (<) \ 0$, there exists positive (negative) duration dependence and the probability that a spell will end instantaneously increases (decreases) with the length of the spell. It makes sense to choose the parametrization in such a way that it allows the *hazard* function to behave as preconceived. Using the *distribution* function instead may fail to reveal any such properties that are undesirable for the analysis of durations - despite the distribution's appeal in other applications. For example, the Normal and Lognormal distributions are unable to comprise the constant hazard as a special case; some other distributions have monotonically increasing or decreasing hazard rates.

The rest of this section is divided into three subsections. We start by analyzing the duration of loans until failure in Section 3.1. In Section 3.2 we model the transition from active loans to dormancy and relate this transition to the characteristics of the applicant. We conclude the empirical analysis with a subsection that addresses some conceptual problems that arise when modelling consumer credits by econometric duration models, as well as practical problems in revealing the functional relationship between the transition and the covariates.

[4] A formal definition is $\lambda(t) = \lim_{h \to 0} Pr(t \leq T < t + h \mid T \geq t) \ / \ h.$

3.1. Bad loans

Traditionally, default risk has been modelled by means of discrete choice models, letting the outcome variable be failure or not. For an adequate analysis of the risk involved in bank lending one needs to consider not only the probabilities of default but even the durations of loans. Failures may occur after a long period of regular installments and interest payments. These revenues will partly or completely make up for the losses on bad loans, particularly because these will be heavily discounted in addition. By starting with an investigation of the time to failure we create a unified framework for the analysis of both risk and return in bank lending and make it possible to contrast the behavior of the defaultness and dormancy models later on.

We denote the duration of the i:th bad loan by $t_{d,i}$, the elapsed time in weeks from the approval to the time point it fails. We attempt to fit a regression model which can be expressed as,

$$\ln t_{d,i} = m(x_i, \beta) + \epsilon_i, \tag{3.1}$$

where $m(\cdot)$ is the function which links the duration variable to the applicant specific covariate vector x_i and β is a parameter vector assumed common to all loans. The functional form of $m(\cdot)$ is selected on basis of the results from non-parametric regressions as outlined in Section 3.3.

Table 3 shows the least squares estimates and the standard errors of the unknown parameters in the vector β. The purpose of fitting the regression model is purely descriptive, and our focus is on the sign, magnitude and standard errors of the parameter estimates. In obtaining the estimates, a considerabe effort has been devoted to the elimination of variation that arises from regional factors and the blend of credit products that the data set is composed of. We have, however, excluded the parameters pertaining to these factors from Table 3 because they are of limited relevance for the present investigation. Except for the regional and product related variables, having a *COAPPLIC*ant or a large number of registered loans are found to significantly increase the time to default. Having capital income or lacking experience in borrowing money reduce duration.

As a minor point of succes, we further note that the residuals closely follow the Normal distribution. Overall, however, we judge that the regression model fails to capture the relationship between survival time and applicants characteristics. We attribute this to two different causes. Firstly, the variety in loan types present in this sample blurs many behavioral relations. Although all loan types have similar

Table 3: LS regression on survival time of bad loans.
Asymptotic standard errors are presented within parentheses. Parameters that are significant at the 10 percent level or higher are printed in bold style.

Variable	Parameter	Standard error
CONSTANT	**2.615**	(.375)
AGE	−.000	(.001)
MALE	−.048	(.033)
MARRIED	.047	(.036)
DIVORCE	−.118	(.094)
SCAND	.009	(.091)
FOREIGN	−.050	(.074)
HOUSE	.050	(.037)
ENTREPR	.099	(.101)
INCOME	−.009	(.036)
ZEROINC	−.247	(.256)
DIFINC	−.000	(.001)
CAPINC	**−.172**	(.078)
NRQUEST	−.023	(.034)
LIMIT	−.013	(.020)
ZEROLIM	**−.268**	(.150)
NRLOANS	**.132**	(.049)
LIMUTIL	−.000	(.001)
BALANCE	.014	(.013)
COAPPLIC	**.210**	(.054)
LOANSIZE	.033	(.033)

$N = 1103$, $R^2 = .185$, $\sigma^2 = .230$. Dummy variables pertaining to the regional belonging of applicants or the type of credit are included in the model but suppressed in this table. They contribute with 7 percentage points to the R^2 measure of fit.

terms, small changes in terms are regularly associated with large shifts in the clientele, because different kinds of loans generally are sold by different types of stores. Any differences in clientele behavior are unlikely to be captured by a single product related dummy variable, as we have attempted here. Secondly, and more importantly, the transition to default cannot be modelled as a univariate process in the way dormancy can, because of the large share of incomplete observations in these data (around 99 percent of the loans have not completed a default spell). In principle, failure and dormancy are competing risks and should thus be modeled simultaneously, unless their error processes are independent. Ignoring this relationship when modelling dormancy will turn out to be a rather harmless act, because the defaulted loans that are disregarded constitute only one percent of the data. Even if the error processes would be strongly correlated, the dormancy model's parameter estimates will hardly be affected because of the disregarded observations negligible weight in the likelihood function. When modelling failure, however, the same strategy will imply overlooking 99 percent of the data set. Obviously, this is likely to have an important influence on the parameters that are estimated. As a consequence, the resulting model will not lend itself for prediction of the time to default. In theory, this problem can be solved by estimating a bivariate duration model with two competing risks: default and dormancy.[5] Because of the extreme imbalance between the shares of dormant and defaulted loans in the population this turned out to be practically unfeasible.

3.2. Dormant loans

The aim of this section is to disclose the transition process by which a loan moves from an active to a dormant status and to relate this process to the applicants characteristics. Unlike the model of the time to default, however, the purpose of the models in this section is not only descriptive but even to predict, i.e. to calculate the expected durations for all granted loans and compare these with the survival time required for a break-even.

We denote the elapsed time (in weeks) from approval of the i:th loan to the time point the loan becomes dormant by T_i. Due to the sampling scheme, the actual duration t_i can be either entirely or partially observed. In the former case we refer to the observation as complete, in the latter case it is called incomplete.

[5] See Carling and Jacobson [11] for an explanation of competing risk models.

The data set consists of the following complete and incomplete observations:

T_i	$= t_i$	$< t_i$	$> t_i$
no.	751	3605	377 .
c_i	-1	0	1

The incomplete observations with the coding variable $c = 1$ are right-hand censored observations. The observations with coding $c = 0$ are less common in duration analysis: they are a consequence of the bank's partial failure to record dormancy[6]. Table 2 contains descriptive statistics on the rich background information on the applicants, x_i, which will be used in the duration analyses. The covariates are exclusively time-invariant. Detailed discussions on how to include time-varying covariates in duration models are provided by Lancaster [22] and Narendranathan and Stewart [24].

Although duration analysis is better suited to model the duration of the active loans, we will begin with fitting a regression model for the complete observations, i.e. those for which $c = -1$. Table 4 presents as Model I the results from a least squares regression of the natural logarithms of the duration on a set of covariates (see the model in (3.1)). There are several reasons for studying the complete observations first. Firstly, least squares regression is a robust estimation method, thus eliminating the need of guessing the error distribution. Secondly, advanced univariate and multivariate statistical and graphical techniques exist for investigating the functional form of $m(\cdot)$ (see Section 3.3, where the principles and the techniques for selecting $m(\cdot)$ is discussed). Finally, estimation is fast and simple.

Model I has a reasonably good fit, the R^2 is 25.1 percent, and the distribution of the residuals follows closely the Normal distribution. However, the omission of the incomplete observations is clearly inappropriate. The censored observations with $c_i = 1$ are the loans with the longest durations and excluding them leads to biased parameter estimates and underestimation of the expected duration. We therefore turn to consider duration models.

As we mentioned above, the hazard function is the most natural entity to model duration variables since our attention is on the transition rate from one state to another. We consider models derived from the two most common clas-

[6] The presence of incomplete observations does not invalidate a consistent estimation of the unknown parameters. The drawback is a reduced precision in the estimates, manifested by larger standard errors.

Table 4: Estimated models of duration to dormancy.
Model I is an OLS regression of the complete observations (c = -1). The parametrization of models II-III conforms to that of model I in that the parameter estimates can be interpreted as the relative contribution to the duration of a loan. Asymptotic standard errors (within parentheses) are obtained from the cross-product of the gradient matrix.

Variable	OLS regression model I	Lognormal duration model II	Log-logistic duration model III	Semi-parametric dur. model IV
CONSTANT	**5.189** (1.546)	**5.584** (2.438)	**5.489** (1.546)	–
AGE	.003 (.003)	.003 (.005)	.002 (.002)	.001 (.002)
MALE	−.107 (.073)	−.053 (.128)	−.081 (.052)	−.023 (.046)
MARRIED	−.092 (.068)	−.184 (.127)	**−.127** (.050)	**−.101** (.046)
SEPARATE	**.422** (.211)	.180 (.358)	.234 (.152)	.144 (.138)
NORDIC	.128 (.215)	.147 (.391)	.098 (.154)	.054 (.155)
FOREIGN	**−.543** (.294)	−.528 (.488)	**−.407** (.206)	−.109 (.161)
HOUSE	.093 (.337)	.004 (.576)	.163 (.233)	.022 (.216)
HOUSEVAL	−.001 (.057)	−.014 (.099)	−.035 (.040)	−.005 (.037)
ENTREPRE	−.235 (.209)	−.417 (.383)	−.246 (.164)	−.179 (.123)
INCOME	**−.423** (.197)	−.445 (.311)	**−.402** (.132)	**−.282** (.104)
DUMMY180	**−3.161** (1.585)	−3.022 (2.493)	**−2.837** (1.063)	**−1.942** (.835)
INCOM180	**.409** (.203)	.410 (.326)	**.379** (.139)	**.262** (.109)
DIFINC	**−.015** (.009)	−.014 (.137)	−.063 (.057)	−.015 (.045)
CAPINC	−.082 (.094)	−.237 (.176)	**−.125** (.075)	**−.090** (.058)
NRQUEST	**.319** (.071)	**.521** (.145)	**.192** (.052)	**.110** (.047)
LIMIT	**.116** (.055)	−.075 (.155)	.021 (.051)	.077 (.053)

Continued on next page.

Table 4 - continued.
Coefficients significant at the 10 percent level or higher are printed in bold style.

Variable	OLS regression model I	Lognormal duration model II	Log-logistic duration model III	Semi-parametric dur. model IV
ZEROLIM	.661	−.051	.208	**1.036**
	(.466)	(1.355)	(.450)	(.432)
NRLOANS	.126	.255	**.354**	**.304**
	(.106)	(.234)	(.088)	(.086)
LIMUTIL	−.002	.009	**.004**	**.006**
	(.002)	(.006)	(.002)	(.002)
BALINC	**.201**	−.007	.110	.042
	(.103)	(.222	(.087)	(.081)
BALINCSQ	**−.043**	.020	**−.032**	−.017
	(.017)	(.039)	(.015)	(.015)
BALANCE	−.005	.041	−.002	−.002
	(.016)	(.032)	(.013)	(.011)
COAPPL	**.053**	.285	**.184**	**.157**
	(.010)	(.186)	(.070)	(.073)
LOANSIZE	−.005	**.046**	**.022**	**.020**
	(.009)	(.019)	(.007)	(.007)
LIMBAL	−.023	−.053	−.015	.036
	(.021)	(.094)	(.020)	(.030)
DUMMY20	−2.757	**−.565**	−.133	−.077
	(1.870)	(.193)	(.082)	(.070)
LIMBAL20	.260	**.100**	−.010	**−.022**
	(.180)	(.033)	(.009)	(.011)
REGION1	**.205**	**−.014**	−.044	−.008
	(.071)	(.003)	(.052)	(.048)
REGION2	.048	−.162	**−.074**	**−.086**
	(.057)	(.109)	(.042)	(.040)
REGION3	.079	.105	.052	.029
	(.094)	(.180)	(.071)	(.064)
REGION4	.141	**.428**	**.282**	**.249**
	(.096)	(.185)	(.073)	(.068)
σ^2	.620	3.820	.850	
R^2	.251			

ses of conditional hazard models.[7] The first is the location-scale model (also known as the Accelerated Failure Time Model) which can be written in regression form as

$$\ln t_i = m(x_i, \beta) + \sigma \epsilon_i. \tag{3.2}$$

Taking $m(x_i, \beta) = h(x_i) \cdot \beta$, the parameters can be interpreted as the absolute change in the expected value of the logarithm of the duration or, approximately, the relative change in the expected duration.

An extension of Model I yields the Lognormal model with hazard function,

$$\lambda(t|x) = \frac{\phi((\ln t - x'\beta)/\sigma)}{t\sigma\overline{\Phi}((\ln t - x'\beta)/\sigma)} \tag{3.3}$$

and survival function

$$S(t|x) = \overline{\Phi}((\ln t - x'\beta)/\sigma). \tag{3.4}$$

We hereafter refer to this as Model II.

The Lognormal model belongs to the class of location-scale models and can thus be specified as regression model, we estimate $\ln t_i = x_i'\beta + \sigma\epsilon_i$, where the residuals are assumed to follow the standard Normal distribution. Thus, the maximum likelihood estimates of Model II in Table 4 can be contrasted to the estimates of Model I. The Maximum Likelihood estimates are obtained by maximizing the log likelihood function[8],

$$\begin{aligned} \ln L(\beta, \sigma) &= \textstyle\sum_{c=-1} \ln \frac{\phi((\ln t - x'\beta)/\sigma)}{t\sigma} + \\ &\textstyle\sum_{c=0} \ln\left(1 - \overline{\Phi}((\ln t - x'\beta)/\sigma)\right) + \sum_{|c|=1} \ln \overline{\Phi}((\ln t - x'\beta)/\sigma), \end{aligned}$$

by an iterative gradient method.

The Log-logistic model, Model III, is an alternative to the Lognormal model. The hazard function is given as

$$\lambda(t|x) = \frac{\exp(x\gamma)\,\alpha t^{\alpha-1}}{(1 + \exp(x\gamma)\,t^{\alpha})}. \tag{3.5}$$

[7] See Lancaster [22] and Kiefer [21].

[8] Starting values were obtained from the Approximate Maximum Likelihood method (Carling [9]), as well as from Model I, and used in conjunction with the BHHH algorithm (see Carling and Söderberg [12]). The same procedure was adopted for Model III and IV as well.

and the log likelihood function to be maximized is

$$\begin{aligned}\ln L(\beta,\sigma) &= \sum_{c=-1}\ln\frac{\exp(x\gamma)\,\alpha t^{\alpha-1}}{(1+\exp(x\gamma)\,t^{\alpha})}+\\ &\quad \sum_{c=0}\ln\frac{\exp(x\gamma)\,t^{\alpha}}{(1+\exp(x\gamma)\,t^{\alpha})}-\sum_{|c|=1}\ln(1+\exp(x\gamma)\,t^{\alpha}).\end{aligned}$$

The Log-logistic model also belongs to the class of location-scale models described in (3.2) where the residuals follow the logistic distribution, and hence there is a simple relation between the parameters of Model II and III: $\beta=-\gamma/\alpha$. Table 4 presents the resulting parameter estimates after reparametrization in order to simplify a comparison of the estimates.

However, Model II and III are not nested and there is no obvious way to discriminate between them. Yet, they impose rather strong restrictions on the error distribution and thus an implicit assumptions on the hazard function. We therefore, by considering a fourth model, relax the assumption on the form of the hazard function. Model IV is a semi-parametric Proportional Hazard (PH) model of the form [9]

$$\lambda(t|x)=\exp(-x'\beta)\,\lambda(t)$$

where $\lambda(t)$ is the base-line hazard function stipulated to be equal for all loans. The Proportional Hazard model can also be formulated as a regression model

$$\ln\int_0^{t_i}\lambda(u)\,du=x_i'\beta+\zeta_i$$

where ζ follows the Extreme Value distribution with expectation $E[\zeta]=-0.5772$ and variance $V[\zeta]=\pi^2/6$. The interpretation of the parameters in the PH-model is different from the location-scale model: a parameter shows, approximately, the relative change in the hazard function. The non-parametric baseline hazard $\lambda(t)$ and the linking function $\exp(-x'\beta)$ are estimated simultaneously by Maximum Likelihood.[10,11] The estimates of Model IV have been reparameterized so that

[9]The somewhat unusual specification of the linking function, $\exp(-x'\beta)$, is given in the interest of making the estimates between Model I-III and Model IV directly comparable.

[10]See Carling et al [10].

[11]We considered four weeks intervals to restrict the number of parameters in the baseline hazard. As a consequence, we assume that the hazard is constant within a four weeks interval, an assumption justified by preceding analysis which showed a minuscule change in the baseline hazard by extending the time unit. Furthermore, there are 40 parameters in the baseline hazard, hence there were no attempt, due to sample restrictions, to estimate the hazard function beyond the 160:th week.

they can be compared with those of models II and III and are shown in Table 4.

The attempt to fit the models to the data has been met by a reasonable success; the question is which model provides the best description of the data. Table 5 shows the fitted probability distributions. Figure 1 shows the hazard function produced by Model II-IV for an average loan-applicant, i.e. an individual with covariate vector equal to the sample average. Table 5 shows that the tail of the lognormal hazard function is obviously too fat on the right hand side. It is less apparent whether one should prefer the log-logistic or the semi-parametric hazard function. In Figure 1, however, we immediately see that both parametric models provide a poor description of the hazard function. The lognormal hazard function is monotonically decreasing with time and fails to register any of the peaks that the semi-parametric model captures at three and twelve months. The loglogistic hazard function does a somewhat better job. It is unimodal, however, and in its attempt to capture the double peakedness it reaches its maximum in between them, at 30 to 40 weeks. It would have been possible to augment the class of location-scale models, as described in (3.2), by even more flexible distributions.[12] Because of the larger number of free parameters, the semi-parametric model will, however, always produce the most accurate description of the data. Since the sample size of our data set permits us to estimate a larger number of free parameters at a relatively low cost (in terms of the relative change in the degrees of freedom that are lost), we prefer the semi-parametric representation as the best way to model the duration to dormancy.

Unlike the model in Section 3.1, the semi-parametric model captures a number of interesting relationships between borrowers characteristics and duration to dormancy. When controlling for income and credit history, age, gender and citizenship no longer have any bearing on duration. Married applicants tend to pay back their loans faster. This may be a result of Swedish couples generally having two wage earners in the family which leads to a stable flow of income. Alternatively, it could reflect the fact that married couples are simply more diligent. Higher income also reduces the duration of loans until dormancy. For incomes over SEK 180k. this effect flattens out. Intuitively this makes sense, considering that the median gross monthly income is about SEK 15k. People with below-median income take a longer time to pay their debts; all others behave alike. Owning assets further reduces the duration of any loans people hold. Such people will

[12] See Bergstöm and Edin [6].

Table 5: The fitted probability distribution for the time to dormancy (in weeks). For models II-IV, the distribution has been calculated for an average applicant, i.e. all covariates equal their sample means. To derive the expected duration by model IV, the hazard rate has been assumed constant beyond the 160th week.

Model	mean	percentiles 5	10	25	50	75	90	95
Lognormal (II)	89	<1	<1	3	13	50	163	331
Log-logistic (III)	61	8	12	21	38	66	117	175
Semi-parametric (IV)	58	8	12	15	44	80	125	160

likely use liquid assets for repayment of a debt with a 30 percent interest rate. Somewhat surprising is the fact that the size of *LIMIT* does not affect duration, whereas increasing the loan size or the number of loans delays repayment. It is interesting to contrast this with the results in Jacobson and Roszbach [20] and Roszbach [25]. In two independent studies of the determinants of default, they find that *LIMIT* has a significant effect on both the risk of default and the duration to default, but *LOANSIZE* does not. Inexperience with servicing debt, as indicated by *ZEROLIM*, slows down the amortization. Similarly, *NRQUEST* is considered a good measure of a persons eagerness to obtain additional credit and as such expected to lead to a longer payback time. *LIMUTIL* has a comparable effect. Finally, having a guarantor also makes it more likely that the loan will be paid back over a longer period of time. This is rather interesting in view of the fact that a guarantor is generally asked for to reduce the risk of what is considered a likely failure.

Model IV can now be used to examine the lending policy of the bank. An averagely sized (SEK 6,300) loan needs to stay active for at least 5 months (of which 3 are free of interest) for the bank to break even. Hence, one would expect the bank to approve loans which are likely to remain active and be restrictive towards loan applications that are likely to turn dormant shortly after the free 13 weeks. For each loan in the sample, we can derive the the conditional expected duration from Model IV. Figure 2 shows by a histogram the distribution of the expected duration of the active loans for all 4,733 sample applicants. If we compare the expected durations with the break-even duration, 4,671 (98.7 %) out of the

4,733 granted loans are found to have positive expected profits. With the help of the expected durations that have been computed to construct Figure 2, we can even calculate the expected profit or return on each loan. Appendix A describes how these profits are derived. Figure 3 contains a plot of the resulting distribution of expected profits for the loans in the sample. The expected profits range from -105 to 7924 kronor, with a mean of 786 and a standard deviation of 616.

Ideally, we would have wanted to carry out the same analysis on the rejected loan applications. Due to way in which the data were sampled, we were however unable to do so. The institution that provided the data extracted loans that were granted between September 1993 and August 1995 from its customer base but only collected applications during the last of these two years. The type of loan that we study was hardly offered during 1994-95 and we thus lack information on the rejected applicants. There are no practical obstacles, however, that prevent banks from doing so for future loan applications. The semi-parametric model that we have estimated for our sample data illustrates how a duration model can be used to evaluate loan applicants by calculating their expected returns.

3.3. Econometric issues

During the completion of this analysis, we have had reason to consider some conceptual and practical problems which we address here. First, in duration analysis it is assumed that eventually the event will occur, e.g. that an active loan eventually turns dormant. Formally, this assumption arises since it is required $\lim_{t\to\infty} P[T > t] = 0$ for the survival function to be well-defined. In practice, an active loan can either become dormant or it can fail. The existence of the two mutually exclusive events seems to suggest that the condition is violated. As we already mentioned in Section 3.1, we have taken a pragmatic point of view in this matter, by claiming that the bad loans are too few to invalidate the analysis of the transition to dormancy. We found support for this claim in an additional estimation of models II-IV on an expanded dataset that includes the bad loans as independent observations that are right-hand censored at the time point of default. The results confirmed the parameter estimates in Table 4. Nevertheless, this condition did prevent us from carrying out a duration analysis of the bad loans. As a consequence, we failed to give a meaning to the expected duration of a bad loan in an environment with a high transition rate to dormancy.

Other conceptual issues in duration analysis are unobserved heterogeneity and multiple spells.[13] We refer to Narendranathan and Stewart [24] and to their arguments for neglecting unobserved heterogeneity in semi-parametric proportional hazards models. Multiple spells arise when an individual has several loan spells. In the sample, there were hardly any such individuals. However, it is likely that the lending history of an individual can be very helpful in explaining the future behaviour. Whenever multiple spells are present, it seems wise to augment the duration model accordingly.

An important practical problem is how to disclose the function $m(x, \beta)$, i.e. how to perform preliminary data analysis. Duration data and duration models are quite difficult in this sense.[14] We prefer to perform the preliminary data analysis on the complete observations. Selecting the covariates and linking these to the duration variable in Table 4 has been done with the help of exploratory tools for ordinal and categorical variables and a number of non-parametric regressions for the continuous variables.[15] This decision hase been a compromise between the desire to reach effective linking and to restrict the number of parameters. We can take the variable *NRLOANS* as an example. In the sample there were applicants with 0 to 14 registered loans. A dummy-variable for each number of questions would imply 14 parameters. The transformation $y = \ln(1 + x)$ implied that the number of parameters could be reduced by 13, at a negligible reduction of the goodness of fit.

4. Discussion

A bank that lends money to a household faces two types of risk. Most commonly mentioned is the risk of default. Hardly ever referred to is the risk of an early redemption of the loan. As a result of the period of regular installments and interest payments that generally preceed a default, loan losses may be partially or completely made up for by revenues in addition to being heavily discounted. Moreover, defaulted loans generally make up only a few percent of the loan portfolio. Together, these facts suggests that dormancy is the economically relevant phenomenon to study - despite widely spread views that banks need to focus on default risk.

[13] See Blossfeld and Hamerle [7] for a clear exposition of the unobserved heterogeneity problem.

[14] Altman and de Stavola [2] provide a careful discussion on available techniques.

[15] Exploratory tools for ordinal and categorical variables are treated by Hoaglin, Mosteller, and Tukey [18]. For literature on non-parametric regressions, see Cleveland [13], Cleveland, Devlin, and Grosse [14], and Härdle [19].

In principle, failure and dormancy are competing risks and should therefore be modelled simultaneously unless their error processes are independent. If default rates had been somewhat higher than the one percent in our data set, estimation of a bivariate duration model with default and dormancy as competing risks would be required. This would be a straightforward extension of the analysis in this paper.[16] Because the defaulted loans make up only one percent of our sample data, we can ignore defaultness as a competing risk when modelling dormancy.

We model the transition by consumer loans' from an active to a dormant state by means of four different univariate representations. For each duration model, we investigate its ability to describe the hazard rates observed in the data. A semi-parametric model is found to best capture the relationship between applicants' characteristics and duration to dormancy. Among other things, the model shows that people tend to pay back their loans faster as their income increases. This effect disappears, however, when individual income reaches the median level in the sample. Bigger loans mean longer pay-back periods but the amount of already outstanding loans does not affect duration.

For the preferred model, we derive the distribution of conditional expected durations of loans. In combination with information provided by the lending institution, these durations enable us to calculate a distribution of expected profits for the sample loans. The duration model of dormancy is thus well suited for use by banks as a tool to evaluate new loan applications. However, unlike credit scoring models, which merely predict default probabilities, it is based on an evaluation of expected profitability. An interesting application of the model would be to examine the expected durations for rejected applications and compare these with the durations of granted loans. This would give us better insight into the efficiency of current bank lending and could help increasing the future profitability and efficiency of bank lending.

[16] See Carling and Jacobson [11].

References

[1] Altman, E.I., Avery, R.B., Eisenbeis, R.A., and Sinkey, J.F., (1981), *Application of classification techniques in business, banking and finance*, JAI Press, Greenwich, CT.

[2] Altman, D.G., and de Stavola, B.L., (1994), Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates, *Statistics in Medicine* 13, 301-341.

[3] Arminger, G., Enache, D., and Bonne, T.,(1997), Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis and feedforward networks, *Computational Statistics* 12, 293-310.

[4] Baek I., and Bandopadhyaya, A., (1996), The determinants of the duration of commercial bank debt renegotiation for sovereigns, *Journal of Banking and Finance* 20, 673-685.

[5] Bandopadhyaya, A., (1994) An estimation of the hazard rate offirms under Chapter 11 protection, *Review of Economics and Statistics*, May, 346-350.

[6] Bergström, R., and Edin, P-A., (1992), Time aggregation and the distributional shape of unemployment duration, *Journal of Applied Econometrics* 7, 5-30.

[7] Blossfeld, H., and Hamerle, A., (1992), Unobserved heterogeneity in event history models, *Quality and Quantity* 26, 157-168.

[8] Boyes, W.J., Hoffman, D.L., and Low, S.A., (1989), An econometric analysis of the bank credit scoring problem, *Journal of Econometrics* 40, 3-14.

[9] Carling, K., (1995), Starting values for the iterative maximum likelihood estimator in survival analysis, *Journal of Applied Statistics* 22, 515-519.

[10] Carling, K., Edin, P-A., Harkman, A., and Holmlund, B., (1996), Unemployment duration, unemployment benefits, and labor market programs in Sweden, *Journal of Public Economics* 59, 313-334.

[11] Carling, K., and Jacobson, T., (1995), Modeling unempoyment duration in a dependent competing risks framework: Identification and estimation, *Lifetime Data Analysis* 1, 111-122.

[12] Carling, K., and Söderberg, H., (1998), An experimental comparison of gradient methods in econometric duration analysis, *Computational Statistics & Data Analysis* 27, 83-97.

[13] Cleveland, W.S., (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* 74, 829-836.

[14] Cleveland, W.S., Devlin, S.J., and Grosse, E., (1988), Regression by local fitting: Methods, properties, and computational algorithms, *Journal of Econometrics* 37, 87-114.

[15] Diebold, F.X., and Rudebusch, G.D., (1990), A non-parametric investigation of duration dependence in the American business cycle, *Journal of Political Economy* 98 (3), 596-616.

[16] Dionne, G., Artis, M., and Guillen, M., (1996), Count data models for a credit scoring system, *Journal of Empirical Finance* 3, 303-325.

[17] Henley, W.E., and Hand, D.J., (1996), A k-nearest-neighbor classifier for assessing consumer credit risk, *The Statistician* 45 (1), 77-95.

[18] Hoaglin, D.C., Mosteller, F., and Tukey, J.W., (1985), *Exploring data tables, trends, and shapes*, Wiley, New York.

[19] Härdle, W., (1990), *Smoothing techniques: with implementation in S*, Springer-Verlag, New York.

[20] Jacobson, T. and Roszbach, K.F., (1998), *Bank lending policy, credit scoring and Value at Risk,* Working Paper No. 260, Working Papers Series in Economics and Finance, Stockholm School of Economics.

[21] Kiefer, N.M., (1988), Economic duration data and hazard functions', *Journal of Economic Literature*, 26, 646-679.

[22] Lancaster, T., (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

[23] Lane, W., Looney, S., and Wansley J., (1986) An application of the Cox proportional hazards model to bank failure, *Journal of Banking and Finance* 10 (4), 511-532.

[24] Narendranathan, W., and Stewart, M.B., (1993), 'Modelling the probability of leaving unemployment: competing risks models with flexible baseline hazard', *Applied Statistics* 42, 63-83.

[25] Roszbach, K.F., (1998), *Bank lending policy, credit scoring and the survival of loans,* Working Paper no. 261, Working Paper Series in Economics and Finance, Stockholm School of Economics.

A. Return on loans

In this appendix we show how the returns on defaulted and non-defaulted loans were calculated. The average survival time for non-defaulted loans is 11 months and has been computed from the 751 observations that we have full information on. This means that we excluded the remaining 3982 censored observations from this calculation. Considering that the expected survival time implied by the non-parametric model (Table 5) is 58 weeks, this estimate appears on the conservative side. The average survival time on bad loans is 13 months and has been computed from the 53 loans that had defaulted in the sample (of 4,786 observations) from which the data for the dormancy model was drawn. Between September 1993 and October 1996, people on average borrowed SEK 6,300 (median 5,400) and paid an interest rate of 31.36 percent per annum, or 2.3 percent on a monthly basis. The maximum yield on 6 months government bonds during this period was 9.63 percent per annum. We therefore apply a discounting factor for the bank's costs and revenue of 10.63 percent.

The loans are free of interest charges during the first three months. If the loan is not paid back after three months, interest starts accumulating. Borrowers receive an invoice every month for which they are charged a fee of 17 kronor. We assume that the loan is paid back as a annuity, the amount being determined by the average survival time and the interest rate. For defaulted loans the annuity will even be a function of the average nominal loss on the principal. Because amortization is slow and only partial, bad debtors pay the bank twice as much interest as good ones do. Moreover, because loan losses are realized in the future, discounting reduces their net present value.

On the cost side, the bank incurs a fixed cost for evaluating an applicant, printing a credit card and entering the new loan into its computer system. Variable costs consist of maintenance of the computer system and sending a monthly invoice to the customer.

The above can be summarized in the following expressions. The return on good loan can thus be expressed as $R_G = \left(\sum_{t=4}^{11} \frac{x_t + 17 - vc}{(1+r)^t} - fc \right) / s$. For a defaulted loan, the expression is only slightly different: $R_D = \left(\sum_{t=4}^{13} \frac{z_t + 17 - vc}{(1+r)^t} - fc - \frac{ll}{(1+r)^{13}} \right) / s$. Here r_t is the discount rate, x_t and z_t are the interest payments at time t on a good and bad loans respectively, fc and vc are the fixed and variable cost the bank incurs for each loan, ll is the average loss of principal on a defaulted loan and s is the average loan size. Entering all parameters into the above two expressions gives that $R_G = 8.1$ and $R_D = -36.7$ percent. The break-even point for both the mean and the median loan lies at 5 months (of which 3 are exempt of interest charges).