Bank lending policy, credit scoring and the survival of loans.*

Kasper Roszbach†

17 September 1998

Abstract

To evaluate loan applicants, banks use a large variety of systems. The objective of such credit scoring models typically is to minimize default rates or the number of incorrectly classified loans. Thereby they fail to take into account that loans are multiperiod contracts. From a utility maximizing perspective it is not only important to know *if* but also *when* a loan will default. In this paper a Tobit model with a variable censoring threshold and sample selection effects is estimated for (1) the decision to provide a loan or not and (2) the survival of granted loans. The model is shown to be an affective tool to separate applicants with short survival times from those with long survivals. The bank's loan provision process is shown to be inefficient. Loans are granted in a way that conflicts with both default risk minimization and survival time maximization. There is thus no trade-off between higher default risk and higher return in the policy of banks.

JEL Classification: C34, C35, D61, D81, G21

Keywords: Banks, lending policy, credit scoring, survival, loans.

*I would like to thank Marcus Asplund, Kenneth Carling, Luigi Ermini, Lennart Flood, Dennis Hoffman, Tor Jacobson, Sune Karlsson, Jesper Lindé, Rickard Sandin, Patrik Säfvenblad, Paul Söderlind, Anders Vredin and seminar participants at the Stockholm School of Economics for their helpful comments and Yngve Karlsson and Björn Karlsson at Upplysningscentralen AB for providing and discussing the data. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

†Department of Economics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden. *Email:* nekr@hhs.se.

1. Introduction

Consumer credit has come to play an increasingly important role as an instrument in the financial planning of households. When current income falls below a household's permanent level and assets are either not available or not acccessible for dissaving, credit is a means to maintain consumption at a level that is consistent with permanent income. People expecting a permanent increase in their income but lacking any assets, like students, have a desire to maintain consumption at a higher level than their current income allows. Borrowing can assist them in doing that. Those who accumulate funds in a pension scheme but are unable to get access to them when they experience a temporary drop in current income can also increase their welfare by bridging the temporary fall in income with a loan.

The quantitative importance of consumer credit may be illustrated by the fact that 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. That is equivalent to 12% of Swedish GDP or 22.7% of total private consumption. Viewed from the perspective of financial institutions, consumer credit also constitutes a significant part of their activities, making up 25 percent of total lending to the public. If one includes residential loans in total lending, this figure drops to 11 percent. When looking at the risk involved in these loans instead of their volume, their importance is even greater, however. BIS rules stipulate an 8 percent capital requirement on consumer credit compared to, for example, 4 percent on residential loans.

From these numbers, it may be clear that a lending institution's decision to grant a loan or not and its choice for a specific loan size can greatly affect households' ability to smooth consumption over time, and thereby even households' welfare. At a more aggregate level, consumer credit makes up a significant part of financial institutions' assets and the effects of any loan losses on lending capacity will be passed through to other sectors of the economy that rely on borrowing from the financial sector. For this reason, the properties and efficiency of banks' credit granting process are of interest not merely because the factors determining the optimal size of financial contracts can be examined. At least as important are the implications these contracts have for the welfare of households and the stability of financial markets.

The starting point of every loan is the application. When lending institutions receive an application for a loan, the process by which it is evaluated and its degree of sophistication can vary greatly. Most continue to use rather naïve, subjective

evaluation procedures. This could be a non-formalized analysis of an applicant's personal characteristics or 'scoring with integer numbers' on these characteristics. Some banks, however, have started to use a statistical 'credit scoring' model to separate loan applicants that are expected to pay back their debts from those who are likely to fall into arrears or go bankrupt.

By far the most commonly used methods are discriminant analysis and logistic regression. Altman, Avery, Eisenbeis and Sinkey [1] contains a good review of this literature. Both models have been fit to separate good loans from bad ones among *approved* applications. The estimated parameters are thus subject to sample selection bias when these models are applied to *all* applicants. More recent studies have employed k-nearest-neighborhood and [10] count data models [7], classification trees and neural networks [3]. These methods tend to suffer from problems with either the calibration, estimation or interpretation of their parameters *in addition to* the sample selection bias I mentioned earlier. All the above mentioned models, however, fail to account for the multiperiod character of an optimal debt contract and the implications this has for the credit-granting decision. In financial markets with perfect information, any optimal multiperiod financial contract can be obtained by a sequence of one-period loan agreements [12]. Loan applicants will be willing to pay the competitive interest rate that corresponds to their idiosyncratic risk and choose a first-best loan size. Under asymmetric information things become more intricate. In the literature that studies credit markets and the form of optimal financial contracts in the presence of adverse selection or costly state verification, credit rationing - the unequal treatment of ex-ante equal people - is a recurring phenomenon. See for example Stiglitz and Weiss [11] or Williamson [13]. When rationing is the mechanism that equilibrates credit markets, some applicants will be excluded from credit despite being equally creditworthy as those granted a loan. The allocation of resources will thus be inefficient.

Let us assume that *single-period* agreements are optimal[1] and that only the probability of default is unobservable to the lender. Gale and Hellwig [8] show that the optimal one-period debt contract consists of a pair (l, θ) where l is the size of the loan and θ the level of the endowment shock below which the debtor will be declared bankrupt. Under these circumstances traditional credit scoring models - by enabling a lending institution to rank potential customers according to their default risk - could improve the allocation of resources, from a second best towards the first best equilibrium. In a more general context, however, this

[1]Since the default probabilities are not observed, this would be in a second best sense.

does not solve a lender's profit maximization problem because financial contracts typically stretch out over several time periods. Townsend [12] proves that, in the presence of asymmetric information between the borrower and lender, ex-ante optimal contracts can only be created by multiperiod debt contracts because they allow payoffs to be dependent on past and present behavior of the borrower.

A loan, being a multi-period contract, generates a flow of funds until it either is paid off or defaults, in which case a part of the principal may still be recovered. The net present value of a loan is thus not determined by whether it's paid off in full or not, but - if it is not - by the duration of the repayments, amortization scheme, collection costs and possible collateral value. It may, for example, still be profitable to provide a loan, even if the lender is certain that it will default. Since the goal of financial institutions is to maximize profit (or utility), not to rank potential customers according to default risk, credit scoring models leave much room for subjective factors in the loan approval process. In a sense, banks use statistical models to forecast bankruptcy, but - conditional on this forecast - resort to ad-hoc methods to predict profitability.

Boyes, Hoffman and Low [5] address this deficiency and investigate if the provision of credit currently takes place in an efficient way. For this purpose they estimate a bivariate probit model with two sequential events as the dependent variables: the lender's decision to grant the loan or not, and - conditional on the loan having been provided - the borrower's ability to pay it off or not. If the lending institution is minimizing credit risk, we ought to find opposite signs for the parameter of one particular explanatory variable in the two different equations. This would imply that variables that increase the probability of positive granting decision also decrease the likelihood of a default, or vice versa. They find, however, that variables like duration of job tenure, education and credit card ownership carried equal signs, indicative of a policy that conflicts with default risk minimization. As we noted earlier, lenders may nevertheless prefer such a policy of supplying loans with a higher default risk because they have a higher expected rate of return (either the interest rate is higher or the default is expected to occur after a long period with regular installments and interest payments). Moreover, Boyes et al. show that unexplained tendencies to extend credit are positively correlated with default frequencies - another fact consistent with a policy that trades off default risk against profitability.

This paper deals with two issues. First, in order to improve upon the currently available methods for evaluating loan applications, I construct and estimate a Tobit model with sample selection and variable censoring thresholds. The model

can be used to predict the expected survival time on a loan to any potential applicant. This allows for a more realistic evaluation of the return on a loan than an estimate of the default risk associated with an individual with a traditional credit scoring model does.

Secondly, I take up the question about the efficiency of banks' loan provision process that is raised by the results in Boyes et al. [5]. Those suggest that the fact that some variables increase the probability of a positive granting decision while at the same time increasing the likelihood of a default is a consequence of profit maximizing behavior by the lender. Here, it will be investigated if a similar relationship continues to exist when one models the survival time of a loan instead of the probability of its default. If variables that increase the likelihood of an applicant obtaining a loan also increase the expected survival and vice versa, then this would constitute further evidence of banks' behaving in a way that is consistent with profit-maximization.

The rest of this paper is organized as follows. Section 2 describes the data set and its sources. In Section 3, I derive the econometric model. Section 4 contains the empirical results and section 5 concludes the paper with a discussion of the results and possibilities for future research.

2. Data

The data set consists of 13,337 applications for a loan that were processed by a major Swedish lending institution between September 1994 and August 1995. All applications were submitted in stores where potential customers applied for instant credit to finance the purchase of a consumer good. Out of 13,337 applications, 6,899 were rejected and 6,438 were approved. The dataset includes 127 second attempts by individuals that had applied once before.

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 analysed the applicant with the help of a database with personal characteristics and credit 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. If approval was given, the store's salesman filled out a loan contract and submitted it to the lending institution. The loan is revolving and administered by the lending institution as any other credit facility. It is provided in the form of a credit card

Table 1: Definition of variables.

Variable	Definition
SURVIVAL	days between granting of loan and its default
MALE	dummy, takes value 1 if applicant is male
MARRIED	dummy, takes value 1 if applicant is married
DIVORCE	dummy, takes value 1 if applicant is divorced
HOUSE	dummy, takes value 1 if applicant owns a house
BIGCITY	dummy, takes value 1 if applicant lives in one of the three greater metropolitan areas around Göteborg, Malmö and Stockholm.
NRQUEST	number of requests for information on the applicant that the credit agency received during the last 36 months
ENTREPR	dummy, takes value 1 if applicant has taxable income from a registered business
INCOME	annual income from wages as reported to Swedish tax authorities (in 1000 SEK)
DIFINC	change in annual income from wages, relative to preceding year, as reported to Swedish tax authorities (in 1000 SEK)
CAPINC	dummy, takes value 1 if applicant has taxable income from capital
ZEROLIM	dummy, takes value 1 if applicant has no collateral-free loans outstanding
LIMIT	total amount of collateral free credit facilities already outstanding (in 1000 SEK)
NRLOANS	number of collateral free loans already outstanding
LIMUTIL	percentage of $LIMIT$ that is actually being utilized
LOANSIZE	amount of credit granted (in 1000 SEK)
COAPPLIC	dummy, takes value 1 if applicant has a guarantor

that can only be used in a specific store. Some fixed amount minimum payment by the borrower is required during each month. However, since the loan is revolving, there is no predetermined maturity of the loan. Earnings on the loan come from three sources: a one-time fee paid by the customer; a payment by the store that is related to total amount of loans granted through it; and interest on the balance outstanding on the card.

Table 2: Descriptive statistics for all loan applicants ($N = 13337$).
The table splits up the sample into rejected and approved applications.

	Rejected ($N = 6899$)				Granted ($N = 6438$)			
Variable	mean	stdev	min	max	mean	stdev	min	max
MALE	.62	.48	0	1	.65	.48	0	1
MARRIED	.47	.50	0	1	.47	.50	0	1
DIVORCE	.13	.34	0	1	.14	.35	0	1
HOUSE	.34	.47	0	1	.47	.50	0	1
BIGCITY	.41	.49	0	1	.37	.48	0	1
NRQUEST	4.69	2.60	1	10	4.81	2.68	1	19
ENTREPR	.04	.21	0	1	.02	.16	0	1
INCOME	129.93	70.38	0	737.9	189.47	75.70	0	1093.0
DIFINC	5.37	34.06	-438.5	252.6	9.03	34.63	-6226.0	5006.0
CAPINC	.12	.32	0	1	.07	.25	0	1
ZEROLIM	.15	.36	0	1	$<.01$	.05	0	1
LIMIT	79.89	93.69	0	1703.0	50.33	49.83	.0	627.0
NRLOANS	2.99	2.42	0	18	3.65	2.04	0	16
LIMUTIL	64.34	38.88	0	278.0	53.22	33.93	0	124.0
COAPPLIC	.16	.36	0	1	.14	.35	0	1

For this study, the lending instutution provided 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 or bad) on October 9, 1996, and the date on which bad loans gained this status.

Although one can think of several different definitions of a 'bad' loan, I classify a loan as bad once it is forwarded to a debt-collecting agency. I do not study what factors determine the differences in loss rates, if any, among bad loans. An alternative definition of the set of bad loans could have been 'all customers who have received one, two or three reminders because of delayed payment'. However, unlike 'forwarded to debt-collecting agency", one, two or three reminders were all transient states in the register of the financial instititution. Once customers

Table 3: Descriptive statistics for granted loans.
The table splits up the subsample of granted applications into defaulted and non-defaulted loans.

	Defaulted loans ($N = 388$)				Good loans ($N = 6050$)			
Variable	mean	stdev	min	max	mean	stdev	min	max
*SURVIVAL**	400.09	151.07	130	789	632.79	93.99	34	795
MALE	.67	.47	0	1	.65	.48	0	1
MARRIED	.24	.43	0	1	.48	.50	0	1
DIVORCE	.20	.40	0	1	.14	.35	0	1
HOUSE	.28	.45	0	1	.48	.50	0	1
BIGCITY	.41	.49	0	1	.36	.48	0	1
NRQUEST	6.15	2.85	1	14	4.72	2.64	1	19
ENTREPR	.02	.13	0	1	.03	.16	0	1
INCOME	165.36	82.35	0	1093.0	191.02	75.00	0	1031.7
DIFINC	3.52	39.01	-135.0	439.7	9.38	34.30	-622.6	500.6
CAPINC	.04	.20	0	1	.07	.26	0	1
ZEROLIM	.04	.20	0	1	<.01	.02	0	1
LIMIT	41.44	57.98	0	511.5	50.90	49.21	0	627.0
NRLOANS	2.34	1.64	0	11	3.74	2.04	0	16
LIMUTIL	75.69	33.37	0	124.0	51.78	33.47	0	112.0
LOANSIZE	7.08	3.95	3.0	24.5	7.12	3.83	3	30.0
COAPPLIC	.07	.26	0	1	.14	.35	0	1

* For good loans these are censored survival times.

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 observations are censored and which are not.

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

Table 4: Descriptive statistics for survival time.
Percentiles for survival time and the natural logarithm of survival time.
The sample has been split up into defaulted and non-defaulted loans.

		Percentiles							
Sample	min	5	10	25	50	75	90	95	max
t, bad loans	130	156	192	278	403	514	606	648	789
t, good loans	34	470	497	564	652	704	746	767	795
ln(t), bad loans	4.87	5.05	5.26	5.63	6.00	6.24	6.41	6.47	6.67
ln(t), good loans	3.53	6.15	6.21	6.34	6.48	6.56	6.61	6.64	6.68

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. Overall, the database includes a total 60-70 variables. The major part consists of publicly available, governmentally supplied information such as sex, citizenship, marital status, postal code, taxable income, taxable wealth, house ownership. The remaining variables, like the total number of inquiries made about an individual, the number of unsecured loans and the total amount of unsecured loans, are reported to Upplysningscentralen by the Swedish banks. Table 1 contains definitions of all variables that are used in the analysis in Section 4. Some descriptive statistics on the explanatory variables are provided in Tables 2 and 3.

Of the applicants, 6,899, or 51.7 percent, were refused credit. The remaining 6,438 obtained a loan ranging from 3,000 to 30,000 Swedish kronor (approximately US$ 375 - 3750) . The lending institution's policy was that no loans exceeding 30,000 kronor were supplied. Although there is an indicated amortization scheme, the loans have no fixed maturity - they are revolving.

On 9 October 1996, all people in the sample were monitored by the lending institutution. At that moment 388 (6.0 %) of those who had obtained a loan had defaulted and been forwarded to a debt collection agency. All other borrowers still fulfilled their minimum repayment obligations at that time. The survival time in the sample, calculated as the number of calendar days between the date of application and the date of default, ranged from 130 days (a defaulted loan) to 795 days (a censored observation). Descriptive statistics for survival time are provided in Tables 4. Because the statistical model that will be presented in Section 3 will be estimated with the natural logarithm of survival time as a dependent variable,

Table 4 also contains descriptive statistics on logaritmized survival time.

3. Econometric model

Under ideal conditions 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 cost structure, one can then in principle calculate an estimate of the return on each loan as a function of survival time.

The econometric model consists of two simultaneous equations, the first one for the binary decision to provide a loan or not, y_i, and the second one for the *natural logarithm* of survival time of a loan (in days), for notational simplicity denoted by t_i. Because the bank from which we obtained our data merely considered whether it would accept an application or not, all people who were granted a loan received the amount of credit they applied for at the going rate of interest. The first equation therefore models a binary decision. I do not model how individuals determine the amount of credit they apply for.

I use the superscript $*$ to indicate an unobserved variable and let y_i^* and t_i^* follow

$$\begin{array}{rcl} y_i^* & = & \mathbf{x}_{1i}.\boldsymbol{\beta}_1 + \varepsilon_{1i} \\ t_i^* & = & \mathbf{x}_{2i}.\boldsymbol{\beta}_2 + \varepsilon_{2i} \quad \text{for } i = 1, 2, ..., N \end{array} \tag{3.1}$$

where the disturbances are assumed to be bivariate normal distributed.

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{matrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix}$$

As mentioned, y_i, is a binary choice variable that takes value 1 if the loan was granted and 0 if the application was rejected:

$$y_i = \begin{cases} 0 & if \quad y_i^* < 0 \\ 1 & if \quad y_i^* \geq 0 \end{cases} \tag{3.2}$$

For loans that turn bad, one can observe the exact survival time. For loans that are still performing on the day of monitoring, survival is censored because we

do not know if and when they will turn bad. Because all loans are monitored on October 9, 1996, but are granted anywhere between September 1994 and August 1995, the good loans' survival times will be censored at varying thresholds. For example, a loan granted on September 1, 1994, has a censoring threshold of 768 days. For a loan granted on August 31, 1995, this is 434 days. A loan's censoring threshold for survival time will be denoted as $\overline{t_i}$. The above can be summarized in the following censoring rule:

$$t_i = \begin{cases} t_i^* & if \quad t_i^* < \overline{t_i} \\ \overline{t_i} & if \quad t_i^* \geq \overline{t_i} \end{cases} \tag{3.3}$$

Due to the fact that one only observes survivals for loans that are actually granted, there is not only a censoring rule for t_i but even an *observation* rule:

Figure 1: Observation rule for y_i and t_i.
Entries in the 2×2 table show pairs $(y_i,\ t_i)$ that are observed for all ranges of y_i^* and t_i^*.

	$t_i^* \leq \overline{t_i}$	$t_i^* > \overline{t_i}$	
$y_i^* < 0$	$(0, .)$	$(0, .)$	
$y_i^* \geq 0$	$(1, t_i^*)$	$(1, \overline{t_i})$	

A dummy variable d_i splits up the sample of granted loans into good ones and bad ones. If a loan's survival is uncensored, $t_i^* \leq \overline{t_i}$, it must be a defaulted one. If survival is censored, it must be a good loan.

$$d_i = \begin{cases} 0 & if \quad t_i^* \leq \overline{t_i} \\ 1 & if \quad t_i^* > \overline{t_i} \end{cases}$$

Because we have three types of observations: no loans, bad loans with survival t_i, and good loans with survival $\overline{t_i}$, the likelihood function will take the following form:

$$\ell \ = \ \prod_{no\ loans} \mathrm{pr}\,(no\ loan) \cdot \prod_{bad\ loans} \mathrm{pr}\,(t_i\ \cap\ bad\ loan) \times \prod_{good\ loans} \mathrm{pr}\,(\overline{t_i}\ \cap\ good\ loan) \tag{3.4}$$

Combining (3.2), (3.3) and Table 1, equation (3.4) becomes

$$\begin{aligned} \ell \;\; = \;\; & \prod_{i=1}^{N} \mathrm{pr}\left(y_i^* < 0\right)^{(1-y_i)} \cdot \prod_{i=1}^{N} \mathrm{pr}\left(y_i^* \geq 0,\; t_i\right)^{y_i \cdot (1-d_i)} \times \\ & \prod_{i=1}^{N} \mathrm{pr}\left(y_i^* \geq 0,\; t_i^* \geq \overline{t_i}\right)^{y_i \cdot d_i} \end{aligned} \tag{3.5}$$

In appendix **A.1.** it is shown that (3.5) implies the following loglikelihood:

$$\begin{aligned} \ln\ell \;\; = \;\; & \textstyle\sum_{i=1}^{N} (1-y_i) \cdot \ln\left[1 - \Phi\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1\right)\right] + \\ & \textstyle\sum_{i=1}^{N} y_i \cdot (1-d_i) \;\; \left\{ \ln\Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1 - \frac{\sigma_{12}}{\sigma_2^2}\left(t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right)}{\sqrt{(1-\rho^2)}} \right) + \right. \\ & \left. -\tfrac{1}{2}\ln 2\pi + \ln\left(\tfrac{1}{\sigma_2}\right) - \tfrac{1}{2}\left(\tfrac{t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2} \right)^2 \right\} + \\ & \textstyle\sum_{i=1}^{N} y_i \cdot d_i \;\; \ln\Phi_2\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\; \frac{\mathbf{x}_{2i}\boldsymbol{\beta}_2 - \overline{t_i}}{\sigma_2};\; \rho \right) \end{aligned} \tag{3.6}$$

where $\Phi(\cdot)$ and $\Phi_2(\cdot,\cdot,\rho)$ represent the univariate and bivariate standard normal c.d.f., the latter with correlation coefficient ρ.

4. Empirical results

To find out which of the 60-70 variables in the dataset have sufficient explanatory power to become significant in the estimation of the final model, I went through several steps. First, I picked out the defaulted loans, that have completed spells of survival time, from the dataset. For this subsample I calculated simple correlation coefficients for the candidate explanatory variables, both with each other and with survival time. For categoric variables I also compared mean survival times between categories. These steps gave me a preliminary impression of what variables were close substitutes. Variables that were strongly correlated were added to the set of potential covariates. For the variables with significant correlation coefficients, I estimated linear regressions and inspected the parameter estimates and coefficients of determination. For the remaining variables, I proceeded by estimating Lowess regressions.

Next, I split up the dataset into rejected and approved applications and contrasted the mean values of all candidate explanatory variables for these two subsamples. I also compared the distribution of some of the explanatory variables between subsamples, both to check if the failure to discover any difference in means could be ascribed to the presence of outliers and to see if there was any apparent gain from transforming the variable.

If no relationship between a variable and either survival time or the loan granting decision could be discovered in any of the above steps, then the variable was deleted from the set of candidate explanatory variables. This procedure led to the selection of just over 20 variables from the dataset. Among the variables that were scrapped are the number of months since the most recent change in marital status, number of houses a person owns (partially), ownership of a registered company, a large number of entries on the two most recently submitted income-tax return forms, like total taxes due, back tax, and a number of transformations of these variables. *Taxable* wealth, although likely to be positively correlated with survival time, was excluded as an explanatory variable because assets up to a value of SEK 900,000 are exempted from taxes. This makes the group of people with *taxable* wealth too small to be useful. Instead I have exploited that income from capital is taxed and therefore registered from the first krona and have created a dummy explanatory variable 'income from capital'.

A number of variables, although they *were* selected in the above procedure, are not part of the final model because they are too strongly correlated with others that measure approximately the same thing but have greater explanatory power. The numerous income measures in the dataset and *BALANCE* were eliminated in this way in favor of *INCOME* and *LIMIT*. Some other variables were omitted from the final model because the bivariate relation with the dependent variable(s) turned insignificant when estimating the model with multiple explanatory variables. Age, citizenship (Swedish, nordic, non-nordic), the number of months since immigration, the combined value of all real estate a person has (partial) ownership in, and *BALANCE/INCOME* were removed in this way.

Finally, it is worthwhile to make some remarks on the distribution of survival time. Table 4 may have created the impression that the distribution of logarithmized survival time for bad loans is more skewed than untransformed survival. A QQ normality graph (not shown here), that compares a variable's sample distribution with a normal distribution with equal mean and variance, shows, however, that the transformation reduces the skewedness of survival time and improves the match with the normal distribution slightly.

After selecting the explanatory variables, the parameters of (3.6) are estimated with the following procedure. First, I calculate starting values for $\boldsymbol{\beta}_1$ from a univariate probit on the first equation in (3.1). These are consistent although not efficient, because the covariation between ε_1 and ε_2 is not taken into account. The starting values for $\boldsymbol{\beta}_2$ and σ_2 come from a Tobit model with variable censoring bound on the survival time of the granted loans. This model implicitly assumes that $\rho = 0$. Under the restriction that $\rho = 0$, one can estimate the second equation in (3.1) separately. Because one ignores the rejected loan applications, these parameter estimates suffer from a sample selection bias and are inconsistent if $\rho \neq 0$ - which is the case here, as we will see below. In all tests of the the model with simulated data, however, these estimates were found to be close (plus minus a decimal) to the true parameter values. The iterative procedure on the full model *with* sample selection converged rather easily when using these estimates as starting values. By comparison, when I let either an OLS or a Heckman's two-step procedure generate the starting values for $\boldsymbol{\beta}_2$ and σ_2 - thus taking the sample selection effect into account while ignoring the censoring in t_i - it was more time-consuming or even impossible to find a maximum for the loglikelihood function (3.6).

With these starting values and letting $\rho^{start} = 0$, I then estimate $\boldsymbol{\beta}_2$, σ_2 and ρ simultaneously by maximizing (3.6) under the restriction that $\boldsymbol{\beta}_1 = \widehat{\boldsymbol{\beta}}_1^{probit}$. These estimates of $\boldsymbol{\beta}_2$, σ_2 and ρ are consistent and are in their turn used as starting values in the last step. Estimating $\boldsymbol{\beta}_2$, σ_2 and ρ first and then estimating $\boldsymbol{\beta}_2$, σ_2, ρ *and* $\boldsymbol{\beta}_1$ by FIML saves a lot of time compared to doing FIML directly. The FIML iterations provide consistent and efficient estimators of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, σ_2 and ρ and a consistent estimator of the variance-covariance matrix. The FIML parameter estimates, their standard errors and t-statistics are presented in Tables 5, 6 and 7.

Table 5 contains two sets of parameter estimates for the loan granting decision: the first one from estimation as a single equation and the second from estimation together with the the survival equation. There appears to be no clear gain in efficiency in the estimate of β_1 from estimating the two equations in (3.1) simultaneously. Remember that *LOANSIZE* could not be used as an explanatory variable because no data on this variable were available for rejected applications.

Table 5: Univariate probit and full information probit MLE of β_1.
The univariate estimators come from separate estimation of the first equation in (3.1) ; the bivariate estimators come from estimation of the complete model (3.1) - (3.4).

	Univariate			Bivariate		
Variable	$\widehat{\beta}_1$	std. error	t-stat.	$\widehat{\beta}_1$	std. error	t-stat.
CONSTANT	-.3361	.05125	-6.56	-.3277	.05122	-6.40
MALE	-.2068	.02812	-7.36	-.1955	.02794	-7.00
MARRIED	-.2416	.02969	-8.13	-.2328	.02951	-7.89
DIVORCE	-.1859	.03953	-4.70	-.1792	.03946	-4.54
HOUSE	.1103	.02802	3.93	.1025	.02820	3.63
BIGCITY	-.2321	.02676	-8.67	-.2223	.02655	-8.37
NRQUEST	-.007228	.005114	-1.41	-.004293	.005058	-.85
ENTREPR	.5697	.06386	8.92	.5703	.06423	8.88
INCOME	.009098	.0001817	50.06	.008863	.0001823	48.63
DIFINC	-.002429	.0003505	-6.93	-.002366	.0003480	-6.80
CAPINC	-.2837	.05098	-5.56	-.2717	.04995	-5.44
ZEROLIM	-2.2529	.1062	-21.22	-2.2180	.1135	-19.54
LIMIT	-.008609	.0001870	-46.04	-.008476	.0002082	-40.71
NRLOANS	.08621	.006834	12.62	.08641	.006961	12.41
LIMUTIL	-.007465	.0004487	-16.67	-.007587	.0004491	-16.89
COAPPLIC	.1559	.03413	4.57	.1463	.03432	4.26

Critical values are 1.645, 1.96,and 2.575 for the 10, 5, and 1 percent significance levels.

The effect of most variables on the probability of obtaining a loan is as one might have expected. *INCOME* and *HOUSE* confirm their role as important factors that contribute positively, while *LIMIT, LIMUTIL* and *DIVORCE* have the traditional negative effects. More surprising are the coefficients on *MARRIED, DIFINC* and *CAPINC.* The parameter on *MARRIED* may be capturing the positive correlation between age and marriage. In preliminary regressions where age was one of the explanatory variables, it consistently had a negative effect on the probability of being granted a loan. Its parameter estimate failed to gain significance, though.

Table 6: Univariate and bivariate Tobit MLE of β_2.
The univariate estimates are computed under the hypothesis that $\rho = 0$; the bivariate estimation takes the sample selection effect into account and estimates ρ.

	Univariate			Bivariate		
Variable	$\widehat{\beta}_2$	std. error	t-stat.	$\widehat{\beta}_2$	std. error	t-stat.
CONSTANT	8.2464	.1555	53.05	9.0647	.1925	47.08
MALE	-.1060	.06085	-1.74	.02395	.05844	.41
MARRIED	.1869	.06823	2.74	.3449	.06611	5.21
DIVORCE	-.1237	.07820	-1.58	-.008730	.06936	-.13
HOUSE	.06070	.06114	.99	-.02330	.06199	-.38
BIGCITY	-.1284	.05808	-2.21	.2325	.05250	.44
NRQUEST	-1.1547	.1218	-9.48	-.9673	.1090	-8.87
ENTREPR	.1355	.1809	.75	.1623	.1595	1.02
INCOME	.03790	.04189	.90	-.2862	.05010	-5.71
DIFINC	.1469	.07490	1.96	.1846	.07856	2.35
CAPINC	-.05713	.1233	-.46	.1948	.09965	1.95
ZEROLIM	-2.2441	.4006	-5.60	-.3277	.1403	-2.34
LIMIT	.005818	.05886	.10	.5614	.04979	11.27
NRLOANS	3.2884	.2560	12.85	2.5869	.2249	11.50
LIMUTIL	-.1295	.01203	-10.76	-.1223	.01156	-10.58
LOANSIZE	-.06863	.07300	-.94	-.06995	.06974	-1.00
COAPPLIC	.5091	.1087	4.68	.3736	.1037	3.60
σ_2	.9187	.04482	20.50	1.0961	.05684	19.28
ρ	$\equiv$0.00	-	-	-0.9855	.02137	-46.11

Critical values are 1.65, 1.96,and 2.58 for the 10, 5, and 1 percent significance levels.

Table 6 compares two different estimators of $\boldsymbol{\beta}_2$ and σ_2. The parameter estimates in the first column are obtained from a Tobit model with a variable censoring threshold that ignores the sample selection effect one generates when disregarding the rejected loan applications. This is equivalent to estimating $\boldsymbol{\beta}_2$ and σ_2 in (3.1) under the hypothesis that $\rho = 0$. One is, in other words, assuming that the likelihood of a survival of a certain length is not affected in any systematic way by the inferences one can make from observing y_i and x_{1i}. If the hypothesis is true, then the parameters in the first and second equation in (3.1)

can be estimated separately from each other. However, if the disturbances ε_1 and ε_2 *are* correlated, these estimators of $\boldsymbol{\beta}_2$ and σ_2 will be biased.

The second set of coefficients in Table 6 are the consistent parameters estimates of $\boldsymbol{\beta}_2$ and σ_2 obtained by estimating the complete model (3.1) – (3.3).

The purpose of comparing these two estimators is to investigate to what extent any misunderstandings about the relation between people's characteristics and financial discipline may have originated in an incorrect way of sampling data for profitability analyses by financial institutions. A comparison of the two estimators will help us determining if inconsistencies in bank lending policy may find their origin in a sample selection bias. If there is any such sample selection bias, we will also want find out whether it is also quantitatively importantant.

In the final model, all explanatory variables enter the model linearly. I have checked for the presence of non-linear effects by adding quadratic terms of all continuous variables. Their coefficients were never significant, however. Out of 16 explanatory variables, four lose or reduce their significance and three turn significant or increase their level of significance when disregarding the sample selection effect. Of the parameters for the remaining 9 variables, 4 are insignificant while the remaining 5 are significant and have identical signs in both models. So although accounting for the sample selection effect never reverses the sign of any of the coefficients, it does clearly affect their magnitude. The influence of the variable *ZEROLIM*, for example, would be badly overestimated if one did not account for the sample selection effect. A look at Tables 2 and 3 may help us understand this phenomenon. Although having no loans outstanding is rather uncommon among the granted loans, it is stronger associated with defaulting than with proper repayment behavior. However, this overlooks the fact that 15% of all rejected applicants did not have any loan yet. If rejected applications are not so much different from approved ones, then the actual impact of having a zero limit may well be much smaller than one would expect by merely looking at granted loans.[2] Similarly, *INCOME* is not significant in the column with biased estimators, whereas the consistent parameter estimate has a significantly negative coefficient. Although one should be careful not to rationalize each counter-intuitive finding, we can look for a tentative explanation. Tables 2 and 5 clearly showed that people with higher incomes are more likely to be granted a loan. This may well lead us to infer - if we disregard the rejected applicants, who have low incomes, and consider only approved ones - that income does not influence a loan's default risk. Suppose,

[2] Rejected applications will differ very little from approved ones if the lending institution grants loans to applicants on the basis of characteristics that have little impact on survival.

however, that it is actually the case that other factors than *INCOME* determine a loan's survival. Then the selection of applicants may be taking place on the basis of a negative bivariate relation between *INCOME* and defaults (see Table 3) that disappears when one controls for both the sample selection effect and the correlation with other variables. It may, for example, be the case that people with higher income also take greater risks.

It is also wortwhile to take notice of the sign of some other parameter estimates in the fourth column of Table 6. *NRQUEST* is considered to be quite good an indicator of a person's efforts to obtain additional credit and as such expected to contribute negatively to survival. Not having any loan at all, as indicated by *ZEROLIM*, is a sign of inexperience with servicing debt and has a negative effect on survival. The reverse holds for *NRLOANS* and *LIMIT*. The positive effect on survival of two granted loan evens out the negative effect of five questions. Although one might expect *LIMIT* to have a negative influence, one should keep in mind that it is merely the ceiling of the credit facility that a person disposes of. *LIMUTIL* captures the extent to which he or she actually uses it, while *LIMIT* proxies for experience with servicing debt in the same way as *NRLOANS* does.[3] A rise in income between years increases expected survival while a higher utilization degree of the available credit facility by an applicant decreases survival. Finally, it is worth commenting the value of the correlation coefficient. Its value of -.98 may create the impression that the algorithm had problems converging. In extensive tests of the model with different sets of explanatory variables and varying sample sizes, rho took values between -.55 and -.98. In tests with the bivariate probit model, the final parameter estimates of which are reported in Tables 5 and 7, ρ ranged from approximately -.65 to -.93. Boyes et al. report -.35. As is the case with most models with limited dependent variables (see Bermann [4]), the computations for the tobit and probit models did not converge for some configurations of explanatory variables. When the computations broke down, divergence always took place after relatively few iterations, however, with rho breaking its constraint before any of the other parameters had stabilized around a final value. In the estimation of the final model, all parameters settled down around their final values rather quickly.

[3]Strong correlation between the variables $BALANCE$ and $LIMIT$ tended to create numerical problems when trying to use both as explanatory variables. Some test regressions indicated that $LIMIT$ and $BALANCE$ have opposite effects on $SURVIVAL$, the former a positive and the latter a negative. The coefficient on $LIMIT$ in tables 5 and 6 is approximately equal to the net effect of $LIMIT$ minus $BALANCE$.

Table 7: Bivariate probit and Tobit MLE with sample selection.
The probit estimator α_2 is the parameter in the equation that models the probability of a default; the tobit estimator β_2 comes from (3.1) - (3.4). Both estimators take the sample selection effect into account.

	Bivariate probit			Tobit with sample selection		
Variable	$\widehat{\alpha}_2$	std. error	t-stat.	$\widehat{\beta}_2$	std. error	t-stat.
CONSTANT	2.4546	.1132	21.69	9.0647	.1925	47.08
MALE	-.02338	.05916	-.40	.02395	.05844	.41
MARRIED	.2527	.06563	3.85	.3449	.06611	5.22
DIVORCE	-.07018	.07427	-.95	-.008730	.06936	-.13
HOUSE	-.01004	.06001	-.17	-.02330	.06199	-.38
BIGCITY	-.04068	.05480	-.74	.02325	.05250	.44
NRQUEST	-.1036	.01026	-10.10	-.9673	.1090	-8.87
ENTREPR	.1347	.1565	.86	.1623	.1595	1.02
INCOME	-.002266	.0005015	-4.52	-.2862	.05010	-5.71
DIFINC	.002049	.0007326	2.80	.1846	.07856	2.35
CAPINC	.1477	.1265	1.17	.1948	.09965	1.95
ZEROLIM	-.6796	.2982	-2.28	-.3277	.1403	-2.34
LIMIT	.004822	.0005693	8.47	.5614	.04979	11.27
NRLOANS	.2704	.01947	13.89	2.5869	.2249	11.50
LIMUTIL	-.01208	.0009290	-13.00	-.1223	.01156	-10.58
LOANSIZE	-.006581	.006850	-.96	-.06995	.06973	-1.00
COAPPLIC	.4189	.09789	4.28	.3736	.1037	3.60
σ_2	-	-	-	1.0961	.05684	15.68
ρ	-.9110	.05624	−16.20	-0.9855	.02137	-41.80

Critical values are 1.65, 1.96,and 2.58 for the 10, 5, and 1 percent significance levels.

Overall, the conclusion one can draw from the results in Table 6 is that ignoring rejected applicants in an analysis of the duration of loans leads to large

biases in the parameter estimates. Although the signs of parameters are never reversed, some of the variables that are generally thought to be among the most important determinants of creditworthiness, like income, outstanding loans, and income from assets, appear to have no relationship whatsoever with survival time when disregarding the sample selection effect. Such misunderstandings may well be the origin of inefficient lending policies at financial institutions.

Finally, in Table 7, I present parameters of the bivariate probit model, as presented in Boyes et al. [5] but re-estimated with the data used in Table 6. The first observation one can make when comparing the probit parameters that determine the probability of a loan *not* defaulting ($\boldsymbol{\alpha}_2$) with those that determine logged survival time ($\boldsymbol{\beta}_2$) is that each variable has coefficients with identical signs in both models.[4] [5] Variables that increase (decrease) the probability of a default thus also decrease (increase) the expected survival time of a loan and thus - since survival time proxies for return - reduce (raise) its expected return.

Moreover, variables like *MALE, DIVORCE, HOUSE, BIGCITY* and *ENTREPR* that are given significant weights in the loan granting decision actually do not affect default risk and survival. *NRQUEST* on the other hand does have a significant effect on survival but is not given any weight in the decision process. For variables like *MARRIED, INCOME, DIFINC, CAPINC* and *LIMIT*, the parameter estimates for the loan granting decision and the probability of a loan not defaulting have opposite signs. These variables are thus used in such a way by the bank in the loan granting process that they increase (decrease) the likelihood of a loan being granted although they in fact increase (decrease) the risk of default. Because the parameters in the survival equation have the same sign as the bivariate probit parameters, these variables also reduce (raise) expected survival and return on the loan. In other words: if the bank is not minimizing default risk in its loan granting policy, it is not doing so because loans with higher default risk have higher expected returns. Moreover, the negative values of ρ in Table 7 indicate that any non-systematic propensity to grant loans is associated with shorter survival times and higher default risk. This is consistent with the above observation that the bank does not appear to trade off risk against return. Rather, the loan

[4] As a matter of fact, all variables with significant parameters in the survival equation of the Tobit model with sample selection also have significant coeffficients in the 'probability that loan doesn't default' equation of the bivariate probit model. The reverse, however, does not hold! Variables that would have been sigificant in a bivariate probit model but are not in the Tobit model like (3.1) have therefore been omitted.

[5] The bivariate probit model implicitly assumes that loans which are still good, will not turn bad later on.

granting policy appears to be inefficient and contain non-systematic components that are strongly negatively correlated with survival.

In the estimation of $\boldsymbol{\alpha}_2$ and $\boldsymbol{\beta}_2$, we controlled for the size of the loan. Table 7 shows that neither default risk nor survival is affected by *LOANSIZE*. This has two implications. First, bigger loans do not carry greater default risk nor do they imply either shorter or longer survivals. Secondly, greater default risk is associated with shorter survival, not with *longer* survival as Boyes et al. suggest. Riskier loans thus have *lower* expected returns. The lending institution that we study, however, always extended loans with size equal to the amount applied for - independent of the risk associated with the applicant - and was thus indifferent between alternative loan sizes. Such behavior is not consistent with the hypothesis that the financial institution is trading off higher default risk against higher expected earnings (that supposedly come with bigger loans). The lending institution's behavior is neither compatible with return maximization due to a (previously assumed) positive relation between loan size and rate of return nor is it in agreement with the maximization of survival time in general. If the lending institution is minimizing default risk, it would be strictly better off granting either nothing at all or the maximum amount possible for the type of loan in question. After all, granting a bigger loan does not increase risk but it raises the revenue. For the same reason, the lending institution would also be better off with this corner solution policy if it is maximizing survival time. It raises revenues without changing the riskiness.

Model (3.1) – (3.3) can now be used to examine the lending policy of the bank. Loans with the longest expected survival time also have the greatest gross returns. The estimated model from Tables 6 and 7 can be used to calculate the expected survival time for all loan applicants. In Table 8, we show the outcome from an experiment where all loan applications are ranked and approved according to their predicted survival time $E\left[t_i^* \middle| \mathbf{x}_{2i}\right]$. The first column in the table shows that only 3,156 out of the 6,438 granted loans (49 percent) would have been approved if selection had taken place according to expected survival time. This strongly suggests that the current lending policy is not efficient, because it selects loans with shorter survivals. These results could, however, also be indicative of an inability by the empirical model to separate good from bad loans. If we look at some other ways to evaluate the reliability of the results in Table 8, then this seems to confirm that the model is an effective tool to evaluate loan applicants. The loans that would be granted with a survival time selection criterion contain

Figure 8: Selecting applications by predicted survival times.
Entries in the 2×2 table show how many applicants that were granted a loan would even be so if applicants were ranked according to predicted survival time $E\left[t_i^* \middle| \mathbf{x}_{2i}\right]$ and the same number of loans were granted as in the data set.

p		actual			# failed loans
r		granted	rejected	sum	among predicted
e	granted	3,156	3,282	6,438	39
d	rejected	3,282	3,617	6,899	349
i	sum	6,438	6,899	13,337	388
c					
t					

merely 39 of the 338 bad loans that are present in the data set. The predicted (logarithm of) survival time for defaulted loans was 10 percent shorter than for all granted loans and all other applicants.

5. Discussion

Traditionally, the objective of credit scoring models used by financial institutions is to minimize default rates or the number of loans that is incorrectly classified as defaulted or non-defaulted. From a profit or utility maximizing perspective, however, it is not only important to know *if* but also *when* a loan will default. Traditional credit scoring models predict default risk and therefore fail to take into account this multiperiod nature of loans contracts. To allow for a more realistic evaluation of the return on a loan, a Tobit model with sample selection and variable censoring thresholds has been constructed and estimated in this paper. This model is shown to be a useful tool to predict the expected survival time on a loan to any kind of applicant. A comparison with a nested model that disregards rejected applications - as has been common in studies of creditworthiness - shows that ignoring the sample selection effect leads to a large bias in the parameters estimates.

From the empirical results we gain several insights. They confirm the findings in Boyes et al. that financial institutions' lending policies are not compatible with default risk minimization. At the same time, though, the results also conflict

with the notion that the financial institution would be trading off higher default risk against higher returns. The lending policy does not favor people that survive longer and thus have a higher rate of return. Firstly, some of the variables that increase (decrease) applicants' odds of obtaining a loan reduce (raise) the expected survival time (and thus return) on a loan and raise (reduce) the likelihood of a default. Secondly, the financial institution is found to be indifferent between loans of different sizes, given its expected survival time. There is thus no evidence of banks' behaving in a way that is consistent with profit-maximization. This impression is strengthened by an experiment in which expected survival times are calculated for all loan applicants, including those who were rejected. In that experiment, only 49 percent of the actually granted loans would have been granted if a survival time criterion had been handled. Moreover, lending to 349 of the 388 defaulted applicants in the sample would have been avoided.

Lending behavior by banks must thus be either a symptom of an inefficient lending policy or the result of some other type of optimizing behavior. The current level of technology in the banking industry generally does not yet allow for the pursuit of composite objectives such as the return on a range of products or revenues from several sources of income. But banks may, for example, be maximizing some other objective like provision income from the turnover on credit cards, the number of customers or lending volume subject to a minimum return constraint. None of these suggestions agree, however, with the practices reported to us by the lending institution who provided our data. Rather, the results bear strong evidence of a lending institution that has attempted to minimize risk or maximize a simple return function without success.

Censoring of data, as is the case with the non-defaulted loans in the sample, increases the uncertainty in the parameter estimates of the survival function. Appropriate changes in sampling methods can improve their accuracy. A longer period of observation of the loans would reduce the regression error. Even better would be to set up an experiment where a predetermined number of applicants is granted a loan without consideration of their personal characteristics. If each loan is monitored at least at termination of the contract then separate survival time functions for good and bad loans can be estimated. An ideal model of bank profitability or bank efficiency will have to be built on time series data for fees, interest payments and amortizations on loans, personal characteristics, macroeconomic indicators and all costs involved.

References

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

[2] Amemiya, Y., (1985), *Advanced Econometrics*, Harvard University Press, Cambridge MA.

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

[4] Bermann, G., (1993), *Estimation and inference in bivariate and multivariate ordinal probit models*, dissertation, Department of statistics, Uppsala University.

[5] Boyes, W.J., D.L. Hoffman and S.A. Low, (1989), An Econometric Analysis of the Bank Credit Scoring Problem, *Journal of Econometrics* 40, 3-14.

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

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

[8] Gale, D., and M. Hellwig, (1985), Incentive-compatible debt contracts: The one-period problem, *Review of Economic Studies* LII, 647-663.

[9] Greene, W.E., (1993), *Econometric Analysis*, 2nd edition, Macmillan, New York.

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

[11] Stiglitz, J.E., and A. Weiss, (1981), Credit rationing in markets with imperfect information, *American Economic Review* 71, 393-410.

[12] Townsend, R.M., (1982), Optimal multiperiod contracts and the gain from enduring relationships under private information, *Journal of Political Economy* 90 (61), 1166-1186.

[13] Williamson, S., (1987), Costly monitoring, loan contracts and equilibrium credit rationing, *Quarterly Journal of Economics* 102 (1), 135-145.

A. Likelihood function and gradient

A.1. Likelihood function

The likelihood function

$$\begin{array}{rcl} \ell & = & \prod_{i=1}^{N} \text{pr}\left(y_i^* < 0\right)^{(1-y_i)} \cdot \prod_{i=1}^{N} \text{pr}\left(y_i^* \geq 0,\ t_i\right)^{y_i \cdot (1-d_i)} \times \\ & & \prod_{i=1}^{N} \text{pr}\left(y_i^* \geq 0,\ t_i^* \geq \overline{t_i}\right)^{y_i \cdot d_i} \end{array} \tag{A.1}$$

implies that

$$\begin{array}{rcl} \ln\ell & = & \sum_{i=1}^{N} (1-y_i) \cdot \ln\left[\text{pr}\left(\varepsilon_{1i} < -\mathbf{x}_{1i}\boldsymbol{\beta}_1\right)\right] + \\ & & \sum_{i=1}^{N} y_i \cdot (1-d_i)\ \ln\ \left[\text{pr}\left(\varepsilon_{1i} \geq -\mathbf{x}_{1i}\boldsymbol{\beta}_1 \cap \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right)\right] + \\ & & \sum_{i=1}^{N} y_i \cdot d_i \ln\left[\text{pr}\left(\varepsilon_{1i} \geq -\mathbf{x}_{1i}\boldsymbol{\beta}_1 \cap \varepsilon_{2i} \geq \overline{t_i} - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right)\right] \end{array} \tag{A.2}$$

If we use that $\varepsilon_{1i}|\varepsilon_{2i} \sim N\left(\frac{\sigma_{12}}{\sigma_2^2}\varepsilon_{2i},\ (1-\rho^2)\right)$ for $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim \mathbf{N}(\mathbf{0},\ \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, then we can simplify by expressing the second line in terms of a univariate normal cdf and - pdf. As a result

$$\begin{array}{l} \text{pr}\left(\varepsilon_{1i} \geq -\mathbf{x}_{1i}\boldsymbol{\beta}_1 \cap \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right) \Leftrightarrow \\ \text{pr}\left(\varepsilon_{1i} \geq -\mathbf{x}_{1i}\boldsymbol{\beta}_1 |\ \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right) \text{pr}\left(\ \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right) \Leftrightarrow \\ \text{pr}\left(\varepsilon_{1i} < \mathbf{x}_{1i}\boldsymbol{\beta}_1 |\ \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right) \text{pr}\left(\ \varepsilon_{2i} = t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2\right) \Leftrightarrow \\ \Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1 - \frac{\sigma_{12}}{\sigma_2^2}(t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2)}{\sqrt{(1-\rho^2)}}\right) \frac{1}{\sigma_2}\phi\left(\ \frac{t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2}\right) \end{array} \tag{A.3}$$

Taking natural logarithms we get

$$\ln\Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1 - \frac{\sigma_{12}}{\sigma_2^2}(t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2)}{\sqrt{(1-\rho^2)}}\right) - \frac{1}{2}\ln 2\pi + \ln\left(\frac{1}{\sigma_2}\right) - \frac{1}{2}\left(\ \frac{t_i - \mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2}\right)^2 \tag{A.4}$$

The last line in $(A.2)$ can be rewritten in terms of a bivariate normal cdf:

$$\begin{array}{l} \text{pr}\left(\varepsilon_{1i} \geq -\mathbf{x}_{1i}\boldsymbol{\beta}_1 \cap \frac{\varepsilon_{2i}}{\sigma_2} \geq \frac{\overline{t_i} - \mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2}\right) \Leftrightarrow \ ^{6} \\ \Phi_2\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \frac{\mathbf{x}_{2i}\boldsymbol{\beta}_2 - \overline{t_i}}{\sigma_2};\ \rho\right) \end{array} \tag{A.5}$$

[6] See Greene (1993) p.661 for a summary of results on the bivariate normal cdf.

Consequently, the loglikelihood function can be written as

$$\begin{array}{rcl} \ln\ell & = & \sum_{i=1}^{N}(1-y_i)\cdot\ln\left[1-\Phi\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1\right)\right]+ \\ & & \sum_{i=1}^{N} y_i\cdot(1-d_i)\ \left\{\ln\Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1-\frac{\sigma_{12}}{\sigma_2^2}(t_i-\mathbf{x}_{2i}\boldsymbol{\beta}_2)}{\sqrt{(1-\rho^2)}}\right)+\right. \\ & & \left.-\frac{1}{2}\ln 2\pi+\ln\left(\frac{1}{\sigma_2}\right)-\frac{1}{2}\left(\frac{t_i-\mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2}\right)^2\right\}+ \\ & & \sum_{i=1}^{N} y_i\cdot d_i\ \ln\Phi_2\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \frac{\mathbf{x}_{2i}\boldsymbol{\beta}_2-\overline{t_i}}{\sigma_2};\ \rho\right) \end{array} \tag{A.6}$$

After further simplification, by setting

$$\begin{array}{rcl} \boldsymbol{\alpha}_2 & = & \boldsymbol{\beta}_2/\sigma_2 \\ lng2 & = & \ln\left(\frac{1}{\sigma_2}\right) \\ e_{2i} & = & -\left(\frac{t_i-\mathbf{x}_{2i}\boldsymbol{\beta}_2}{\sigma_2}\right) \\ & = & \mathbf{x}_{2i}\boldsymbol{\alpha}_2-t_i\cdot\exp\left(lng2\right) \\ \overline{e_{2i}} & = & \mathbf{x}_{2i}\boldsymbol{\alpha}_2-\overline{t_i}\cdot\exp\left(lng2\right) \\ \delta & = & \frac{1}{(1-\rho^2)^{1/2}} \end{array}$$

we get that

$$\begin{array}{rcl} \ln\ell & = & \sum_{i=1}^{N}(1-y_i)\cdot\ln\left[1-\Phi\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1\right)\right]+ \\ & & \sum_{i=1}^{N} y_i\cdot(1-d_i)\ \left\{\ln\Phi\left(\delta\left[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\ \rho\cdot e_{2i}\right]\right)+\right. \\ & & \left.-\frac{1}{2}\ln 2\pi+lng2-\frac{1}{2}e_{2i}^2\right\}+ \\ & & \sum_{i=1}^{N} y_i\cdot d_i\ \ln\Phi_2\left(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \overline{e_{2i}};\ \rho\right) \end{array} \tag{A.7}$$

The parameters with respect to which we maximize $\ln\ell$ are $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_2$, $lng2$, and ρ.

A.2. Gradients

The gradients corresponding to each observation in the above loglikelihood function are:

$$
\begin{aligned}
\tfrac{\partial \ln \ell_i}{\partial \boldsymbol{\beta}_1} &= (1-y_i)\,\tfrac{-\phi(\mathbf{x}_{1i}\boldsymbol{\beta}_1)}{1-\Phi(\mathbf{x}_{1i}\boldsymbol{\beta}_1)}\cdot \mathbf{x}_{1i} + \\
&\quad y_i\,(1-d_i)\cdot \tfrac{\phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}{\Phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}\cdot \delta\mathbf{x}_{1i} + \\
&\quad y_i\cdot d_i\;\frac{\phi(\mathbf{x}_{1i}\boldsymbol{\beta}_1)\Phi\left(\frac{\overline{e_{2i}}-\rho\mathbf{x}_{1i}\boldsymbol{\beta}_1}{(1-\rho^2)^{1/2}}\right)}{\Phi_2(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \overline{e_{2i}};\ \rho)}\cdot \mathbf{x}_{1i} \\
\tfrac{\partial \ln \ell_i}{\partial \boldsymbol{\alpha}_2} &= y_i\cdot(1-d_i)\ \left\{\tfrac{\phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}{\Phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}\delta\,(-\rho)\,\mathbf{x}_{2i} - e_{2i}\mathbf{x}_{2i}\right\} + \\
&\quad y_i\cdot d_i\frac{\phi(\overline{e_{2i}})\Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho\overline{e_{2i}}}{(1-\rho^2)^{1/2}}\right)}{\Phi_2(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \overline{e_{2i}};\ \rho)}\cdot \mathbf{x}_{2i} \\
\tfrac{\partial \ln \ell_i}{\partial lng2} &= y_i\cdot(1-d_i)\left\{\tfrac{\phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}{\Phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}\delta\,(-\rho)\,(-t_i)\exp(lng2) + \right. \\
&\qquad \left. +1 + e_{2i}\cdot t_i\cdot\exp(lng2)\right\} + \\
&\quad y_i\cdot d_i\left\{\frac{\phi(\overline{e_{2i}})\Phi\left(\frac{\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho\overline{e_{2i}}}{(1-\rho^2)^{1/2}}\right)}{\Phi_2(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \overline{e_{2i}};\ \rho)}\cdot(-\overline{t_i}\cdot\exp(lng2))\right\} \\
\tfrac{\partial \ln \ell_i}{\partial \rho} &= y_i\cdot(1-d_i)\ \tfrac{\phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}{\Phi(\delta[\mathbf{x}_{1i}\boldsymbol{\beta}_1-\rho e_{2i}])}\,(-\delta e_{2i} + \rho\delta^3\,[\mathbf{x}_{1i}\boldsymbol{\beta}_1 - \ \rho e_{2i}]) + \\
&\quad y_i\cdot d_i\cdot\tfrac{\phi_2(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\overline{e_{2i}};\ \rho)}{\Phi_2(\mathbf{x}_{1i}\boldsymbol{\beta}_1,\ \overline{e_{2i}};\ \rho)}
\end{aligned}
\tag{A.8}
$$

After convergence of the iterative procedure, a consistent estimator of the variance-covariance matrix is obtained by applying the delta method.[7] If we define $\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\\ \sigma_2\\ \rho\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\alpha}_2\,[\exp(lng2)]^{-1}\\ [\exp(lng2)]^{-1}\\ \rho\end{pmatrix} \equiv \mathrm{f}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\alpha}_2\\ lng2\\ \rho\end{pmatrix}$, then

$$
\boldsymbol{\Gamma} \equiv \frac{\partial \mathrm{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \mathbf{I}_{K_1} & \mathbf{0}_{\mathbf{K}_1\times\mathbf{K}_2} & \mathbf{0}_{\mathbf{K}_1\times 1} & \mathbf{0}_{\mathbf{K}_1\times 1} \\ \mathbf{0}_{\mathbf{K}_2\times\mathbf{K}_1} & \frac{\mathbf{I}_{K_2}}{[\exp(\mathrm{lng2})]} & \frac{-\boldsymbol{\alpha}_2}{[\exp(\mathrm{lng2})]} & \mathbf{0}_{\mathbf{K}_2\times 1} \\ \mathbf{0}_{1\times\mathbf{K}_1} & \mathbf{0}_{1\times\mathbf{K}_2} & \frac{-1}{[\exp(\mathrm{lng2})]} & 0 \\ \mathbf{0}_{1\times\mathbf{K}_1} & \mathbf{0}_{1\times\mathbf{K}_2} & 0 & 1 \end{bmatrix}
\tag{A.9}
$$

In some of the iterations, I also made the additional transformation $\rho = \frac{\exp x - 1}{\exp x + 1}$ to assure that $0 \le \rho \le 1$. In those cases the gradient $\partial \ln\ell/\partial\rho$ needs to be

[7] A description of the delta method is provided in Greene [9], pp. 297.

multiplied by the term $\partial\rho/\partial x = \frac{2\exp x}{(\exp x\ +1)^2}$ while the 4th diagonal element in $\mathbf{\Gamma}$ needs to be set equal to $\partial\rho/\partial x$.