Customer revolving credit – how the economic conditions make a difference

Natasa Sarlija
e-mail: natasa@efos.hr
University of J.J. Strossmayer in Osijek
Faculty of Economics
Gajev trg 7, 31000 Osijek, Croatia

Mirta Bensic
e-mail: mirta@mathos.hr
University of J.J. Strossmayer in Osijek
Department of Mathematics
Gajev trg 6, 31000 Osijek, Croatia

Zoran Bohacek
e-mail: zbohacek@hub.hr
Croatian Banking Association
Nova Ves 17, 10000 Zagreb, Croatia

ABSTRACT:

The aim of this paper is to discuss credit scoring modeling of a customer revolving credit depending on customer application data and transaction behavior data influenced by specific economic conditions that exist in Croatia. Since Croatia is a country in transition with war consequences, changes in political, institutional and social systems and, above all, with specific economic conditions characterized by slow economy, a high unemployment rate and a relatively low personal income, it is assumed that this influences credit behavior of customers and, as a consequence, a composition of credit scoring models. For instance, it has been shown that a small business credit scoring model developed in Croatia is influenced by economic conditions. The data set for our research consisted of 50,000 customer accounts (application data and transaction data) in Croatia over the period of 12 months. We have developed a logistic credit scoring model and a survival-based credit scoring model in order to assess the relative importance of different variables in predicting default as well as profitability of a customer. The paper analyzes influences of economic conditions on credit scoring modeling.

KEYWORDS: credit scoring modeling, logistic regression, survival analysis, revolving credit

INTRODUCTION

Credit scoring is a process of determining how likely applicants are to default with their repayments (Hand, Henley, 1997). It includes various decision models and tools that assist credit managers in making credit decisions. The aim is to assess the risk of default associated with a credit product/decision. For several decades credit scoring models have been used in commercial and consumer lending and recently in small business lending. Scoring models can be divided into two types: (i) credit scoring in targeting and application, which deals with new applicants, and (ii) credit scoring in managing existing account called behavior scoring. Behavior scoring models deal with existing accounts and are used for making on-going decisions on open accounts. They are based on dynamic account performance information, and risk decisions are applied to account management, collections and recoveries areas of credit cycle (McNab, Wynn, 2000).

Selection of characteristics that are used in developing a credit scoring model for consumer credit depend on the type of credit, the amount of credit and the purpose of credit. If we talk about consumer application credit scoring, characteristics usually used in different consumer credit scoring models are the following: time at present address; home status; postcode; telephone; applicant's annual income; owing a credit card; type of bank account; age; type of occupation; purpose of loan; marital status; time with bank; time with employer; credit bureau rating; monthly debt as a proportion of monthly income; time at current job; number of dependents (Hand, Henley, 1997; Caouette, Altman, Narayanan, 1998; Desai et al. 1996; Koch, MacDonald, 2000). Behavior scoring models include application data as well as behavior data. Typical characteristics that are used in different behavior scoring models are: (i) delinquency history, (ii) usage history, (iii) application data, (iv) payment history, (v) collection activity, (vi) revolving credit transactions, (vii) customer service contacts, (viii) promotions history, (ix) credit bureau data (McNab, Wynn, 2000).

In this paper we deal with personal open-end accounts where clients can make different kinds of payments to the accounts, withdrawals from the accounts and can use revolving credits. Credit scoring models concerning the unsecured open-ended consumer credit accounts are typically based entirely on information found in credit reporting agency (Avery, Calem, Canner; 2004). It should be noticed that there is a lack of published papers that reveal composition of the models and if there is some, usually they are oriented on the developed counties. There is a problem when there is no credit bureau in the country we are developing a scoring model for, in which case there is no data and there is almost no previous research and experience about the credit scoring models. This paper should present some ideas about these issues. The main purpose of our work is to study customer behavior and to discover determinants of default in circumstances that exist in Croatia today – no credit bureau, no previous research and specific economic and social condition. There were some attempts in Croatia of applying scoring models developed in the foreign country, but they were unsuccessful. We believe that specific economic conditions in countries in transition also influence the credit scoring modeling. Being one of the post-communist transitional countries itself, Croatia shares all the typical transitional characteristics with other countries of that type. These common characteristics are: economic changes focused to the market economy development, then political, institutional, and social changes. Economic changes are marked with a rapid fall of economic activity, although the falling trend has been stopped in all countries, and even changed to the growth in some of them. Institutional changes include the development of institutions needed for enabling market economy, then liberalization of prices and foreign trade, and finally restructioning and privatization of businesses (Mervar, 2002). Croatia is a post-communist country in which the process of democratic transition coincided with the process of creating the state by means of a war (Kasapovic, 1996), therefore its economy is additionally featured by some specific war-caused characteristics. Since previous research (Selowsky and Martin,1997. in Mervar, 2002) proved a negative influence of war to economic development, it can be expected that the specific conditions caused by war will also influence credit scoring modeling on a Croatian dataset. Also, basic economic indicators: GDP per capita of 7732$, net wage and salary of 4137 kunas (600 EUR) and unemployment rate of 18% (HGK, 2004), lead us to assume that slow economy, high unemployment rate and relatively low personal income influence composition of consumer credit scoring model.

Data set for this research comes from one bank in Croatia. In order to achieve our goal, we apply two methods: (i) logistic regression which helps us to determine the probability that a client will default in the next 6 months and to reveal determinants of default; (ii) survival analysis which describes the probability that the loan defaults in the next instant period conditioned to the fact that it survived at least to the time t.

The structure of the paper is the following: a brief overview of logistic regression and survival analysis methodology used in the research is given in the next section. Then the previous research results are described. Data and variable section provides a description of data and variables used in our research, followed by the results of logistic regression and survival analysis. Using these results, important features for credit scoring models are discussed.

METHODOLOGY

Traditionally, different parametric models are used for classifying input vectors into one of two groups, which is the main objective of statistical inference on the credit scoring problem. Here we used a logistic regression as a classical one and survival-based model as a model which allow us to add a dynamic element of clients' behavior and at the end (among other advantages of such type of modeling) again answer the same question: classify client into one of the two groups (see eg. Banasik, Crook, Thomas, 1999).

Logistic regression modeling is widely used for analyzing multivariate data involving binary responses that we deal with in credit scoring modeling. It provides a powerful technique analogous to multiple regression and ANOVA for continuous responses. Since the likelihood function of mutually independent variables $Y_1,\ldots,Y_n$ with outcomes measured on a binary scale is a member of the exponential family with $\left(\log\left(\frac{\pi_1}{1-\pi_1}\right),\ldots,\log\left(\frac{\pi_n}{1-\pi_n}\right)\right)$ as a canonical parameter (π_j is a probability that Y_j becomes 1), the assumption of the logistic regression model is a linear relationship between a canonical parameter and the vector of explanatory variables $\mathbf{x}_j$ (dummy variables for factor levels and measured values of covariates):

$$\log\left(\frac{\pi_j}{1-\pi_j}\right) = \mathbf{x}_j^{\tau}\boldsymbol{\beta}$$

This linear relationship between the logarithm of odds and the vector of explanatory variables results in a nonlinear relationship between the probability of Y_j equals 1 and the vector of explanatory variables:

$$\pi_j = \exp(\mathbf{x}_j^\tau \boldsymbol{\beta}) / (1 + \exp(\mathbf{x}_j^\tau \boldsymbol{\beta}))$$

Survival-based model is made using the length of time before a loan defaults (let us denote it by T). The main interest in survival modeling is in description the probability that T is bigger then the given time t (the survivor function $S(t)$). The most interesting object for this is a hazard function, $h(t)$, which we can explain by the equation:

$$h(t)*t=\text{Prob}\{t \leq T \leq t+*t \mid T \geq t\}$$

i.e. which describes the probability that the loan defaults in the next instant period conditioned to the fact that it survived at least to the time t. (Frank, Harrell, 2001)

In fact, in the proportional hazard models, it is a hazard function that is modeled using the explanatory variable. The main assumption of a proportional hazard model is that the explanatory variables have a multiplier effect on the hazard rate which does not depend on the time t:

$$h(t) = \exp(\mathbf{x}^\tau \boldsymbol{\beta}) h_0(t)$$

Here we used the Cox modeling procedure because with this model it is possible to estimate the regression parameters without knowing the type of baseline hazard function $h_0(t)$. As we are not, in this moment, interested in the form of the hazard function but only in the variables that significantly influence the modeling procedure and in survival-based credit scoring model, it was enough for our purpose. (Frank, Harrell, 2001)

In order to extract important variables we used forward selection procedure available in SAS software, with standard overall fit measures in both methods.

REVIEW OF PREVIOUS RESEARCH RESULTS

There is some previous research that investigate influences on default in cardholders revolving credit. Although one of the most important variable usually is debt-to-income ratio, Dunn and Hill (1999) find that the default on household credit card debt is also increased with the increase of the following variables: (i) the ratio of total minimum required payment from all credit cards to household income; (ii) the percentage of total credit line which has been used by consumer; (iii) the number of credit cards on which consumer has reached the borrowing limit. They also find that default risk is inversly related to the age of cardholders, default is less likely for married cardholders but its likelihood increases with number of children. Calem and Mester (1995) find that calholders with higher balances have a higher probability of default. Dey, Mumy (2005) find that the higher creditworhiness of the borrower, the lower is their likelihood to default. Black and Morgan (1998) reveal that the default is increased by higher income, debt payments/income ratio and total debt/income ratio and by lower liquid asset. Delinquency rate is lower with executive and managers then with operators and laborers. Also, the default is lower if a client is more educated, older, married, owns a house and if she/he stays longer with one firm and at the same home address. Curtis (2003) investigates important variables in behaviour scoring. He founds that cash advances, convenience cheques and balance transfer are characteristics that make a significant improvement to the scorecard performance. Hamilton, Khan (2001) find that the most important discriminating variables between those who defaulted and didn't default are behavior characteristics: cash advances, minimum payment due and interest paid in previous period.

There are some papers that emphasize importance of incorporating economic conditions into credit scoring models either by developing different scoring models for different economic conditions or by buildling scoring models that will incorporate some economic indicators, Thomas (2000). Avery, Calem, Canner (2004) say that a failure to consider situational circumstances in credit scoring may infulence accuracy of credit scoring model in quantifying individual credit risk. They suggest situational information about the economic and personal circumstances of individuals should be included into the scoring models. Their results show that unemployment rate is positively associated with estimated likelihood of default, that there is higher probabilty of default to individuals who have resided in areas that are recovering from a local economic downturn. Marriage status also infulence probability of default – for example long term married individuals have lower probabilty of default compared to never married individuals. Also, likelihood of default is higher in lower income tracts compared with higher income tracts.

Another group of papers deals with the methodology used in credit scoring modeling. One of the most frequently used method is logistic regression (Hand, Henley, 1997). In behavior scoring modeling survival analysis is also used very often. Narain (1992) was one of the first who used survival analysis in credit scoring. Banasik, Crook, Thomas (1999) compare logistic regression to survival analysis in analysing personal loan data set. They suggest that the proportional hazard models are competetive with the logistic regression approach in identifying clients who default in the first year and may be superior to that approach for looking at who will pay-off early in the first year. Stepanova, Thomas (2001) use Cox proportional hazard model and logistic regression in personal behavior scoring modeling. It has been shown that Cox proportional hazard model improves over time compared to logistic regression scores. After about two years Cox outperform logistic regression scores. They also use survival analysis to estimate survival probability over time which enables them to estimate the profit for the loan. Andreeva, Ansell, Crook (2003) analyze revolving credit in retail cards where they used survival analysis (proportional hazard models) in order to find out when the second purchase will happen. Andreeva, Ansell, Crook (2004) applied logistic regression in revolving credit and then Andreeva (2004) compared logistic regession and survival analysis in revolving credit for retail card on the same data set from three European countries. It has been shown that survival analysis is competitive with the logistic regression and there is a little difference in classification accuracy between the parametric models, non-parametric Cox PH model and logistic regression (error rate on hold-out sample for all models is arround 16%). Baesens, Gastel, Stepanova, Vanthienen (2003) used statistical and neural network survival analysis for predicting early payment and loan default in analyzing personal loan data. In predicting default in the first 12 months, logistic regression and Cox model gave the same results, neural networks were better but not significantly. In predicting loan default between 12 and 24 months logistic regression has hit rate 78.24%, Cox model 77.50% and neural networks 78.58%. Neural networks is significantly better then Cox model. Hand (2005) describes methods that are still to be investigated in retail banking: generalized additive models, multivariate adaptive regression splines, survival analysis, causal models and belief networks, longitudinal data analysis, nonparametric methods, indirect classification methods and improved criteria for assessing performance of models.

VARIABLES AND DATA

Data for this research was collected randomly in one bank in Croatia for the period of 12 months – in year 2004. An observation point is settled in the middle of the period, on 30 June. A period preceding this point is the performance period and the characteristics of the performance in this period are used in developing scoring models. On the basis of client's performance in the period of 6 months after the observation point, client is defined as good or bad (Thomas, Ho, Scherer 2001). A client is "bad" if she/he exceeds a contracted overdraft for more then 35 days during the period of 6 months. Otherwise, a client is considered to be "good". This definition is used for developing a scoring model using logistic regression. In order to get output variable for survival analysis, we calculate a time before a client went to default. If the account was closed or survived beyond the observation period, it was considered to be censored.

All data used in the research come from the Bank since there is no credit bureau in Croatia. The whole data sample consisted of 50000 but it was reduced to 44089 after data cleansing. The data set is divided to a development sample and a validation sample. The development sample consisted of 34879 accounts: 1134 (3.25%) bad and 33745 (96.75%) good accounts. The validation sample consisted of 9210 accounts: 278 (3.02%) bad and 8932 (96.98%) good accounts. Structure of the sample data is equal to the structure of the whole data set in the Bank.

Input variables deal with personal open-end accounts. They are divided into three main groups: (i) demographic data; (ii) socio-economic data; (iii) behavior data –repayment and usage (average values for the period of 6 months). List of variables together with their explanation and descriptive statistics for the development data sample, separate for good and bad clients, are given in table 1.

Table 1. Input variables and their statistical distribution

Variable code	Variable description	Frequencies, Means, Standard deviations
Group 1	**Demografic data**	
GEN	Client's gender	Female G: 50,96% B: 46,47% Male: G: 49,04% B: 53,53%
AGE	Client's age	G: Mean= 51,13 (σ=16,56); B: Mean= 41,87 (σ=13,17)
Group 2	**Socio-economic data**	

WS	Working status	Retired G: 27,45% B: 13,93% Employed G: 21,51% B: 22,49% Unemployed G: 3,22% B: 10,41% No answer G: 47,82% B: 53,17%
CITY	Does the client live in a town	Town G: 65,78% B: 58,11% Village G: 42,22% B: 41,89%
JOB+	Does the client have a second job	Second job G: 33,19% B: 36,07% No second job G: 66,81% B: 63,93%
CARD	Does the client have credit cards	Credit card G: 1,50% B: 1,41% No credit card G: 98,5% B: 98,59%
ACAGE	Age of the client's account	G: Mean= 9,98 (σ=5,13); B: Mean= 7,69 (σ=4,77)
SAVE	Does the client have savings acc.	No savings G: 65,23% B: 71,78% ; Savings in kunas G: 9,62% B: 14,28% ; Savings in foreign currency G: 21,2% B: 11,46% ; Savings in kunas and foreign currency G: 3,95% B: 2,47%
Group 3	**Behavior data** (average values)	
LIM	Contracted overdraft	G: Mean=2564,25 kunas (σ=1950,53); B: Mean= 1513,66 kunas (σ=1653,20)
CASH	Cash payments to the account	G: Mean=154,97 kunas (σ=4623,10); B: Mean= 168,97 kunas (σ=572,36)
PON	Payments to the account from other accounts	G: Mean=24,08 kunas (σ=400,06); B: Mean= 9,17 kunas (σ=135,24)
PNR	Irregular payments to the account	G: Mean=593,08 kunas (σ=3325,64); B: Mean= 644,16 kunas (σ=2499,76)
SLR	Regular payments (salaries)	G: Mean= 2414,05 kunas (σ=1744,27); B: Mean= 1417,75 kunas (σ=1555,65)
INT+	Interests paid to the client	G: Mean= 1,49 kunas (σ=6,97); B: Mean= 0,20 kunas (σ=1,49)
STOR	Withdrawals with standing order	G: Mean= 17,72 kunas (σ=90,79); B: Mean= 13,15 kunas (σ=83,8)
CHE	Withdrawals with cheques	G: Mean= 133,89 kunas (σ=299,37); B: Mean= 224,24 kunas (σ= 376,93)
CRE	Withdrawals with credit and debit cards	G: Mean= 84,68 kunas (σ=279,09); B: Mean= 87,81 kunas (σ=343,75)
FCAS	Cash withdrawals from the account	G: Mean=2889,66 kunas (σ=6775,36); B: Mean= 2023,68 kunas (σ=3216,79)
INT-	Interests paid from the client	G: Mean= 9,77 kunas (σ=14,29); B: Mean= 19,29 kunas (σ=19,22)
TIN	Total payments to the account	G: Mean= 3187,49 kunas (σ=6017,13); B: Mean= 2240,29 kunas (σ=3109,22)
INLI	Total payments to the account/ contracted overdraft	G: Mean= 1,13 (σ=1,83); B: Mean= 0,91 (σ=1,48)
TOUT	Total withdrawals from the account	G: Mean= 3161,82 kunas (σ=6853,05); B: Mean= 2399,85 kunas (σ=3399,53)
OUTL	Total withdrawals from the account/ contracted overdraft	G: Mean= 1,13 (σ=1,9); B: Mean= 1,05 (σ=1,5)
INOUT	Total payments to the account minus total withdrawals from the account	G: Mean= 25,66 kunas (σ=1723,04); B: Mean= - 159,57 kunas (σ=1374,41)
BAL	Balance	G: Mean= 1021,47 kunas (σ=8343,46); B: Mean= -1105,26 kunas (σ=2507,22)
LBAL	Balance plus contracted overdraft	G: Mean= 3598,80 kunas (σ=8793,43); B: Mean= 514,4 kunas (σ=3036,19)
BALD	Contracted overdraft/balance	G: Mean= 15,56 (σ=858,85); B: Mean=1,4 (σ=52,8)
RAT	Total payments to the account/ total withdrawals from the account	G: Mean= 12,84 (σ=86,07); B: Mean= 6,89 (σ=26,69)
USE	Percentage of usage of contracted overdraft	Did not use contracted overdraft G: 48,92% B: 41,38% Used contracted overdraft G: 46,47% B: 47,88% Overdraft exceeded contracted G: 4,61% B: 10,74%
NEG	How many times client's overdraft exceeded contracted overdraft	G: Mean= 0,33 (σ=0,84); B: Mean= 2,08 (σ=1,75)

KN	Amount of money that exceeded contracted overdraft	G: Mean= - 26,49 kunas (σ=193,7); B: Mean= -371,93 kunas (σ=689,08)
DAYS	Number of days the client's overdraft exceeded contracted over.	G: Mean= 0,75 (σ=2,93); B: Mean= 9,25 (σ=10,97)
CONT	Number of continuous months with overdraft exceeded	G: Mean= 0,27 kunas (σ=0,7); B: Mean= 1,78 kunas (σ=1,64)
LAST	Number of days since the last time when the client exceeded her/his contracted overdraft	G: Mean= 153,89 (σ=58,16); B: Mean= 51,53 (σ=76,3)

1 EUR ~ 7,4 kunas

Our intention is to see whether some of the variables that reflect specific circumstances in Croatia appear in the model or not. If they do, this would mean they are important in consumer behavior and worth including into the modeling. Some of the variables we expect to see in our scoring model are:

- *working status of the client*: we expect that clients without a job are more likely to be delinquent
- *does the client live in a town*: we expect that clients who live in villages have higher probability of default since rural areas are poorer and with higher unemployment rate then in towns
- *a second job*: we expect that clients with a second job would have lower probability of default
- *salary*: clients with higher salaries should have lower probability of default
- *all kinds of different payments to the client's account (CASH, PON, PNR)* should be important in lowering probability of default since salaries are pretty low (average monthly salary is 600 EUR)
- we also expect that behavior during the first six months will effect behavior during the second six months (variables: NEG, KN, DAYS, CONT, LAST)

Since there is no credit bureau in Croatia we weren't able to examine the influence of some other variables found significant in previous research (e.g. especially those that create credit bureau score).

RESULTS OF CREDIT SCORING MODELS

Logistic regression scoring model

The reason of using logistic regression is to find determinants of default in revolving credit in personal open-end accounts. It will also give us a probability that a client will default in the next 6 months. The total number of input variables that entered the model was 34. The final logistic regression scoring model using forward selection ended with the number of 18 variables. The model is given in the table 2.

Table 2: Estimated parameters of the logistic regression scoring model

Variable code	**Variable**		**Estimation**	**p value**
WS	Working status of the client:	Retired:	-0.6806	<.0001
		Employed:	-0.3056	0.0162
		No answer:	-0.3043	0.0106
JOB+	Does the client have a second job:	Yes:	-0.1445	0.0398
LIM	Contracted overdraft		-0.00111	<.0001
PON	Payments to the account from other accounts		-0.00055	0.0212
SLR	Regular payments (salaries)		-0.00022	<.0001
INT+	Interests paid to the client		-0.3530	<.0001
CHE	Withdrawals with cheques		0.000729	<.0001
FCAS	Cash withdrawals from the account		-0.00005	<.0001
INT-	Interests paid from the client		0.0203	<.0001
INOUT	Total payments on the account minus total withdrawals from the account		-0.00038	<.0001
BAL	Balance		-0.00081	<.0001
RAT	Total payments to the account/ total withdrawals from the account		-0.00477	0.0019
KN	Amount of money that exceeded contracted overdraft		-0.00019	0.0284
DAYS	Number of days the client's overdraft exceeded contracted overdraft		0.0804	<.0001
CONT	Number of continues months with overdraft exceeded		0.1250	0.0003
LAST	Number of days since the last time when the client exceeded her/his contracted overdraft		-0.00818	<.0001
ACAGE	Age of the client's account		-0.0410	<.0001
AGE	Client's age		-0.0113	<.0001

The logistic regression model regarding standard overall fit measures for logistic regression [e.g. Score= 8309.4855, (p <.0001), Wald= 2919.9757 (p <.0001)] on the holdout sample showed the total hit rate of 81,16%, good hit rate of 82,4% and bad hit rate of 74,46%. The percentage of the good clients estimated as bad ones is 17,6% and the percentage of bad clients estimated as good clients is 25,54%.

On the basis of the logistic regression model results it can be seen that default is increased: (i) with higher withdrawals with cheques; (ii) with higher number of days client's overdraft exceeds contracted overdraft in the previous period; (iv) with higher number of continues months with overdraft exceeded; (vi) if the client doesn't have a permanent job; (vii) if the client doesn't have a second job. Opposite influence on probability of default, which means lowering it, has been found in the following variables: (i) higher amount of contracted overdraft; (ii) higher salaries; (iii) higher amount of cash paid from the account; (iv) higher amount of payments to the account from other account; (v) higher differences between total payments to the account and total withdrawals from the account; (vi) higher balance; (vii) higher amount of money that exceeded contracted overdraft in the previous period; (viii) higher number of days from the last time the client exceeded contracted overdraft in the previous period; (ix) older age of the client and older age of her/his account.

Cox's regression scoring model

Cox's regression is used in order to find determinants of default in personal open-end accounts including time to default. Also, it will give us likelihood to default in the period of next 6 months. The total number of input variables that entered the model was 34. The final Cox regression scoring model using forward selection ended with the number of 10 variables. The model is given in the table 3.

Table 3: Estimated parameters of the Cox's regression scoring model

	Variable		**Estimation**	**Hazard ratio**	**p value**
GEN	Gender	Female:	0.15769	1.171	0.0097
WS	Working status	Retired:	-0.40249	0.669	<.0001
WS	Working status	Unemployed:	0.53939	1.715	<.0001
JOB+	Second job	No:	0.12424	1.132	0.0525
LIM	Contracted overdraft		-0.0000630	1.000	0.0001
KN	Amount of money that exceeded contracted overdraft		-0.0003090	1.000	<.0001
DAYS	Number of days when the client's overdraft exceeded contracted overdraft		0.05852	1.060	<.0001
CONT	Number of continues months with overdraft exceeded		0.12961	1.138	<.0001
LAST	Number of days since the last time when the client exceeded her/his contracted overdraft		-0.01152	0.989	<.0001
ACAGE	Age of the client's account		-0.05925	0.942	<.0001
AGE	The client's age		-0.00604	0.994	0.0320

The Cox's logistic regression model has the following standard overall fit measures: Score=8777.4851 ($p<.0001$), Wald=3407.2117 ($p<.0001$).

From looking at the estimated coefficients the model indicates that as the amount of contracted overdraft increases, the rate of default decreases. The same applies for amount of money that exceeded contracted overdraft and for number of days since the last time when the client exceeded her/his contracted overdraft. Also, if the client is older as well as his account, it will also decrease probability of default. On the contrary, as the number of days the client's overdraft exceeded contracted overdraft increase, the rate of default will be increased. Female clients are more likely to have an account in default compared to male clients. If the client is without a permanent job it is more likely that the account will be in default compared to employed clients. Also, if the client doesn't have a second job it is more likely he/she will be in default compared to the client with an extra job.

Conclusion and discussion

The paper was aimed to identify important features for the credit scoring model for open-end accounts using logistic regression and survival analysis on a Croatian data set. Since Croatia is a country in transition with war consequences, changes in political, institutional and social systems and, above all, with specific economic conditions characterized by slow economy, a high unemployment rate and a relatively low personal income, it was assumed that this influences credit behavior of customers and, as a consequence, a composition of credit scoring models.

Both credit scoring models developed by logistic regression and Cox's regression identify a group of the same variables being important in consumer behavior with the difference that in logistic regression scoring model some additional variables were extracted. Common variables are: (i) working status; (ii) having a second job; (iii) contracted overdraft; (iv) amount of money that exceeded contracted overdraft; (v) number of days the client's overdraft exceeded contracted overdraft; (vi) number of continues months with overdraft exceeded; (vii) number of days since the last time the client exceeded her/his contracted overdraft; (viii) client's age; (ix) age of a client's account. So, Cox's regression scoring model extracted delinquency history as well as working status, and logistic regression scoring model also extracted usage history and payment history besides the already mentioned variables. The variables common to both models show the same conclusions – it is important that the client has a permanent job and a second job and that her/his delinquent history is clean. It is a little bit surprising that with the increase of amount of money that exceeded contracted overdraft in the first 6 months, the probability of default is decreased. Maybe it is because such clients have been warned about their behavior and then they changed their behavior.

Some of our assumptions about how specific economic transitional conditions influence the variable selection have been confirmed and some haven't. Although we have assumed that living place would affect behaviour, it hasn't been found relevant in both of our models. Also, we have assumed that all different payments to the account, besides salaries, would be important for consumer behaviour but it hasn't been completely confirmed. Payments to the account from other accounts are found significant but cash payments to the account and irregular payments aren't. Looking at data, we realized that payments to the account usually come from other client's account rather then as cash or irregular payments. Other assumptions have been confirmed. Comparing our research to other previous research in studying consumer behaviour in open-end accounts it could be noticed that common issues are delinquency history as well as payment and usage history and the different is importance of client having a permanent job and a second job that we included in our research because of high unemployment rate in Croatia.

The results provide some new information about credit scoring modeling in a transitional country with war consequences such as Croatia. By including more datasets from different transitional countries, it could be possible to provide more generalized results.

As guidelines for further research we therefore suggest to apply the methodology on more datasets from transitional countries, especially by adding credit bureau data that were not available in our country, but were found relevant for credit scoring by other authors.

References:

1. Andreeva, G., Ansell, J., Crook, J.N., Credit scoring in the context of the European integration: Assesing the performance of the generic models, Working paper, Credit Research Centre, University of Edinburgh, 2004
2. Andreeva, G., Ansell, J., Crook, J.N., Analysing revolving credit, Working paper, Credit Research Centre, University of Edinburgh, 2004
3. Andreeva, G., European generic scoring models using survival analysis, Working paper, Credit Research Centre, University of Edinburgh, 2004
4. Avery, R.B., Calem, P.S., Canner, G.B., Consumer credit scoring: Do situational circumstances matter?, Journal of Banking and Finance 28; 2004, p. 835-856
5. Baesens, B., Gastel, T., Stepanova, M., Vanthienen, J., Neural network survival anaysis for personal loan data, Credit Scoring and Credit Control 2003, Edinburgh, 2003.
6. Black, S.E., Morgan, D.P., Risk and the democratization of credit cards, Federal Reserve Bank, Working Papers, 1998
7. Caouette,J.B., Altman,E.I., Narayanan,P., Managing credit risk, John Wiley & Sons, New York, 1998.
8. Curtis, C. (2003) . Use transaction data within behavioural scoring, Credit Scoring and Control VIII Edinburgh, 2003
9. Desai, V.S., Crook, J.N., Overstreet, G.A., A comparison of neural network and linear scoring models in Credit Union environment, European Journal od Operational Research 95 (1996), p. 24-35.
10. Dey, S.; Mumy, G., Determinants of borrowing limits on credit cards, Bank of Canada Working Paper 2005-7, 2005.
11. Dunn, L.F., Kim, T., An empirical investigation of credit card default, The Ohio State University, Working papers, 1999
12. Frank, E., Harrell, Jr., Regression Modeling Strategies, Springer-Verlag, New York, 2001.
13. Hamilton, R., Khan, M., Revolving credit card holders: who are they and how can they be identified?, The Service Industries Journal, Vol.21, No.3, 2001, p.37-48

14. Hand, D., New uses of statistics in retail banking, The final version of this paper appeared in Journal of Mathematical and Management Sciences, 2005
15. Hand, D.J., Henley, W.E., Statistical classification methods in consumer Credit scoring: A review, Journal Royal Statistical Society A, 160, 1997, p. 523-541.
16. Harrel FE Jr. 2001. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. Springer: Berlin.
17. Kasapovic, M., Democratic transition and political institutions in Croatia, Politicka misao, **33**: No. 2-3, 1996, p. 84-99
18. Koch, T.W., MacDonald, S.S., Bank management, Harcourt College Publishers, Fortworth, Tokyo, 2000.
19. Mays, E., editor, Handbook of credit scoring, Glenlake Publishing Company, Ltd., Chicago, 2001.
20. McNab, H., Wynn, A., Principles and practice of consumer credit risk management, CIB Publishing, Canterbury, 2000.
21. Mervar, A.. Ekonomski rast i zemlje u tranziciji. *Privredna kretanja i ekonomska politika.* **12**:, 2002, p. 53-87.
22. Narain, B., Survival analysis and the credit granting decision, Credit Scoring and Credit Control (L.C.Thomas, J.N. Crook & D.B. Edelman, eds.). Oxford: Oxford University Press, 1999, p. 109-122
23. Thomas, C.L., A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, International Journal of Forecasting 16, 2000, p. 149-172
24. Thomas, L.C., Ho, J., Scherer, W.T., Time will tell: behavioural scoring and the dynamics of consumer credit assessment, IMA Journal of Management Mathematics 12, 2001, pp. 89-103
25. Till, R.J., Hand, D.J., Behavioural models of credit card usage, Journal of Applied Statistics, 2002.