Credit Risk Model: Assessing Default Probability of Mortgage Loan Borrower

Aleksandre Ergeshidze[1]

Abstract

Over the past decade, as a result of rapid growth of the loan portfolio and the financial crisis, importance of credit risk analysis has increased worldwide. After the global financial crisis, more attention has been paid to loan granting process by various researchers and financial market participants. New regulations forced commercial banks to improve credit risk management and existing statistical models. This paper, based on data obtained from three major banks of Georgia, develops logit model to examine mortgage loan borrowers' characteristics that determine their default probability. Similar data is rarely available for developing countries, therefore findings of this study can be useful for those countries as well. According to the research, main characteristics that determine borrowers' creditworthiness are payment to income ratio, loan to value ratio, credit history and borrower's type (whether borrower receives income in that bank). Average prediction accuracy of the model within the sample equals to 93.4%. Findings of this study will enable commercial banks in Georgia to improve their credit risk assessment and make decisions on loan approval cost efficiently. In addition, acquired results can be used by the National Bank of Georgia to estimate the adequacy of loan loss provisions, to assess commercial banks credit portfolio used as collateral for monetary operations and to enhance collateral base to support de-dollarization policy.

Keywords: Credit risk; Logit model, Default Probability, PTI, LTV.

JEL Codes: C14, G21, G33, E58.

1. Introduction

Credit risk is one of the main risks for the banking sector and to the financial stability, which arises from the inability of the borrower to service their financial obligations. Over the past decade, statistical methods examining borrowers' default probability have improved significantly, but they still suffer from low prediction accuracy. As the global financial crisis showed, wrong assessment of mortgage loans' credit risk can lead to severe distress and cause the financial crisis. To reduce such risks, it is crucial for commercial banks to assess borrowers' creditworthiness, analyze tradeoff between default risk and expected return from loan approval and determine profit-maximizing amount of loan to approve. Therefore, it is important to improve credit risk models and their forecasting power.

Moreover, banking sector's credit risk analysis and management are essential for financial sector supervisor, which periodically checks banks' loan portfolios and determines the adequacy of loan loss reserves. In addition, commercial banks use granted domestic currency mortgage loans as a collateral for monetary operations. Commercial banks are able to borrow money from the central bank, but they are required to put appropriate collateral. If the commercial bank does not repay the loan, the national bank can initiate execution of pledged

[1] *PhD Student, Ivane Javakhishvili Tbilisi State University (TSU), Faculty of Economics and Business. E-mail: aergeshidze@gmail.com*

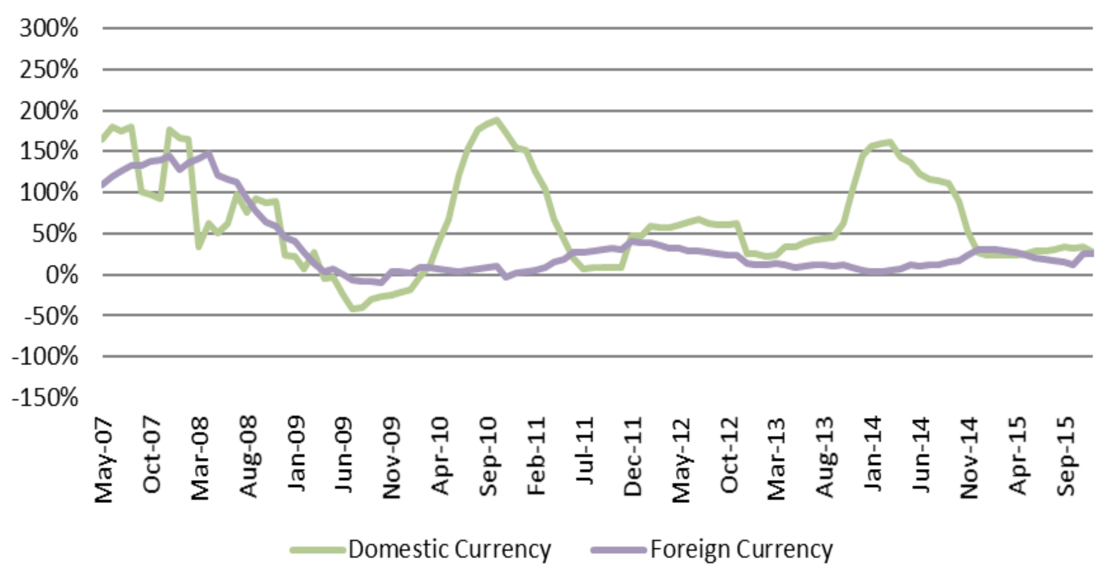
ISBN: 978-605-307-583-7

collateral. Therefore, analysis of loans used as collateral is important for the central bank to minimize possible risks, manage collateral requirements' accordingly and support de-dollarization strategy, which is viewed as the main challenge for monetary policy transmission mechanism.

The research aims to develop logit model to identify important individual characteristics at the time of loan approval that determine mortgage loan borrowers' default probability and evaluate its impact. This will help commercial banks to improve credit risk assessment methodology and reflect respective risks in the interest rates. Moreover, acquired results can be used by the National Bank of Georgia to evaluate the adequacy of loan loss reserves and manage collateral requirements for monetary operations.

The study focuses on mortgage loans because they occupy a significant portion (40%) of individual loans in Georgia, they have grown rapidly over the past decade (Figure 1) and borrowers' characteristics for these types of loans are more affordable and reliable. At the same time, In addition, mortgage loans are used as a collateral base for monetary operations.

Figure 1. Growth rate of mortgage loans



Source: National Bank of Georgia

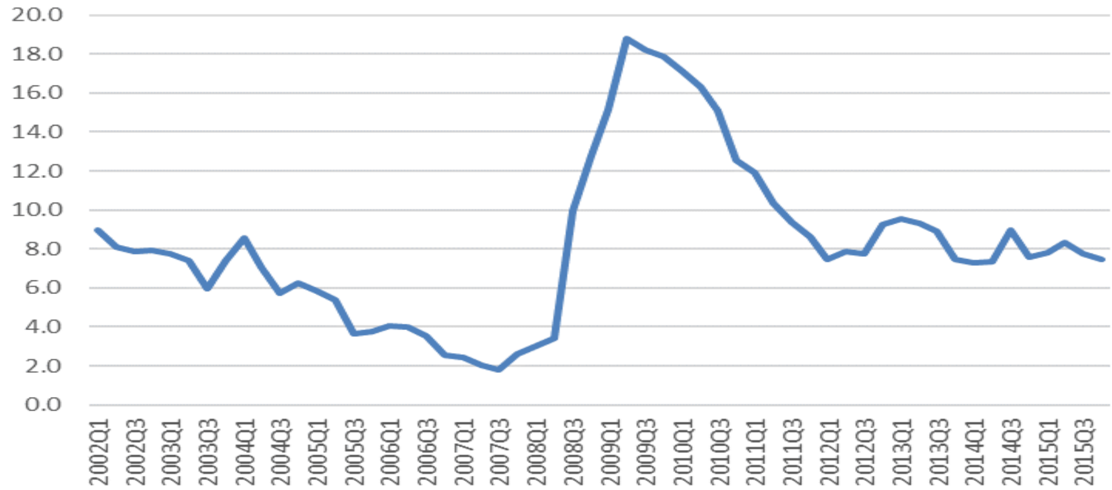
Currently, in Georgia mortgage loans are mostly granted based on expert judgment on client's default risk. To reduce costs of expert valuation, increase efficiency, and reduce the time required for loan approval (gain a competitive advantage), statistical models with high prediction accuracy are required. Therefore, it is crucial for commercial banks to have a comprehensive database, which will enable them to assess borrowers' risk based on their characteristics and periodically update the econometric model to take into account changes in key factors determining client's default probability. Improved credit risk management is profitable for banks (lower risk premium and improved financial results), for borrowers (adequate level of indebtedness and less financial difficulties in the future) and for the regulators (low probability of financial distress and financial crisis).

Credit risk is usually assessed based on dynamics of non-performing loans, but this indicator is backward-looking and does not depict future risks. In Georgia, the share of non-performing loans to total loans increased significantly in 2008 as a result of the twin shock (financial crisis and the August war). During that period, the share of non-performing loans to total loans increased by 16 percentage points and amounted to 18.8%. However, afterward it had a downward trend. The share of non-performing loans in mortgage loans had similar dynamics, but in 2015, as a result of domestic currency depreciation against US dollar, the share of non-performing loans increased by 1 percentage point to 3.2%. In order to maintain the low level of non-performing loans, it is essential to correctly estimate borrowers default probability while approving the loan.

Figure 2. Share of non-performing loans to total loans

DOI: 10.19275/RSEPCONFERENCES055

ISBN: 978-605-307-583-7



Source: National Bank of Georgia

Models predictability, developed in this study, can be further improved by enhancing database and by taking into account other economic, demographic and behavioral variables. In addition, similar model can be developed for other retails and corporate loans.

2. Literature Review

A rapid growth of the loan portfolio, increased competition and improved technologies led to the development of statistical models and improvement in their forecasting power. Over the past decade, various researchers have devoted their studies to statistical models concerning credit risk and their development. Particular attention was paid towards a relationship between borrowers' characteristics and their default risk. Statistical methods used to analyze credit risk are mainly based on Logit regression, discriminant analysis, decision tree analysis and linear regressions. In addition, it is quite common to transform coefficients into"scores" and to develop scoring models.

The study by Durand (1941) was one of the first studies, which assessed borrowers credit risk based on data collected from 37 firms. Afterward, statistical methods have improved significantly and currently, major banks use similar models to quantify credit risk. Based on research of Tsai et al. (2009), available literature in this field can be divided into two parts: the first includes studies about variables that determine borrowers' solvency (Dinh, Thanh, and Kleimeie, 2007; Avery, Calem, and Canner, 2004; Thomas, 2000; Desai, Crook, and Overstreet, 1996; Steenackers and Goovaerts, 1989). It should be noted that social, demographic and economic factors affecting borrowers creditworthiness differ by studies. The second part focuses on developing optimal credit risk model (Bellotti, and Crook, 2009; Crook, Edelman, and Thomas, 2007; Lee, Chiu, Chou, and Lu, 2006; Baesens, Gestel, Stepanova, Poel, and Vanthienen, 2005; Ong, Huang, and Tzeng, 2005; Rohb, and Hana, 2005; Lee and Chen, 2005; Jones, and Hensher, 2004; Chen and Huang, 2003; Lee, Chiu, Lu, and Chen, 2002; Malhotra and Malhotra, 2002 ; Noh, West, 2000). As these studies illustrate the relevance of the model and its forecastingpower depends on the country characteristics. Therefore, there is no unique model, which will be optimal worldwide.

Avery, Calem, and Canner (2004) suggest accounting for country's economic situation in the credit model. Because it is expected that the credit model will overestimate default probability in regions, where the economy is getting stronger, while in regions, where the economic conditions are getting worse, default probability will be underestimated.

Some researchers emphasize the importance of behavioral variables in evaluating the creditworthiness of the borrower (Lim, Teo, and Loo, 2003; Roberts and Jones, 2001; Roberts and Sepulveda, 1999; Hayhoe, Leach, and Turner, 1999; Lim and Teo, 1997). According to Roberts and Sepulveda (1999) and Tsai et al. (2009) forecasting

DOI: 10.19275/RSEPCONFERENCES055

ISBN: 978-605-307-583-7

power of credit model will improve by accounting borrower's spending behavior and attitude toward money. Hayhoe et al. (1999) had similar conclusions and demonstrated that the students' attitude toward money affected their debt repayment.

It should be noted that credit risk models suffer from several drawbacks that needs to be taken into account while assessing borrower's creditworthiness. First, the models need a constant update, because over time the population evolves and characteristics determining bankruptcy risk might change (Hand, D. J., & Henley, W. E. 1997). Characteristics change more rapidly in developing countries, where the financial education and income are low. Another shortcoming is that statistical models are based on the incomplete database. The data includes only the borrowers, who took loans and does not take into account the characteristics of people, who were denied to take loans.

3. Data and Methodology

Finding reliable data for credit risk assessment is one of the main challenges for researchers, especially in developing countries. To collect data for this study, three major banks of Georgia were requested to send mortgage loans database that was approved during the second half of 2011. This time was chosen because before that complete database of borrowers' characteristics were not available. From mortgage loans database 600 borrowers were chosen randomly and respective banks were asked to send complete information about them. Finally, after obtaining and filtering the data, 503 borrowers were left with complete information. Out of which in 2015 7.8% had a non-performing loan or were already defaulted. For the purpose of this study, these borrowers were classified as defaulted. If during this period borrower took an additional loan, then borrower's characteristics at the time of last loan approval are reported in the model, because new loan influences borrower's default probability.

In the database borrowers characteristics include loan volume in national and foreign currency, number of loans, payment to income ratio (PTI), loan to value ratio (LTV), value of collateral, credit history (positive, was negative, negative), employment sector (public, private, self-employed and unemployed), income in national and foreign currency, age, sex, marital status, borrowers type (receives salary in that bank or somewhere else), date of opening of first account according to the respective bank and the credit bureau. The last factor could be used as a proxy for financial education.

To examine factors determining borrowers' probability of default logit model is used. Because the dependent variable is binary (bankruptcy, not bankruptcy) one of the most widely used and appropriate statistical method for this type of analysis is logit regression. Some empirical studies (Wiginton 1980) have shown that logit regression predicts borrowers default better than discriminant analysis and linear regressions. The advantage of logit models is its ability to analyze the impact of continuous independent variables on the binary dependent variable. Accordingly, the results of the logit regression represent the probability of an event (in this case probability of default) to occur.

Logit regression, instead of minimizing the sum of squared residuals, estimates parameters based on the maximum likelihood of expected outcomes. In order to estimate logit model, a binary dependent variable is transformed into a continuous variable by taking a logarithm of odds ratio (y*). The odds of the event are measured based on predictors' characteristics. Afterward, logit transformation function enables to use linear regression.

$$y^* = \frac{p}{1-p}$$

$$\ln(y^*) = X_i{}'\beta + u_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_1 x_{ki} + u_i$$

DOI: 10.19275/RSEPCONFERENCES055

Where p is the probability of default and X -borrowers characteristics. To evaluate the probability of default simple mathematical operations are required to transform logit transformation function:

$$P(y_i = 1) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_1 x_{ki}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_1 x_{ki}}}$$

Obtained result from this transformation will be between 0 and 1. Using this method one can determine borrowers' probability of default based on their characteristics and influence of changes in characteristics on the probability.

4. Results

Based on econometric analysis four significant variables affecting credit risk are revealed: payment to income ratio (PTI), loan to value ratio (LTV), credit history and the type of the client (receives a salary in that bank or somewhere else). PTI, LTV and credit history are significant variables at 1% confidence intervals (p-value <0.01), while the type of the borrower is significant at 5% confidence interval (p-value <0.05). Obtained coefficients are not significantly altered in case of adding other variables that underlines robustness of acquired results. Model's explanatory power (R-square) equals to 26.5%. To improve R-square further studies are needed to expand the database and account for other demographic, economic and behavioral variables.

Table 1. Regression results

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-8.175176	0.916555	-8.919455	0.0000
PTI	5.606510	1.045242	5.363841	0.0000
LTV	2.140303	0.686947	3.115673	0.0018
CREDIT_INFO	1.078996	0.279755	3.856927	0.0001
INCOME_INBANK	1.042027	0.448515	2.323283	0.0202
McFadden R-squared	0.264864	Mean dependent var		0.077535
S.D. dependent var	0.267704	S.E. of regression		0.239780
Akaike info criterion	0.420834	Sum squared resid		28.63228
Schwarz criterion	0.462788	Log likelihood		-100.8397
Hannan-Quinn criter.	0.437292	Deviance		201.6794
Restr. deviance	274.3430	Restr. log likelihood		-137.1715
LR statistic	72.66360	Avg. log likelihood		-0.200477
Prob(LR statistic)	0.000000			
Obs with Dep=0	464	Total obs		503
Obs with Dep=1	39			

In order to interpret acquired coefficients and determine borrowers' default probability logit transformation function has been used:

$$P(y_i = 1) = \frac{\exp(-8.2 + 5.6 * PTI_{1i} + 2.14 * LTV_{1i} + 1.08 * Credit\_History_{1i} + 1.04 * Income\_inbank_{1i})}{1 + \exp(-8.2 + 5.6 * PTI_{1i} + 2.14 * LTV_{1i} + 1.08 * Credit\_History_{1i} + 1.04 * Income\_inbank_{1i})}$$

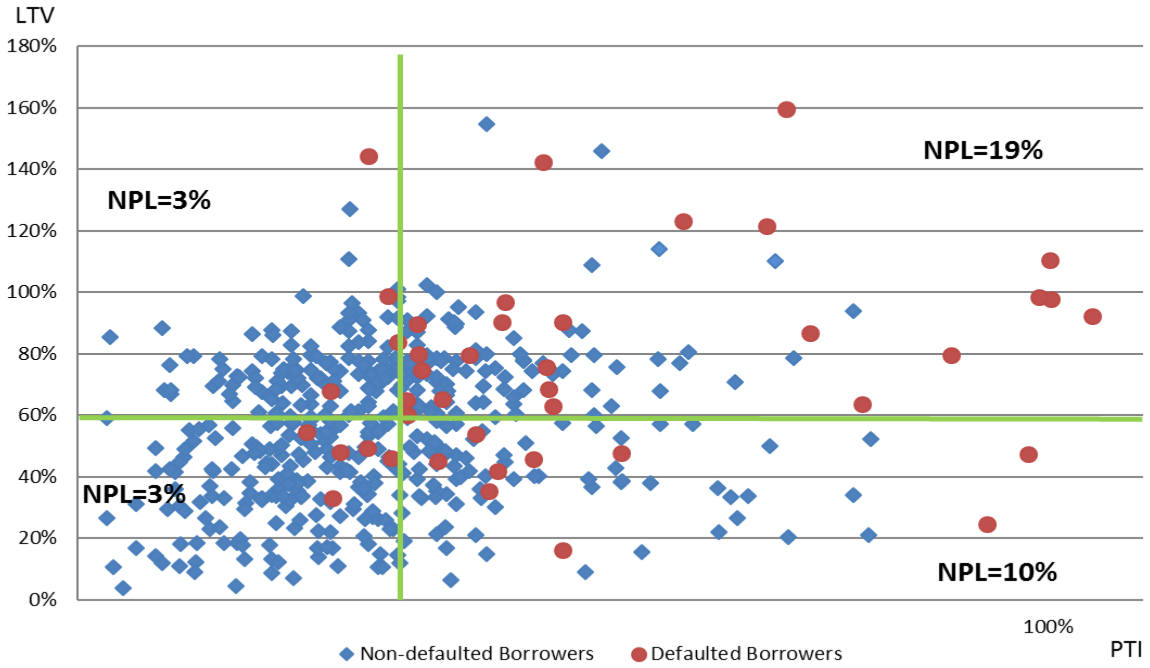
Afterward, based on these coefficients, default probabilities can be calculated. For example, for an average borrower with following characteristics: PTI = 0.3, LTV = 0.6, positive credit history and salary received in a different bank, default probability equals to 4.4%. In this case, increase in PTI ratio by 10-percentage points increases default probability by 3.1 percentage points to 7.5%. If LTV ratio increases by10-percentage points

ISBN: 978-605-307-583-7

default probability rises by 1-percentage point to 5.4%. Deterioration of credit history increases default risk to 11.9%, and if the client would receive a salary in that bank,the probability to default would decline by 2.8 percentage points to 1.6%.

It should be noted, that PTI and LTV are intuitively very important variables and Basel committee and International Monetary Fund (IMF) suggest using them more actively as themacro-prudential tools. Currently, the National Bank of Georgia uses these indicators only for loans used as a collateral for monetary operation at NBG. Figure 3 shows that when PTI and LTV exceed their average ratios than the share of defaulted borrowers' increase from 3% to 19%. According to the same figure, PTI is more important variable in assessing borrowers' creditworthiness than LTV.

Figure 3. Distribution of borrowers based on PTI and LTV



The impact of other two significant variables (credit history and borrowers type) on credit risk is also intuitive. If the borrower has previously violated the terms of the contract and therefore has a negative credit history, probability that he/she will have solvency problems in future will be higher. The type of the client (receives a salary in that bank or somewhere else) shows how closely is a borrower linked to the commercial bank and how familiar is the bank with the client's behavior. When a bank is closely linked to the borrower and knows his/her spending behavior, the probability of default for approved loan will be lower. The study showed that all other variables are insignificant in determining credit risk. Accordingly, age, sex, employment sector, marital status, income and loan amount and their currency do not have a significant impact on the borrowers' default probability.

If we take default threshold of 10% (consider all borrowers defaulted, whose model estimated default probability is more than 10%), then the model correctly predicts 86% of cases for borrowers who did not default and 64% of borrowers who defaulted on their loans in given sample. Average correct forecast rate equals to 84.7%. If we lower threshold, so that all non-defaulted borrowers will be correctly identified, then correct forecast rate in defaulted people drops to 15%. Average correct forecast rate amounts to 93.4%

DOI: 10.19275/RSEPCONFERENCES055

ISBN: 978-605-307-583-7

Table 2. Forecast evaluation

	Estimated Equation		
	Dep=0	Dep=1	Total
P(Dep=1)<=...	401	14	415
P(Dep=1)>C	63	25	88
Total	464	39	503
Correct	401	25	426
% Correct	86.42	64.10	84.69
% Incorrect	13.58	35.90	15.31

	Estimated Equation		
	Dep=0	Dep=1	Total
P(Dep=1)<=...	464	33	497
P(Dep=1)>C	0	6	6
Total	464	39	503
Correct	464	6	470
% Correct	100.00	15.38	93.44
% Incorrect	0.00	84.62	6.56

Obtained results can be used by the National Bank of Georgia to enhance collateral base in monetary operations without increasing its risk. Currently, for loans used as collateral upper bound requirements are PTI equal to 40% and LTV equal to 75%. Based on acquired results loans with upper bound of PTI=50% and LTV=50%, or PTI=30% and LTV=90% have the same credit risk. Therefore, National Bank can enhance collateral base in monetary operations without increasing its risk, which will help de-dollarization process and improve the effectiveness of monetary policy transmission mechanism.

Also, the collateral base can be expanded by covering following types of loans:

PTI	LTV	Credit History	Borrower receives salary in that bank
30%	90%	Positive	No
50%	50%	Positive	No
50%	90%	Positive	Yes
30%	50%	Was negative	No
40%	70%	Was negative	Yes
15%	40%	Negative	No
25%	60%	Negative	Yes

5. Conclusion

By developing logit model this study reveals borrowers characteristics that mainly determine their default probability and estimates impact of those characteristics on borrowers' credit risk. Research finds out that main characteristics determining borrowers' creditworthiness are: payment to income ratio (PTI), loan to value ratio (LTV), credit history and customer type (receives a salary in that bank or somewhere else). The average rate of models' correct prediction within the sample amounted to 93.4%.

This model can be used as a complementing tool to existing expert judgment for assessing credit risk. If commercial banks will improve this model by expanding the database and adding other behavioral, demographic and economic variables, they will be able to rely more heavily on the model outcome, lower costs of decision-making and time of credit approval.

DOI: 10.19275/RSEPCONFERENCES055

In addition, acquired results can be used by the National Bank of Georgia to assess commercial banks credit portfolio and adequacy of loan loss reserves. Moreover, National Bank can enhance collateral base in monetary operations without increasing its risk, which will help to de-dollarization process and improve the effectiveness of monetary policy transmission mechanism.

Further studies are needed to account for demographic, economic and behavioral variables. In addition, the models' predictability can be further improved by expanding the database. In addition, a similar model can be developed for other retail and corporate loans. Therefore, commercial banks should pay more attention to improve the database, which will be beneficial for them to improve profitability, for borrowers to have a manageable level of debt and for the regulator to lower probability of the crisis.

7. References

Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." The journal of finance 23, no. 4 (1968): 589-609.

Altman, Edward I. "Predicting financial distress of companies: revisiting the Z-score and ZETA models." Stern School of Business, New York University(2000): 9-12.

Avery, Robert B., Paul S. Calem, and Glenn B. Canner. "Consumer credit scoring: do situational circumstances matter?." Journal of Banking & Finance 28, no. 4 (2004): 835-856.

Baesens, Bart, Tony Van Gestel, Maria Stepanova, Dirk Van den Poel, and Jan Vanthienen. "Neural network survival analysis for personal loan data."Journal of the Operational Research Society (2005): 1089-1098.

Bellotti, Tony, and Jonathan Crook. "Credit scoring with macroeconomic variables using survival analysis." Journal of the Operational Research Society 60, no. 12 (2009): 1699-1707.

Chen, Mu-Chen, and Shih-Hsien Huang. "Credit scoring and rejected instances reassigning through evolutionary computation techniques." Expert Systems with Applications 24, no. 4 (2003): 433-441.

Crook, Jonathan N., David B. Edelman, and Lyn C. Thomas. "Recent developments in consumer credit risk assessment." European Journal of Operational Research 183, no. 3 (2007): 1447-1465.

Desai, Vijay S., Daniel G. Conway, Jonathan N. Crook, and GEORGE A. OVERSTREET. "Credit-scoring models in the credit-union environment using neural networks and genetic algorithms." IMA Journal of Management Mathematics 8, no. 4 (1997): 323-346.

Dinh, Thi, Huyen Thanh, and Stefanie Kleimeier. "A credit scoring model for Vietnam's retail banking market." International Review of Financial Analysis16, no. 5 (2007): 471-495.

Durand, David. "Risk elements in consumer instalment financing." NBER Books (1941).

Hand, David J., and William E. Henley. "Statistical classification methods in consumer credit scoring: a review." Journal of the Royal Statistical Society: Series A (Statistics in Society) 160, no. 3 (1997): 523-541.

Hayhoe, Celia Ray, Lauren Leach, and Pamela R. Turner. "Discriminating the number of credit cards held by college students using credit and money attitudes." Journal of Economic Psychology 20, no. 6 (1999): 643-656.

Jones, Stewart, and David A. Hensher. "Predicting firm financial distress: A mixed logit model." The Accounting Review 79, no. 4 (2004): 1011-1038.

Lee, Tian-Shyug, and I-Fei Chen. "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines."Expert Systems with Applications 28, no. 4 (2005): 743-752.

Lee, Tian-Shyug, Chih-Chou Chiu, Chi-Jie Lu, and I-Fei Chen. "Credit scoring using the hybrid neural discriminant technique." Expert Systems with applications 23, no. 3 (2002): 245-254.

ISBN: 978-605-307-583-7

Lee, Tian-Shyug, Chih-Chou Chiu, Yu-Chao Chou, and Chi-Jie Lu. "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines." Computational Statistics & Data Analysis 50, no. 4 (2006): 1113-1130.

Malhotra, Rashmi, and D. K. Malhotra. "Differentiating between good credits and bad credits using neuro-fuzzy systems." European journal of operational research 136, no. 1 (2002): 190-211.

Ong, Chorng-Shyong, Jih-Jeng Huang, and Gwo-Hshiung Tzeng. "Building credit scoring models using genetic programming." Expert Systems with Applications 29, no. 1 (2005): 41-47.

Roberts, James A., and Cesar J. Sepulveda M. "Money attitudes and compulsive buying: an exploratory investigation of the emerging consumer culture in Mexico." Journal of International Consumer Marketing 11, no. 4 (1999): 53-74.

Roberts, James A., and Eli Jones. "Money attitudes, credit card use, and compulsive buying among American college students." Journal of Consumer Affairs 35, no. 2 (2001): 213-240.

Srinivasan, Venkat, and Yong H. Kim. "Credit granting: A comparative analysis of classification procedures." The Journal of Finance 42, no. 3 (1987): 665-681.

Steenackers, A., and M. J. Goovaerts. "A credit scoring model for personal loans." Insurance: Mathematics and Economics 8, no. 1 (1989): 31-34.

Thomas, Lyn C. "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers." International journal of forecasting 16, no. 2 (2000): 149-172.

Tsai, Ming-Chun, Shu-Ping Lin, Ching-Chan Cheng, and Yen-Ping Lin. "The consumer loan default predicting model–An application of DEA–DA and neural network." Expert Systems with applications 36, no. 9 (2009): 11682-11690.

West, David. "Neural network credit scoring models." Computers & Operations Research 27, no. 11 (2000): 1131-1152.

Westgaard, Sjur, and Nico Van der Wijst. "Default probabilities in a corporate bank portfolio: A logistic model approach." European journal of operational research 135.2 (2001): 338-349.

Wiginton, John C. "A note on the comparison of logit and discriminant models of consumer credit behavior." Journal of Financial and Quantitative Analysis 15, no. 03 (1980): 757-770.

DOI: 10.19275/RSEPCONFERENCES055