PressAcademia

Journal of Economics, Finance and Accounting

Year: 2017 Volume: 4 Issue: 2

CREDIT SCORING BY USING GENERALIZED MODELS: AN IMPLEMENTATION ON TURKEY'S SMEs

DOI: 10.17261/Pressacademia.2017.438
JEFA- V.4-ISS.2-2017(3)-p.98-105

Aysegul Iscanoglu Cekic, Kasirga Yildirak
Trakya University, Faculty of Economics & Administrative Sciences, 22030 Edirne, Turkey. aysegulcekic@trakya.edu.tr
Author 2. Hacettepe University, Faculty of Science, 06800, Beytepe, Ankara, Turkey. kasirga@hacettepe.edu.tr

To cite this document
Cekic, A.I and K. Yildirak. (2017). Credit scoring by using generalized models: an implementation on Turkey's SMEs. Journal of Economics, Finance and Accounting (JEFA), V.4, Iss.2, p.98-105.
Permemant link to this document: http://doi.org/10.17261/Pressacademia.2017.438

ABSTRACT
Purpose - In this study, we make an empirical research and a comparison study on econometric models used with logistic link functions. We compare the predictive powers of models in credit granting process.
Methodology - We collected data belonging to 87 medium sized companies. 21 of these companies are defaulted. The data set includes 15 continuous financial ratios for estimation of the models. We implement three models which are Logistic Regression, Generalized Partially Linear Models(GPLM) and Generalized Additive Models(GAM). For each model the best fitted model is selected according to AIC criteria.
Findings- GPLM have pointed out that the equity turnover ratio has a significant nonparametric effect. On the other hand GAM pointed out that (total liability)/(total assets) and Increase in Sales have significant nonparametric effects. Comparison of the models have implemented according to their accuracy ratios, Type I and Type II errors. Results show that generalized additive model with logistic link outperforms both Logistic Regression and generalized partially linear model in terms of three performance measures.
Conclusion- After 1980s as a result of the financial crises the default events become a main issue of the credit agencies. For this reason, a credit agency' objective is to determine whether a credit application should be granted or refused. Here, the problem is to learn default some time before the default event occurs. The empirical studies in this area have indicated that commonly used classification methods are good to detect signals of defaults. Especially the models which allow logistic link function are good choices for modeling default risk. In this study we mainly focused on the generalized linear models and its semi- and non-parametric extensions with logistic link function. We compare their performances in a credit granting procedure. We use a real data belonging to Turkish SMEs. Our results show that the GAM outperforms the other two models and it will be a good choice for credit granting procedure.

Keywords: Credit scoring, logistic regression, GPLM, GAM.
JEL Codes: C14, C38, C52

1. INTRODUCTION

In a credit granting procedure, the first step is to decide which credit applications should be accepted or rejected. The credit agency takes this decision by following an internal credit scoring procedures which is in fact based on the measured default risk of the applicant. If these procedures work properly then the risk of losing for a credit agency will not be a problem. Therefore, the main problem of a credit agency should be the selection of the best method to detect risky application before accepting the application.

The empirical studies in this area has indicated that commonly used classification methods are good canditates to detect signals of defaults especially the logistic link function is a good choice to estimate default risk. In this study, we make an empirical research on logistic type models. We will mainly focus on statistical methodologies with logistic link function. These are classic logistic regression, generalized partially linear models (semi-parametric models) and generalized additive models. Our aim is two fold. One is to decide which model is performed best and second is to decide which variables are important in explaining the default risk of Turkish SME's. This study is important because credit scoring applications are very limited in Turkey.

The organization of the study is as follows. In Section 2 we give an overview of the literature. Section 3 gives a brief overview of the generalized models with logistic link function used in credit scoring. In Section 4 we follow a comparison study on logistic type models including logistic regression, generalized partial linear models, and generalized additive models with an implementation on credit data of Turkish SMEs. We also give performance results for the methods. Finally, in Section 5, we give a conclusion.

2. LITERATURE REVIEW

The correct classification of defaultable applicant has been an attractive issue for the researchers since the 1930's. The studies in this area have considered various methodologies and variables. The first studies have been based only on simple ratio analysis. In these studies financial ratios of the defaulted companies have been compared that of the non-default companies. This methodology gives only an intuition to the credit agencies. In other words, this methodology does not have predictive power. These studies can be listed as follows: Ramser and Foster(1931), Fitzpatrick (1931), Smith and Winekor (1935), Merwin (1942).

The turning point of credit scoring is the first use of discriminant analysis. This is a turning point because the classification power of the model can be calculated. The discriminant analysis has been firstly applied by Beaver (1966). This study was followed by Altman (1968). In this study the Altman's famous z-score was introduced including 5 variables which are (MV of equity) / (book value of debt), (net sales / total assets), (operating income) / (total assets), (retained earnings) / (total assets), (working capital) / (total assets). The results of this study showed that 94% classification accuracy in detecting defaulted and and 97% classification accuracy in detecting non-default companies and 95% overall classification accuracy (Altman (1968)). The other studies in this area can be listed as follow: Deakin (1972), Awh and Waters (1974), Sinkey (1975), Altman and Lorris (1976), Altman, Haldeman and Narayan (1977), Dambolena and Khory (1980) and, Pantalone and Platt (1987).

The main drawback of discriminant analysis is the strong assumptions on variables. Because of this the prediction performance is not very promising. Moreover, relative performances of the variables cannot be calculated.

In 1970's the regression methodologies applied to credit scoring procedures. The linear regression was firstly applied regression analysis in credit scoring and was implemented by Orgler (1970). Then, Fitzpatrick(1976) applied a multivariate regression analysis. However, the linear regression has an important drawback because it assumes normality and could predict the default probabilities out of the interval [0, 1]. In order to overcome this major drawback Olhson (1980) implemented a logistic regression analysis. He used data belonging to the period of 1970-76 and calculated the type one and type two types of errors in different cut points. Then, Pantalone and Platt (1987) compared the logistic regression with the discriminant analysis. The results showed that logistic regression has 98% classification accuracy in detecting defaulted and 92% accuracy in detecting non-default companies. The implementation of logistic regression in credit scoring is important because it has no normality assumptions on variables, allows predictions and interpretation of coefficients and the estimated default risk is on the interval [0, 1]. The studies in which the logistic regression is implemented are as follows: Tam and Kiang (1992), Laitinen and Kankaanpaa (1999), Shi and Jin (2004), Miyamoto (2014).

The innovations in computer technologies make the use of modern optimization techniques and complex methodologies in credit scoring possible. Therefore, after 1990s numerous studies can be found in this area. These are extended from neural networks to fuzzy based methods including combination of methodologies like fuzzy logistic, neural network logistic, etc. Some can be listed as follows: Odom and Sharda (1990) , Cadden, Coats and Fant (1991), Tam and Kiang (1992), Coats and Fant (1993), Kiviluoto (1993), Laitinen, Sere and Wesel (1996), Laitinen and Kankaanpaa (1999), Muller and Kanz (1999), McKee and Greenstein (2000), Xiao and Fei (2006), Galindo and Tamayo (2006), Yang, Zhu and Cheng (2013), Huo, Chen and Chen (2017).

3. METHODOLOGY

In the study Xi ∈ R represents a financial ratio of the applicant, e.g., X_1 may be the current assets/current liabilities ratio and so on. We identify each applicant by a tuple of p random variables by the vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)$. Moreover, we define the realization of the indicator variables for a particular applicant is $\mathbf{x} = (x_1, x_2, \ldots, x_p) \in \mathbf{X} \in R_n$.

In the problem the response variable takes only two values: 0 and 1. We represent defaulted companies (D) by Y=1 and non-defaulted companies (ND) by Y=0.

3.1. Logistic Regression

Logistic regression is a form of generalized linear model. It is designed to model problems where the dependent variable is binary or dichotomous and the explanatory variables are of any type. The model aims to estimate the expected value of the dependent variable under the known explanatory variables, i.e., $E(Y|\mathbf{x})$ (Hosmer D.W., and Lemeshow (2000))

Let us use the notation $p(x_i) = E(Y|x_i)$ being the probability of default for the i[th] firm. Then the logistic regression is defined as follows

$$p(x_i) = G(x_i, w) = \frac{e^{w_0+w_1x_{i1}+w_2x_{i2}+\cdots++w_px_{ip}}}{1+e^{w_0+w_1x_{i1}+w_2x_{i2}+\cdots++w_px_{ip}}} = \frac{e^{x_i w}}{1+e^{x_i w}}, \tag{1}$$

with the corresponding logit transformation given as follows

$$\ln\left(\frac{p(x_i)}{1-p(x_i)}\right) = w_0 + w_1x_{i1} + w_2x_{i2} + \cdots + +w_px_{ip} + \varepsilon_i = \boldsymbol{x_i}\,\boldsymbol{w} + \varepsilon_i \tag{2}$$

3.2. Generalized Partial Linear Models

Partial linear models are composed of two parts, a linear and non-parametric part. With a known link function G(•), a generalized partial linear model (GPLM) is represented by

$$E(Y\,|U,T) = G(U^T\beta + m(T)), \tag{3}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_k)$, U is n × k matrix including categorical variables, T is n × (p − k) matrix of numerical variables in X and m(•) is a smooth function.

The model with logistic link function is defined as follows:

$$\ln\left(\frac{p(x_i)}{1-p(x_i)}\right) = U^T\beta + m(T), \tag{4}$$

In this study we use the methodology stated in Muller (2000).

3.3. Generalized Additive Models

Generalized additive models (GAMs) are developed by Hastie and Tibshirani (1987). GAM is a more general version of generalized models. Therefore, it allows not only the estimation of parametric models but also the semi and non-parametric models.

The generalized additive models are defined as follows:

$$G(E(Y\,|X)) = \beta_0 + \sum_{i=1}^{p} f_i(\boldsymbol{x_i}), \tag{5}$$

where f_i are unknown smooth functions, G is the link function.

In this paper we apply the generalized additive logistic model. Generalized additive logistic model is defined as follows:

$$\ln\left(\frac{p(x_i)}{1-p(x_i)}\right) = \beta_0 + \sum_{i=1}^{p} f_i(\boldsymbol{x_i}), \tag{6}$$

In estimation we follow the procedure described in Hastie (1991):

4. APPLICATION

In this study, we make an empirical research on logistic type models. We collected data from A Turkish bank, a set of annual reports, news and internet. We are able to gather a sample of 87 medium sized companies' information. 21 of these companies are defaulted at some time in their history. We collect the information belonging to these defaulted companies from their last financial statements before the default event had occurred. Moreover, for the non-defaulted companies we collect information from their available last balance sheets. The data set includes 15 continuous explanatory variables which are listed as follows:

I. Liquidity Ratios:
- X1 : current ratio,
- X2 : liquidity ratio,
- X3 : cash ratio.

II. Activity Ratios:

- X4 : receivables turnover ratio,
- X5 : inventory turnover ratio,
- X6 : equity turnover ratio,
- X7 : equity/(total liability),
- X8 : (total liability)/(total assets).

III. Profitability Ratios:

- X9 : gross merchandise margin,
- X10 : equity profitability ratio,
- X11 : active profitability ratio,
- X12 : net profit margin.

IV. Growth Ratios:

- X13 : increase in sales,
- X14 : equity growth rate,
- X15 : active growth rate.

Table 1 summarizes the basic statistics of the corresponding 15 variables for defaulted and non-defaulted companies, separately. Accordingly, we see that nearly all the ratios show differences between defaulters and non- defaulters. The mean values show that before the default the liquidity, activity and profitability of the companies are decreasing and the ratios significantly differ from those for non-defaulters. However, defaulted companies show a growth before default. We also observe that the distribution of the ratios for non-defaulters are more skewed and leptokurtic than that of defaulters.

Table 2 illustrates the correlations between the quantitative variables. Accordingly, we see strong correlations between some variables which are listed as follows:

- current ratio, liquidity ratio and cash ratio,
- equity profitability ratio, active profitability ratio and (total liability)/(total assets),
- net profit margin and equity growth rate,

Therefore, in the estimation we exclude the variables X2, X3, X8, X10 and X12 from the list of quantitative variables.

Table 1: The Descriptive Statistics of the Quantitative Variables for Defaulted and Non-Defaulted Companies

	Default							Non-Default					
Variables	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis		Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
x_1	0.93	0.41	0.19	1.98	0.36	1.04		1.92	1.74	0.22	14.37	5.88	41.39
x_2	0.35	0.25	0.00	0.78	0.60	-0.62		1.16	1.15	0.09	8.96	4.97	32.30
x_3	0.06	0.08	0.00	0.29	1.73	3.08		0.40	0.85	0.01	6.40	5.78	39.82
x_4	0.07	0.12	0.00	0.52	3.29	12.46		0.10	0.14	0.01	0.98	4.61	25.82
x_5	0.08	0.16	0.00	0.74	4.16	18.21		0.13	0.43	0.01	3.51	7.79	62.27
x_6	0.16	0.32	-0.11	1.37	3.07	11.02		0.04	0.06	-0.21	0.29	0.82	11.38
x_7	0.30	0.54	-0.28	2.04	1.94	4.72		0.20	0.63	-0.09	3.67	4.23	18.93
x_8	1.07	0.85	0.33	4.54	3.71	15.63		0.55	0.21	0.04	1.12	-0.02	-0.22
x_9	1.58	2.52	-2.84	6.80	0.81	-0.16		0.16	0.36	0.00	2.60	5.30	32.82
x_{10}	0.02	0.67	-1.96	1.08	-1.39	2.96		0.18	0.34	-1.56	0.75	-2.49	11.38
x_{11}	-0.18	0.86	-3.89	0.25	-4.41	19.85		0.10	0.12	-0.18	0.41	0.60	0.61
x_{12}	-0.63	1.82	-6.01	0.27	-2.82	6.90		0.15	0.51	-0.25	4.00	6.91	51.77
x_{13}	1.82	3.45	-0.87	14.82	2.94	10.36		1.19	1.49	-0.42	12.41	6.73	50.75
x_{14}	-1.11	7.06	-20.01	14.65	-1.06	3.07		0.97	1.21	-5.51	5.48	-1.20	15.24
x_{15}	2.35	4.70	-0.06	22.31	4.19	18.46		0.97	0.51	0.17	3.08	1.37	4.03

Table 2: The Correlations between Quantitative Variables

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$	$X10$	$X11$	$X12$	$X13$	$X14$	$X15$
$X1$	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$X2$	0.957	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-
$X3$	0.868	0.870	1.000	-	-	-	-	-	-	-	-	-	-	-	-
$X4$	-0.139	-0.142	0.000	1.000	-	-	-	-	-	-	-	-	-	-	-
$X5$	-0.001	0.044	0.037	0.191	1.000	-	-	-	-	-	-	-	-	-	-
$X6$	-0.134	-0.131	-0.107	-0.015	-0.017	1.000	-	-	-	-	-	-	-	-	-
$X7$	-0.005	-0.023	0.046	-0.114	-0.045	-0.087	1.000	-	-	-	-	-	-	-	-
$X8$	-0.348	-0.321	-0.231	-0.159	-0.085	0.191	-0.003	1.000	-	-	-	-	-	-	-
$X9$	-0.157	-0.131	-0.108	0.007	0.025	0.259	0.026	0.081	1.000	-	-	-	-	-	-
$X10$	0.162	0.191	0.139	0.066	0.128	0.002	-0.041	-0.536	0.073	1.000	-	-	-	-	-
$X11$	0.208	0.217	0.142	0.065	0.093	-0.027	-0.063	-0.872	0.018	0.581	1.000	-	-	-	-
$X12$	0.420	0.467	0.432	0.047	0.044	0.015	-0.163	-0.125	-0.028	0.146	0.178	1.000	-	-	-
$X13$	-0.170	-0.141	-0.083	0.077	0.028	-0.093	-0.150	0.207	-0.163	0.097	-0.179	0.098	1.000	-	-
$X14$	0.076	0.065	0.053	0.126	0.035	-0.430	0.024	-0.163	0.015	0.221	0.094	0.560	0.143	1.000	-
$X15$	-0.024	-0.119	-0.033	-0.033	-0.029	0.036	0.261	-0.051	0.060	0.027	0.010	0.057	0.040	-0.011	1.000

4.1. Credit Scoring Application Results

In this section we apply three versions of generalized models mentioned in Section3. We firstly apply the logistic regression. In the analysis the parameters of the logistic regression are obtained by using maximum likelihood estimation and the best model is selected by using a backward selection method. The coefficients of reduced model are summarized in Table 3. According to Wald-statistics of the reduced model intercept and the variables X1, X14 and X15 are all statistically significant at 0.05. On the other hand X4 is statistically significant at 0.10 and it is included in the reduced model. Therefore we accept the reduced model as final model.

Table 3: Coefficients of Logistic Regression

	Estimate	Std. Error	Wald	p.value
(Intercept)	4.1323	1.9434	2.126	0.03348 *
X_1	-5.4169	1.645	-3.293	<0.01 **
X_4	-5.8674	3.2167	-1.824	0.068145
X_{14}	-0.3973	0.1682	-2.362	0.018198 *
X_{15}	1.2024	0.5367	2.24	0.025077 *

Logit transform of our model is written as follows:

$$\ln\left(\frac{\mathrm{p}(\mathrm{x}_i)}{1-\mathrm{p}(\mathrm{x}_i)}\right) = 4.1323 \; - \; 5.4169\,\mathrm{X}_1 - \; 5.8674\,\mathrm{X}_4 \; - \; 0.3973\,X_{14} + 1.2024\,X_{15}$$

For model diagnostics we firstly apply Hosmer-Lemeshow goodness of fit test. The test statistic is 5.8546 and corresponding p.value is 0.6635 which indicates a good fit. Secondly, we calculate McFadden Pseudo R2 as 0.5802. According to Louviere et. all. (2000) if the value of the McFadden Pseudo R2 is between 0.2 and 0.4 then the model fits good. Our value is above this level but it is not too much above therefore we can say that our model fits very well.

Secondly, we implement the generalized partially linear models with logistic link function. We try 63 different models including nonparametric individual and interaction effects of insignificant variables in reduced parametric logistic model. In estimation we use ‘uniform’ kernel and ‘backfitting’ algorithm. Among the models the best fitted generalized partially linear model is selected according to AIC values. The model with AIC value 49.084 is as follows:

$$\ln\left(\frac{\mathrm{p}(\mathrm{x}_i)}{1-\mathrm{p}(\mathrm{x}_i)}\right) = \mathrm{m}(\mathrm{X6}) - \; 4.8978\mathrm{X}_1 \; - \; 6.0855\mathrm{X}_4 \; - 0.1602\,X_{14} + 1.5362\,X_{15}$$

The summary of the statistics of coefficients is listed in Table 4. Accordingly the parametric effect of the variable X4 and X14 are not significant however as in the logistic regression we left this model as a final partially linear model with logistic link function.

Table 4: Coefficients of Partially Linear Models with Logistic Link

Variables	Coeff.	Std. Errors	t	p-value
$m(X_6)$	–	–	–	–
X_1	-4.8978	1.8623	-2.6299	0.0043**
X_4	-6.0855	4.3250	-1.4071	0.0797
X_{14}	-0.1602	0.1886	-0.8494	0.1978
X_{15}	1.5362	0.7125	2.1561	0.0155*

We also check the goodness of fit of the model we refer to Muller's study again (Muller (2000)). In the paper in order to test the effect of generalized partially linear model the following hypothesis is stated

H_0: G (XT β + β0)

H_a: G(UT β + m(T)).

Then we use test statistic which is derived by Muller (2000). The test statistic is as follows:

$$R^{\mu} = 2\sum_{i=1}^{n} L(\hat{\mu}_i, \hat{\mu}_i) - L(\bar{\mu}_i, \hat{\mu}_i),$$

where $\hat{\mu}_i$ represents the GPLM fit and $\bar{\mu}_i$ represents a `biased' parametric estimate instead of parametric one. The result shows that test value is 29.672 and the corresponding alpha is < 0.01. Therefore, we can conclude that the model is correctly fitted.

Finally, we apply the generalized additive models with logistic link. We implement 63 different models including the individual effects of the insignificant variables in parametric logistic regression model. The best model is selected according to AIC values. The best model with AIC value 40.009 is as given in Table 5. X_5 and X_7 become significant when it is estimated nonparametrically.

Table 5: Coefficients of Generalized Additive Models with Logistic Link

	Estimate	Std. Error	F Stat	p.value
(Intercept)	23.7996	-	-	
X_1	-42.4559	0.1942	45.720	<0.01**
X_4	-20.2379	0.1132	26.634	<0.01**
X_{14}	-2.6369	3.4386	809.379	< 0.01**
X_{15}	9.9362	4.0534	954.082	<0.01**
	Estimate	Std. Error	Chi Sq.	p.value
s(X_7)	35.2975	6.1379	4.0479	0.2563**
s(X_{13})	0.2685	0.2228	3.2987	0.3478**

Logit transform of the model is as follows:

$$\ln\left(\frac{\mathrm{p(x}_i)}{1-\mathrm{p(x}_i)}\right) = 23.7996 + 35.2975\ \mathrm{s}(X_7) + 0.2685\ \mathrm{s}(X_{13})\ - 42.4559\mathrm{X}_1\ - 20.2379\ \mathrm{X}_4\ - 2.6369\ X_{14} + 9.9362\ X_{15}.$$

4.2. Comparison of Model Performances

In credit scoring classification power of the methods is an important issue. In this part of the paper we implement a comparison analysis. For this purpose we calculate Type I, Type II errors and accuracy ratios for Turkish SME's data. The results are given in Table 6.

Table 6: Type I, Type II Errors and Accuracy Ratio of Generalized Models with Logistic Link

	Type II	Type I	Accuracy Ratio
Logistic Regression	0.3334	0.2121	0.8736
GPLM(Logit Link)	0.1905	0.0152	0.9425
GAM(Logit Link)	0.0476	0	0.9885

Accordingly the accuracy ratio of generalized partially linear model and generalized additive model are higher than that of Logistic Regression. Both models perform comparably well in terms of accuracy ratio. However, in terms of Type II errors generalized additive model with logistic link function provides better results. Both logistic regression and generalized partially linear model show considerably high Type II errors and this is particularly undesirable feature in credit scoring.

5. CONCLUSION

In this paper, we have implemented a comparison analysis of generalized models and their application to credit scoring of Turkish SMEs data. In order to improve the reliability of the comparison the parametric part of the models has been remained unchanged for all models. The GPLM and GAM models have pointed out that different variables have significant nonparametric effects. In other words, GPLM uses the equity turnover ratio for nonparametric part on the other hand GAM shows significant nonparametric effect of (total liability)/(total assets) and Increase in Sales. Comparison of the models have implemented according to their accuracy ratios, Type I and Type II errors. Results show that generalized additive model with logistic link outperforms both Logistic Regression and generalized partially linear model in terms of three performance measures. Therefore, GAM model can be used to take decisions on credit granting process.

REFERENCES

Altman, E. (1968). Financial Ratios, Discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance,* 589-609.

Altman, E.I., and Loris, B. (1976). A financial early warning system for over- the counter broker-dealers, *The Journal of Finance,* 31,4 1201-1217.

Altman, E., Haldeman, R. and Narayanan, P. (1977). ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations, *Journal of Banking and Finance*, 29-54.

Awh, R. Y., & Waters, D. A. (1974). Discriminant analysis of economic, demographic, and attitudinal characteristics of bank charge card holders: A case study, *The Journal of Finance*, 29(3). 973-980.

Back B., Laitinen T., Sere K. and van Wezel M. (1996). *Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms*, Turku Centre for Computer Science Technical Report 40

Beaver, W. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research,* Empirical Research in Accounting: Selected Studies, 1966, supplement to vol.5, 71-111.

Coats, K.P. and Fant, L.F. (1991). A neural network approach to forecasting financial distress, *The Journal of Business Forecasting*, Vol. 10(4). 9-12.

Coats P. and Fant L. (1993). Recognizing financial distress patterns using a neural network toll, *Financial Management,* 22, 142-155.

Dambolena, I.G., and Khoury, S.J. (1980). Ratio stability and corporate failure, *The Journal of Finance*, 1017-1026.

Deakin,E.B. (1972). Discriminant analysis of predictors of business failure, *Journal of Accounting Research,* 10-1, 167-179.

Fitzpatrick, P. J. (1931). Symptoms of Industrial Failures as Revealed by an Analysis of the Financial Statements of the Failed Companies, Washington, D. C; Catholic University of America.

Fitzpatrick, D. B. (1976). An analysis of bank credit card profit, *Journal of Bank Research*, 7, 199- 205.

Hastie, T. J. (1991). Generalized additive models. Chapter 7 of Statistical Models in S eds. J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Hastie, T. J. and Tibshirani, R. (1987). Generalized additive models: some applications, *Journal of the American Statistical Association*, Vol. 82, No. 398 pp. 371-386.

Hosmer D.W., and Lemeshow Jr.S. (2000). Applied Logistic Regression, New York, Wiley.

Huo, Y., Chen, H., and Chen J. (2017). Research on Personal Credit Assessment Based on Neural Network-Logistic Regression Combination Model, *Open Journal of Business and Management*, 5, 244-252.

Galindo, J. and Tamayo, P. (2006) Evaluation of Neural Networks and Data Mining Methods on a Credit Assessment Task for Class Imbalance Problem, *Nonlinear Analysis: Real World Applications*, 7, 720-747.

Laitinen, T. and Kankaanpaa, M. (1999). Comparative analysis of failure prediction methods: the finnish case, *The European Accounting Review* 8:1, 67-92.

Louviere, J. J., Hensher, D. A. and Swait, J. (2000). Stated choice methods: analysis and application. Cambridge University Press.

McKee , T.E. and Greensten M. (2000). Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 19: pp. 219-230.

Mears, P. K., Discussion of financial ratios as predictors of failures, Jour- nal of Accounting Research, Emprical Research in Accounting: Selected Studies 4 (1966) 119-122.

Merwin, C. L. (1942). Financing Small Corporations: In Five Manufacturing Industries, 1926-36, National Bureau of Economic Research.

Miyamoto, M. (2014). Credit Risk Assessment for a Small Bank by Using a Multinomial Logistic Regression Model*, International Journal of Finance & Accounting*, 3, 327-334.

Muller, M. (2000). Semiparametric extensions to generalized linear models, Humboldt Universitat zu Berlin. Working Paper.

Muller M. and Ronz B. (1999). Credit scoring using semiparametric methods, Humboldt Universitaet Berlin in its series Sonderforschungsbereich 373 with number 1999-93.

Odom, M.D., and Sharda, R. (1990). A Neural network model for bankruptcy prediction, Neural Networks, IJCNN International Joint Conference on 17-21, 2, 163-168.

Orgler, Y.E. (1970). Credit scoring model for commercial loans, *Journal of Money, Credit and Bank,* 2: 435-445.

Pantalone, C. and Platt, M. (1987). Predicting commercial bank failure since deregulation, *Federal Reserve Bank of Boston New England Economic Review,* (July/ August):37-47.

Ramser, J. & Foster, L., A (1931). Demonstration of Ratio Analysis, Bulletin 40, Bureau of Business Research, University of Illinois, Urbana.IL.

Shi, Q. and Jin, Y. (2004). A Comparative Study on The Application of Various Personal Credit Scoring Models in China *Statistical Research*, 21, 43-47.

Smith, R. and Winakor, A. (1935). Changes in Financial Structure of Unsuccessful Industrial Corporations. Bureau of Business Research, Bulletin No. 51. Urbana:University of Illinois Press,

Sinkey, J.F. (1975). A Multivariate analysis of the characteristics of problem banks, *The Journal of Finance*, 30: 21-36.

Xiao, W. and Fei, Q. (2006). A Study of Personal Credit Scoring Models on Support Vector Machine with Optimal Choice of Kernel Function Parameters, *Systems Engineering Theory & Practice*, 26, 73-79.

Yang, S., Zhu, Q. and Cheng, C. (2013). The Building of the Combined Model for Personal Credit Rating: A Study Based on the Decision Tree-Neural Network, *Financial Forum*, 27, 57-61.