ACADEMY OF ECONOMIC STUDIES
DOCTORAL SCHOOL OF FINANCE AND BANKING
May 2015

Early Warning Systems for Bankruptcy Prediction

MSC STUDENT: DUMITRAN RALUCA ALEXANDRA

SCIENTIFIC COORDINATOR: PHD PROFESSOR MOISA ALTAR

Contents:

1. Problem overview

- The bankruptcy of a firm not only affects the company itself, but every economic agent in interaction with the aforementioned company, due to the systemic character of risk;

- Banks need to understand the credit risk and default probability related to their portfolio of clients, in order to improve the quality of their portfolios, and increase their profits, mitigating the risk of their own default due to bad loans;

- Having a better understanding of the insolvency peril, the management of a company can take the necessary measures in avoiding bankruptcy, given there are no strong macroeconomic influences that cause the company’s situation;

2. Objectives:

- The paper aims to develop an early warning system model based on logistic regression approach, that predicts the default probability of Romanian stock exchange companies;

  - Asses model’s accuracy of prediction;
  - Asses model’s stability;
  - Compare models considering different time frames;

3. Literature review:

- The first credit risk assessments are dated from 1910s, with the development of Risk Rating Agencies (Moody's , S&P, Fitch);

- The pioneers of bankruptcy prediction models are Beaver(1966), who developed a univariate financial ratios analysis and Altman(1968), who developed a multivariate discriminant analysis (Z-score);

- The first to use the Logit model was Ohlson(1980), followed by Andrew Lo(1985), Altman and Sabato(2007), Bartual et al(2012), Li & Wang(2014);

- Among other methods are probit (Zmijewski -1984), and soft computing/artificial intelligence methods (Artificial neural networks: Wilson & Sharda-1994, Genetic Algorithms: Min and Jeong-2008, decisional trees: Lin and McClean-2001);

4. Methodology

4.1 Database

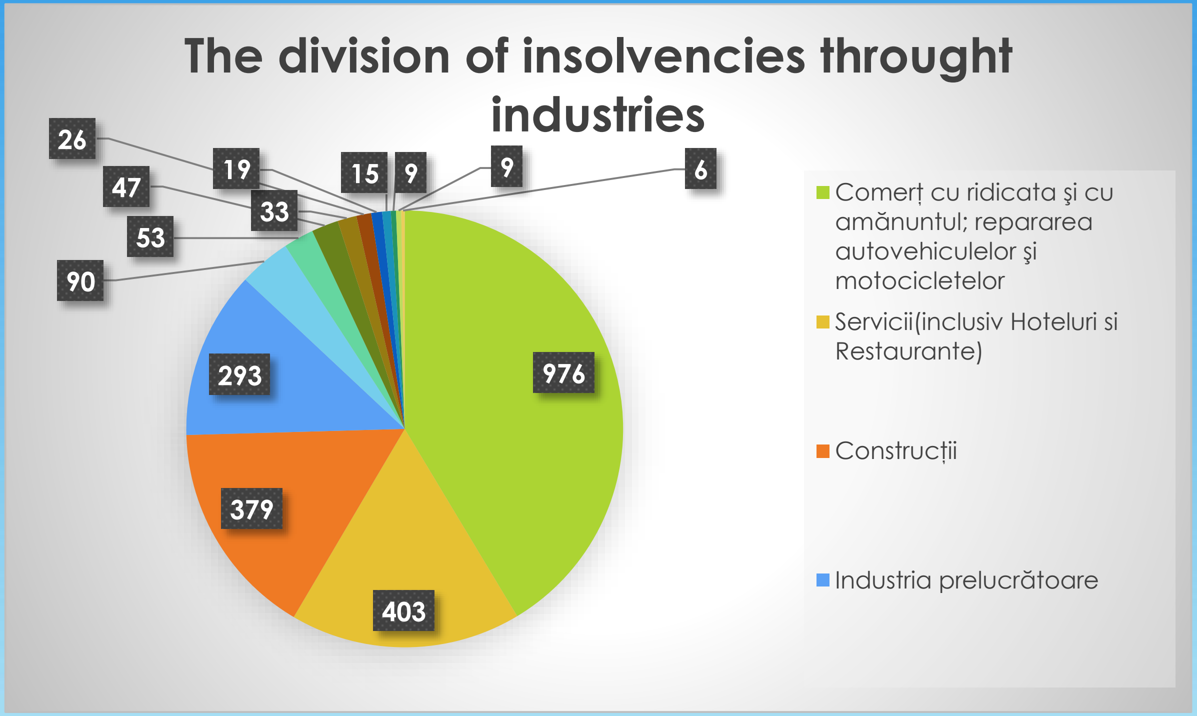
- 445 stock exchange Romanian companies, traded both on BVB and OTC markets;
- 390 solvent companies and 55 insolvent one, established based on two criteria:
  - Legally declared insolvency;
  - Negative net worth for at least three consecutive years (technical bankruptcy);
  - The model that has also taken into account technical bankruptcy yielded much better results;
- Companies are from the following industries (SIC code classification): Retail and Wholesale trade, Construction, Agriculture, Forestry and Fishery, Services and Manufacturing;

Source of data : Thomson Reuters and Duns & Bradstreet;

- I have chosen for my analysis the 5 industries that had the most insolvencies in 2015 Q1 (legally declared):

Industry	01.01.2015 - 31.03.2015	01.01.2014 - 31.03.2014	Dynamics
Comerţ cu ridicata şi cu amănuntul; repararea autovehiculelor şi motocicletelor	976	1801	-45.81%
Servicii(inclusiv Hoteluri si Restaurante)	403	797	-49.44%
Construcţii	379	736	-48.51%
Industria prelucrătoare	293	517	-43.33%
Agricultură, silvicultură şi pescuit	90	134	-32.84%
Tranzacţii imobiliare	53	78	-32.05%
Informaţii şi comunicaţii	47	81	-41.98%
Distribuţia apei; salubritate, gestionarea deşeurilor, activităţi de decontaminare	33	40	-17.50%
Activităţi de spectacole, culturale şi recreative	26	27	-3.70%
Intermedieri financiare şi asigurări	19	33	-42.42%
Producţia şi furnizarea de energie electrică şi termică, gaze, apă caldă şi aer condiţionat	15	14	7.14%
Industria extractivă	9	25	-64.00%
Sănătate şi asistenţă socială	9	12	-25.00%
Învăţământ	6	12	-50.00%

Source: ONRC

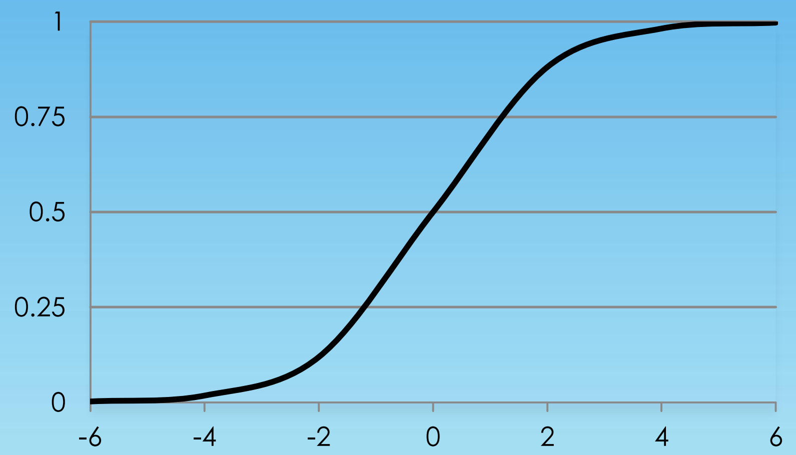


Source: Excel computation

4.2 Logit Model:

- Logistic function was first introduced by Pierre Verhulst, in his study related to population growth(1845);
- Logistic regression was developed by D.R Cox(1958);

Standard logistic function



Source: own computation

$$f(x) = \frac{1}{1+e^{-x}}$$

Logistic function, the cumulative distribution function of the logistic distribution

$$logit(p) = \log(\frac{p}{1-p}),$$

Logit function, the inverse of logistic function, a measure of entropy for the Bernoulli process

- For the bankruptcy prediction case:

$$Pr(y_i = 1) = \frac{1}{1+e^{-\Sigma\beta_i x_i}} \qquad \leftrightarrow \qquad ln\frac{\Pr(yi=1|xi)}{\Pr(yi=0|xi)} = \Sigma\beta_i x_i$$

- The logit regression is developed in SAS Enterprise Guide 4.3, based on Fischer's Scoring method

- The analysis is realized on whole sample, as well as three subsamples, to establish model's stability;

5 Data and Results:

5.1 1 year model

- 2013 financial data, used to predict bankruptcy for 2014-2015 time frame;
- 18 financial indicators are entered into a stepwise selection model, after which only the most statistically significant, with the highest R squared, are used for model prediction:

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	h3		1	1	213.9144		<.0001
2	d3		1	2	17.7707		<.0001
3	a3		1	3	10.8790		0.0010
4	m3		1	4	4.2749		0.0387
5	l3		1	5	7.1957		0.0073
6		l3	1	4		1.3081	0.2527

Where:
a3=cash/total assets(liquidity measure)
d3=operating income/total assets(profitability measure)
h3=total liabilities/total assets(solvency measure)
m3=Equity/total assets(capital adequacy/solvency measure)

5.1.1 Model statistics:

Indicator		a3	d3	h3	m3
Prob	1	0.04389423	-0.1675551	2.21160544	0.274877
	0	0.05843952	-0.0076768	0.40062612	0.363264

R-Square	0.4596	Max-rescaled R-Square	0.8725

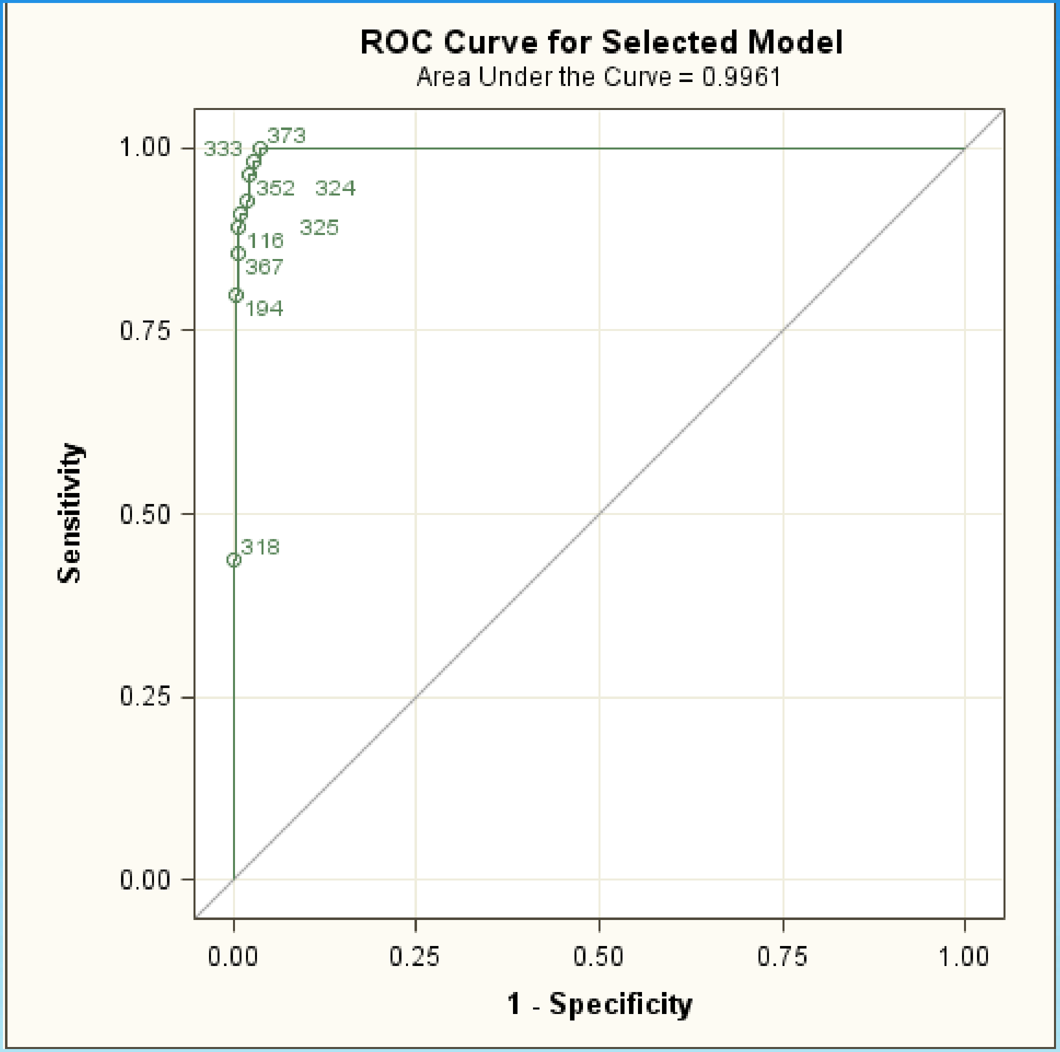
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
2.0387	7	0.9577

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	334.885	69.032
SC	338.983	89.522
-2 Log L	332.885	59.032

- A clear difference can be depicted for the average indicators, especially in terms of debt ratio, however, there is no qvasi or complete separation for the indicators;
- Rescaled R-Square is 0.8725, showing the proportion in which the depended binary variable is explained by the chosen independent variables
- The goodness of fit test shows that the observed event rates match the predicted event rates
- Information criterion Akaike and Schwartz show that the quality of the model given the dataset is highest for the 4 selected variables

Source: SAS computation

Model statistics(2):

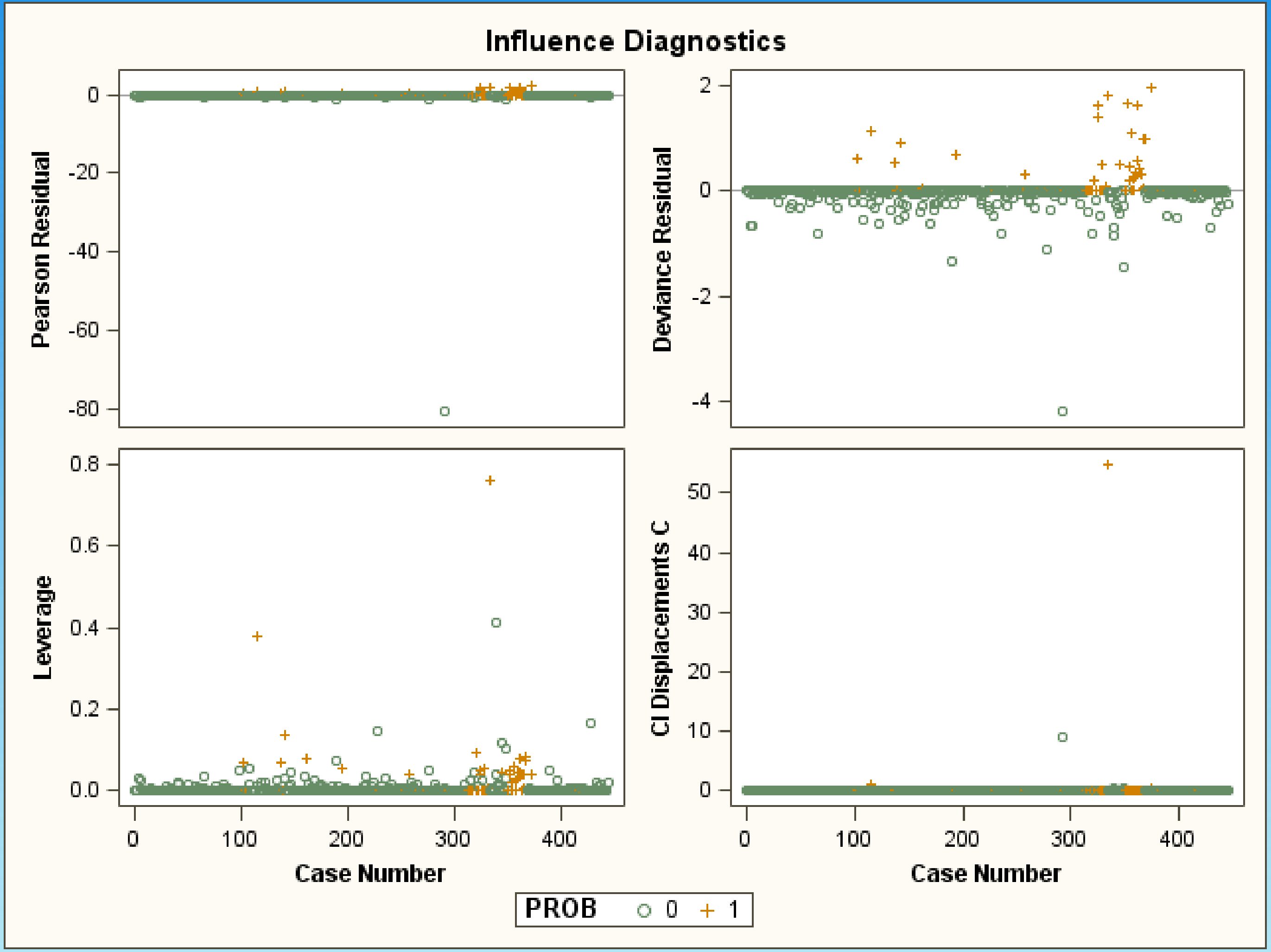


Source: SAS computation

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.900	37	389	1	18	95.7	67.3	99.7	2.6	4.4
0.800	43	389	1	12	97.1	78.2	99.7	2.3	3.0
0.700	44	388	2	11	97.1	80.0	99.5	4.3	2.8
0.600	44	387	3	11	96.9	80.0	99.2	6.4	2.8
0.500	48	386	4	7	97.5	87.3	99.0	7.7	1.8
0.200	53	379	11	2	97.1	96.4	97.2	17.2	0.5

Source: SAS computation: Prediction accuracy under different cutoff points

Model statistics(3):



- The **Pearson** and **Deviance Residuals** show that only case 300 is poorly accounted for by the model;
- The **Leverage**(diagonal elements of the **influence matrix**) show three extreme points, for which the observed values are not fitted with the predicted values;
- The furthest outlier is also depicted by **CI Displacement C** graphic (between 300 and 350);

5.1.2 Stability:
1 year results without Manufacturing Companies:

Response Profile		
Ordered Value	PROB	Total Frequency
1	0	227
2	1	46

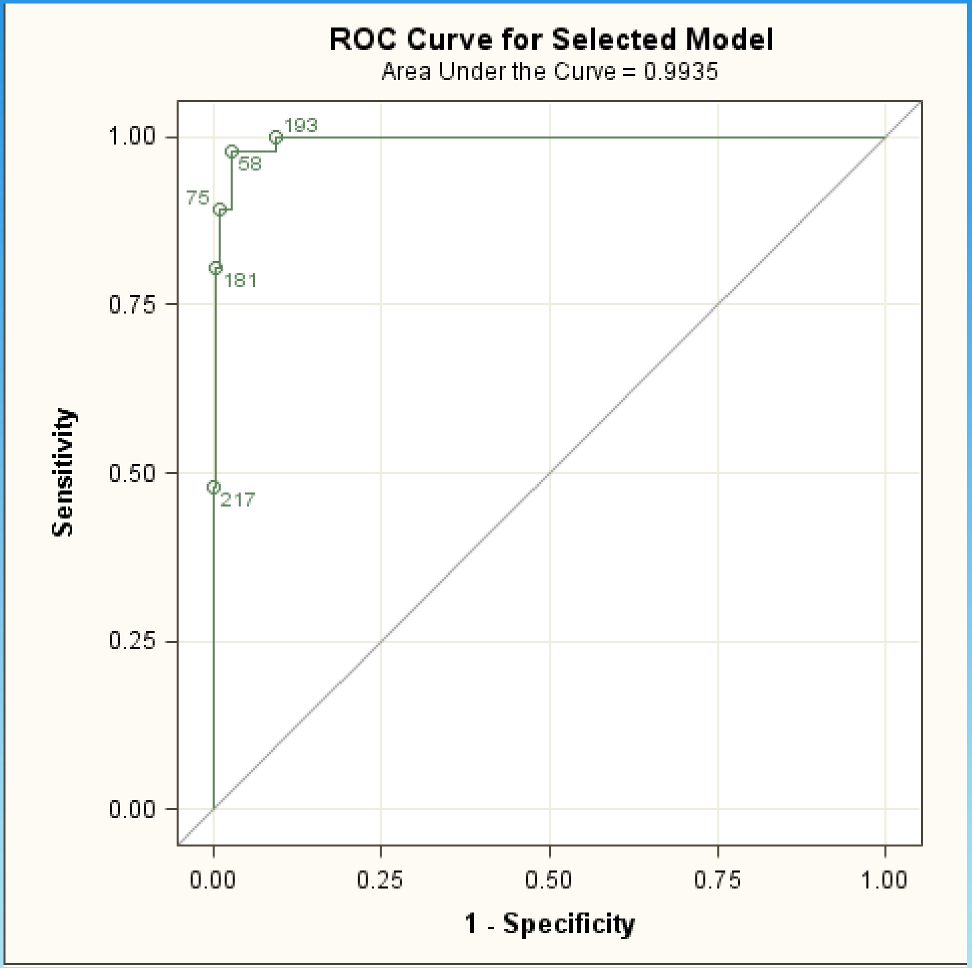
Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	249.609	56.468
SC	253.219	78.125
-2 Log L	247.609	44.468

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	h3		1	1	133.7348		<.0001
2	d3		1	2	17.3545		<.0001
3	a3		1	3	9.6848		0.0019
4	m3		1	4	2.2355		0.1349
5		m3	1	3		3.6740	0.0553

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.700	38	225	2	8	96.3	82.6	99.1	5.0	3.4
0.500	38	223	4	8	95.6	82.6	98.2	9.5	3.5
0.200	45	217	10	1	96.0	97.8	95.6	18.2	0.5

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.9225	8	0.9987

Source: SAS Computation



1 year results without Trade Companies:

Response Profile		
Ordered Value	PROB	Total Frequency
1	0	338
2	1	43

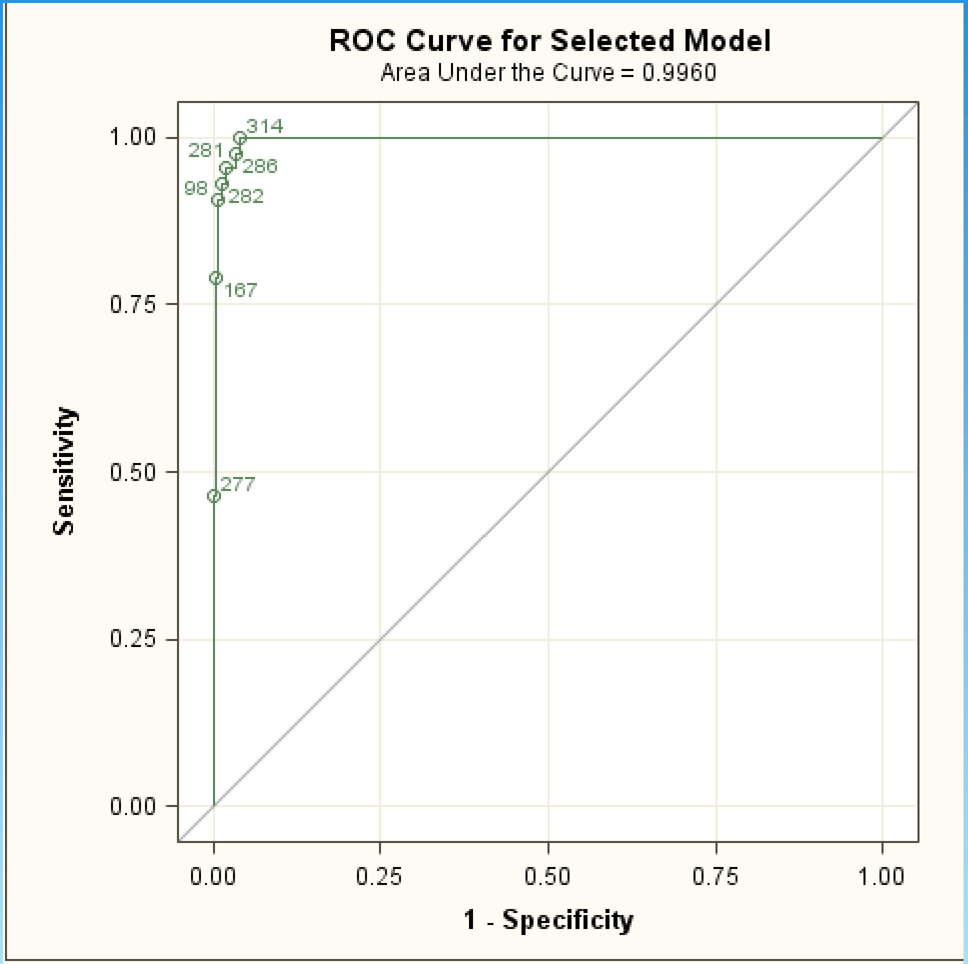
Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	270.571	56.729
SC	274.514	80.386
-2 Log L	268.571	44.729

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	h3		1	1	181.1295		<.0001
2	d3		1	2	17.4352		<.0001
3	a3		1	3	14.0409		0.0002
4	k3		1	4	8.2632		0.0040
5	m3		1	5	6.9691		0.0083
6	l3		1	6	7.1317		0.0076
7		l3	1	5		1.4180	0.2337

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.700	34	335	3	9	96.9	79.1	99.1	8.1	2.6
0.500	38	335	3	5	97.9	88.4	99.1	7.3	1.5
0.200	41	329	9	2	97.1	95.3	97.3	18.0	0.6

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.8184	6	0.9916

Source: SAS Computation



1 year results without Services Companies:

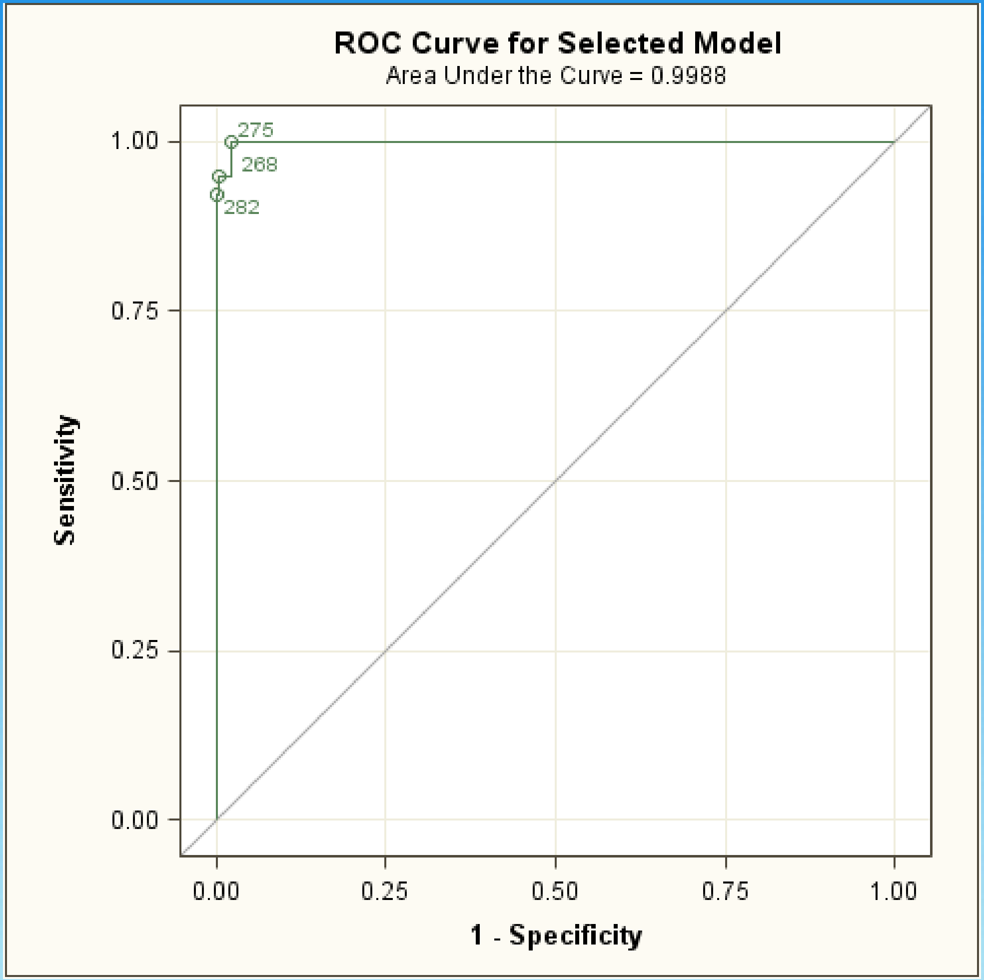
Response Profile		
Ordered Value	PROB	Total Frequency
1	0	311
2	1	39

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	246.644	21.674
SC	250.502	33.248
-2 Log L	244.644	15.674

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	h3		1	1	189.0655		<.0001
2	d3		1	2	11.3336		0.0008
3	c3		1	3	3.3185		0.0685
4		c3	1	2		2.4243	0.1195

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.700	35	309	2	4	98.3	89.7	99.4	5.4	1.3
0.500	36	309	2	3	98.6	92.3	99.4	5.3	1.0
0.200	37	303	8	2	97.1	94.9	97.4	17.8	0.7

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.1344	1	0.7139



Source: SAS Computation

5.2 2 and 3 year Models:

- I have first considered two models based on data from 2011-2013 period, 2012-2013 respectively, with all indicators, for all years, included in the stepwise selection
- Secondly, I have developed two models that take into account the average of the indicators from the aforementioned time-frames

Model		Variables	R-Square	AIC	SC	Hosmer-Lemenshow	ROC area	Overall prediction accuracy(min/max)
Average model	2 years time frame	ava avd avh	85.35%	73.194	85.488	41.41%	0.9934	96.4%; 97.5%
	3 years time frame	avgd avgh	83.19%	82.309	94.693	87.55%	0.993	96%; 97.3%
Full model	2 years time frame	a3 d3 h3 a2 d2 f2	74.90%	77.403	107.868	75.70%	0.9921	96.6%; 97.3%
	3 years time frame	h1 h3	85.20%	73.864	86.158	94.38%	0.9939	96%; 96.6%

Where:
a=cash/total assets
d=operating income/total assets
f=financial result/total assets
h=total liabilities/total assets

5.3 Summary of findings

- There is no correlation between the past industry indicators and the percentage of insolvencies on each industry for the considered time frame: Industry indicators yielded p-values greater than 0.1 in the logistic model; therefore, they have no significance as predictors, but a future direction of the study could take into account their current influence(scenario analysis), for which data is not yet available;

Year	2011	2012	2013	2014	% of insolvencies 2014-2015
Construction	2.90	1.40	-0.60	-6.7	11.86
Manufacturing	8.20	3.30	7.80	7.9	5.23
Trade	5.50	8.40	2.00	6.7	18.75
Services	9.20	5.60	4.30	0.2	17.70

Source: Eurostat

- The choice for the cut-off point is important- even if the model's overall accuracy of prediction is high for all three considered cutoff points, it is of greater interest to minimize the type 1 errors, for which lower cutoff point have given better results-overall, the **0.5** threshold proved to be the most efficient in establishing a balance between type 1 and type 2 errors.

Summary of findings(2)

- For the **0.5** cutoff point, sensitivity ranges between **82.6%** and **92.3%,** being highest for the sample without Service companies, followed by model based on whole database(**87.3%),** and lowest for the subsample without the Manufacturing industry;

- For the same cutoff point, Specificity ranges between **98.2%** and **99.7%**, lowest for the subsample without Manufacturing industry and highest for the subsample without Services industry;

- The highest Sensitivity levels are registered for the **0.2** cutoff point, however, the false positive rate is very high at this level;

- For the models based only on 2013 data, there is a need for more predictors(4 or 5) to maximize the accuracy, predictors that take into account the liquidity, profitability and solvability of the company; for the 2 and 3 years' models, there is a greater accent put of the solvability of the company(translated into its total debt ratio);

- The lowest -2LOGL statistics, AIC, SC info criteria, and the highest accuracy is obtained by the subsample without the Services industry, which may mean that the Services companies are more unpredictable;

6.Conclusions:

- The model has a high prediction accuracy (over 90%) in all analyzed cases, considering both different time frames and different data samples;
- Therefore, the model can be successfully implemented on the Romanian market, on different companies portfolios, to determine their default probability;
- The model did not take into account financial sector, due to the different analysis performed for this type of companies;

Future directions of the study:

Given the fact that the analysis is performed on a single country, the macroeconomic influence could not be captured. Therefore, it would be of interest to realize the analysis on different markets, both developed and emerging, in order to capture the macroeconomic influence upon the companies' bankruptcy probabilities.

References:

- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1), 109–131;
- Beaver, W. H. (1966): Financial ratios as predictors of failures. Empirical Research in Accounting , Supplement to Journal of Accounting Research: 71-111;
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589–609.
- Altman, E. I., Sabato, G.(2007),Modelling Credit Risk for SME: Evidence from The US Market;
- Andrew W. Lo (1986). Logit versus discriminant analysis: A specification test and application to corporate bankruptcies, Journal of Econometrics, pages 151-178
- Y. Wu, C. Gaunt, S. Gray(2009), A comparison of alternative bankruptcy prediction models, Journal of Contemporary Accounting & Economics

References:

- Shuangjie Li , Shao Wang (2014), A financial early warning logit model and its efficiency verification approach, www.elsevier.com Knowledge-Based Systems;
- Concepción Bartual et al (2012), Credit Risk Analysis: Reflections on the Use of the Logit Model, Journal of Applied Finance & Banking, vol. 2, no. 6, 2012, 1-13
- Tomasz Korol (2013), Early warning models against bankruptcy risk for Central European and Latin American enterprises, Economic Modelling 31 (2013) 22–30
- www.onrc.ro;
- http://ec.europa.eu/eurostat/
- https://www.dbai.dnb.com/