UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2014:35

On a new logistic regression model for bankruptcy prediction in the IT branch

Elena Belyaeva

Examensarbete i matematik, 30 hp
Handledare och examinator: Silvelyn Zwanzig
December 2014

Department of Mathematics
Uppsala University

Abstract

This work deals with an important topic of today's market research - bankruptcy. In this paper different statistical methods for bankruptcy evaluation were investigated. A new method was implemented based on generalized linear model which includes a trend of economical covariates. The method is applied to IT and telecommunication data of around 250 middle size companies.

The new method showed a good result in comparison to the old literature models.

Aknowledgement

I would like to say my deepest and sincerest thanks to my supervisor Silvelyn Zwanzig for her patience, support and guidance through the long way while this research was creating. She was always open for my ideas and helpful with report and presentation. Furthermore I would like to say thanks to my family who continiously supported me during my studies in Uppsala University.

Contents

Chapter 1

Introduction

Bankruptcy research is the large study area that has been investigated for more than 50 years. A lot of publications has been issued and a lot of results has been represented. New techniques and methodologies are created constantly. This paper was created as an intention to introduce a new method for bankruptcy prediction studies.

Originator of the studies is considered to be Altman, he used statistical methods as a base to investigate bankruptcy and later his work was re-developed and used in the other researches. Research has been continued constantly and a new type of models has been presented based on neural networks, tree-decisions, component models and so on. Some of them had quite good accuracy and probability, but still couldn't be used as universal instrument for bankruptcy prediction.

The aim of this study was to investigate and compare different statistical methods for bankruptcy evaluation and to create an own method based on obtained knowledge and research. Data came from the IT and communication branches. The data set includes a given period of time, namely last 10 years.

The methods were compared based on their ability correctly classify and predict bankruptcy of the given companies during the 10 year period from 2002 to 2012. The study uses as old well-known simple methods like: Altman, Taffler, Ohlson as the new advanced methods: i.e. regression methods.

By inspecting these models we were aimed to answer two questions:

(i) can we present a useful comparison and evaluation of the different methodologies

(ii) which model is a better predictor of the company's future?

During investigation it turned out that the number of parameters considered in the model has a little influence on predictionability. In our research it is arranged between 2 to 21 covariates (broken down in the list below).

Models that are considered as few as 2 parameters have predictive accuracies ranging from 86% to 100% and models with the higher amount of parameters have almost the same result 76% to 95%, according to literature [5].

Combining and playing with different parameters leads to successful result as bankrupt and non-bankrupt companies are classified with a higher probability.

Let's introduce parameters that have been used during research work, all definitions are taken from financial dictionary resource [20]:

NI - Net income. A company's total earnings (or profit). Net income is calculated by taking revenues and adjusting for the cost of doing business, depreciation, interest, taxes and other expenses. Is an important measure of how profitable the company is over a period of time.

TA - Total assets. Sum of liabilities and capital (where Capital for a corporation equals to Owner's Equity).

TL - Total liabilities. The aggregate of all debts an individual or company is liable for.

CA - Current/working assets. Cash or an asset expected to be converted into cash within one year. In addition to cash, current assets include marketable securities, accounts receivable, inventories and prepaid expenses.

CL - Current liabilities. A company's debts or obligations that are due within one year.

NOL - Net operating loss. A NOL is the net loss for the year attributable to business or casualty losses.

TNW - Total net worth. Amount by which assets exceed liabilities.

STL - Short-term liquidity. A company's ability to pay off its short-terms debts obligations.

NCI - No credit interval. Equals to Sales/Assets of company.

FFO - Fund provided by operation. A figure used by real estate investment trusts (REITs) to define the cash flow from their operations. It is calculated by adding depreciation and amortization expenses to earnings.

ND - Net debt. A metric that shows a company's overall debt situation by netting the value of a company's liabilities and debts with its cash and other similar liquid assets.

WC - Working capital. A measure of both a company's efficiency and its short-term financial health.

SV - Sales volumes. A quantity or number of goods sold or services sold in the normal operations of a company in a specified period.

DC - Loan/debt capital. A capital that a business raises by taking out a loan. It is a loan made to a company that is normally repaid at some future date.

REVs - Revenues. Amount of money that a company actually receives during a specific period, including discounts and deductions for returned merchandise.

P/E Ratio - Price-Earnings Ratio. A valuation ratio of a company's current share price compared to its per-share earnings.

DER - Debt equity ratio. A way to determine a company's leverage = Long-term debt / Common stock.

RE - Retained Earnings. The percentage of net earnings not paid out as dividends, but retained by the company to be reinvested in its core business or to pay debt = Beginning RE + NI - Dividends.

MVE - Market value of equity. The total dollar market value of all of a company's outstanding shares. Market value of equity is calculated by multiplying the company's current stock price by its number of outstanding shares.

The main important thing is to define what bankruptcy is, so further in this work we might understand its meaning and easily classify companies in 2 groups: bankrupt and non-bankrupt. Bankruptcy can be defined in many ways but the terms commonly used in this context have been: "financial distress" by Dun and Bradstreet (1985), "corporate failure", "insolvency", "default", "receivership" and "liquidation". Karels and Prakash (1987) explain like that: "bankruptcy is a process which begins financially and is consummated legally" [10].

Lam (1994) [9] defined it as the discontinuity of business enterprise. Foster (1986) [11] defines financial distress as a company facing severe liquidity problems that cannot be resolved without a sizeable resealing of the entity's operations. Liquidity problems leading to insolvency and is a state where the company cannot meet all its current obligations. Default is when debt covenants or other conditions of a financial agreement are violated. Receivership seeks to protect the interest of the secured creditors and liquidation is when the company is wound up under Chapter VII of the Bankruptcy Act. Bankruptcy is a legal term where the entity ceases operations following the filing of a bankruptcy petition which may be due to unpaid debts or voluntarily by shareholders.

As the purpose of the analysis should dictate the appropriate definition, in this study two state definitions: financially stable and bankrupt are used. Bankruptcy is defined as firms that have been delisted from the COMPUSTAT Files which are classified as having gone bankrupt or been liquidated. This definition is used for purposes of simplicity and due to the primarily comparative nature of this research. However, Jones (1987) [12] points out that some companies may file for bankruptcy for other reasons other than experiencing serious financial difficulty.

A similar opinion is expressed by Gilbert et al. (1990) [13] when they say that not all companies in financial distress are motivated to file for bankruptcy. Management may view bankruptcy as one possible strategic decision among others such as mergers, restructuring and voluntary liquidation. This represents a limitation to the study.

Chapter 2

Most common discriminant models

There are two different approaches to model a bankruptcy. The first one is to carry out a discriminant analysis and second one is to find a model to explain the probability of bankruptcy.

2.1 Multivariate discriminant analysis

Discriminant analysis is a problem of classification, where we define some amount of groups (two or more), so called clusters. Some of them are already known while the other ones not and we need to classify them into one of the apriori known populations based on the observations and on some measured characteristics.

Discriminant function is of the form:

$$Z = \sum_{i=1}^{n} V_i X_i = V_1 X_1 + V_2 X_2 + ... + V_n X_n \tag{2.1}$$

here the individual variable is transformed into a single discriminant score or Z value, which is then used to classify an object where $\mathrm{V_1}$, $\mathrm{V_2}$ ... $\mathrm{V_n}$ are discriminant coefficients and $\mathrm{X_1}$, $\mathrm{X_2}$ ... $\mathrm{X_n}$ are independent variables.

According to definition [23] score is defined as "a linear combination of couple of common business ratios, weighted by coefficients, we use to predict company's failure". It is easy to calculate and very usefull when evaluating failure status or financial health of the company.

Discriminant coefficients $\mathrm{V_i}$ were taken as estimation of number of firms declared for bankrutpcy and gathering a sample of stable firms with almost the same size and industry.

Let's consider two groups: bankrupt and non-bankrupt firms. Let $\mathrm{X_0}$ be a Bernulli random variable, where $\mathrm{X_0}$=0 means bankruptcy. Then the result is calculated by fitting a score to a given zone intervals.

In regression models score is used to find a bankruptcy probability applying a logistic regression formula, in this case:

$$P = P(X_0 = 1) = \frac{1}{1 + exp^(- Z)} \tag{2.2}$$

Four most popular models among discriminant analysis were selected to be presented and tested.

2.1.1 Altman Z-score model

Altman "applied a statistical method of discriminant analysis to a dataset of publicly held manufacturers" [23] and predicts that a firm goes bankruptcy within 2 years.

$$Z = V_0 + \sum_{i=1}^{5} V_i X_i \tag{2.3}$$

$$Z = 0.012 * X_1 + 0.014 * X_2 + 0.033 * X_3 + 0.006 * X_4 + 0.999 * X_5 \tag{2.4}$$

where
X_1 = WC/TA;
X_2 = RE/TA.
X_3 = Earnings before interest and taxes/TA.
X_4 = MVE/TL.
X_5 = Sales/TA.
V_i= (0.012, 0.014, 0.033, 0.006, 0.999)
Zones of discrimination:

$$Z \;=\; \begin{cases} > 2.99 & \text{"Safe zone"} \\ 1.8 - 2.99 & \text{"Grey zone(undefined)"} \\ < 1.8 & \text{"Financial distress"} \end{cases} \tag{2.5}$$

Altman reports that his model identified 95% of total first sample correctly, from statements one year before their bankruptcy. When it becomes 2 years, it is 72% accurate.

2.1.2 Ohlson O-score model (generalized additive model)

In 1980 James Ohlson introduced his model. The Ohlson model describes a relationship between the market value and the accounting information. An explanation of the model must begin by pointing out that the market value depends on the future expected dividends, under a clean surplus accounting system and has positive correlation with the abnormal earnings, that is, the earnings above the cost of equity at a risk free rate. The Ohlson model rests, in the neoclassical view, that the price of a stock is a function of present value of the expected dividends discounted at a risk free rate. It includes the following economic indicators:

X_1 = log (TA/GNP), a price-level index. The index assumes a base value of 100 for 2003.
X_2 = TL/TA

X_3 = WC/TA

X_4 = CL/CA

$$X_5 = \begin{cases} 1 & \text{TL exceeds TA} \\ 0 & \text{otherwise} \end{cases}$$

X_6 = NI/TA

X_7 = FFO/TL

$$X_8 = \begin{cases} 1 & \text{NI was negative for the last two years} \\ 0 & \text{otherwise} \end{cases}$$

X_9 = (NI(t)-NI(t-1))/($\|NI(t)\| + \|NI(t-1)\|$), where NI(t) is a net income for the most recent period.
The denominator of X_9 acts as a level indicator, it's a measure of a net income change. The index year is as of the year prior to the year of the balance sheet date.
V_i=(-1.32,-0.407,6.03,-1.43,0.757,-1.72,-2.37,-1.83,0.285,-0.521)
Formula:

$$\begin{aligned} O = -1.32 - 0.407 * X_1 + 6.03 * X_2 - 1.43 * X_3 \\ +0.757 * X_4 - 1.72 * X_5 - 2.37 * X_6 \\ -1.83 * X_7 + 0.285 * X_8 - 0.521 * X_9 \end{aligned} \tag{2.6}$$

P comes from (2.2) and a result of the model is a number. Let's use O instead of P, as an outcome of Ohlson model is a O-score.

$$O = \begin{cases} <= 0.5 & \text{"Safe zone"} \\ > 0.5 & \text{"Financial distress"} \end{cases} \tag{2.7}$$

It means that results larger than 0.5 can define that a company will go bankrupt within two years.

O-Score model was derived after evaluating over 2000 companies, comparison to Altman Z-Score where it was 66 firms. So we may conclude

O-Score is more accurate to predict default within a 2 year. Literature accuracy of O-Score model is above 90%.

2.1.3 Taffler model

Richard Taffler - British scientist, an economist, in 1997, the author proposed model based on an extensive survey of the vast array of data.

Using computer technology 80 financial ratios were calculated. Data were evaluated for a number of solvent and bankrupt enterprises. This information has been processed through a series of statistical methods, was eventually built multivariate discriminant and through it derived model solvency built on private factors.

Leverage, profitability, liquidity, capital adequacy and other parameters were evaluated for model creation. Taken together, the coefficients of the model give an objective picture of the risk of bankruptcy in the future and solvency at the moment.

The model is applicable to companies in the form of joint stock companies, whose shares were subject to public offering and traded on various stock exchanges.

$$Z = \sum_{i=1}^{4} V_i X_i \tag{2.8}$$

Formula:

$$Z = 0.53 * X_1 + 0.13 * X_2 + 0.18 * X_3 + 0.16 * X_4 \tag{2.9}$$

where
X_1 = ProfitBeforeTax/CL or (EBT/SD)
X_2 = CA/TL or CA/ER(External resources)
X_3 = CL/TA or SD/A(Short term debts/Assets)
X_4 = NCI = (ImmediateAssets-CurrentLiabilities/OperatingCost-Depreciation) = Sales/Assets or SR/CBR (SR/Core business revenues)
V_i = (0.53, 0.13, 0.18, 0.16)

$$Z = \begin{cases} > 0.3 & \text{good chances} \\ 0.2 - 0.3 & \text{"grey zone" (undefined)} \\ < 0.2 & \text{almost a bankrupt} \end{cases} \tag{2.10}$$

2.2 Generalized linear models

Let's define what the generalized linear model is: according to statistics definition it is a flexible generalization of a linear regression so that response variables are related to linear models via a link function and allow distribution models other distribution than a normal. Also "the variance of each measurement might be a function of its predicted value" [15].

2.2.1 Logit model

"Logit regression analysis is a uni/multivariate technique which allows for estimating the probability that an event occurs or not, by predicting a binary dependent outcome from a set of independent variables." [4]

2.2.2 Probit model

According to a definition [15] "Probit models offer an alternative to logistic regression for modeling categorical dependent variables. Even though the outcomes tend to be similar, the underlying distributions are different. Probit models are popular in social sciences like economics."

Decision if a company i is a bankrupt depends on unobservable utility index $\mathrm{I_i}$, we may explain following feature of it by the larger the value of index $\mathrm{I_i}$, the greater the probability of the company is stable. Index is comparable with the score from discriminant analysis.

The index $\mathrm{I_i}$ can be expressed as

$$I_i = \beta_1 + \beta_2 X \tag{2.11}$$

Vasisht [4] claims: "For each company there can be a critical or threshold level of the index ($\mathrm{I_i}$*), such that if $\mathrm{I_i}$ exceeds $\mathrm{I_i}$*, the company will be a bankrupt or non. But the threshold level $\mathrm{I_i}$* is also not observable. If it is assumed that it is normally distributed with the same mean and variance, it is possible to estimate the parameters of (2.11) and thus get some information about the unobservable index itself.

In probit analysis, the unobservable utility index $\mathrm{I_i}$ is known as normal equivalent deviate (n.e.d) or simply normit". Since n.e.d. or $\mathrm{I_i}$ will be negative whenever $\mathrm{P_i} > 0.5$, in practice the number 5 is added to the n.e.d and the result so obtained is called the probit i.e:

$$Probit = n.e.d + 5 = I_i + 5 \tag{2.12}$$

In order to estimate β_1 and β_2 (2.11) can be written as

$$I_i = \beta_1 + \beta_2 X + U_i \tag{2.13}$$

Where U_i is an extra value.

2.2.3 Zmijewski model (probit model)

The following definition of Zmijewski method comes from [14] Zmijewski (1984) "used financial ratios that measured firm performance, leverage and liquidity to develop his model. The ratios were not selected on a theoretical basis, but rather, on the basis of their performance in prior studies. Zmijewski estimated the model using probit analysis, which weights the log-likelihood function by the ratio of the population frequency rate to the sample frequency rate of the individual groups, bankrupt and nonbankrupt. Zmijewski (1984) derived his model using 40 bankrupt and around 40 to 800 nonbankrupt firms". Nearby we present this model:

$$P(X_0 = 1) = \frac{1}{1 + exp(-I)} = \frac{1}{1 + exp(-V_0 - V_1X_1 + V_2X_2 - V_3X_3)} \tag{2.14}$$

insert discriminant coefficients:

$$I = -4.3 - 4.5 * X_1 + 5.7 * X_2 - 0.004 * X_3 \tag{2.15}$$

where:
X_1 = NI/TA;
X_2 = TD/TA;
X_3 = CA/CL;
I = overall index.

Like the logit function, the probit function maps the value between 0 and 1. Zmijewski (1984) classified the correct hits on another way than Ohlson did.

$$P \;=\; \begin{cases} >= 0.5 & \text{bankrupt} \\ < 0.5 & \text{non bankrupt} \end{cases} \tag{2.16}$$

The probit model of Zmijewski is preferred in comparison with multiple discriminant analysis because the probit function maps the value to a probability bounded between 0 and 1, this value is easily to interpret. This is also the case for the logit model.

2.2.4 Logit versus probit

The following comparison is taken from [15]: "In the logit model we assume that Z follows a logistic distribution. In the probit model we assume that Z follows a standard normal distribution.

They behave similarly, except that the logistic distribution tends to be slightly flatter tailed (see Figure 2.1). One of the reasons the logit model was formulated was that the probit model was computationally difficult due to the requirement of numerically calculating integrals. Modern computing however has made this computation fairly simple. The coefficients obtained from the logit and probit model are fairly close. However, the odds ratio is easier to interpret in the logit model." An odds ratio (OR) is a measure of association between an exposure and an outcome.

$$OR = \frac{P}{1-P} \tag{2.17}$$

The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.

The main reasons to choose probit model to logit are:

• Underlying distribution is normal for probit

• Event we are inspecting is not a binary output (for instance bankruptcy) but a proportion.

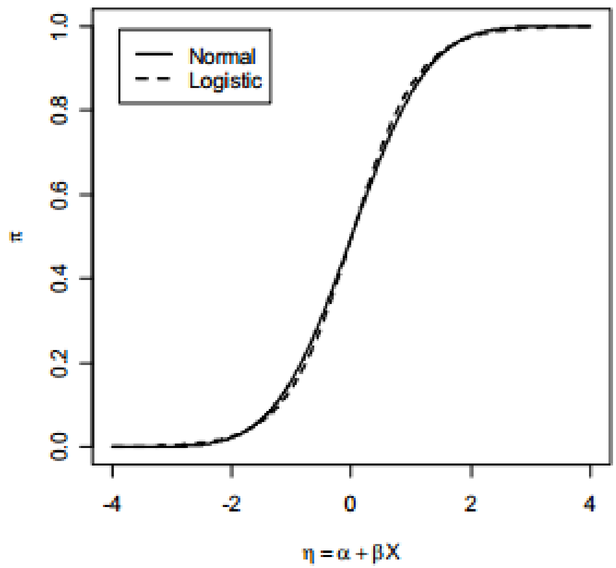


Figure 2.1: Logit and probit graph

The decision is to follow logit approach in construction as it's binary and we get a probability that is appropriate for a new method.

2.2.5 Predictive ability of the models

R program is commonly used to get the results for logit and probit models, to estimate given data set in a simple and effective manner.

The standard way to estimate a logit/probit model is glm() function with family binomial and link logit/probit. See Figure 2.2 for probit code and Figure 2.3 for logit code.

```
mydat <- data.frame(y,x1,x2)
res <- glm(y ~ x1 + x2 , family = binomial(link=probit), data = mydat)
> summary(res)
>
> library("sampleSelection")
> probit(y ~ x1 + x2, data = mydat)
> summary(res)
```

Figure 2.2: Probit code in R

```
> res <- glm(y ~ x , family  = binomial(link=logit))
> summary(res) # results
> confint(res) # confindence intervals
> names(res)
> exp(res$coefficients) # odds ratio
> exp(confint(res)) # Confidence intervals for odds ratio (delta method)
> predict(res) # prediction on a linear scale
> predict(res, type = "response") # predicted probabilities
> plot(x, predict(res, type = "response")) # plot the predicted probabilities
```

Figure 2.3: Logit code in R

2.3 Results from the literature models

There were a big amount of models starting from 60s to our days, we decided to choose the most important and significant ones such as Altman (2 factors models), Taffler, Ohlson, Logit and Probit. First I will provide literature results and then results I got applying that models on the data set, so further we may compare it.

In the following table we may see how many different models were applied to each field [5].

Table 2. Model Types

	Discriminant Analysis	Logit Analysis	Probit Analysis	Neural Networks	Other
1960's	2	0	0	0	1
1970's	22	1	1	0	4
1980's	28	16	3	1	7
1990's	9	16	3	35	11
2000's	2	3	0	4	3
Overall	63	36	7	40	26

If to say about general accuracy [5]:

Table 6. Predictive Ability by Model

	Lowest Accuracy	Highest Accuracy	Studies which obtained Highest Accuracy
MDA	32%	100%	Edmister [1972]; Santomero and Vinso [1977]; Marais [1980]; Betts and Belhoul [1982]; El Hennawy and Morris [1983]; Izan [1984]; Takahashi et al. [1984]; Frydman et al. [1985]; Patterson [2001]
Logit analysis	20%	98%	Dambolena and Shulman [1988]
Probit analysis	20%	84%	Skogsvik [1990]
Neural networks	71%	100%	Messier and Hansen [1988]; Guan [1993]; Tsukuda and Baba [1994]; El-Temtamy [1995]

Let's describe literature theoretical model results in more detail: Altman model from 1968 could define correct result with probability 79%, while Altman 1973 showed the better results 83%. Taffler model provided us with the following: failed firms: 96%, non-failed firms - 100%.

Logit method has been applied in different researches and the results differs significantly: so in the Wang research bankruptcy was defined with probability 26% and non-bankruptcy with 90,8%. Probit method gave almost the same result: for the most famous probit method (Zmijewsky) accuracy is: for bankrupt companies - 20%, non-bankrupt - 99.5% Comparing the results in a time frame period (let's take 3 years) the accuracy tendency is going down. That's why we need to compose a new method that responses to the present needs and economics.

Chapter 3

Data

3.1 IT data representation

The field of the studies is IT and telecommunication companies, cause we needed data from more or less homogeneous area for more accurate result. Data set consists of medium-size and large predominantly US financially stable and unstable companies. Period when companies filed in for bankrutcy is 2002–2012. We also included in our analysis some amount of startup companies it would be interesting to follow up their evolution in time and they are non-bankrupt on the beginning of this research.

We got a big bankruptcy dataset from New Generation Research, Inc., which is one of the most full bankruptcy datasets available in web.
One more bankruptcy dataset is available at

http://bdp.law.harvard.edu/fellows.cfm and http://lopucki.law.ucla.edu/, it contains small companies and no debt data. Non-bankrupt data came from Yahoo Finance web resource, which is the most informative and complete source of financial data reports.

Our data set includes annual reports of around 100 public and private firms. Value of firms capital for our research ranges ranges from 50 million to almost 600 billion US dollars.

Multiple financial ratios are also used in this paper, like ratio of net earnings and total assets. Adding more covariates could significantly improve the predictive ability of the models, but we will test it in our research. In general this section devoted to composing a new predictive method and covers around 20 different coefficients (compare a list on a page 4).

The average lifetime of the bankrupt firms is 7-10 years.

The data sample can be found in Appendix A and references to data source in Bibliography section. Further in this work we will refer to this data set as an IT data set. It is also organized as a data frame in R called "Companies data".

Chapter 4

Results

4.1 Training data set

To build up a model we need to start with a training data set. It is when an initial data set is splitted up in two parts: training and validation sets in proportion 20 to 80. One training set consisted of financial covariates of approximately 20 -30 companies. I selected random training sets to compare the models.

Below we represent how an ordinary training set looks like (it's splitted into two tables due to its size):

Table 4.1

B	Reven.	Net Earn.	Tot. Assets	Ret. asset rat.	Cur. rat.
1	77395	122386	49894	77224	45567
1	95612	156868	72959	90941	45667
1	120777	213915	108319	120445	235464
1	130877	230899	110922	123002	457353
1	132676	235877	112746	125678	34654
1	139844	240678	115789	130498	45823
1	458390	589788	612900	632099	458232
1	34890	35892	37892	31892	569245
1	457822	490233	390238	443900	45892
1	238990	230899	110922	123002	679123
1	149000	179389	120349	150344	82368
0	239011	240678	210399	290239	67896
1	34467	56034	48894	89224	459267
0	24920	35894	22192	37903	56930
1	23100	25905	18987	28912	28679
1	348900	379233	310283	390812	568923

Continued on the next page

B	Reven.	Net Earn.	Tot. Assets	Ret. asset rat.	Cur. rat.
1	1389701	1590348	1209384	1495873	56892
0	89405	240678	115789	130498	23482
1	489011	122386	49894	77224	49875
0	97300	156868	72959	90941	5682335

Table 4.2

B	Earn p/share rat.	Ret. on eq. rat.	Debt eq. rat.	Tot. liabil.
1	138185	297847	5377	56788
1	176099	1856	2386	34877
1	252308	34545	23226	56383
1	254899	264336	23661	56724
1	256378	23542	67832	67212
1	263009	34546	233522	456336
1	6702399	345462	78275	78923
1	43092	56837	542443	12868
1	539023	42875	23684	78252
1	254899	345298	36924	59234
1	130230	34784	186269	28695
0	270009	23754	346209	286473
1	110244	19567	24856	2864
0	57434	27842	25845	34879
1	29847	45934	56835	38645
1	412098	56835	68235	65934
1	1798745	45782	45199	12856344
0	263009	568235	45823	12876
1	138185	98723	27564	4867
0	176099	57824	284528	592351

With the help of training sets we applied different combinations of covariates to select appropriate variables for our model. They are: total assets, net earnings, total liabilities, revenues.

When we achieved the best possible performance we started to apply a model to fit to a validation data set (see Appendix A and Conclusion section for results).

4.2 Application of the standard models to IT data set

In this section we record the results of the application of theoretical models from the literature such as Olson, Taffler, Zmijevsky, Altman to our data set.

We had around 100 companies to analyze, both bankrupt and non-bankrupt. For each company I inspected 4 discriminant models by inserting corresponding values into the Altman, Taffler, Ohlson, Zmijevsky formulas.

Numeric results are represented in the table below:

B/NB	ALtman2	Taffler	Olson	Zmijevsky
1	-399.557846	2272.9224	19289.02	490.92355
1	-921.576544	4915.3192	35478.44	1060.42582
1	-4974.606634	6308.4809	37701.73	1238.64008
1	1030.49526	10622.8462	123735.9	3158.41372
0	-3131.427348	6934.1827	48780.53	1446.16165
1	-33139.97114	21856.1027	73436.89	3034.43714
1	-4099.343742	8154.5719	52856.93	1715.50898
1	-1365.978798	7665.6518	77404.83	1805.92039
0	-4615.900294	19074.4026	115067.1	3273.91041
1	-7036.992948	12386.5387	76901.41	2351.00922
1	8061.25711	33233.4503	208346.1	10291.0913
1	-4827.158458	23285.4196	156393.5	5238.10724
0	-1470.108374	3815.9457	27887.62	781.88015
1	-7619.377968	10135.7232	54250.42	1682.87764
0	-3986.408682	11164.9073	127358.2	2865.39957
1	-4839.1564	11710.6483	126405.3	2441.21762
1	-9372.496778	36418.9932	197119.2	5328.1288

Figure 4.1: Numeric results from standard models

First column determines if a company is a bankrupt or not (1 is a bankrupt, 0 is not), the other columns represent the value of discriminant function of the model that has been used. All the parameters listed on pages 4 and 5 are used for discriminant analysis.

The table is a subset of a bigger result randomly chosen.

Now we can interpret every value from the table. For that let's again look at the models definition and find the results scale. Depending on the value we got from the model, bankruptcy probability for every company can be classified as "High" or "Low".

The results are:

Theoretical model results				
Bankrupt	**Altman2**	**Taffler**	**Olson**	Zmijevsky
No	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
Yes	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
Yes	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low
Yes	Low	Low	Low	Low
No	Low	Low	Low	Low
Yes	Low	Low	Low	Low
No	Low	Low	Low	Low
No	Low	Low	Low	Low

Figure 4.2: Description of results for standard models

As we may see for financial stable companies discriminant models give a correct result, they define non bankrupt companies as companies with stable economics (bankruptcy probability is "Low"). Unfortunately we cannot say the same for bankrupt companies – in the table they are also defined with "Low" probability, despite it's false.

From this we may conclude the models are not really helpful when predicting company's stability in the IT field.

In the following we search for the new powerful model for IT companies bankruptcy forecasting.

4.3 Find a model to explain the probability of bankruptcy

We chose generalized linear approach (see 2.2). The main problem is to recognize significant variables which should be inducted in the model. As a graphical tool I produced scatterplots of the variables, where the bankrupt companies are determined as red dots and stable as green.

The analysis is done using glm formula applied on the IT data.

The following formula was applied in R to get a classification depiction: data $< -$ glm(Bankr $\sim$ log(Revenues) + log(NetEarnings) + log(TotalAssets) + log(TotalLiabil),family=binomial, data=companiesdata).

Combinations of different covariates have been tried and after research we choose to pick the ones that have the smallest correlation between each other, i.e: Revenues, Net Earnings, Total Assets, Total Liabilities.

Below I represent the most successful of them.

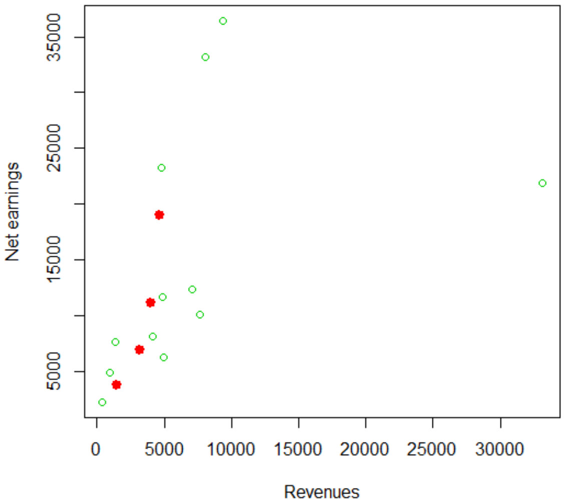


Figure 4.3: Revenues & Net Earnings

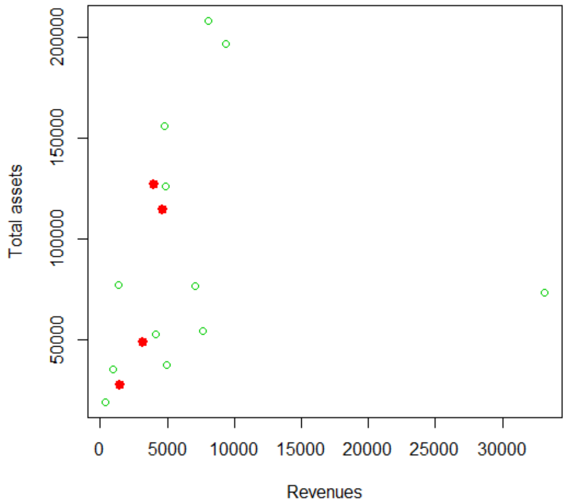


Figure 4.4: Revenues & Total Assets:

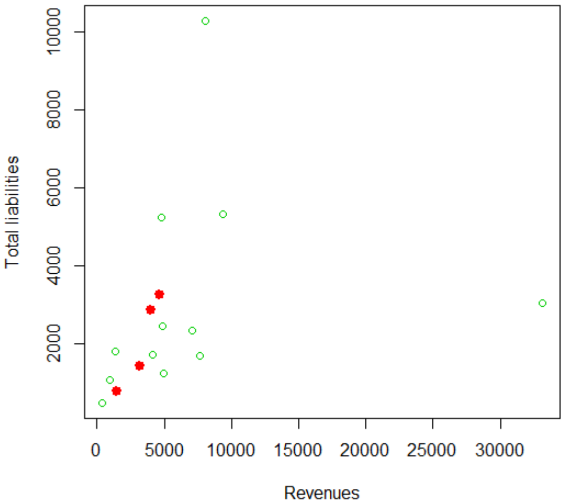


Figure 4.5: Revenues & Total Liabilities

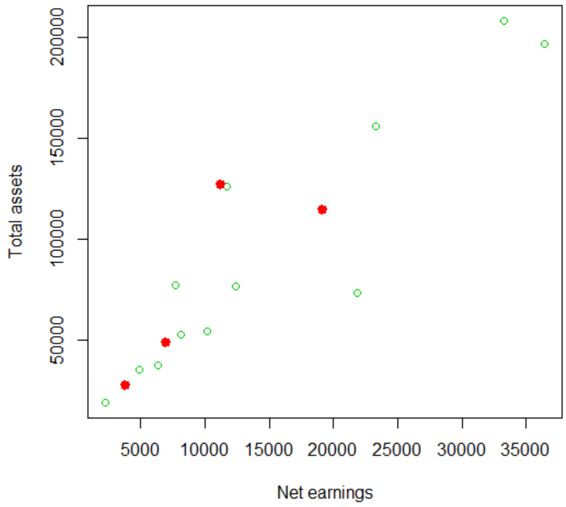


Figure 4.6: Net Earnings & Total Assets

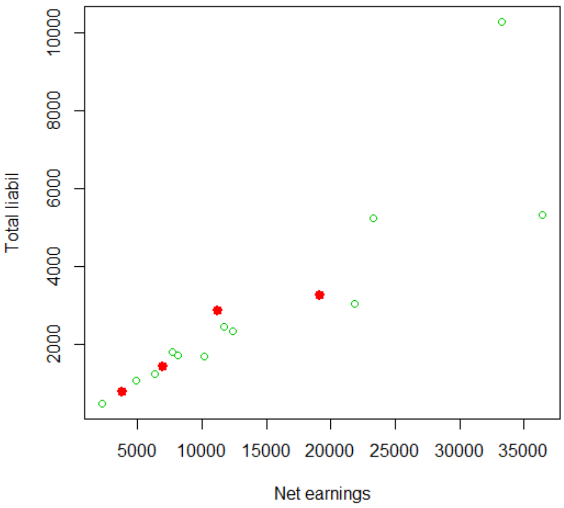


Figure 4.7: Net Earnings & Total Liabilities

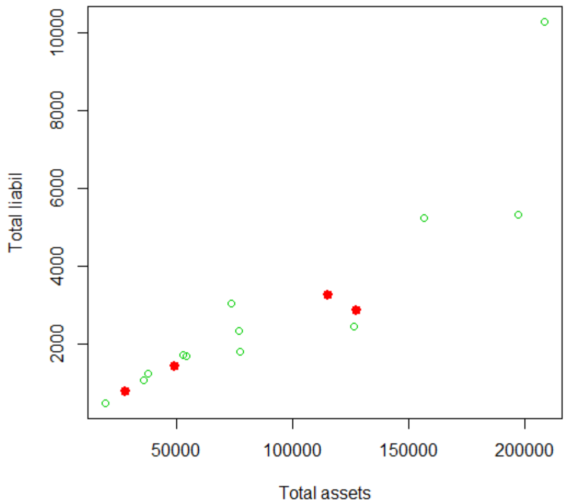


Figure 4.8: Total Assets & Total Liabilities

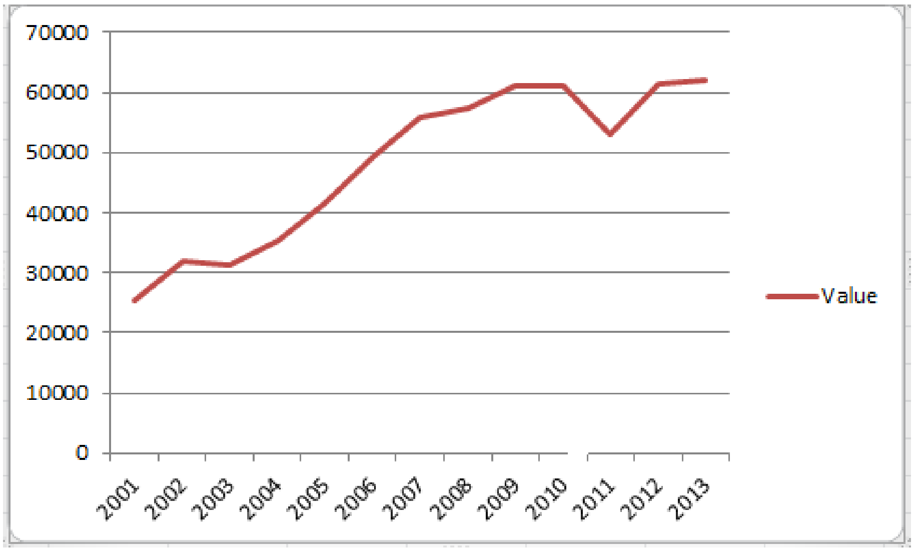


Figure 4.9: Dell time series

4.3.1 Models based on the time series basic characteristics

Bankrupt firms are mixed with non-bankrupt, so it's almost impossible to classify companies by hand, just looking on the picture. There is no strong dependency between "good" and "bad" company's results.

As we may see it's difficult to group elements uniquely by bankruptcy attribute. There are still some values from bankrupt group entering to a non-bankrupt group. So we may not classify companies by bankruptcy feature using only financial characteristics and dependency between them.

The next step is to combine time series method with logistic regression method. First apply MA (moving average - is a calculation to analyze data points by creating a series of averages of different subsets of the full data set) principle to see how close are predictive values to the real ones. This technique will smooth out short-term irregularities in the time series.

Insert our data set values into the MA formula and inspect the results:
A = Actual value in previous k periods

$$k - periodmovingaverage = \sum_{k=1}^{n}(A)/k \quad (4.1)$$

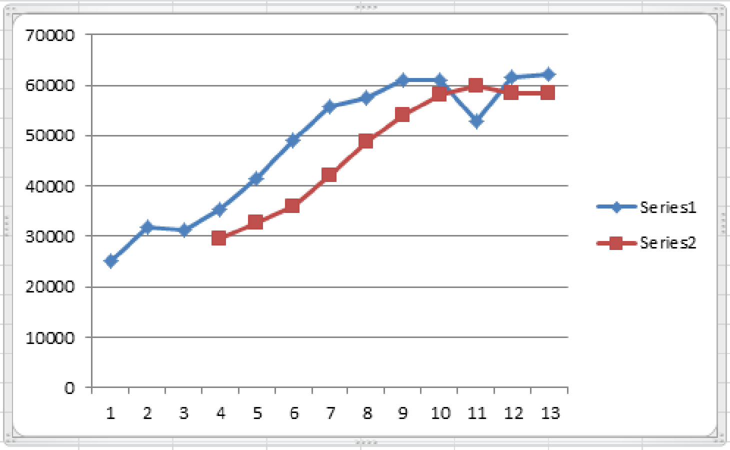
Time series for one of the Dell covariate is represented above.

The first 2 characteristics are taken from Dell financial report and are Revenue and LongTermBorrowings
Let's first look at the time series for Revenue covariate:

Dell:

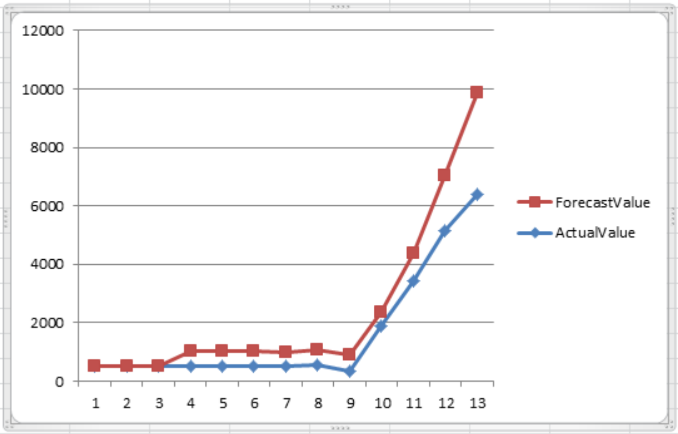
Revenues:

Actual Value	Forecast	Error
25265		
31888		
31168		
35404	29440.33333	5963.666667
41444	32820	8624
49205	36005.33333	13199.66667
55908	42017.66667	13890.33333
57420	48852.33333	8567.666667
61133	54177.66667	6955.333333
61101	58153.66667	2947.333333
52902	59884.66667	-6982.666667
61494	58378.66667	3115.333333
62071	58499	3572



LongTermBorrowings:

ActualValue	ForecastValue	Error
508		
509		
520		
506	512.3333333	-6.333333333
505	511.6666667	-6.666666667
505	510.3333333	-5.333333333
504	505.3333333	-1.333333333
569	504.6666667	64.33333333
362	526	-164
1898	478.3333333	1419.666667
3417	943	2474
5146	1892.333333	3253.666667
6387	3487	2900

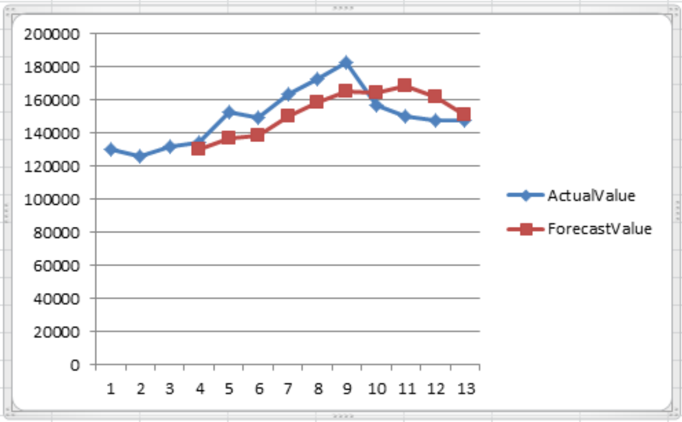


Do the same for the next company:

GeneralElectric:

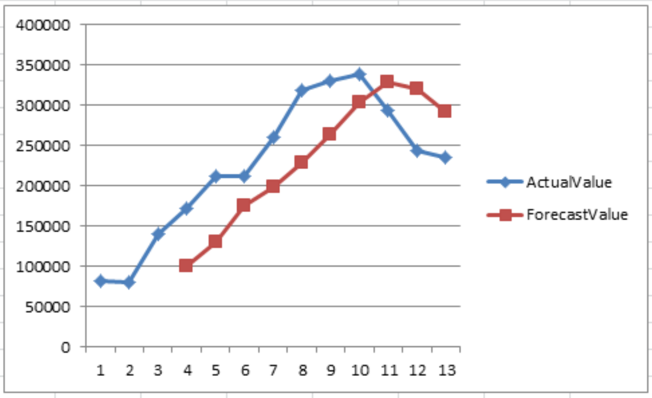
Revenues:

ActualValue	ForecastValue	Error
130385		
126373		
132226		
134641	130466.3333	4174.666667
152866	137155	15711
149702	138856.3333	10845.66667
163391	150299.3333	13091.66667
172738	158435.3333	14302.66667
182515	165202.6667	17312.33333
156783	164304	-7521
150211	168488	-18277
147300	162199.3333	-14899.33333
147359	151451	-4092



LongTermBorrowings:

ActualValue	ForecastValue	Error
82132		
79653		
140401		
171966	100728.6667	71237.33333
212670	130673.3333	81996.66667
212281	175012.3333	37268.66667
260804	198972.3333	61831.66667
319015	228585	90430
330067	264033.3333	66033.66667
338215	303295.3333	34919.66667
293323	329099	-35776
243459	320535	-77076
236084	291665.6667	-55581.66667



The method was quite close in forecasting. Trend is going in the correct direction and graphic lines: practical and predictional are quite close to each other, but still it doesn't give us an answer to our question and can only be applied as a short term prediction method.

Considering the results from MA cannot be so stable and accurate let's use more advanced method: Weighted moving average.

Its characteristics are saying:
- Assumes data from some periods are more important than data from other periods (e.g. earlier periods).
- Use weights to place more emphasis on some periods and less on the others.

Weighted moving average formula:

$$M = \frac{\sum_{t=1}^{n}(W_t + V_t)}{\sum_{t=1}^{n}(W_t)} \tag{4.2}$$

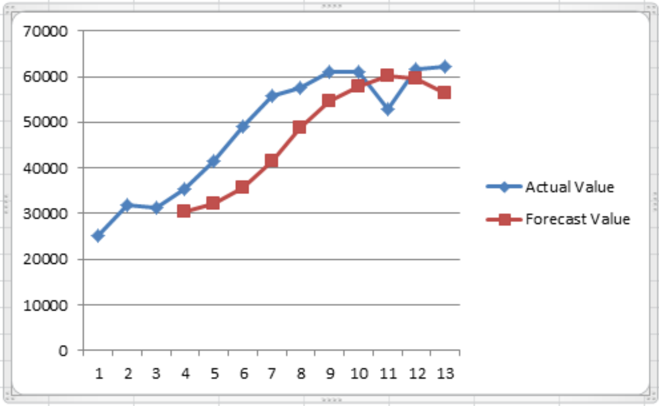
where
M - Average value
V - Actual value
W- Weightning factor
n - number of periods in the weightning group

Calculate results for the given method and represent it graphically:

Dell:

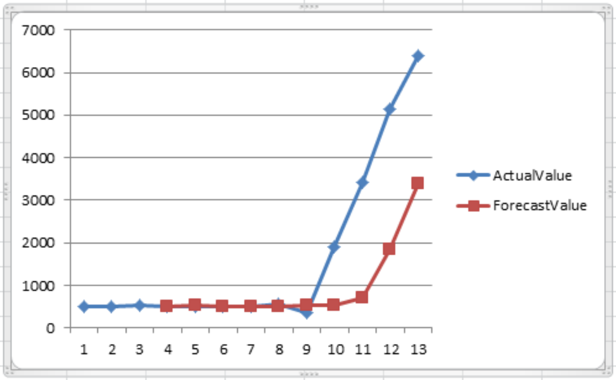
Revenues:

Moving Average			
Actual Value	Forecast Value	Error	Weight
25265			0.222
31888			0.593
31168			0.185
35404	30284.494	5119.506	
41444	32111.5	9332.5	
49205	35581.008	13623.992	
55908	41538.905	14369.095	
57420	48722.113	8697.887	
61133	54699.654	6433.346	
61101	57771.241	3329.759	
52902	60302.794	-7400.794	
61494	59591.289	1902.711	
62071	56311.698	5759.302	



LongTermBorrowings:

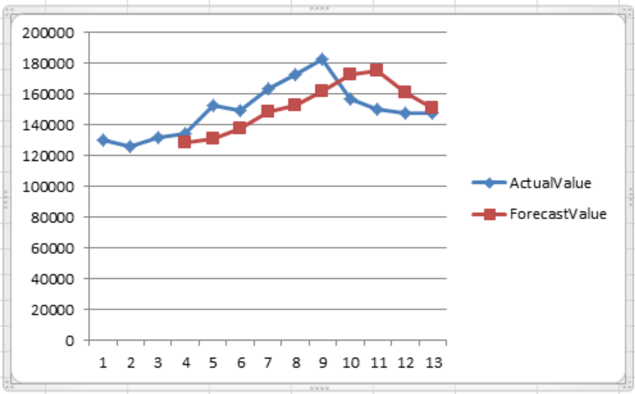
MovingAverage			
ActualValue	ForecastValue	Error	Weight
508			0.222
509			0.593
520			0.185
506	510.813	-4.813	
505	514.968	-9.968	
505	508.923	-3.923	
504	505.222	-1.222	
569	504.815	64.185	
362	516.247	-154.247	
1898	516.275	1381.725	
3417	692.114	2724.886	
5146	1838.023	3307.977	
6387	3399.647	2987.353	



GE:

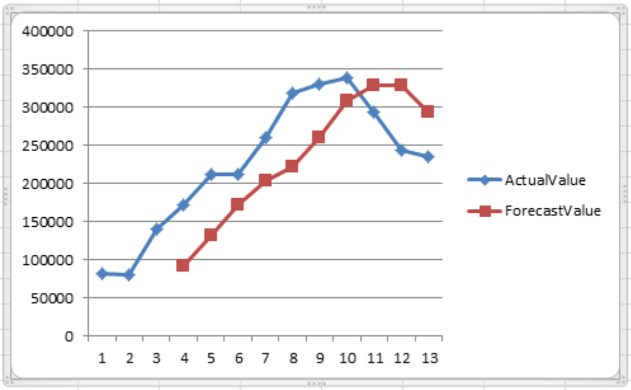
Revenues:

MovingAverage			
ActualValue	ForecastValue	Error	Weight
130385			0.222
126373			0.593
132226			0.185
134641	128346.469	6294.531	
152866	131373.409	21492.591	
149702	137476.495	12225.505	
163391	148234.71	15156.29	
172738	152936.873	19801.127	
182515	162081.237	20433.763	
156783	172471.711	-15688.711	
150211	175584.086	-25373.086	
147300	161279.684	-13979.684	
147359	151131.449	-3772.449	



LongTermBorrowings:

MovingAverage			
ActualValue	ForecastValue	Error	Weight
82132			0.222
79653			0.593
140401			0.185
171966	91441.718	80524.282	
212670	132754.469	79915.531	
212281	172488.81	39792.19	
260804	203561.747	57242.253	
319015	221344.113	97670.887	
330067	260800.929	69266.071	
338215	308136.778	30078.222	
293323	329120.836	-35797.836	
243459	328101.124	-84642.124	
236084	294064.184	-57980.184	



Looking at the pictures we might say given the results are much similar to MA results. Inspecting error values says the same. We can make a conclusion that weighted MA method can't provide us with more significant results, that's why is also counted as a short term prediction method.

4.3.2 New covariance presentation using R tool

To start an investigation first we upload our data into R and apply GLM command:

$$companiesdata <- read.table("C:/CompaniesData.txt", header = T) \quad (4.3)$$

We can decide whether there is any significant relationship between the dependent variables V_i and the independent variables X_i, where i = (1, 2, ..., p) in the logistic regression equation. In particular, if any of the null hypothesis that $H_i = 0$, where i = 1, 2, ..., p is valid, then X_i is statistically insignificant in the logistic regression model.

At .05 significance level, decide if any of the independent variables in the logistic regression model of bankruptcy in data set 'companiesdata' is statistically insignificant.

Solution: We apply the function glm to a formula that describes the bankrupt type by the four covariates. This creates a generalized linear model in the binomial family (Bernoulli distribution).

$$res <- glm(Bankr \sim Revenues + NetEarnings + TotalAssets + TotalLiabil, family = binomial, data = companiesdata) \quad (4.4)$$

We then print out the summary of the generalized linear model and check for the p-values of that covariates. Trend of the variables is estimated by the year and two years difference between the report values.

In the pictures below I represented the following types of research:

1. 1 year difference between covariate's values
2. 2 year difference between covariate's values
3. 1 year difference between covariate's absolute values
4. 2 year difference between covariate's absolute values

We also used log function application on our values to modify a result to get smoother outcome:

$$res < -glm(Bankr \sim log(Revenues)+log(NetEarnings)+log(TotalAssets) + log(TotalLiabil), family = binomial, data = companiesdata) \quad (4.5)$$

Now we present output parameters we should pay attention to. First we give official definitions of what we are going to inspect and work with:

According to the article [21]: P-value is the probability of obtaining a test statistic result at least as extreme or as close to the one that was actually observed, assuming that the null hypothesis is true. A researcher will often "reject the null hypothesis" when the p-value turns out to be less than a predetermined significance level, often 0.05 or 0.01. Such a result indicates that the observed result would be highly unlikely under the null hypothesis.

and according to this source in Wikipedia [22] "Z-value is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Because of the central limit theorem, many test statistics are approximately normally distributed for large samples. For each significance level, the Z-test has a single critical value (for example, 1.96 for 5% two tailed)". Small values are interpreted as important for the model.

AIC - The Akaike information criterion (AIC) is a measure of the relative quality of a statistical model for a given set of data.

It deals with the trade-off between the goodness of fit of the model and the complexity of the model. With AIC criteria we may estimate what is the information loss when using a corresponding model. Small AIC are considered as appropriate.

Applying the same type of scatterplots on our new covariates we obtain the following pictures, which shows that companies can be grouped into bankrupt and non-bankrupt clusters.

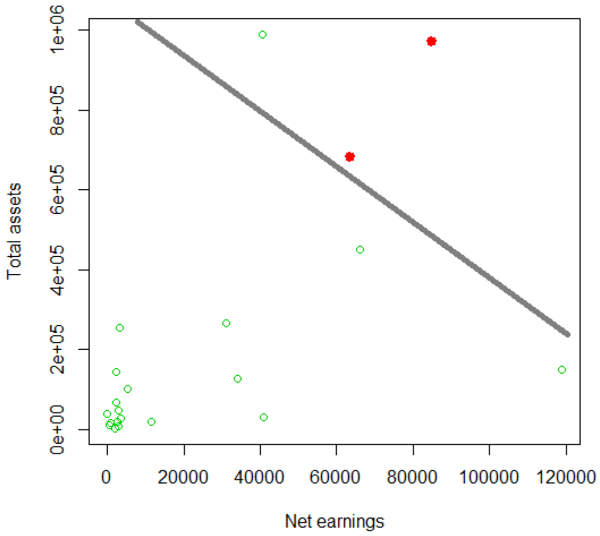


Figure 4.10: TotalAssets/NetEarnings

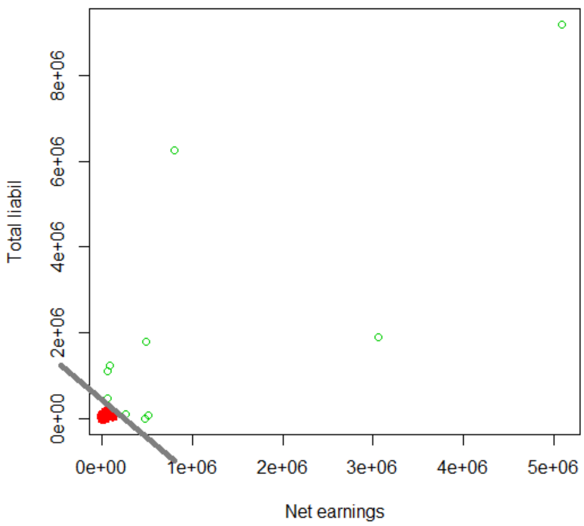


Figure 4.11: TotalLiabil/NetEarnings

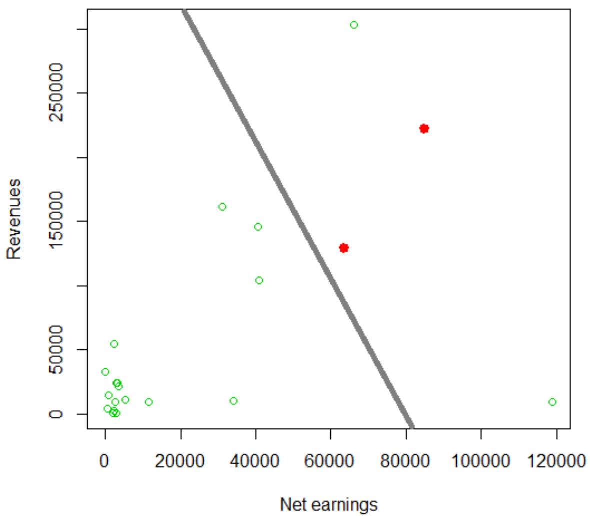


Figure 4.12: Revenues/NetEarnings

Application on 1-year difference data

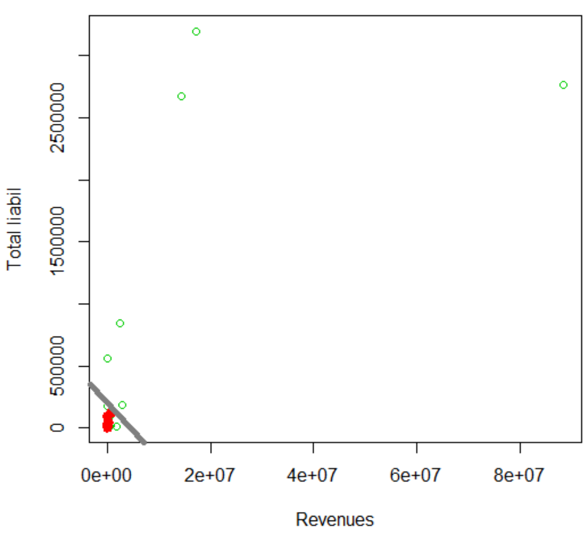


Figure 4.13: TotalLiabil/Revenues

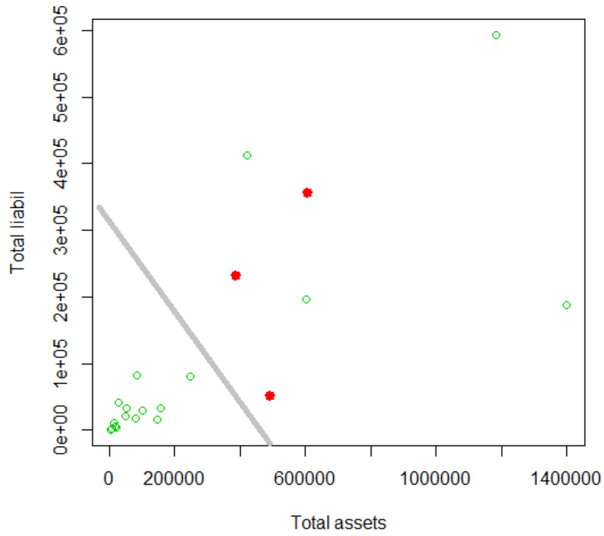


Figure 4.14: TotalLiabil/TotalAssets

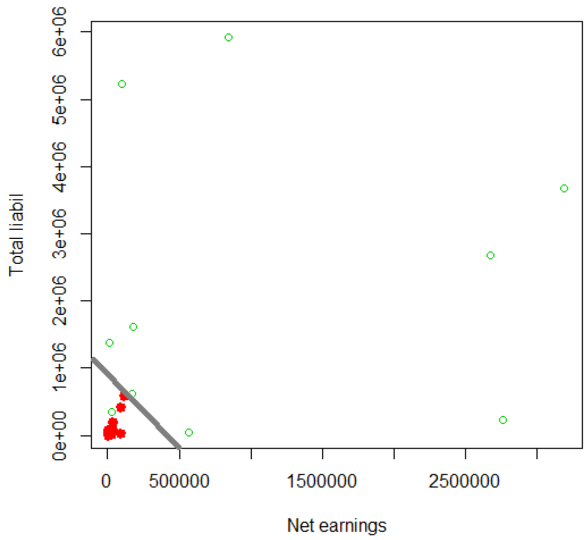


Figure 4.15: TotalLiabil/NetEarnings

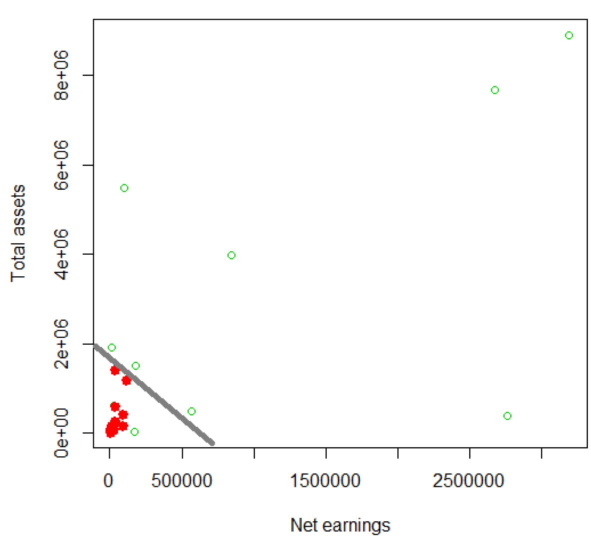


Figure 4.16: TotalAssets/NetEarnings

Application on 2-year difference data

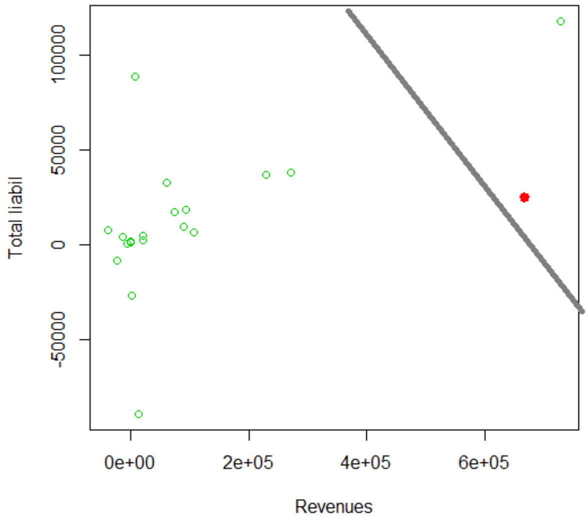


Figure 4.17: TotalLiabil/Revenues

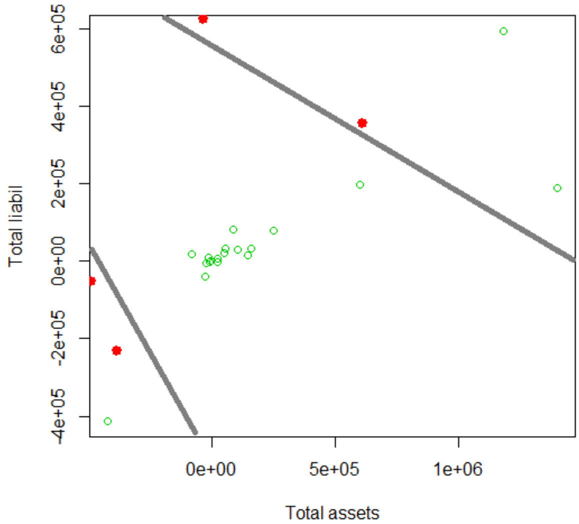


Figure 4.18: TotalLiabil/TotalAssets

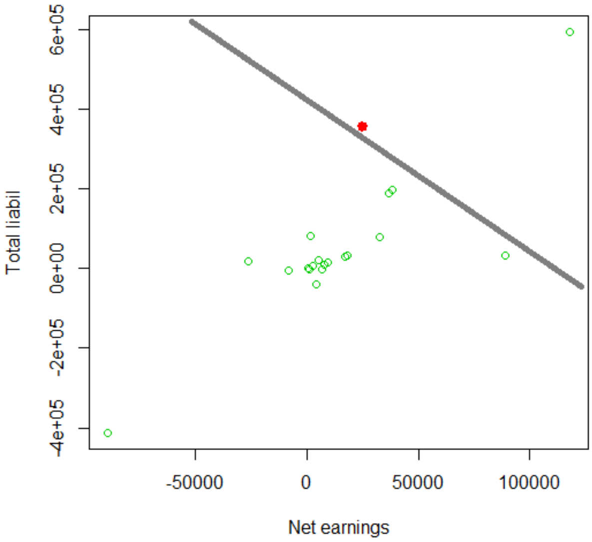


Figure 4.19: TotalLiabil/NetEarnings

Application on log of 2-year difference data

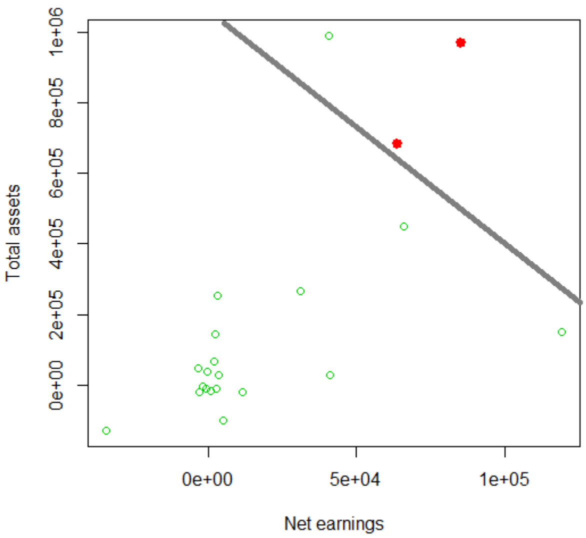


Figure 4.20: TotalAssets/NetEarnings

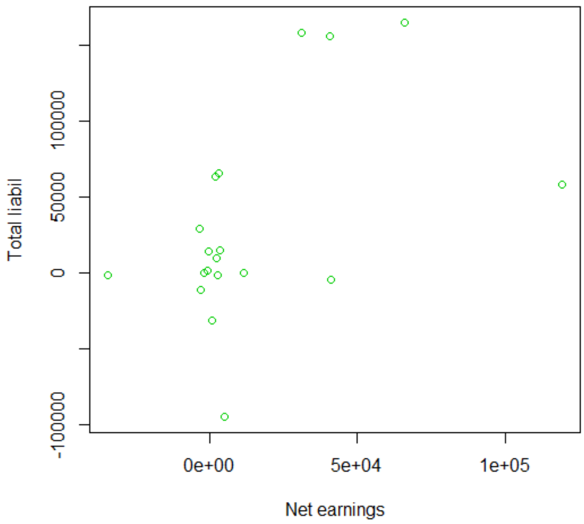


Figure 4.21: TotalLiabil/NetEarnings

Application on 2-year difference data with absolute values

As we may see from the above scatterplots elements are classified and separated with a high accuracy, that means we may go further to investigate this method.

Next, let's look at its values from glm formula applied:

2 year difference (abs. values + log):

```
        Coefficients:
                        Estimate Std. Error   z value Pr(>|z|)
        (Intercept)    4.202e+14  1.460e+07  28779122   <2e-16 ***
Diff2   Revenues      -1.153e+07  1.791e+00  -6437616   <2e-16 ***
Diff2   NetEarnings   -1.352e+09  5.133e+01 -26347034   <2e-16 ***
Diff2   TotalAssets    7.068e+08  2.027e+01  34872720   <2e-16 ***
Diff2   TotalLiabil   -1.653e+09  1.714e+01 -96405931   <2e-16 ***
        ---
        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AIC:15.17

This model with 2 year difference values looks much better. All the criterias are met. AIC is not high. Z-values are quite small. We can consider this model as the best.

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        36.1525    27.8002   1.300    0.193
log(Revenues)       0.8924     1.0045   0.888    0.374
log(TotalLiabil)   -3.8067     3.0444  -1.250    0.211
```

AIC:11.957

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         3615.8  1713098.7   0.002    0.998
log(NetEarnings)    -189.6    99250.8  -0.002    0.998
log(TotalAssets)    -114.6    54934.6  -0.002    0.998
```

AIC:6

This model has P values close to 1, that means we couldn't count this model as appropriate.

2 year difference:

```
        Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
        (Intercept)  9.525e-01  4.583e-01   2.078   0.0377 *
Diff2   Revenues     4.812e-07  5.275e-07   0.912   0.3616
Diff2   NetEarnings  6.230e-07  1.542e-06   0.404   0.6862
Diff2   TotalAssets -8.833e-08  1.539e-06  -0.057   0.9542
Diff2   TotalLiabil  9.846e-08  1.148e-06   0.086   0.9317
        ---
        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AIC:14.455

1 year difference:

```
     Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
     (Intercept)    1.299e+00  5.660e-01   2.295   0.0217 *
Diff. Revenues     -1.069e-07  1.118e-07  -0.956   0.3390
Diff. NetEarnings   4.900e-06  3.071e-06   1.595   0.1106
Diff. TotalAssets   4.250e-06  2.070e-06   2.053   0.0401 *
Diff. TotalLiabil  -4.282e-06  2.106e-06  -2.033   0.0420 *
     ---
     Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AIC:32.241

The model with 1-year difference seems also good, but AIC criteria could be smaller.

4.3.3 AIC comparison

First calculations of AIC from GLM formula with log application for 1 and 2 years data difference. AIC for different data combinations in the beginning of research are represented in the graphs below:

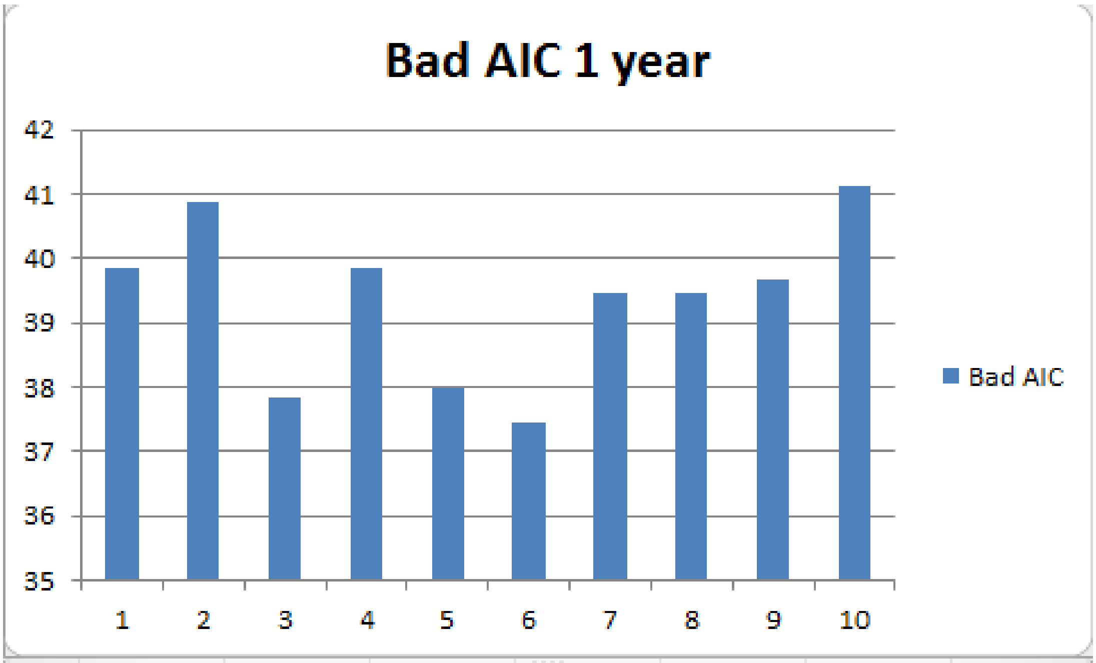


Figure 4.22: bad AIC for 1 year difference

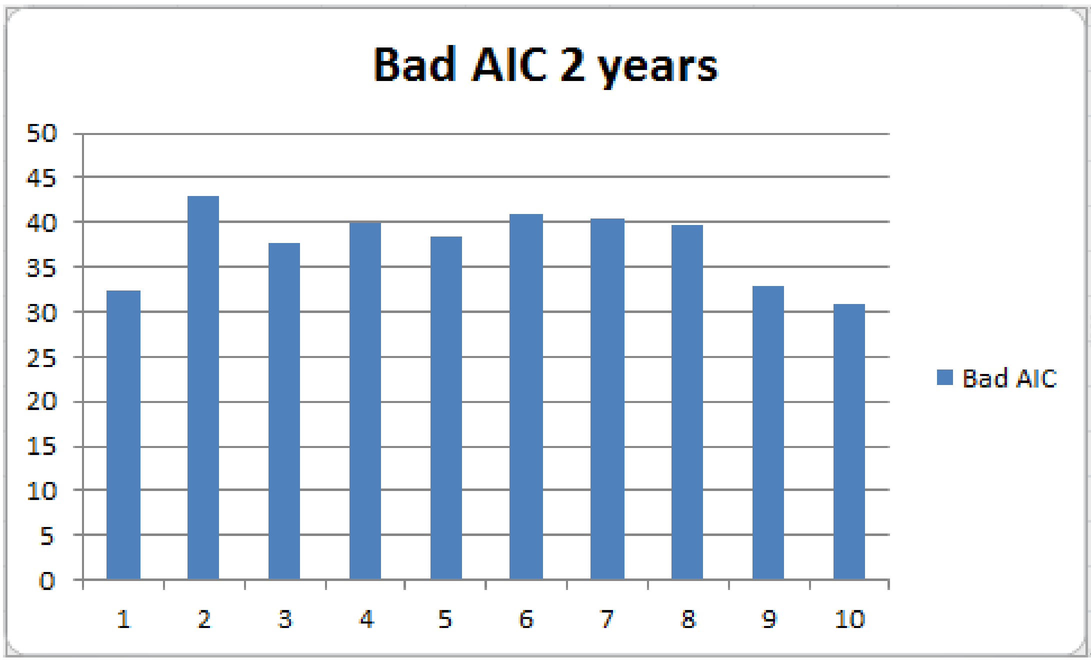


Figure 4.23: bad AIC for 2 year difference

As we may see AIC is very high for all of the experiments. Its interval is between 31 and 43, that means we cannot rely on that methods and use them to predict bankruptcy.

Now represent AIC for the model that was driven after a long research and investigation:

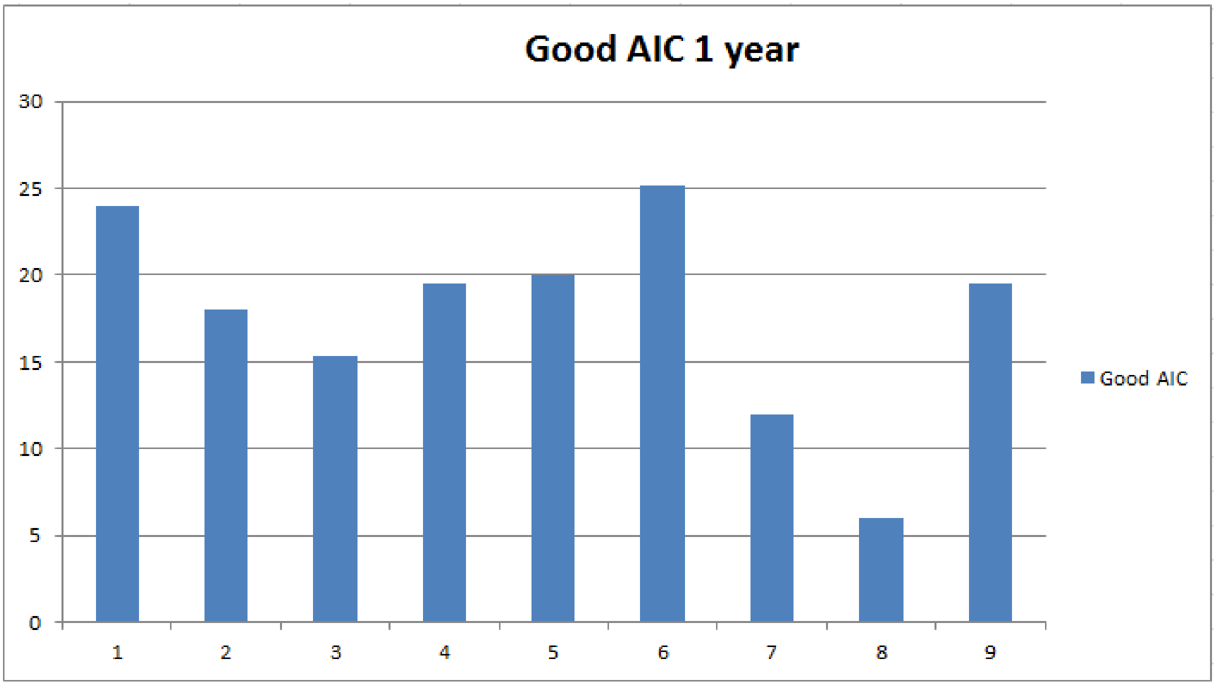


Figure 4.24: AIC for 1 year difference

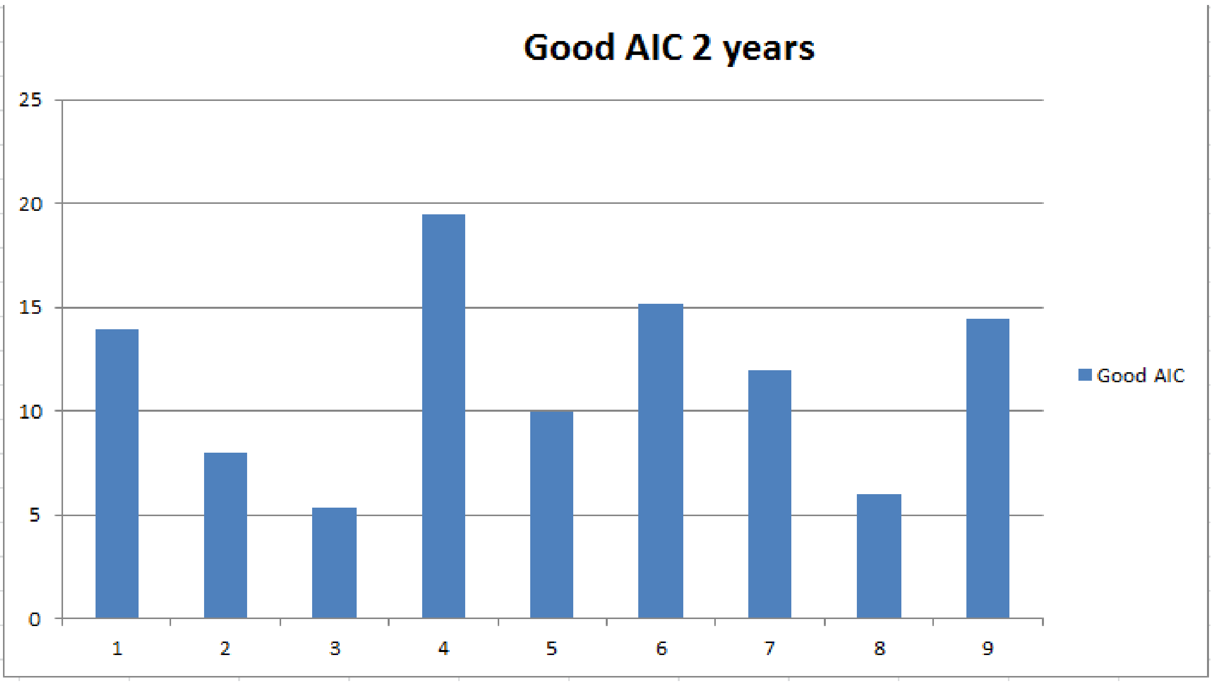


Figure 4.25: AIC for 2 year difference

Improvements are significant, AIC decreased and the smallest value is 5 and the highest is 25. That means we may use one of the represented methods (namely sixth model) as a tool for bankruptcy prediction. Let's represent that model in conclusion.

Chapter 5

Conclusion

As we may see from our graphical results and AIC diagrams, the most effective method was detected for a model where log function was applied to 2 year difference between financial covariates. The smaller AIC the more model fits. The same with Z-values, we need to pick up those whose values are closer to 0. By the end of the study we may tell the following: old models didn't show good results: bankrupt companies were not fully discovered. So time series methods might be used in the future as a new approach for bankruptcy investigating.

To prove this compare theoretical models results with the results we got using newly derived method:

B/NB	ALtman2	Taffler	Ohlson	Zmijevsky	My Method
1	-399,557846	2272,9224	19289,02	490,92355	14,64
1	-921,576544	4915,3192	35478,44	1060,42582	14,37
1	-4974,606634	6308,4809	37701,73	1238,64008	16,68
1	1030,49526	10622,8462	123735,9	3158,41372	15,45
0	-3131,427348	6934,1827	48780,53	1446,16165	22,36
1	-33139,97114	21856,1027	73436,89	3034,43714	9,76
1	-4099,343742	8154,5719	52856,93	1715,50898	5,56
1	-1365,978798	7665,6518	77404,83	1805,92039	16,23
0	-4615,900294	19074,4026	115067,1	3273,91041	20,06
1	-7036,992948	12386,5387	76901,41	2351,00922	17,4
1	8061,25711	33233,4503	208346,1	10291,0913	12,77
1	-4827,158458	23285,4196	156393,5	5238,10724	13,66
0	-1470,108374	3815,9457	27887,62	781,88015	19,03
1	-7619,377968	10135,7232	54250,42	1682,87764	11,93
0	-3986,408682	11164,9073	127358,2	2865,39957	20,53
1	-4839,1564	11710,6483	126405,3	2441,21762	14,49
1	-9372,496778	36418,9932	197119,2	5328,1288	10,05

Figure 5.1: Comparison of theor. results with a new method

Theoretical model results					
Bankrupt	**Altman2**	**Taffler**	**Olson**	Zmijevsky	**My model**
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
Yes	Low	Low	Low	Low	Medium
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
Yes	Low	Low	Low	Low	High
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low
Yes	Low	Low	Low	Low	High
No	Low	Low	Low	Low	Low
Yes	Low	Low	Low	Low	High
No	Low	Low	Low	Low	Low
No	Low	Low	Low	Low	Low

Figure 5.2: Explanation of theor. results with a new method

As the p-values and standard error are quite small and less than 0.05, so neither TOA, TOL, Rev nor NetE are insignificant in the logistic regression model. Also Z variables are appropriate for our research. Scatterplots with covariates are also showing us good results. With a small error we may easily classify and separate stable companies from non-stable just by hand.

So in the following we may present a new model for IT data bankruptcy forecast:

$$Z = -1.15 * X_1 - 1.35 * X_2 + 7.06 * X_3 - 1.65 * X_4 \tag{5.1}$$

where
X_1 = log(REVs(t) - REVs(t-2))
X_2 = log(NI(t) - NI(t-2))
X_3 = log(TA(t) - TA(t-2))
X_4 = log(TL(t) - TL(t-2))

$$Z = \begin{cases} >= 0.5 & \text{bankrupt} \\ < 0.5 & \text{non-bankrupt} \end{cases} \tag{5.2}$$

A future research can be continued to improve and extend this method, to help to predict bankruptcy with a new inspection tool and new technology used. As this work was intended to show there are exist other algorithms and techniques to handle bankruptcy forecasting.

Chapter 6

Appendix A

Table 6.1

Bankruptcy	Revenues	Net Earnings	Total Assets	Total Liabilities
0	7350000	518000	30000	74000
0	10402000	3063000	5519000	1905000
0	6687000	5092000	6907000	9183000
0	1711000	490000	1533000	1784000
0	43125	261717	178182	103191
0	40000	801000	3776000	6247000
0	1122773	63692	1397495	1091461
0	222538	84844	971795	1245775
0	574500	471100	352500	800
0	129100	63400	683200	469300
1	995	2943	8599	1386
1	4535	502	10862	1077
1	8975	11612	17935	8
1	32733	27	37337	14067
1	11041	5295	100126	94853
1	2559	2240	66212	63582
1	21609	3562	28269	14574
1	14558	957	15782	31334
1	24337	3066	48011	29338
1	161181	31143	265571	158328
1	54282	2445	143875	9301
1	23838	3085	255126	65622
1	104234	40890	29620	4886
1	10352	34073	127773	1789
1	145584	40711	989492	155784
1	1060	1900	2551	29

Continued on the next page

Table 6.1 – continued from previous page

Bankruptcy	Revenues	Net Earnings	Total Assets	Total Liabilities
1	9661	118969	150625	57963
1	9117	2711	19008	11421
1	303000	66000	450000	165000
1	6873	14213	22458	13179
1	61000	4247	196778	23425
1	130524	17943	70970	29773
1	159693	36508	351804	63647
1	119521	5585	1135345	1084874
1	23279	4012	472855	437268
1	136783	4011	114624	41499
1	361033	12583	135130	40970
1	64514	6049	160285	85763
1	329032	46746	689529	211433
1	137567	19113	187696	22034
1	314487	203989	2600627	254873
1	737255	51596	674598	246655
1	88743	6927	193565	113107
1	288200	68055	1732352	230519
1	17963	3086	18529	5211
1	37329	121287	204993	81314
1	115385	9015	61724	29784
1	2665000	419000	3066000	1287000
1	6858000	1153000	10536000	5354000
1	156508000	41733000	176064000	57854000
1	127434000	7264000	272315000	179953000
1	3797000	98000	5070000	2803000
1	12435655	1007284	14979848	11258714
1	2318000	43000	14328000	6628000
1	329032	46746	689529	211433
1	110061	2344	1132590	801036
1	14351000	1838000	9243000	5748000
1	7346472	1051263	6521571	1667188
1	836860	54597	2024339	1130117
1	16006000	1560000	89409000	63257000
1	413552	9032	268158	122073
1	970	522	2840	713
1	5063000	491000	6198000	6851000

Continued on the next page

Table 6.1 – continued from previous page

Bankruptcy	Revenues	Net Earnings	Total Assets	Total Liabilities
1	7974	15161	9521	16336
1	235	1384	770	3328
1	32681000	298000	38338000	23606000
1	427145	38148	218832	119843
1	120357000	12650000	108768000	86332000
1	96996	3950	785092	508232
1	7581000	1413000	15133000	2549000
1	8699923	303594	26494407	23470812
1	113626000	5565000	208693000	95854000
1	8234000	1032000	14328000	6628000
1	601000	42900	2406400	1258600
1	169510	10480	384847	218103
0	280820	148703	396299	231745
0	10695652	77111	120260	131547
0	6839005	157806	3732536	2443631
0	830638	10174	392794	300794
0	76090	72240	301140	48260
0	11073000	646000	7101000	3705000
0	39784000	4995000	27526000	27030000
0	72215000	457000	38721000	127481000
0	21386000	2155000	49809000	42530000
0	2417501	352020	3340948	2347596
0	1293000	139000	4535000	2032000
0	13028985	265640	16606335	13507462
0	13998888	241584	5852454	3278394
0	10717100	42500	7382500	6165400
0	5671400	583100	6795000	4077600
1	6873	14213	22458	13179
1	61000	4247	196778	23425
1	130524	17943	70970	29773
1	159693	36508	351804	63647
1	119521	5585	1135345	1084874
1	23279	4012	472855	437268
1	136783	4011	114624	41499
1	361033	12583	135130	40970
1	64514	6049	160285	85763
1	329032	46746	689529	211433

Continued on the next page

Table 6.1 – continued from previous page

Bankruptcy	Revenues	Net Earnings	Total Assets	Total Liabilities
1	137567	19113	187696	22034
1	314487	203989	2600627	254873
1	737255	51596	674598	246655
1	88743	6927	193565	113107
1	288200	68055	1732352	230519
1	17963	3086	18529	5211
1	37329	121287	204993	81314
1	115385	9015	61724	29784
1	2665000	419000	3066000	1287000

Table 6.2: Time series data for unstable companies

Comp. Name	Year	2013	2012	2011
BBRY	Revenues	11,073,000	18,423,000	19,907,000
	Net Earnings	646,000	1,164,000	3,411,000
	Total Assets	7,101,000	7,071,000	7,488,000
	Long Term Borr.	3,705,000	3,631,000	3,937,000
Nok	Revenues	39,784,000	50,186,000	56,944,000
	Net Earnings	4,995,000	1,932,000	1,802,000
	Total Assets	27,526,000	33,045,000	36,417,000
	Long Term Borr.	27,030,000	28,935,000	30,711,000
SNE	Revenues	72,215,000	78,902,000	86,647,000
	Net Earnings	457,000	5,549,000	3,132,000
	Total Assets	38,721,000	45,628,000	46,381,000
	Long Term Borr.	127,481,000	136,664,000	124,805,000
TWC	Revenues	21,386,000	19,675,000	18,868,000
	Net Earnings	2,155,000	1,665,000	1,308,000
	Total Assets	49,809,000	48,276,000	45,822,000
	Long Term Borr.	42,530,000	40,746,000	36,612,000
DYN	Revenues	1,293,000	1,333,000	2,059,000
	Net Earnings	139,000	940,000	242,000
	Total Assets	4,535,000	8,311,000	10,013,000
	Long Term Borr.	2,032,000	8,279,000	7,267,000
CYH	Revenues	13,028,985	11,906,212	11,092,422
	Net Earnings	265,640	201,948	279,983
	Total Assets	16,606,335	15,208,840	14,698,123

Continued on the next page

Table 6.2 – continued from previous page

Comp. Name	Year	2013	2012	2011
	Long Term Borr.	13,507,462	12,416,001	12,121,187
SPWR	Revenues	2,417,501	2,374,376	2,219,230
	Net Earnings	352,020	613,737	178,724
	Total Assets	3,340,948	3,519,130	3,379,331
	Long Term Borr.	2,347,596	2,244,405	1,721,897
SHI	Revenues	13,998,888	14,221,426	10,940,693
	Net Earnings	241,584	156,740	424,053
	Total Assets	5,852,454	4,880,659	4,354,908
	Long Term Borr.	3,278,394	2,032,619	1,670,497
AVP	Revenues	10,717,100	11,291,600	10,862,800
	Net Earnings	42,500	513,600	606,300
	Total Assets	7,382,500	7,735,000	7,873,700
	Long Term Borr.	6,165,400	6,164,600	6,217,200
SNE	Revenues	5,671,400	5,542,300	5,003,900
	Net Earnings	583,100	519,700	558,200
	Total Assets	6,795,000	6,111,800	6,187,800
	Long Term Borr.	4,077,600	3,608,300	3,721,500

Table 6.3: Time series data for stable companies

Comp. Name	Year	2013	2012	2011
YOD	Revenues	6,873	7,868	7,649
	NetEarnings	14,213	11,270	12,904
	TotalAssets	22,458	31,057	30,633
	LongTermBorr	13,179	14,565	14,692
SUPX	Revenues	61,000	65,535	83,172
	NetEarnings	4,247	4,749	12,282
	TotalAssets	196,778	207,640	220,596
	LongTermBorr	23,425	22,348	27,677
HWG	Revenues	130,524	139,499	168,354
	NetEarnings	17,943	6,331	9,880
	TotalAssets	70,970	88,905	85,277
	LongTermBorr	29,773	29,765	19,806
CNIT	Revenues	130,524	139,499	168,354
	NetEarnings	17,943	6,331	9,880
	TotalAssets	70,970	88,905	85,277

Continued on the next page

Table 6.3 – continued from previous page

Comp. Name	Year	2013	2012	2011
	LongTermBorr	29,773	29,765	19,806
AMAP	Revenues	159,693	126,960	85,765
	NetEarnings	36,508	36,535	19,135
	TotalAssets	351,804	314,467	248,623
	LongTermBorr	63,647	49,580	33,678
ARL	Revenues	119,521	108,480	106,505
	NetEarnings	5,585	290	94,747
	TotalAssets	1,135,345	1,235,471	1,557,275
	LongTermBorr	1,084,874	1,179,727	1,496,658
RIVR	Revenues	23,279	20,720	22,536
	NetEarnings	4,012	1,772	2,320
	TotalAssets	472,855	406,643	386,609
	LongTermBorr	437,268	373,686	355,141
KTEC	Revenues	136,783	115,174	116,328
	NetEarnings	4,011	449	1,454
	TotalAssets	114,624	86,355	94,405
	LongTermBorr	41,499	26,925	35,631
KTCC	Revenues	361,033	346,475	253,846
	NetEarnings	12,583	11,626	5,736
	TotalAssets	135,130	150,912	112,364
	LongTermBorr	40,970	72,304	44,341
KNDI	Revenues	64,514	40,177	42,880
	NetEarnings	6,049	9,115	951
	TotalAssets	160,285	112,274	109,615
	LongTermBorr	85,763	56,425	65,141
QIHU	Revenues	329,032	167,851	57,665
	NetEarnings	46,746	15,603	8,508
	TotalAssets	689,529	423,958	87,808
	LongTermBorr	211,433	53,105	15,393
YELP	Revenues	137,567	83,285	47,731
	NetEarnings	19,113	16,668	9,566
	TotalAssets	187,696	43,821	41,015
	LongTermBorr	22,034	12,733	6,658
AKAM	Revenues	314,487	290,649	254,277
	NetEarnings	203,989	200,904	171,220
	TotalAssets	2,600,627	2,345,501	2,352,676
	LongTermBorr	254,873	189,251	175,071

Continued on the next page

Table 6.3 – continued from previous page

Comp. Name	Year	2013	2012	2011
HIMX	Revenues	737,255	633,021	642,692
	NetEarnings	51,596	10,706	33,206
	TotalAssets	674,598	644,978	619,620
	LongTermBorr	246,655	251,541	213,742
STV	Revenues	88,743	99,095	87,123
	NetEarnings	6,927	41,000	33,421
	TotalAssets	193,565	321,338	273,642
	LongTermBorr	113,107	114,896	95,142
YOKU	Revenues	288,200	142,616	58,743
	NetEarnings	68,055	27,344	31,061
	TotalAssets	1,732,352	742,860	332,362
	LongTermBorr	230,519	74,735	42,290
KOOL	Revenues	17,963	19,023	23,400
	NetEarnings	3,086	4,986	2,567
	TotalAssets	18,529	21,080	24,399
	LongTermBorr	5,211	5,182	4,306
SRPT	Revenues	37,329	46,990	29,420
	NetEarnings	121,287	2,318	32,177
	TotalAssets	204,993	54,368	45,976
	LongTermBorr	81,314	23,351	48,793
ZHNE	Revenues	115,385	124,502	129,036
	NetEarnings	9,015	11,726	4,781
	TotalAssets	61,724	80,732	90,111
	LongTermBorr	29,784	41,205	40,696
TDC	Revenues	2,665,000	2,362,000	1,936,000
	NetEarnings	419,000	353,000	301,000
	TotalAssets	3,066,000	2,616,000	1,883,000
	LongTermBorr	1,287,000	1,122,000	694,000

Bibliography

1. Leslie Lamport, *LaTeX: A Document Preparation System.* Addison Wesley, Massachusetts, 2nd Edition, 1994.
2. Boris Podobnik, *Bankruptcy Risk Model and Empirical Test.* Davor Horvatic, Zagreb, Alexander M. Petersen, Boston, Branko Urosevic, Belgrade, Eugene Stanley, Boston 1994.
3. Eivind Bernhardsen, *A model of Bankruptcy Prediction.* Eivind Bernhardsen, Norway, Norges Bank, 1994.
4. A.K.Vasisht, *Logit and Probit analysis.* A.K.Vasisht, New Delhi, 1994.
5. Jodi Bellovary, *A review of Bankruptcy Prediction Studies: 1930 to Present.* Jodi Bellovary, Don Giacomino, Michael Akers, Marquette University, 2007.
6. Ruey S. Tsay, *Analysis of financial time series.* University of Chicago 2005.
7. Lidia Mandru, *The Diagnosis of Bankruptcy Risk Using Score Function.* Adnan Khashman, Claudia Carstea, Nicoleta David, Lucian Patrascu
8. Martins Blums, *D-Score: Bankruptcy Prediction Model for Middle Market Public Firms.* Macalester College 2003.
9. Lam, L.Y., *Empirical Research in Financial Distress Prediction, Chapter 3, Masters Thesis,*, Department of Accountancy, Victoria University of Wellington, pp. 22-51, 1994.
10. Karels, G.V. and Prakash, A.J., *Multivariate Normality and Forecasting of Business Bankruptcy*, Journal of Banking and Finance, Vol 14(4), (Winter 1987), pp. 573-593
11. Foster, G., *Financial Statement Analysis*, Second edition (Prentice-Hall International Editions, 1986), pp. 533-572.
12. Jones, F.L., *Current Techniques in Bankruptcy Prediction*, Journal of Accounting Literature, Vol 6 (1987), pp. 131-164

13. Gilbert, L., Menon, K. and Schwartz, K. *Predicting Bankruptcy for firms in Financial Distress*, Journal of Finance and Accounting, Vol 17(1), (Spring 1990), pp. 161-171

14. Zmijewski, M. Methodological issues relating to the estimation of financial distress prediction models, Journal of Accounting Research (Supplement) (1984), pp. 59-82

15. Jeffrey S. Strickland. *Predictive modelling and analytics*, Simulation Educators (2014), pp. 38, Colorado Springs CO

16. [Online], available: `http://en.wikibooks.org/wiki/R_Programming/Binomial_Models`, Probit and Logit estimation in R, [accessed, 16 December 2014]

17. [Online], available: `http://www.ats.ucla.edu/stat/r/dae/probit.htm`, Probit method in R, [accessed, 16 December 2014]

18. [Online], available: `http://www.ats.ucla.edu/stat/r/dae/mlogit.htm`, Logit method in R, [accessed, 16 December 2014]

19. [Online], available: `http://en.wikipedia.org/wiki/Predictive_analytics`, Predictive analysis, [accessed, 16 December 2014]

20. [Online], available: `http://www.investopedia.com/dictionary/`, Financial definitions, [accessed, 16 December 2014]

21. [Online], available: `https://lsc.cornell.edu/Sidebars/Stats%20Lab%20PDFs/Topic7.pdf`, P value definition, [accessed, 16 December 2014]

22. [Online], available: `http://en.wikipedia.org/wiki/Z-test`, Z value definition, [accessed, 16 December 2014]

23. [Online], available: `http://en.wikipedia.org/wiki/Altman_Z-score`, Altman model definition, [accessed, 16 December 2014]

24. [Online], available: `http://www.annualreports.com/`, [accessed, 16 December 2014]

25. [Online], available: `http://bdp.law.harvard.edu/fellows.cfm`, [accessed, 16 December 2014]

26. [Online], available: `http://lopucki.law.ucla.edu/`, [accessed, 16 December 2014]

27. [Online], available: `http://finance.yahoo.com/`, [accessed, 16 December 2014]