STUDY OF INVESTMENT ATTRACTIVENESS OF RUSSIAN COMPANIES ON THE BASIS OF THEIR MARKET CHARACTERISTICS AND PERFORMANCE REPORTING

Dmitriy Bogachev

***Abstract**: In the paper different well-known market characteristics and indicators of financial accounting (net income, revenue, revenue growth, etc.) are considered. We propose new characteristics, which could be useful for company assessment. We also describe a classification technique based on Naive Bayes method to identify the most attractive companies.*

***Keywords**: classification, Naive Bayes, financial market, stock's returns.*

***ACM Classification Keywords**: I.2 ARTIFICIAL INTELLIGENCE*

Introduction

It is widely known that the utilization of accounting and other performance information of comparing business firms is highly subjective. Investment services provide a wide variety of lists of recommended securities classified into groups, where companies within a given grouping are perceived by the analyst as "similar" with respect to anticipated price appreciation, yield, and risk.

The purpose of this research is a quantitative analysis of similarity between business activity and the utilization of this similarity for grouping related firms.

Algorithm and Data

The classification algorithm used in this article is a Naive Bayes. This algorithm was selected because of its high speed in comparison with other algorithms (support vector machines, trees, nearest neighborhood) [Wang, 2008]. But it supposes the independence of object parameters. We cannot say that the indicators, which will be mentioned hereinafter, are completely independent. In reality they are mutually dependent. Naive Bayes often is not able to give good estimates of probability of correct class. He makes a correct estimate of the class, while the corresponding class is more likely than the others, regardless of how probabilistic assessment corresponds to reality. The classifier is sufficient robust at the ignoring the independents all the [Jensen, 1991].

Data were collected from 102 Russian companies from two industries: Manufacturing and Production of electricity, gas and water. One economic cycle during the period of 1999-2008 was considered. There are 51 companies in each industry divided by the level of total return for shareholders (TRS):

$$TRS = \frac{Price_{end} - Price_{begin} + Div}{Price_{begin}} \quad (1)$$

They are divided into four classes:

1. With high returns (greater than 20%),
2. With average returns (from 10% to 20%),
3. With low returns (from 0% to 10%),
4. With negative returns (less than 0%).

The next parameters were used in classification: Equity, Revenue, Revenue's growth rate, EBIT, Net profit, ROIC, ROE, TIE, Capitalization, P/E, P/S [Aslinger, 2004; Chen, 1998; Varaiya, 1987].

The indicators were considered in all possible combinations. The sum of all possible combinations is 2047 at n=11

$$\sum_{k=1}^{11} \frac{n!}{k!(n-k)!} = 2^n - 1 \quad (2)$$

Firms from each industry were divided into training and test samples. Because the number of companies is not so great the cross-validation procedure with 5 folds was implemented. To analyze the results of the classification the following ratios were introduced: accuracy, sensitivity and specificity. The accuracy of the classifier was defined as the proportion of companies whose classes are predicted correctly:

$$Accuracy = \frac{Number\ of\ companies\ whose\ classes\ were\ predicted\ accurately}{Total\ number\ of\ predictions} \quad (3)$$

Sensitivity is the proportion of underestimated companies. For example, a company with high returns was attributed to a class of low returns:

$$Sensttivity = \frac{Number\ of\ companies\ whose\ classes\ were\ underestimated}{Total\ number\ of\ predictions} \quad (4)$$

Specificity is the proportion of overestimated companies. For example, a company with negative returns was attributed to the class of average returns:

$$Specificity = \frac{Number\ of\ companies\ whose\ classes\ were\ overestimated}{Total\ number\ of\ predictions} \quad (5)$$

Sensitivity and specificity are analogues to error of the first kind and error of the second kind respectively.

The average difference between the true and predicted classes was calculated for all predictions as well. Obviously, the best combinations would be those having the greatest accuracy. With the equal precision you should choose a combination of parameters with greater sensitivity, and lesser specificity. The sum of accuracy, sensitivity and specificity is equal 1. It is better to find high profitable company, to consider it unattractive and not to invest in it, than to invest in a company with negative returns.

Results

3.1 Manufacturing industry

There are 51 companies in the manufacturing sector. They were ordered by TRS. This is reflected in the Figure 1:

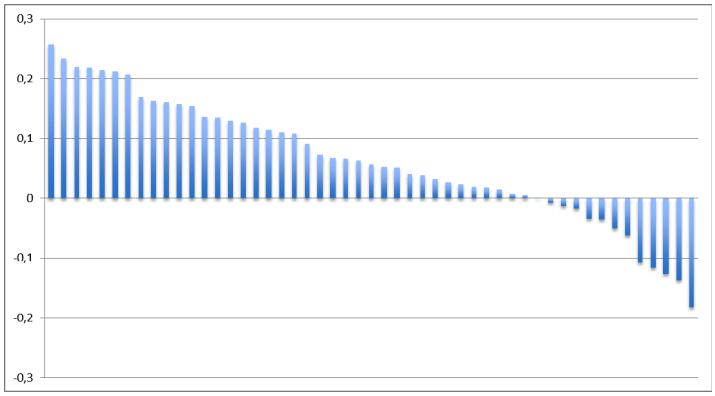

Fig. 1. Returns of companies in the industry of manufacturing

Below the five best combinations of accuracy indicators are given (Table 1).

Table 1. The results of the classification of companies in the industry of manufacturing

Parameters	Accuracy, proportion	Sensitivity, proportion	Specificity, proportion	Average distortion	F-measure
Revenue, EBIT, Net profit, ROE, Capitalization	0,471	0,275	0,255	0,020	0,546
Revenue, Net profit, ROE, Capitalization	0,471	0,255	0,275	-0,059	0,536
Equity, Capitalization	0,471	0,137	0,392	-0,451	0,410
Equity, Revenue, Revenue's growth rate, EBIT, Net profit, ROE, Capitalization	0,451	0,392	0,157	0,490	0,482
Equity, Revenue, EBIT, Net profit, ROE, Capitalization	0,451	0,275	0,275	0,039	0,523

The table shows that some combinations of parameters allow you to achieve accuracy in almost 50%. For example, a set of Revenue, EBIT, Net profit, ROE and Capitalization or Revenue, Net profit, ROE and Capitalization are equally accurate in 47.1%, proportions of the sensitivity and specificity are almost equal, and in the these sets do not distort the value of the class.

3.2 Industry of production of electricity, gas and water

Companies from the sectors of production and distribution of electricity, gas and water have also been ordered by TRS:

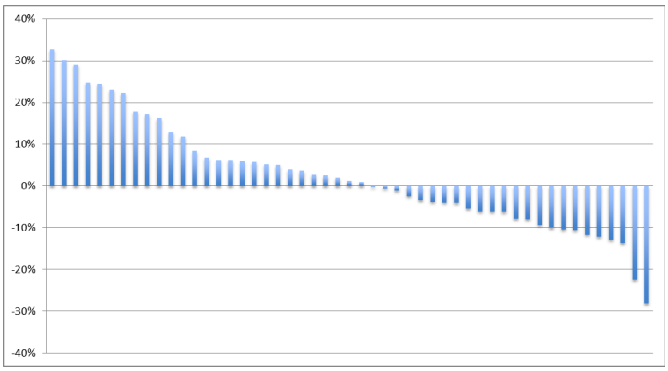

Fig. 2. Returns of companies in the industry of production of electricity, gas and water

Further classification was carried out similarly to the previous field. The table below shows the best five 5 sets of the accuracy.

The table shows that the maximum accuracy in the industry is slightly lower and is 43.1%. It is achieved in sets of Revenue, ROIC, P/S and Revenue, ROIC, TIE. There is low sensitivity (13.7% and 11.8% respectively) and high specificity (43.1% and 45.1% respectively) in these sets.

Table 2. The results of the classification of companies in the industry of production electricity, gas and water

Parameters	Accuracy, proportion	Sensitivity, proportion	Specificity, proportion	Average distortion	F-measure
Revenue, ROIC, P/S	0,431	0,137	0,431	-0,608	0,422
Revenue, ROIC, TIE	0,431	0,118	0,451	-0,529	0,378
EBIT, Net profit	0,412	0,235	0,353	-0,294	0,455
Equity, Revenue, Revenue's growth rate, ROIC, Capitalization, P/E, P/S	0,412	0,196	0,392	-0,294	0,324
EBIT, ROIC, P/S	0,412	0,157	0,431	-0,569	0,308

Conclusion

The purpose of this study is to propose a technique of finding companies with high returns for shareholders over the long term. This technique is based on splitting companies into classes of similarity of parameters. There is the permanent interest from investors and the virtual absence of research to this topic.

The technique involves the several stages:

a) The rationale for the initial parameters of the sample is provided;

b) The rationale for the choice of parameters is provided;

c) The companies are divided into groups by level of returns;

d) A comparison of predicted and a priori selected classes of companies is made. The sets of parameters are found, which ensure the greatest accuracy in predicting.

Empirical testing of the hypothesis showed that high returns companies could be identified not openly, but with the help of known values of parameters and returns for other companies.

This conclusion is valid for Russian companies in industries of manufacturing and production of electricity, gas and water in the period 1999-2008, including one economic cycle.

Bibliography

[Aslinger, 2004] P. Aslinger, T. Breene. Redefining high performance. Journal of appled corporate finance, Vol.16, No.1, Winter 2004

[Chen, 1998] N. Chen and F. Zhang. Risk and Return of Value Stocks. The Journal of Business, Vol. 71, No. 4 (October 1998), pp. 501-535

[Jensen, 1991] R.E. Jensen. A Cluster Analysis Study of Financial Performance of Selected Business Firms. The Accounting Review, Vol. 46, No. 1 (Jan., 1991), pp. 36-56

[Varaiya, 1987] N. Varaiya, R. Kerin, D. Weeks. The relationship between growth, profitability and firm value. Strategic Management Journal, 1987, 8

[Wang, 2008] Y.J. Wang, H.S. Lee. A clustering method to identify representative financial ratios. Journal: Information Sciences, 2008, 178, pp. 1087–1097

Authors' Information

Dmitriy Bogachev *– Bachelor, The Russian Presidential Academy of National Economy and Public Administration, Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; Moscow Institute of Physics and Technology (State University), Institutslii per. 9, Dolgoprudny, Moscow Region, 141700, Russia, e-mail: myyy48@gmail.com*

Major Fields of Scientific Research: data mining, stock markets