Advances in Business-Related Scientific Research Conference 2013 in Rome
October 2-4, 2013, Rome, Italy

CLASSIFICATION OF BORSA ISTANBUL FIRMS BASED ON MARKET PERFORMANCE DATA: COMPARISON OF CLASSIFICATION AND REGRESSION TREES

Şebnem ER
Nuray TEZCAN

Agenda

- Introduction-Purpose
- Literature Review
- Methodology
- Data
- Results
- Conclusion

Introduction-Purpose

This study aims to classify Borsa Istanbul firms according to their return levels with one of the tree-based approaches known as Classification and Regression Trees (C&RT) using market performance data such as price to earnings ratio, market to book value ratio, risk measure of beta as well as firm level performance data such as debt ratio and profitability ratios. in order to produce an accurate classifier and to understand what variables or interactions of variables drive to that classification

Literature Review

- EKLENECEK

Classification

Classification methods differ from each other based on the fact that the classes are predefined or not.

if the classes are not predefined

cluster analysis

if the classes are known in advance,

discriminant analysis or a nonparametric alternative of tree-based approaches.

Methodology:Tree-based Approaches

Nonparametric approaches that do not require data distribution specification like normality of the explanatory variables

They can be used for both classification and regression.

Variables that are numerical and categorical can be analysed together.

They offer a visual representation of the classification structure.

The final results of tree-based approaches are summarized in a logical if-then format.

Tree-Based Approaches

- Chi-squared automatic interaction detection (CHAID),
- Random forests and boosted trees,
- Artificial neural networks and
- Support vector machines.

Here we focus on classification and regression trees (C&RT) method and compare the results of classification and regression trees.

The main difference of a C&RT algorithm is that it is a binary splitting algorithm. By "binary" we mean that the algorithm splits the tree into only two branches

C&RT

C&RT algorithm is a form of a decision tree that can be used for either classification or regression estimation.

Classification trees are mainly used when the predicted outcome variable is categorical and regression trees are used when it is numerical.

Our predicted outcome is the return levels of Borsa Istanbul firms. Firstly, we classified the firms as those with negative returns and as those with positive returns. Since in this case the predicted outcome variable is categorical having two categories (positive-negative return), we applied classification trees to investigate which of the market performance rates play a crucial role in this distinction. Secondly, we used the numerically measured return levels of the firms and applied regression trees since in this case the predicted outcome is a numerical variable.

Data

- The data set includes 306 firms that are quoted in Borsa Istanbul in 2012.
- Data has been collected from Finnet commercial website that collects and arranges firm level data (Finnet URL).
- We have excluded the outliers from our dataset.
- All of the analysis is applied using **rpart** R package

Distribution of the firms according to the sectors

Name of the Sector	Number of Firms	Percentages
Mining	2	0.7%
Construction and Public Works	3	1.0%
Electricity, Gas and Water	4	1.3%
Education, Health, Sports And Other Social Services	4	1.3%
Transportation, Telecommunication And Storage	6	2.0%
Technology	11	3.6%
Wholesale And Retail Trade, Hotels And Restaurants	20	6.5%
Financial Institutions	99	32.4%
Manufacturing Industry	157	51.3%

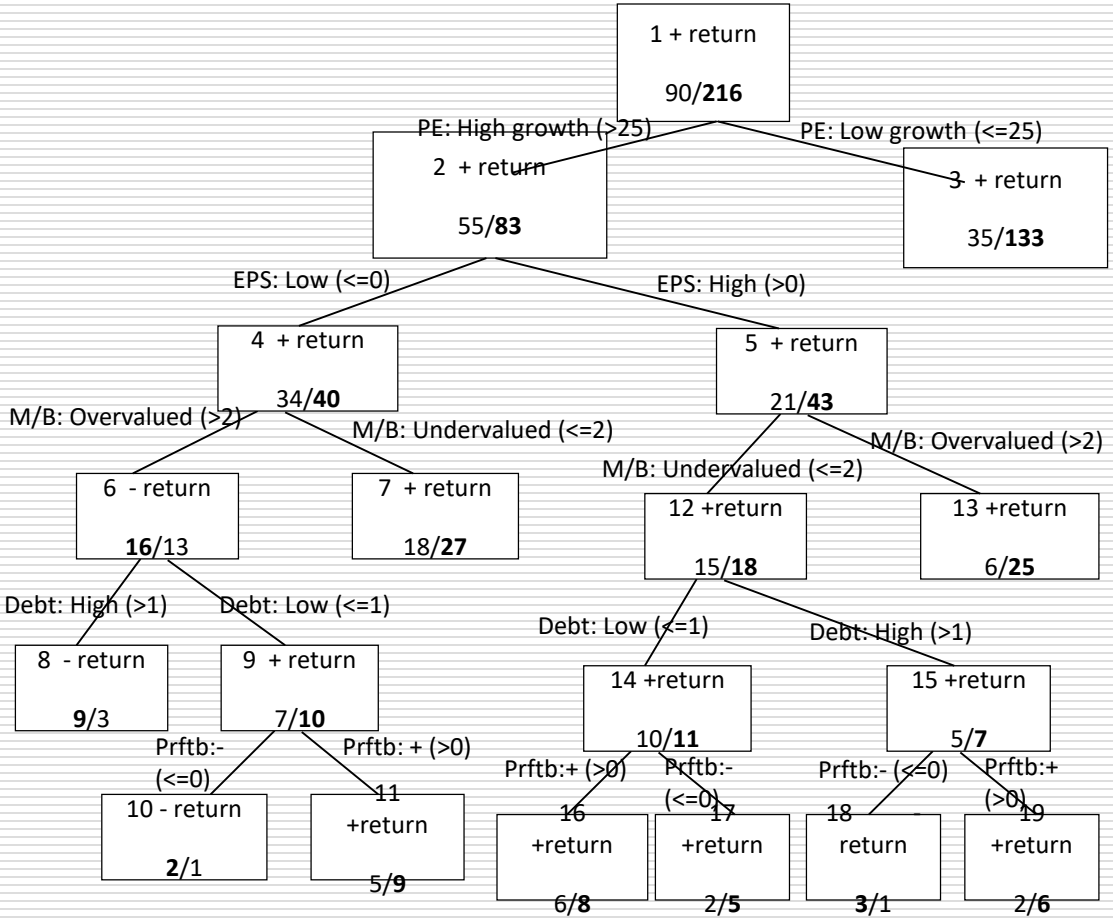
List of Variables Used

Abbreviation	Variables
return	Return
Beta	Beta
M/B	Market to book value
PE	Price to earnings ratio
EPS	Earnings per share
Debt	Debt ratio
Prftb	Profitability

Recoding of the variables used in the analysis

Variable		Categories	N
Return	>=0	(positive return)	216
	<0	(negative return)	90
Beta	>1	(risky)	15
	<=1	(less risky)	291
Market to book value	>2	(overvalued)	100
	<=2	(undervalued)	206
Price to earnings ratio	>25	(high growth)	138
	<=25	(low growth)	168
Earnings per share	>0	(high profit)	232
	<=0	(low profit)	74
Debt ratio	>1	(high debt)	136
	<=1	(low debt)	170
Profitability	>=0	(positive profit)	232
	<0	(negative profit)	74

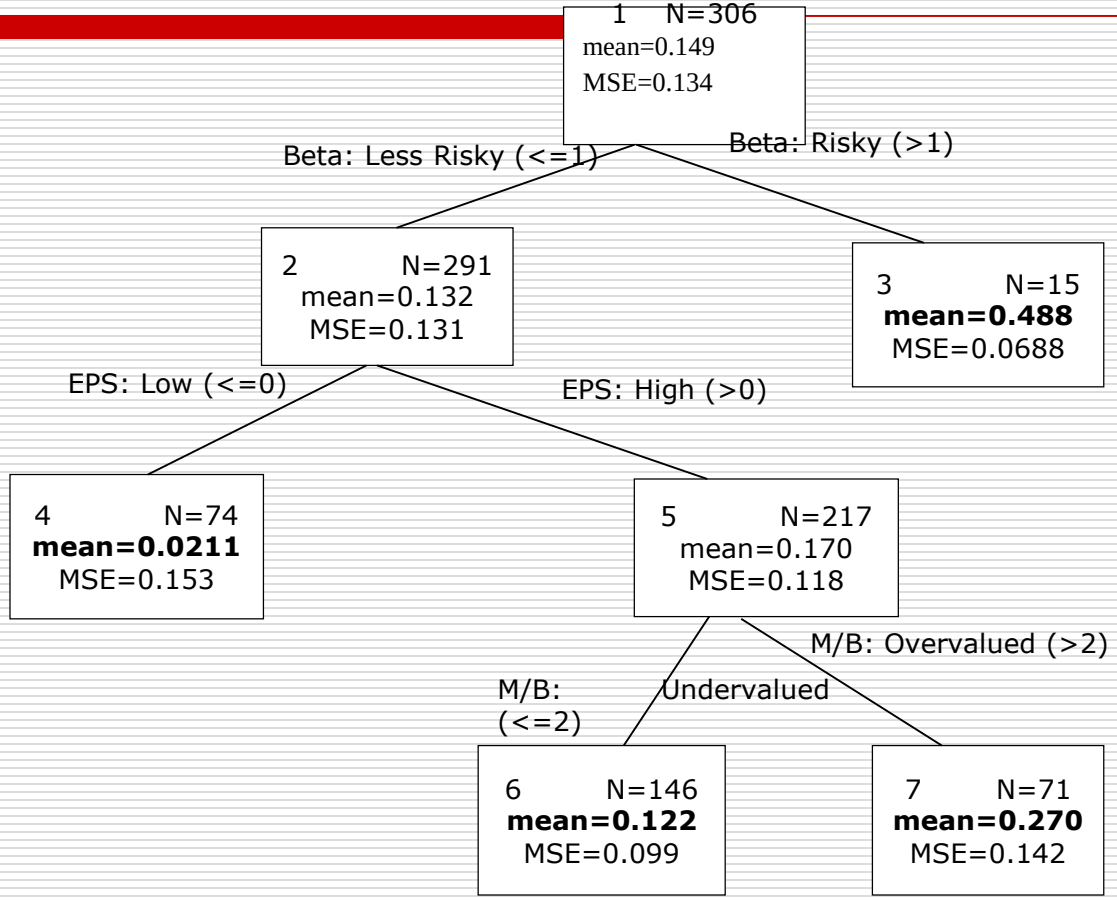
Classification Tree Result



Correct Classification Table with Classification Trees

		Observed Categories		
		+ Return	- Return	Total
Predicted Categories	+Return	**213**	74	287
	- Return	5	**14**	19
	Total	218	88	306

Regression Tree Result



Concluding Remarks

Both of the trees obtained from each of the algorithms were tested against overfitting problems using the cross validation cost rule. In around 20% of the simulations the classification algorithm provided nine splits whereas around 80% of the simulations suggested zero split. This indicated us that the results obtained with Classification Trees are not consistent. Moreover, the misclassification rate with Classification Trees for the negative return category is very high with a rate of 84% (=74/88).

Concluding Remarks

- On the other hand, Regression Tree algorithm provides more consistent results. Regression trees suggest that price to earnings and beta values play a crucial role. The most important decision rule that can be extracted from the regression tree is that if a stock's beta value is greater than one indicating that it is a more risky stock, then the average return level in this case is significantly different than all the other averages provided with the rest of the terminal nodes. It is also found that those companies that are less risky with high EPS values have higher average returns compared to companies with low EPS values. Moreover, if a company is less risky but has a high EPS value, then the average return for those undervalued companies will be lower than the overvalued ones. However, these findings should be interpreted with caution since terminal nodes provided with regression trees have high variation.

Concluding Remarks

- In summary, classification of Borsa Istanbul companies with classification trees did not provide satisfying results, however the regression trees were better in terms of pruning and overfitting problems. The analysis can be extended using the more advanced classification algorithms such as random forests, neural networks and support vector machines to find a more reliable classification method.

- Thanks for your kind attendance…

Dr. Nuray TEZCAN

nuraytezcan@hotmail.com