A novel approach to exploring company’s financial soundness: Investor’s perspective

Zivile Kalsyte, Antanas Verikas

1. Introduction

Financial distress or bankruptcy is the most often considered valuation when assessing company’s financial soundness. Various criteria are used to attribute a company into a financially distressed or non-distressed group. For example the Chinese Stock Exchange considers a company as being financially distressed if the company has had negative net profit in two consecutive years (Sun and Li, 2011; Sun et al., 2011).

Bankruptcy prediction is usually considered as a two class classification task. A variety of techniques have been applied to predict bankruptcy or financial distress. Multilayer perceptron, self-organizing maps, support vector machines, probabilistic neural networks, decision trees, Bayesian networks, fuzzy decision trees, case-based reasoning, fuzzy logic, random forests (RF), rough sets, hybrid systems, ensembles of predictors comprise a list of the most popular techniques applied (Cao, 2012; Kumar and Ravi, 2007; Rada, 2008; Vellido et al., 1999; Wong et al., 2000). There is a large number of examples demonstrating that hybrid and ensemble-based systems, when properly designed, outperform a one predictor based system designed for solving a prediction task. A recent review on hybrid and ensemble-based soft computing techniques in bankruptcy prediction can be found in Verikas et al. (2010).

However, considering companies financial soundness as a two-class classification task we do not get much information to characterize company’s future performance even if the company is deemed as financial sound. There are several studies indicating that company’s future performance prediction accuracy can be increased substantially by including non-financial features into the modeling process, for example macroeconomic indicators and qualitative variables, in addition to financial ratios (Kumar and Ravi, 2007; Rada, 2008; Verikas et al., 2010). Commercial and academic risk measures for detecting and predicting accounting irregularities (Price et al., 2011) make one more group of important indicators. For example, Commercial Accounting and Governance Risk (AGR) measure computed by Audit Integrity aims at detecting suspicious patterns in accounting.

This paper presents a novel technique for exploring company’s financial soundness based on analysis of a large number of...
accounting-based factors, market-based factors, and factors reflecting accounting irregularities. Bearing in mind investor's perspective, we formulate the following main objectives for this work.

1. To develop a technique for predicting company's stage in its life cycle for a near future. Many researchers argue that not only ability to remain profitable and to avoid financial distress is important for the survival of a company. The survival is also very dependent on company's ability to recognize its stage in the life cycle and make decisions relevant for that stage (Dickinson, 2011; Gupta and Chin, 1993; Whitaker, 1999).

2. Using accounting-based factors, market-based factors, and factors reflecting accounting irregularities, to create an ordered 2D map allowing to explore companies' financial soundness. Companies rated by known rating agencies, like Moody's will serve as landmarks of the map. Moody's ratings reflect the likelihood of default and financial loss in the case of default. Such maps can be generated once a quarter, for example, and used to reveal the direction of company's financial soundness.

3. To develop models for predicting trends of main valuation attributes usually used by investors. The models should consider a short prediction horizon from the situation reflected in the 2D map.

2. The data

The data used in this study concern health care industry of the United States. Below given is a list of industry sectors used in the study. There more sectors in the health care industry, however, several sectors, having to few observations, were excluded from the analysis.

i. Diagnostic Substances (DS).
ii. Drug Manufactures, Major (DMM).
iii. Drug Manufactures, Other (DMO).
iv. Drug Generic (DG).
v. Drug Delivery (DD).
vi. Health Care Plans (HCP).
vii. Hospitals (H).
viii. Medical Instruments and Supplies (MiAS).
ix. Medical Laboratories and Research (MiAR).
x. Medical Appliances and Equipment (MAaE).
xi. Specialized Health Services (SHS).
xii. Biotechnology (B).

There are 198 data vectors in total available for modelling companies of the health care industry from the period 2007–2012. The US housing bubble resulted into frozen credit markets and liquidity problems in the United States banking system during 2008–2009. Numerous failures of businesses and financial institutions as well as downturns in stock markets were observed. Companies from the health care industry were also affected by these conditions in different ways.

2.1. Input features

Each company was characterized from five different perspectives with a variety of variables, namely:

i. Balance. There are 175 variables characterizing each company from the balance perspective, \( x_{ij} \), \( j = 1, 2, \ldots, 175 \). The variable set is made of results of vertical analysis (see iii) of all company's balance sheet items with respect to Total current liabilities, Total current assets, Total liabilities, and Total assets.

ii. Income. There are 125 variables in the Income subset, \( x_{ij} \), \( j = 1, 2, \ldots, 125 \). The subset is obtained by performing vertical analysis of all items of company's profit/losses statement with respect to Gross Profit, Operating income, Pretax Income, Income after tax, and Net income.

iii. Cash flow. All items from company's cash flow report were included into this subset of variables, 23 variables in total, \( x_{ij} \), \( j = 1, 2, \ldots, 23 \). A much larger set of cash flow variables, \( x_{5j} \), was constructed when performing companies' life cycle studies. Six indicators from company's cash flow statements, namely: Net Income (NI), Net Cash from Operating Activities (NC\(\alpha\)), Net Cash from Investing Activities (NC\(\alpha\)), Net Cash from Financing Activities (NC\(\alpha\)), Cash at the beginning of period (C\(\alpha\)), and Cash at the end of period (C\(\alpha\)) together with results of horizontal and vertical analysis of the indicators were used. The indicators and the analysis results were registered at 7 different time moments: the end of 2007, 2008, 2009, 2010, and the end of last three quarters. Horizontal analysis is done by subtracting indicator values registered at previous time moment from corresponding current indicator values. Vertical analysis is given by the following 5 ratios: NI/(C\(\alpha\)), NC\(\alpha\)/(C\(\alpha\)), NC\(\alpha\)/(C\(\alpha\)), NC\(\alpha\)/(C\(\alpha\)), and C\(\alpha\)/C\(\alpha\). Since we have 6 indicators and 7 time moments, the cash flow variable set for life cycle studies consists of 239 variables: 162 from the horizontal analysis, 35 from the vertical analysis, and 6 \( \times \) 7 indicator values.

iv. Stock price. Four parameters of the \( \alpha \)-stable distribution, \( \chi \)-characteristic exponent, \( \beta \)-skewness, \( \gamma \)-scale, and \( \delta \)-location are used to characterize stock prices. For each company, five distributions: Open, High, Low, and Close price and Volume are computed separately for each of 2007–2011 years. Thus, we have 4 \( \times \) 5 \( \times \) 5 \( \times \) 5 \( \times \) 100 variables. We also use values of Open, High, Low, Close, and Volume at the end of period and probabilities of getting these values. The probabilities are estimated using the \( \alpha \)-stable distributions of Open, High, Low, Close, and Volume variables using data of all available companies. Thus, we have 110 stock price related variables, \( x_{ij} \), \( j = 1, 2, \ldots, 110 \), in total.

v. Risk indicators. There are 41 variables, \( x_{ij} \), \( j = 1, 2, \ldots, 41 \), in this variable subset, where 22 variables are those used in academic risk assessment models (Price et al., 2011) and the remaining 19 are variables used by Audit Integrity to compute the AGR measure.

Two sets of variables are used from the balance, income and cash flow perspectives. The first set is given by historical mean values of the variables, to eliminate seasonal and various random fluctuations. The second set is composed of differences between current and historical mean values. The total number of variables in all five perspectives is large. However, when building models, we use variables of only perspective at a time. Moreover, variable selection applied usually results into a rather small set of relevant variables.

2.2. Attributes characterizing company's performance

To get a comprehensive characterization of company's financial soundness, we predict eight groups of attributes listed below.

i. Valuation, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 9 \);
ii. Dividends, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 6 \);
iii. Growth rates, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 5 \);
iv. Financial strength, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 5 \);
v. Profitability, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 6 \);
vi. Management effectiveness, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 6 \);
vii. Efficiency, \( y^{\text{vr}}_{ij} \), \( j = 1, 2, \ldots, 5 \);
viii. AGR measure, \( y^{\text{vr}}_{j} \); \( j = 1 \);
There are 43 attributes in total. In addition to these attributes, we also predict company’s life cycle given by one of five possible stages.

3. The approach

3.1. Life cycle stage prediction

We consider the task of predicting company’s stage in its life cycle as a classification problem. Dickinson suggests using five life cycle stages: Introduction, Growth, Mature, Shake-Out, and Decline (Dickinson, 2011). To determine a stage, accounting information is classified as a classification problem. Thus, eight combinations are possible, which are grouped into five stages, as shown in Table 1. We use the same approach to definition of life cycle stages in this work.

To predict stage of company’s life cycle, we create a random forest (Section 4.1) with five aforementioned output classes/stages. A random forest, a general data mining tool proposed by Breiman (2001), is used by the forest contribute to the proximity values.

To predict stage of company’s life cycle, we create a random forest with five output classes/stages. A random forest, a general data mining tool proposed by Breiman (2001), is used by the forest contribute to the proximity values.

3.2. A 2D map for exploring companies financial soundness

Apart from a predictor, the random forest software also produces a data proximity matrix II. To obtain the matrix, for each tree grown, the data are run down the tree. If two observations xi and xj occupy the same terminal node of the tree, p(ij) is increased by one. When a random forest is grown, the proximities are divided by the number of trees in the random forest. Data proximities possess an important property meaning that only variables used by the forest contribute to the proximity values.

We used data proximity matrices for exploration of financial data by mapping them onto the 2D space. The t-distributed stochastic neighbor embedding (t-SNE) algorithm (Borg and Groenen, 1997) was used to perform the mapping in this work.

The data exploration procedure is encapsulated in the following steps.

i. Train an RF using input variables of one perspective to predict a financial attribute yj. Apply feature selection, see Section 4.3, retrain the RF, and save the proximity matrix. ii. Do Step i for all five perspectives, (see Section 2.1) and all financial attributes (see Section 2.2); iii. For each proximity matrix compute the average silhouette value S(i) for rated companies. Assume that companies having the same Moody’s rating come from the same cluster. The silhouette value S(i) of a data point i measures how similar the point i is to points in its own cluster versus points in other clusters. S(i) is given by:

\[ S(i) = \frac{1}{K} \sum_{j=1}^{K} \frac{b(i) - a(i)}{\max[b(i), a(i)]} \] (1)

where K is the number of data points, a(i) is the average dissimilarity (the reciprocal of proximity) from the data point i to all points within its own cluster, and b(i) is the lowest average dissimilarity from i to any other cluster of which i is not a member.

iv. Eliminate the proximity matrices exhibiting negative average silhouette values and average the remaining ones.
v. Map the average proximity matrix onto the 2D space using the t-SNE algorithm. We expect companies with the same Moody’s rating to be placed close together on the map.

3.3. Models to predict trends of main valuation attributes

The models consider a short prediction horizon from the situation reflected in the 2D map, discussed in the previous section. We distinguish three classes of trends, namely increase, decrease, and unchanged and consider six valuation attributes: P/E Ration, Price to Sales, Price to Book, Price to Tangible Book, Price to Cash Flow, and Price to Free Cash Flow. These valuation attributes are the main concern of investors.

We use RF to predict the trends and consider the task as a classification problem. Three models, revealing consequences of different actions of company’s management team, are built. Each model consists of six sub-models (classifiers), one for each valuation attribute. Each sub-model has three outputs, reflecting trends of the attribute: positive (+), unchanged (0), or negative (−).

The three models differ in input information provided to the models. The following procedure is applied to create the models.

1. Predict each of the 43 financial attributes (see Section 2.2) by each of five RF created using input variables of one (out of the five available) perspectives.
2. For each financial attribute, aggregate predictions of the five RF into a committee prediction by weighted averaging, using weights proportional to the prediction accuracy estimated on OOB (see Section 4.1) data.
3. Use variables \( y^i \), \( j = 1, 2, \ldots, 6 \) and \( y^f \), \( j = 1, 2, \ldots, 5 \), predicted in Step 2, together with \( x^i \), \( j = 1, 2, \ldots, 23 \) as input feature set to create Model 1. Variables \( y^p \) and \( y^p \), thus the model as well, reflect actions of company’s management team concerning dividends paid to shareholders.
4. Use variables \( y^m \), \( j = 1, 2, \ldots, 6 \) and \( y^m \), \( j = 1, 2, \ldots, 5 \), predicted in Step 2, together with \( x^f \), \( j = 1, 2, \ldots, 23 \) as input feature set to create Model 2. Variables \( y^m \) and \( y^m \) reflect how well the company is able to exploit opportunities occurring in the industry sector and how effective is the management of company’s internal resources.
5. Use variables \( y^j \), \( j = 1, 2, \ldots, 5 \) and \( y^j \), \( j = 1, 2, \ldots, 6 \), predicted in Step 2, together with \( x^G \), \( j = 1, 2, \ldots, 23 \) as input feature set to create Model 3. Variables \( y^j \) and \( y^j \) reflect company’s management team choices concerning the tradeoff between high profit and company’s growth.

4. Basic components

4.1. The basic model

Random forest, a general data mining tool proposed by Breiman (2001), is used as a basic predictor in this work. Random forest is a committee of decision trees, i.e. CART (Breiman et al., 1993), see Fig. 1.

As the number of trees in RF increases, the test set error rates converge to a limit, meaning that there is no over-fitting in large RFs (Breiman, 2001). Low bias and low correlation are essential for accuracy. To get low bias, trees are grown to maximum depth. To achieve low correlation, randomization is applied:

<table>
<thead>
<tr>
<th>Cash/cycle</th>
<th>Introduction</th>
<th>Growth</th>
<th>Mature</th>
<th>Shake-Out</th>
<th>Decline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+/−</td>
<td>−</td>
</tr>
<tr>
<td>Investing</td>
<td>+/−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+/−</td>
</tr>
<tr>
<td>Financing</td>
<td>+</td>
<td>+</td>
<td>+/−</td>
<td>+/−</td>
<td>+/−</td>
</tr>
</tbody>
</table>
i. Each tree of RF is grown on a bootstrap sample of the training set.

ii. When growing a tree, at each node, n variables are randomly selected out of the N available. At each node, only one variable, providing the best split, is used out of the n selected.

In RF, n is the only parameter to be selected experimentally. For a tree grown on bootstrap data, the out-of-bag (OOB) data can be used as a test set for that tree. As the number of trees increases, RF provides an OOB data-based unbiased estimate of the test set error. OOB data are also used to estimate importance of variables. Random forests have been successfully applied in a variety of fields (Coussement and Van den Poel, 2009; Verikas et al., 2011; Xie et al., 2009).

4.2. t-Distributed stochastic neighbor embedding

t-distributed stochastic neighbor embedding has shown an excellent performance on both real world and artificial data (van der Maaten and Hinton, 2008). t-SNE represents similarities of data points as conditional probabilities. The similarity of \( x_i \) to \( x_j \) in a high-dimensional space is defined as the conditional probability \( p_{ij} \) that \( x_j \) is picked by \( x_i \) as its neighbour when neighbours are picked in proportion to their probability density defined by a Gaussian centered on \( x_i \) (van der Maaten and Hinton, 2008):

\[
p_{ij} = \frac{\exp(-||x_i - x_j||^2/2\sigma^2)}{\sum_{k \neq i} \exp(-||x_i - x_k||^2/2\sigma^2)}
\]

where \( \sigma \) is a parameter and the values of \( p_{ij} \) are set to zero. A similar conditional probability \( q_{ij} \) is computed for low-dimensional counterparts \( y_i \) and \( y_j \) of \( x_i \) and \( x_j \). The joint probability \( q_{ij} = q_{ji} \) is defined as:

\[
q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq j} (1 + \|y_k - y_j\|^2)^{-1}}
\]

where \( q_{ij} \) are also set to zero.

The desired mapping is found by minimizing a Kullback-Leibler divergence between the joint probability distributions \( P \) and \( Q \).

4.3. Variable selection

Variable importance estimates obtained from the RF software are sensitive, to some extent, to the number of randomly selected variables \( n \) used to split a tree node. Therefore, variable importance estimates were done for several different \( n \) values and the results were averaged. Seven different \( n \) values around \( n = \sqrt{N} \) were used. Variable selection was done by applying sequential backward variable elimination based on the average variable importance estimate. At each step, 5% of least important variables were eliminated. The elimination process was continued until we were left with only one variable. The final variable set was that providing the best performance on the OOB data.

5. Experimental investigations

The available data set was split into the learning set \( S_L \) and the test set \( S_T \). The set \( S_L \) was used to train the models and the set \( S_T \) to estimate the prediction/classification accuracy. In all tests involving estimation of the prediction/classification accuracy, we run experiments 10 times using different splits of the data set into learning and test subsets. The results presented here are average values calculated from such 10 runs. To assess model prediction accuracy, we used the average relative (relative to the standard deviation of the target) mean squared prediction error (MSE/STD).

5.1. Life cycle stage prediction

Life cycle stage prediction allows determining if the company will transit to a next life cycle stage in the very near future. The Introduction, Growth, Mature, Shake-Out, and Decline life cycle stages were represented with 9.1%, 4.5%, 6.6%, 20.7%, and 59.1% of all companies used in the experiment. Thus, the classes are rather unbalanced in size. The applied variable selection procedure drastically reduced the initial set of 239 available variables and we were left with 38 selected variables to design a final random forest for life cycle stage prediction. Fig. 2 displays importance of the selected input variables. Most of the selected variables are related to net cash flow from operating activities. The most important variables concern years 2010 and 2011. This is probably due to the fact that the financial crises had a much less impact on cash flow in 2010 and 2011, compared to 2008 and 2009.

The average life cycle stage prediction error for the test data was equal to 27.3%. Bearing in mind fuzziness of the transition moment, the obtained prediction accuracy is rather encouraging. Table 2 presents a confusion matrix for the test data providing deeper insights into the prediction errors. Numbers given in Table 2 are percents of a given class data assigned to the considered class. As can be seen from Table 2, most of the errors are made when distinguishing between neighbouring life cycle stages, except for the Decline class, where 21.3% of the Decline class data were assigned to the Introduction class. This is probably due to the fact that companies in both Decline and Introduction stages generate little cash flow from main activities. Focus of Introduction stage companies is on product development, while the main concern of Decline stage companies is debts.
5.2. Proximity maps

About one third of proximity matrices, generated by RF to predict 43 financial attributes using input data, split into five different perspectives, were eliminated based on negative silhouette values. However, all the five perspectives of input variables were represented amongst the remaining proximity matrices. *Stock price* and *balance* were the most important perspectives according to the study results. Financial attributes reflecting dividend yield, and min and max stock prices were amongst the least important according to the silhouette criterion.

To obtain better insights into a 2D map of the average proximity matrix, we split the map into several sub-maps, where each sub-map contains companies from only one or very few related sectors. Fig. 3 presents an example of mapped average proximities, computed for companies of the "Medical Laboratories and Research" and "Biotechnology" sectors. All four companies from the Biotechnology sector were rated as ’B’, while companies of the other sector were rated as ’Baa’ or ’B’ by Moody’s. As can be seen from Fig. 3, companies of the same industry sector assigned to different categories (reflecting different likelihood of default) by Moody’s are well separated in the 2D space.

Fig. 4 displays a 2D map of companies coming from two other industry sectors, namely “Medical Appliances and Equipment” and “Medical Instruments and Supplies”. The companies are attributed to three categories, ’A’, ’Baa’, and ’Ba’, by Moody’s. As can be seen from Fig. 4, companies of different categories are well separated in the 2D space. The same pattern can also be observed in Fig. 5, presenting a map of rated companies coming from the “Drug Delivery” and “Drug Generic” sectors.

The mapping results presented in Figs. 3–5 substantiate that data proximity matrices obtained from the set of models created to predict a variety of financial attributes using data from the five different perspectives contain enough information to reflect the financial soundness of the companies in an ordered way. The 2D map presented in Fig. 6 contains data from four different industry sectors and does not exhibit a very clear ordering property. However, this uncertainty is due to a big difference between parameter

![Fig. 3. The average data proximity matrix mapped onto the 2D space. The matrix was obtained for companies of the “Medical Laboratories and Research” (MLaR) and “Biotechnology” (B) sectors.](image)

![Fig. 4. The average data proximity matrix mapped onto the 2D space. The matrix was obtained for companies of the “Medical Appliances and Equipment” (MAaE) and “Medical Instruments and Supplies” (MIaS) sectors.](image)

![Fig. 5. The average data proximity matrix mapped onto the 2D space. The matrix was obtained for companies of the “Drug Delivery” (DD) and “Drug Generic” (DG) sectors.](image)

![Fig. 6. The average data proximity matrix mapped onto the 2D space. The matrix was obtained for companies of four sectors: “Drug Manufactors, Major” (DMM), “Drug Manufactors, Other” (DMO), “Drug Delivery” (DD), and “Drug Generic” (DG).](image)

<table>
<thead>
<tr>
<th>Prediction</th>
<th>True</th>
<th>Introduction</th>
<th>Growth</th>
<th>Mature</th>
<th>Shake-Out</th>
<th>Decline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td><strong>83.5</strong></td>
<td>11.1</td>
<td>00.0</td>
<td>04.8</td>
<td>21.3</td>
<td></td>
</tr>
<tr>
<td>Growth</td>
<td>05.5</td>
<td><strong>66.7</strong></td>
<td>22.7</td>
<td>12.1</td>
<td>01.7</td>
<td></td>
</tr>
<tr>
<td>Mature</td>
<td>00.0</td>
<td>11.1</td>
<td><strong>69.7</strong></td>
<td>04.8</td>
<td>09.0</td>
<td></td>
</tr>
<tr>
<td>Shake-Out</td>
<td>00.0</td>
<td>11.1</td>
<td>07.6</td>
<td><strong>68.6</strong></td>
<td>03.4</td>
<td></td>
</tr>
<tr>
<td>Decline</td>
<td>11.0</td>
<td>00.0</td>
<td>00.0</td>
<td>09.7</td>
<td><strong>73.6</strong></td>
<td></td>
</tr>
</tbody>
</table>
values characterizing companies from the different sectors. When studying companies of only one sector, DMM or DG, for example, one can easily notice ordering of the companies according their ratings. All DMM companies rated as ‘A’ are on the left-hand side of the cluster, while those rated as ‘Aaa’ are on the right-hand side of the cluster. The ‘Aaa’ and ‘A’ companies of the DMM sector are quite close together, however, they become more far apart when companies only from the DMM sector are used to create the map.

### 5.3. Predicting trends of main valuation attributes

In this work, we studied two month ahead prediction of trends from a situation reflected on the 2D maps. Two versions of models were built. The first version was built in accordance with the description presented in Section 3.3. In the second version, only three \( x_j \) features were used, instead of all 23. The features used are: net cash flow from operating, investing, and financing activities—the features used to determine company’s life cycle stage. There was no significant difference between the average test set classification errors obtained from the two versions of models. Table 3 presents the average test set classification error (%) obtained from the three models when predicting the short term trends of the six valuation attributes. In Table 3, TTM stands for Trailing Twelve Months, \( \text{P/E} = \text{Price/Earnings} \), and MRQ means Most Recent Quarter. As can be seen from Table 3, the obtained average test set classification error is very low. Thus, the models are able to make a correct trend prediction of main valuation attributes, which are the main concern of investors.

It is interesting to study importance of variables used to create the three models. Figs. 7 and 8 present the importance of full and reduced sets of input variables of Model 1, created to predict the trend of the “Price to Sales” attribute. The first 11 variables, listed in Table 4, are the same in both versions of Model 1 and are given by attributes characterizing Dividends and Financial strength, see Section 3.3. Observe that values of these variables are predicted by attributes characterizing operating, investing, and financing activities. These variables in Fig. 7 are labelled as 19, 24, and 29. As can be seen from Figs. 7 and 8, for both Model 1 versions (with the full and reduced sets of “cash flow” variables), the variables characterizing Dividends and Financial strength are more important than the “cash flow” variables, even though predicted values of “Dividends and Financial strength” variables are used

### Table 3

The average test set classification error (%) obtained from the three models when predicting the short term trends of the six valuation attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/E Ratio (TTM)</td>
<td>2.80</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>Price to Sales (TTM)</td>
<td>2.00</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>Price to Book (MRQ)</td>
<td>2.20</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>Price to Tangible Book (MRQ)</td>
<td>2.00</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Price to Cash Flow (TTM)</td>
<td>3.00</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>Price to Free Cash Flow (TTM)</td>
<td>1.70</td>
<td>0.10</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Table 4

Model 1 variables characterizing Dividends and Financial strength.

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dividend yield (%)</td>
</tr>
<tr>
<td>2</td>
<td>Dividend yield (5 year avg.) (%)</td>
</tr>
<tr>
<td>3</td>
<td>Dividend 5 year growth rate (%)</td>
</tr>
<tr>
<td>4</td>
<td>Payout ratio (TTM)</td>
</tr>
<tr>
<td>5</td>
<td>Sales (MRQ) vs Q 1 year ago (%)</td>
</tr>
<tr>
<td>6</td>
<td>Sales (TTM) vs 1 year ago (%)</td>
</tr>
<tr>
<td>7</td>
<td>Quick ratio (MRQ)</td>
</tr>
<tr>
<td>8</td>
<td>Current ratio (MRQ)</td>
</tr>
<tr>
<td>9</td>
<td>Long term debt to equity (MRQ)</td>
</tr>
<tr>
<td>10</td>
<td>Total debt to equity (MRQ)</td>
</tr>
<tr>
<td>11</td>
<td>Interest coverage (TTM)</td>
</tr>
</tbody>
</table>
The trend of the "P/E Ratio" attribute.

To illustrate the results, Figs. 9 and 10 present importance of input variables of Model 2, created to predict the trend of the "Price to Book" attribute, while Figs. 11 and 12 present importance of input variables of Model 3, created to predict the trend of "P/E Ratio".

The results of experimental investigations substantiate that the developed models are capable of predicting the short term trends of the main valuation attributes with low error. The valuation attributes provide valuable information for investors from the three different perspectives encoded in the models. The models reflect financial soundness of actions taken by company's management team.

6. Conclusions

A novel approach to exploring company's financial soundness from investor's perspective was proposed in this work. Three main techniques enabling: (i) to predict company's life cycle stage in a near future; (ii) to create a 2D map allowing to explore company's financial soundness from a rating agency perspective; and (iii) to predict trend of main valuation attributes usually used by investors; were developed. The developed techniques were experimentally tested using data concerning companies from twelve different sectors of the health-care industry of the United States from the period 2007–2012.

It was found experimentally that company's life cycle stage in a near future can be determined with the average accuracy of 72.7%. Bearing in mind fuzziness of the transition moment, the obtained prediction accuracy is rather encouraging. It is worth noting that most of the errors were made when distinguishing between to neighbouring life cycle stages.

Data from five different perspectives, namely balance, income, cash flow, stock price, and risk indicators were aggregated via proximity matrices of random forests (created to predict a variety of financial attributes) to enable exploration of company's financial soundness from a rating agency perspective. The proposed use of information not only from companies' financial statements but also from the stock price and risk indicators perspectives has proved useful in creating ordered 2D maps of rated companies. The experimental investigations have shown that proximity matrices obtained using information from these two perspectives survived the elimination process. The rated companies were well ordered (according to the credit risk rating assigned by Moody's) on 2D maps created for companies from one or very few related industry sectors.

The results of experimental investigations substantiate that the developed three models are capable of predicting the short term trends of the main valuation attributes with low error. The valuation attributes provide valuable information for investors from three different perspectives encoded in the models. The models reflect financial soundness of actions taken by company's management team. The first model reflects soundness of company's financial attributes; the second model reflects company's management effectiveness and efficiency; and the third model reflects management team actions related to dividends paid to shareholders, while the second model reflects how well the company is able to exploit opportunities occurring in the industry sector and how effective is the management of company's internal resources. The third model reflects company's management team choices concerning the tradeoff between high profit and company's growth. Estimated importance of variables used to create the models indicates that predicted values of variables related to dividends and financial strength (Model 1), management effectiveness and efficiency (Model 2), and growth rates and profitability (Model 3) provide more information than the "cash flow" variables characterizing current situation.
References


