**Martina NOVOTNÁ, PhD**
**Technical University of Ostrava**
**Department of Finance**
**Ostrava**
**E-mail: martina.novotna@vsb.cz.**

# A MULTIVARIATE ANALYSIS OF FINANCIAL AND MARKET-BASED VARIABLES FOR BOND RATING PREDICTION

***Abstract.*** *The analysis of factors which have the strongest influence on rating can contribute to the higher information availability of market participants, and it enables to react on changes and new information sooner and independently from rating agencies. The paper presents an estimation of corporate bond rating models based on both financial and market-based indicators. Multivariate discriminant analysis and logistic regression were used to identify variables with a significant impact on corporate bond rating in oil and gas industry. In addition to common financial variables, the following market-based indicators such as earnings per share, enterprise value, market capitalization and beta are considered in this paper. Among all the variables used in this study, the enterprise value is the most significant variable for bond rating prediction. The practical use of models lies in the area of management decision process and managing credit risk.*

***Keywords:*** *Credit risk; discriminant analysis; logistic regression; prediction; rating model.*

**JEL Classification: C51, C52, C53, C58, C81, D81, E47**

## 1 Introduction

Credit rating is used for credibility rate marking of the issuer or issue that the obligations will be able to carry out in the future. The concept of credit risk is not a new phenomenon; however this area has attracted a huge attention in last decades, especially during the recent financial crisis. An increasing need to actively and effectively manage credit risk across many sectors of the economy has resulted in more sophisticated and readily available tools and techniques. Credit risk can be seen and measured from various perspectives and can be explained as the chance that money owned may not be repaid. The definition of credit risk is not always clear. It is advisable to distinguish between credit risk and default risk, which is a part of credit risk. De Laurentis et al. (2010, pp. 6) describe a default-mode valuation and consider three types of credit risk: default risk, exposure risk and recovery risk. Default risk is also known as the counterparty risk and represents an event related to the borrower's default. Determination of probability of default is a crucial step in any credit risk management approach and can be achieved in different ways. De Laurentis et al. (2010, pp. 6) point out following alternatives:

- The observation of historical default frequencies of borrowers' homogenous classes (for example, assigned ratings and default rates observed ex post per rating class);
- The use of mathematical and statistical tools, models are based on large databases and enable ex ante measure of expected probability;
- Hybrid methods that combine both judgmental and mechanical approaches (quantitative results are corrected by qualitative aspects);
- Entirely different approach that extracts the implicit probability of default embedded in market prices (securities and stock).

The systematic survey on all determinants of default risk is run by rating agencies. Together with financial ratios, rating agencies consider other traditional analytic areas such as management's reputation, reliability, experience, and past performance. De Servigny and Renault (2004, pp. 26) show that some factors may influence industries in different ways and that the influence of business risk and financial risk on the final rating assessment can be different in various sectors (e.g. high financial risk is connected with airlines industry, high business risk with retail, property or pharmaceuticals). The well-known rating agencies such as Moody's, Standard & Poor's and Fitch are considered as very important international players. Moody's focuses on issues ratings, while S&P concentrates on issuers' ratings and Fitch is offering namely an issuer rating. Although the general concept and description of rating is analogical, there are minor differences depending on each rating agency. Thus, rating assessments of rating agencies are not directly comparable, see for example Cantor and Packer (1997). Ong (2002, pp. 213) states that unlike the U.S., agency ratings are not widely adopted in continental Europe and the demand for external ratings comes mostly from financial companies there. As there is a lack of rating information in the financial markets, in particular in emerging markets, a need for parsimonious models arises. Contribution of own credit models is for example evaluated by Rerolle and Rimaud (2009). As they confirm, research in credit risk area and credit models has in comparison with certified rating important value added, because it enables to react on changes and new information sooner than in the case of complete dependency.

The aim of this study is to examine and quantify relationships among rating and other relevant company data. The primary question is whether financial and market variables affect bond rating. If the answer is positive, the next question is what the nature of their relationship is. The study is based on cross-sectional data that provide information on a variety of 155 US companies from oil and gas industry with Moody's rating at the same point in time (December, 31, 2011).

## 2 Research overview

Quantitative models provide rating based on publically available information only. By such means, it is possible to assess unquoted companies, which is significant for countries with a low number of companies with certified ratings. There are many empirical studies dealing with rating models. Some research of bond rating dates back to the first half of the last century, for example Harold (1938), Hickman (1958) or Fisher (1959). A regression analysis became one of the most used methods to estimate rating in this period. An alternative approach to predict bond ratings is multiple discriminant analysis introduced for example by Pinches and Mingo (1973), Ang and Patel (1975), Altman and Katz (1976) and Belkaoui (1980). Subsequent research was concentrated on comparison of particular statistical methods; e.g. Kaplan and Urwitz (1979) compare ordered probit analysis with ordinary least square regression, Wingler and Watts (1980) compare ordered probit analysis with multiple discriminant analysis. These authors examined the performance of various statistical models in corporate bond rating and used agency ratings as the benchmark to predict bond rating. Multivariate discriminant analysis became the most popular and widely accepted method to estimate bond rating models in this period, usually the methods with best results in this area. Recent studies come from the theoretical framework mentioned above and extend statistical methods for new non-conservative approaches such as neural networks, for example Dutta and Shekhar (1988), Surkan and Singleton (1990). Waagepetersen (2010) assesses the relationship between quantitative models and expert rating evaluation. More recently, Altman, Sabato and Wilson (2010) focus on the importance of non-financial information within risk management.
This paper shows one of the approaches of credit rating modelling having been mentioned above. The next paragraph describes the overview of the methodology; the models are estimated in the next chapters of this paper.

## 3 Description of methodology

### 3.1 Discriminant analysis

Discriminant analysis is a common statistical method used for separation of groups, and hence a suitable method for bond rating modelling. The analysis can be used for two major objectives, first the description of group separation and second, the prediction or allocation of observations to groups. In the case of group separation, linear functions of the variables are used to describe the differences between two or more groups. The main objective is to identify the relative contribution of $p$ variables to separation. Rencher (2002, pp. 270) distinguishes between discriminant and classification functions. Discriminant functions are those used to separate groups, while classification functions can be used to assign individual units to one or more gropes. The latter problem is focused on the prediction or

allocation of observations to groups, which is a common goal of discriminant analysis. A prediction rule then consists of a set o linear combinations of predictors, where the number of combinations reflects the number of groups. Discriminant functions are linear combinations of variables that best separate groups, for example the *k* groups of multivariate observations. The description of discriminant analysis and methods for group separation can be found for example in Rencher (2002) or Huberty and Olejnik (2006). The following definitions and equations were taken from Rencher (2002, chapter 8).
Assume that for *k* groups with observations in the group, we transform each observation vector to obtain , i = 1, 2,..., k; j = 1, 2,..., $n_i$, and find the means , where . We seek the vector **a** that maximally separates . The separation criterion among can be expressed in term of matrices,

(1)

where matrix **H** has a between sum of squares on the diagonal for each of the *p* variables, and matrix **E** has a within sum of squares for each variable on the diagonal. Another expression of the separation criterion is

(2)

where SSH (*z*) and SSE (*z*) are the between and within sums of squares for *z*.

The main task of the discriminant analysis is to find a set of weights (**a** values) for the outcome variables to determine a linear composite:

(3)

so that the ratio (2) is maximized. The discriminant analysis follows by assessing the relative contribution of the to separation of several groups and testing the significance of a subset of the discriminant function coefficients. The discriminant criterion (1) is maximized by , the largest eigenvalue of **;** the remaining eigenvalues correspond to other discriminant dimensions. The test of significance is usually based on the Wilks' lambda, , the most widely used criterion. The test statistic at the *m*th, step (m = 2, 3, …, *s*), is

(4)

which is distributed as . The statistic

(5)

has an approximate -distribution with *(p-m+1)(k-m)* degrees of freedom.

**3.2 Multinomial logistic regression**

In finance, logistic regression is usually used in its univariate context. The common problem where logistic regression can be applied is the prediction of default. Most bankruptcy models are based on scoring methodology, where two alternatives of dependent variable can occur. When exploring relationships among rating and firms' indicators, multinomial logistic regression must be applied, since there are more than two categories of dependent variable. In this case, the number of categories comes from the number of rating groups. The simplest case is the situation when there are just two rating categories, for example investment and speculative grade. The outcome rating is dichotomous (or binary) and univariate or multiple logistic regressions can be used to estimate the prediction model. To extend the previous case, now we assume that the outcome variable has more than three levels, or categories. It is a typical example of bond rating, since the outcome includes rating categories. For estimation of the models in this study, we use a modification of logistic regression, which is called multinomial, polychotomous or polytomous logistic regression (Hosmer and Lemeshow, 2000, pp. 260).
To describe the multinomial logistic regression, it is advisable to start with the multiple logistic regression, which is the case of univariate model with more than one independent variable. Hosmer and Lemeshow (2000, pp. 31) define the multiple logistic regression model as follows. We denote a collection of *p* independent variables by the vector , assuming that at each of these variables is at least interval scale, and p + 1 coefficients .
The conditional probability that the outcome is present is denoted by Then, the logit of the multiple logistic regression model is given by the equation

, (6)

and the logistic regression model is expressed by the following formula,

(7)

Based on De LAurentis et al. (2010, pp. 54 – 55), the g (.) function (6) is known as a link function, which links variables $x_j$ and their coefficients $\beta_j$ with the expected value E(Yi) = $\pi_i$ of the *ith* observation of Y. The link function can be defined as the logarithm of the ratio between the probability of event (e.g. default) and the probability of non-event (e.g. non-default). This ratio is known as “odds” and can be formulated as follows:

(8)

The logit function associates the expected value of the dependent variable to a linear combination of the independent variables. The relationship between independent variables and the probability of default is nonlinear, while the relationship between logit ( and independent variables is linear.
If we have a sample of n independent observations , i=1,2,...*n*, then fitting the model requires to estimate vector by the maximum likelihood method.

The likelihood function can be described by the following formula, according to Hosmer and Lemeshow (2000, pp. 8):

, (9)

where is defined as (7). Assume is the solution to the likelihood equations, then the fitted values for the multiple regression model are , the value of the expression (8) computed using and .

The multinomial logistic regression is a modification of the binary alternative. In this case, it breaks the outcome variable down into series of comparisons between two categories. In the analysis below, bond rating is a dependent variable, which has four possible outcomes. The existence of four categories requires three logit functions and determination of the baseline category, which is then compared with other logits (Hosmer and Lemeshow, 2000, pp. 260 – 262). If we for example use the highest rating category as the baseline, then we form three logits comparing Y = 2, Y = 3 and Y = 4 to it. The three logit functions are denoted as

(10)

A general expression for the conditional probability in the four category model is

, (11)

where = for k = 1, 2, 3, 4 and for the baseline category. The likelihood function is then constructed to obtain parameters of equations. The construction is presented for example in Hosmer and Lemeshow (2000, pp. 263).

## 4 Sample description

Companies with Moody's rating assessment have been considered in this study, the relevant data come from Moody's official websites[1], companies' annual reports and Yahoo! Finance websites[2] of business finance, stock market, quotes and news. The whole sample covers 155 companies, however for the purposes of validation; it was split into two sub-samples. Experimental sample (75 % of the data sample) will be use for model estimation and the remaining part (test sample) will be used for validation of models. For the reasons of calculations[3], original rating categories have been re-coded as presented in the Table 1. The first three highest categories have been merged together because of a small number of representative companies, which could negatively affect results and stability of models.

**Table 1 Sample structure**

| Rating category | Rating code | Number of cases | Marginal percentage |
|---|---|---|---|
| Aaa, Aa, A | 1 | 21 | 13.5 % |
| Baa | 2 | 59 | 38.1 % |
| Ba | 3 | 30 | 19.4 % |
| B | 4 | 45 | 29.0 % |
| **Total** | **X** | **155** | **100%** |

The selection of independent variables should be thoroughly considered, because the set of input variables can substantially affect results, specifically predictive ability and stability of final models. The analysts usually stand on their previous results, experience and other research studies. Basically, most models are estimated based on financial statements of companies. Many studies prove that relatively simple rating models containing basic financial indicators provide good classification ability and can be used as a tool to assign a rating classification.
There are various possible **financial indicators** that can be used in the analysis. The selected indicators should reflect profitability, activity, liquidity and capital

[1] http://www.moodys.com/
[2] http://finance.yahoo.com/
[3] PASW Statistics 18

structure of companies and all of them should have a relationship with rating. To use some variables in the analysis, the main assumptions should be met. First, the variables should have a normal distribution; secondly, multicolinearity should be avoided. In this study, the following financial variables are considered initially.

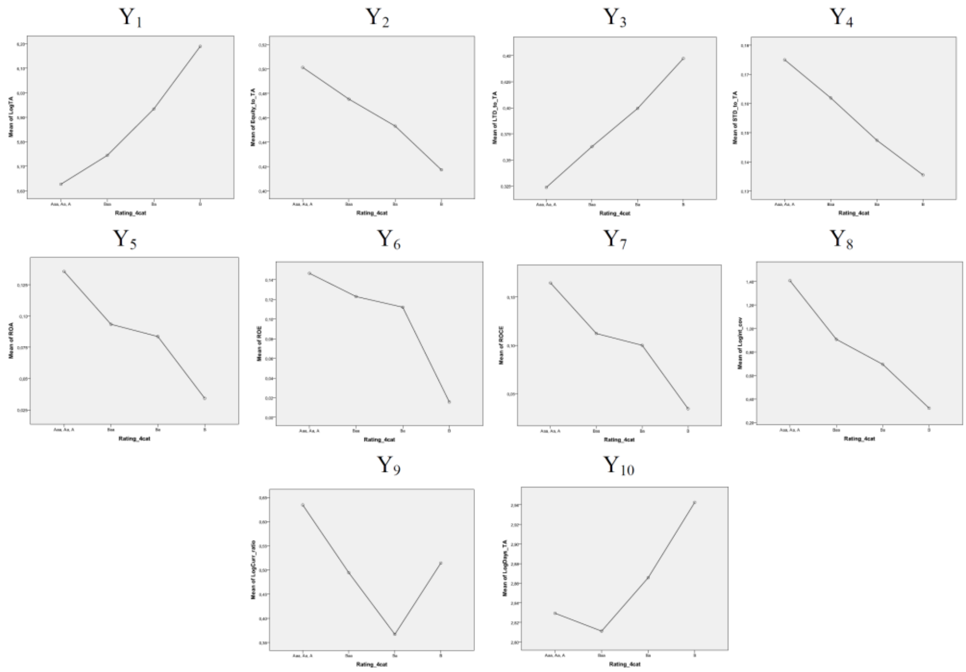
**Table 2 Financial indicators**

| | | |
|---|---|---|
| $\mathbf{Y_1}$ | Total assets | Log*TA* |
| $\mathbf{Y_2}$ | Equity to total assets ratio | *Equity_to_TA* |
| $\mathbf{Y_3}$ | Long term debt to total assets ratio | *LTD_to_TA* |
| $\mathbf{Y_4}$ | Short term debt to total assets ratio | *STD_to_TA* |
| $\mathbf{Y_5}$ | Return on assets | *ROA* |
| $\mathbf{Y_6}$ | Return on equity | *ROE* |
| $\mathbf{Y_7}$ | Return on capital employed | *ROCE* |
| $\mathbf{Y_8}$ | Interest coverage | Log*Int_cov* |
| $\mathbf{Y_9}$ | Current ratio | Log*Curr_ratio* |
| $\mathbf{Y_{10}}$ | Total assets days outstanding | Log*Days_TA* |

The relationship between each variable and rating should have an economic rationale. For example, we can assume that the higher the size of total assets, the higher the protection of company's creditors, and the higher the rating category. Some financial variables had to be transformed to approach a normal distribution, such as TA (Log*TA*), Int_cov (Log*Int_cov*), Curr_ratio (Log*Curr_ratio*), Days_TA (Log*Days_TA*).

The following graphs (Figure 1) present relationships between the ten independent variables (or their logarithmic transformations) and rating category. The positive relationships are apparent for *LTD_to_TA* ($Y_3$) and Log*Days_TA* ($Y_{10}$), which is in line with the economic theory. The higher the proportion of long term debt to total assets and the higher the number of days of assets, the higher the credit risk and the lower rating. The same relationship can be seen for Log*TA* ($Y_1$), which does not support the economic assumption that a higher level of total assets can indicate a higher rating category. The opposite relationships are evident for *Equity_TA* ($Y_2$), *STD_to_TA* ($Y_4$), *ROA* ($Y_5$), *ROE* ($Y_6$), *ROCE* ($Y_7$), Log*Int_cov* ($Y_8$) and Log*Curr_ratio* ($Y_9$). Higher values of these variables are associated with higher rating categories, which support the hypothesis that higher profitability, liquidity and solvency ratios indicate lower default risk.

**Figure 1 Means plots of financial variables**



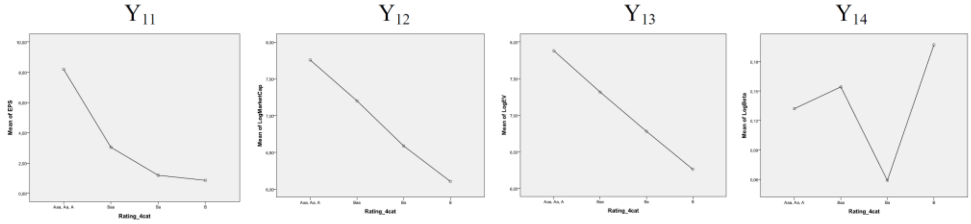
The main task of this study is to investigate the relationship among rating and selected **market-based variables.** Basically, we can expect that companies with higher market evaluation would have a higher rating assessment. For the purposes of this study, four simple market indicators will be included.

**Table 3 Market-based indicators**

| | | |
|---|---|---|
| $\mathbf{Y_{11}}$ | Earnings per share | *EPS* |
| $\mathbf{Y_{12}}$ | Market capitalisation | Log*MarketCap* |
| $\mathbf{Y_{13}}$ | Enterprise value | Log*EV* |
| $\mathbf{Y_{14}}$ | Beta coefficient | Log*Beta* |

To approach a normal distribution, some of these variables have been transformed (*LogMarketCap*, *LogEV*, *LogBeta*). The relationships between these indicators and rating are presented in the next figure (Figure 2). The *EPS* ($Y_{11}$), Log*MarketCap* ($Y_{12}$) and Log*EV* ($Y_{13}$) variables suggest that the higher the value, the higher the rating category. The relationship is ambiguous for Log*Beta* ($Y_{14}$), so it will not be likely a suitable predictor for bond rating.

**Figure 2 Means plots of market variables**




The overview of basic descriptive characteristics of the whole sample is presented in the Table 4.

**Table 4 Description statistics**

| Rating_4cat | | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $Y_6$ | $Y_7$ | $Y_8$ | $Y_9$ | $Y_{10}$ | $Y_{11}$ | $Y_{12}$ | $Y_{13}$ | $Y_{14}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Mean | 5.6268 | 0.5013 | 0.3237 | 0.175 | 0.1357 | 0.1465 | 0.1645 | 1.4074 | 0.6344 | 2.8294 | 0.1321 | 8.2023 | 7.8804 | 7.7564 |
| | Median | 5.2666 | 0.5047 | 0.2928 | 0.1767 | 0.1196 | 0.1367 | 0.1452 | 1.4134 | 0.4775 | 2.8095 | 0.1461 | 7.2600 | 7.9113 | 7.9157 |
| | Std. Deviation | 1.3765 | 0.1448 | 0.1401 | 0.0649 | 0.1071 | 0.0974 | 0.1211 | 0.6127 | 0.5149 | 0.3103 | 0.1485 | 5.9659 | 0.3666 | 0.5178 |
| | Kurtosis | -0.0186 | 0.5442 | 0.2088 | 0.2762 | 8.0502 | 4.0918 | 5.1294 | 0.4629 | 6.8238 | -0.4389 | 1.2232 | 4.8552 | -0.2420 | 2.1572 |
| | Skewness | 0.9983 | -0.5262 | 0.756 | -0.0715 | 2.3671 | 1.4344 | 1.8110 | 0.0863 | 2.2974 | 0.5457 | -1.1742 | 2,02250 | 0.3465 | -0.9343 |
| 2 | Mean | 5.7448 | 0.4753 | 0.3628 | 0.1619 | 0.0934 | 0.1228 | 0.1122 | 0.9078 | 0.4943 | 2.8111 | 0.1543 | 3.0442 | 7.3167 | 7.2022 |
| | Median | 4.8461 | 0.4749 | 0.3575 | 0.1350 | 0.0812 | 0.1124 | 0.0949 | 0.8574 | 0.537 | 2.8665 | 0.1105 | 2.4200 | 7.2716 | 7.1327 |
| | Std. Deviation | 1.7968 | 0.1475 | 0.1379 | 0.1015 | 0.0522 | 0.0789 | 0.0609 | 0.5211 | 0.2916 | 0.3391 | 0.7444 | 2.6704 | 0.5913 | 0.5865 |
| | Kurtosis | 0.3258 | 0.4753 | -0.2671 | 1.7761 | 1.2982 | 1.3400 | 1.1753 | 0.2873 | -0.8336 | -0.7782 | 29.7451 | 5.1563 | 14.9249 | 14.3273 |
| | Skewness | 0.9082 | 0.1222 | 0.0712 | 1.4725 | 1.0681 | 0.1778 | 0.9577 | 0.5088 | 0.2983 | -0.1089 | 5.2090 | 1.9799 | 3.0190 | 2.9266 |
| 3 | Mean | 5.9340 | 0.4532 | 0.3994 | 0.1474 | 0.0836 | 0.1120 | 0.1000 | 0.6949 | 0.3669 | 2.8656 | 0.0589 | 1.1904 | 6.7808 | 6.5886 |
| | Median | 6.4014 | 0.4435 | 0.4212 | 0.1261 | 0.0789 | 0.1023 | 0.0967 | 0.6302 | 0.2877 | 2.9204 | 0.1303 | 1.4300 | 6.7084 | 6.5874 |
| | Std. Deviation | 1.1518 | 0.1583 | 0.1546 | 0.0863 | 0.0544 | 0.1166 | 0.0660 | 0.5002 | 0.3081 | 0.3378 | 0.2737 | 2.9576 | 0.4046 | 0.4984 |
| | Kurtosis | 0.0718 | 1.9643 | 0.6565 | 0.5162 | 0.8045 | 1.6094 | 0.1635 | 0.2621 | 0.1517 | -0.6246 | -0.7332 | 0.7360 | 0.4970 | 0.0252 |
| | Skewness | -1.0941 | 0.4607 | -0.1069 | 1.0897 | 0.0331 | 0.9705 | 0.0229 | 0.6603 | 0.7087 | -0.4581 | -0.6749 | -0.3365 | 0.9090 | 0.0254 |
| 4 | Mean | 6.1896 | 0.4173 | 0.4472 | 0.1355 | 0.0341 | 0.0157 | 0.0345 | 0.3216 | 0.5143 | 2.9426 | 0.1972 | 0.8567 | 6.2631 | 6.1063 |
| | Median | 6.1390 | 0.4292 | 0.4548 | 0.1189 | 0.0394 | 0.0204 | 0.0498 | 0.4161 | 0.5330 | 2.9656 | 0.2500 | 0.8050 | 6.2601 | 6.0899 |
| | Std. Deviation | 0.6881 | 0.1623 | 0.1558 | 0.0815 | 0.0840 | 0.2902 | 0.1187 | 0.5114 | 0.3643 | 0.3679 | 0.2321 | 1.7164 | 0.3313 | 0.5563 |
| | Kurtosis | 8.7820 | 0.1622 | 0.3578 | 4.4224 | 8.4263 | 14.5305 | 17.8013 | 1.5585 | 0.6715 | 0.0369 | 0.0512 | 1.8570 | 0.0503 | 1.5639 |
| | Skewness | 1.1639 | 0.2564 | -0.2401 | 1.8629 | -1.8148 | -2.2847 | -3.3597 | -1.0082 | 0.5278 | -0.6661 | -0.8606 | -0.0172 | 0.1439 | 0.2244 |
| Total | Mean | 5.8946 | 0.4577 | 0.3891 | 0.1532 | 0.0800 | 0.0928 | 0.0944 | 0.8089 | 0.4944 | 2.8623 | 0.1426 | 2.5253 | 6.9213 | 6.7915 |
| | Median | 6.0585 | 0.4575 | 0.3825 | 0.1304 | 0.0753 | 0.1013 | 0.0888 | 0.7110 | 0.4894 | 2.9218 | 0.1553 | 1.8200 | 6.8643 | 6.7903 |
| | Std. Deviation | 1.3752 | 0.155 | 0.1517 | 0.0890 | 0.0787 | 0.1810 | 0.1000 | 0.6243 | 0.3580 | 0.3450 | 0.4785 | 3.7236 | 0.7090 | 0.7852 |
| | Kurtosis | 1.0347 | 0.355 | -0.1504 | 2.0014 | 10.4075 | 33.1121 | 16.5119 | 0.7734 | 4.7038 | -0.6302 | 57.6132 | 11.2276 | 3.0042 | 1.9416 |
| | Skewness | 0.5614 | 0.1038 | 0.0844 | 1.3915 | 0.2485 | -3.4016 | -1.3410 | 0.2382 | 1.3239 | -0.2481 | 6.5373 | 2.3581 | 0.9841 | 0.4884 |

## 5 Bond rating models

Discriminant analysis (DA) and multinomial logistic regression (MLR) will be carried out to identify variables most relevant to rating classification. Two approaches will be used, the method in which all independent variables are included in the model (full), and stepwise method (step), which aims to include only the most significant variables in the model.

### 5.1 Estimation of models

First, bond rating models will be estimated from financial data only. Then, only market-based data will be used and finally, results will be compared and a combination of both previous approaches will be applied.

**Models with financial variables**

The original set of independent variables was modified and three financial variables ($Y_3$, $Y_7$, $Y_{10}$) were removed for the reasons of high correlations with other variables. For example, a large correlation is achieved between *LTD_to_TA* and *Equity_to_TA* (-0.846), or Log*Int_cov* (-0.619). The correlation between *ROCE* and *ROA* is very high, 0.984. The variable Log*Days_TA* is positively correlated with *STD_to_TA* (-0.636).[4] For the purposes of discriminant analysis, it is recommended to use rather lower number of variables than more variables with large interdependencies. Thus, the final set consists of ten financial variables.

**Table 5 Models with financial variables**

| Model | Approach | Number of input variables | Number of variables in the model | Variables included | Classification ability |
|---|---|---|---|---|---|
| (A) | DA Full | 7 | 7 | All | 46.8 % |
| (B) | DA Step | 7 | 1 | Log*Int_cov* | 44.4 % |
| (C) | MLR Full | 7 | 7 | All | 56.8 % |
| (D) | MLR Step | 7 | 3 | *Equity_TA*, Log*Int_cov*, Log*Curr_ratio* | 52.3 % |

**Models with market-based variables**

Analogically to the previous case, both discriminant analysis and multinomial logistic regression were used to estimate models and find the most significant indicators for classification. The results (Table 6) show that with considering companies' market data only, models with much better classification ability can be obtained. The most significant variables are *EPS* and Log*EV*.

**Table 6 Models with market-based variables**

| Model | Approach | Number of input variables | Number of variables in the model | Variables included | Classification ability |
|---|---|---|---|---|---|
| (E) | DA Full | 4 | 4 | All | 64.3 % |
| (F) | DA Step | 4 | 2 | *EPS*, Log*EV* | 62.7 % |
| (G) | MLR Full | 4 | 4 | All | 79.7 % |
| (H) | MLR Step | 4 | 1 | Log*EV* | 70.9% |

**Combination of financial and market-based variables**

When all independent variables enter the analysis, the overall classification ability gently rises, especially in the case of MLR. By using all 11 variables, classification

[4] The coefficients of Pearson correlation are significant at the 0.01 (2-tailed) in all cases.

ability of 89.6 % can be achieved. By applying stepwise methods, the final models contain only two variables, Log*Curr_ratio* and Log*MarketCap* (Table 7).

**Table 7 Combination of financial and market-based variables**

| Model | Approach | Number of input variables | Number of variables in the model | Variables included | Classification ability |
|---|---|---|---|---|---|
| (I) | DA Full | 11 | 11 | All | 62.7 % |
| (J) | DA Step | 11 | 2 | Log*Curr_ratio* Log*Market_Cap* | 66.7 % |
| (K) | MLR Full | 11 | 11 | All | 89.6 % |
| (L) | MLR Step | 11 | 2 | Log*Curr_ratio* Log*Market_Cap* | 76.1 % |

**Modifications and adjustments**

Based on the results above, it is evident that some variables contribute to classification more than the others. The final models would stand on the previous results and use only four predictors with the most significant discriminating power on rating, such as Log*Int_cov*, *EPS*, Log*EV* and Log*MarketCap*. Classification ability of the adjusted models is in the table below (Table 8).

**Table 8 Modification of models**

| Model | Approach | Number of input variables | Number of variables in the model | Variables included | Classification ability |
|---|---|---|---|---|---|
| (M) | DA Full | 4 | 4 | All | 64.3 % |
| (N) | MLR Full | 4 | 4 | All | 76.3 % |

The overall results suggest that market indicators contribute to the discrimination more than financial ratios. In terms of the firm's size, markets capitalisation is more significant that the value of total assets. By adding market data to the original set of financial ratios, the total classification ability of models increases. Both methods, the discriminant analysis and multinomial logistic regression, provide similar results, however models estimated by MLR achieve higher classification ability. The best model from this point of view was estimated by MLR and uses all eleven financial and market variables (Model K). Good classification results are then achieved by MLR models using either 4 market indicators (Model G), or 4 combined variables (Model N). The overall results are surprising because they suggest that earnings per share, enterprise value and market capitalization can give a good signal of a bond investment quality.

## 5.2 Verification and validation

Based on the criterion of classification ability on the original sample, the following three models, (G), (K) and (N) will be examined in more detail. All these models have been estimated by multinomial logistic regression, which allows simpler comparing of results and overall fit of models. To assess the fit of models, we use a log-likelihood statistic, which is based on summing the probabilities associated with the predicted and outcome variables, see Tabachnik and Fidell (2007, pp. 446). The statistic indicates how much unexplained information there is after the model has been fitted. The larger the value, the more unexplained observations there are. The chi-square test tests the decrease in unexplained variance from the baseline model to the final model. All the final models explain a significant amount of the original variability, so they better fit than the original model. The next test tests whether the models predicted values are significantly different from the observed ones. If the statistics (Pearson and Deviance) are not significant, than predicted and observed values are not different, and the model is a good fit. All three models are a good fit based on this test. The significance of predictors to the models was assessed by the likelihood ratio tests. In all models, variable Log*EV* has a significant main effect on rating category classification; it is even the only significant predictor in Model N. Due to a large number of derived models in this study, parameter estimates and odds ratios are not included in this paper, however they can be provided on demand. Verification of the three models is presented in Table 9.

**Table 9 Verification of models**

| **Criterion** | **Model G** | **Model K** | | **Model N** |
|---|---|---|---|---|
| Predictors included:<br><br>(*Likelihood ratio tests of parameters*) | *EPS*<br>Log*MarketCap*<br>Log*EV****<br>Log*Beta**** | Log*TA*<br>*Equity_to_TA****<br>*STD_to_TA***<br>*ROA**<br>*ROE**<br>Log*Int_cov* | Log*Curr_ratio*<br>*EPS*<br>Log*MarketCap*<br>Log*EV***<br>Log*Beta*** | Log*Int_cov*<br>*EPS*<br>Log*MarketCap*<br>Log*EV** |
| Model fitting: *Chi-Square* | 133,286*** (df=12) | 150,528*** (df=33) | | 132,122*** (df=12) |
| Goodness-of-Fit:<br>*Pearson*<br>*Deviance* | <br>184.708 (df=219)<br>78.954 (df=219) | <br>28.304 (df=165)<br>30.030 (df=165) | | <br>157.824 (df=264)<br>114.580 (df=264) |
| Measures of $R^2$:<br>*Cox and Snell*<br>*Nagelkerke* | <br>0.815<br>0.875 | <br>0.758<br>0.816 | | <br>0.894<br>0.959 |

****p<.001, ** p<.01, * p<.05*

The three selected models (G, K, N) were used to predict bond rating of companies other than that used for estimation of models. The test sample covers 25 companies randomly selected from the original sample.

**Table 10 Validation of models**

| Model | Correct classification 4-rating model | Correct classification 2-rating model |
|---|---|---|
| Model G | 16 % | 64 % |
| Model K | 28 % | 60 % |
| Model N | 32 % | 84 % |

As expected, the ratio of correctly classified companies is relatively low, which is likely the result of a small control sample. However, all models contribute significantly to the classification in case of just two rating groups, investment and speculative category. Since there are fourteen models estimated in this study, it is not relevant to present details about each model's parameters. As the validation proved that the Model N (Table 10) provides satisfying predictions, the following text presents explanation of this model's parameters (Table 11), logit functions and conditional probabilities' expressions.

**Table 11 Parameters of Model N**

| Variable | Rating 2 | Rating 3 | Rating 4 |
|---|---|---|---|
| *Intercept* | 53.559 | 86.307 | 117.457 |
| *LogCurr_ratio* | -3.288 | -5.966 | -6.165 |
| *LogMarketCap* | -17.402 | -14.815 | -14.390 |
| *Log EV* | 10.730 | 3.752 | -1.415 |
| *EPS* | -0.108 | -0.381 | -0.295 |

*Rating 1 as a reference category*

The three following logit functions (10),

,

,

,

are then used to determine the conditional probabilities (11) of each rating category in the following way, where = for k = 1, 2, 3, 4 and for the baseline category. The firm is assigned to rating category with the highest value of the conditional probability.

## 6 Conclusion

The paper examined the role of financial and market-based variables on corporate bond rating. The analysis was carried out for 155 US companies in the oil and gas industry having a rating assessment from Moody's rating agency. Multivariate discriminant analysis and logistic regression were carried out to identify variables with a significant impact on corporate bond rating in the selected industry. In addition to common financial variables, the market-based indicators such as earnings per share, enterprise value, market capitalization and beta were considered in this paper. Fourteen bond rating models were estimated by a variety of combinations of variables and a statistical procedure. In this paper it was demonstrated that both approaches discriminant analysis and logistic regression are suitable for bond rating modelling. Although these methods are based on different methodology, they show very similar results and classification ability.
The overall results suggest that market-based indicators contribute to the separation of rating groups more than financial ratios and the total classification ability of models increases with these variables. Among all the variables used in this study, the enterprise value of the company is the most significant variable for bond rating prediction in oil and gas industry. Financial variables such as equity to total assets ratio, interest coverage and current ratio, together with the enterprise value, market capitalisation and earnings per share can give a good signal about the investment quality. These results confirm the importance of profitability, liquidity, solvency and capitalisation for bond rating, which is consistent with the economic rationale. These variables are simply-to-use and often publically available. Thus, we obtained a useful instrument that can be applied in the first stage of accessing companies without certified rating.

The practical use of such models lies within the area of management decision process and managing credit risk. One should focus on these variables in particular. This strategy is relevant when it is compared over time for the firm, to the industry (industry and market averages can be used as benchmarks) and economy-wide measures of performance and financial position. Such models are useful instruments for primary assessing companies without agency rating, especially in countries with less developed capital markets which are usually associated with lack of information. It is evident that this approach has some limitations, including the limitations of accounting data and the fact, that financial statements disclose little about the important willingness to pay. By using this strategy, it is possible to get the first signal about the companies' ability to meet their obligations and the first measure for comparing among companies.

**Acknowledgements**

***The paper is based on research activities sponsored through SGS research project No. SP2012/19. The support is greatly acknowledged.***

## REFERENCES

[1] **Altman E.I., Katz S. (1976),** ***Statistical Bond Rating Classification Using Financial and Accounting Data***. *M. Schiff and G. Sorter, eds.: Topical Research in Accounting* , NYU Press, New York;
[2] **Altman E.I., Sabato G., Wilson N. (2010),** ***The Value of Non-financial Information in Small and Medium-sized Enterprise Risk Management***. *The Journal of Credit Risk*, 6, 95 – 127;
[3] **Ang J.S., Patel K. (1975),** ***Bond Rating Methods: Comparison and Validation***. *The Journal of Finance,* 30, 631 – 640;
[4] **Belkaoui A. (1980),** ***Industrial Bond Ratings: A Discriminant Analysis Approach*****.** Working paper (University of Ottawa);
[5] **Cantor R, Packer F. (1997),** ***Differences of Opinion and Selection Bias in the Credit Rating Industry***. *Journal of Banking & Finance,* 21, 1395 – 1417;
[6] **De Laurentis G, Maino R., Moletni L. (2010),** ***Developing, Validating and Using Internal Ratings. Methodologies and case studies***. *John Wiley & Sons Inc.:* Chichester;
[7] **De Servigny A, Renault O. (2004),** ***Measuring and Managing Credit Risk*****.** *McGraw-Hill*: New York;
[8] **Dutta, S, Shekhar S. (1988), Bond Ratings:** ***A Non-conservative Application of Neural Network***. *IEEE International Conference on Networks, 2, San Diego, California,* 443-450;

[9] **Fisher L. (1959),** ***Determinants of Risk Premiums on Corporate Bonds***. *Journal of Political Economy* , 67, 217-237;

[10] **Gray, S, Mirkovic, A, Ragunathan, V. (2006),** ***The Determinants of Credit Ratings: Australian Evidence***. *Australian Journal of Management*, 31, 333-354;

[11] **Hickman W.B. (1958),** ***Corporate Bond Quality and Investor Experience*****.** *Princeton University Press*: Princeton;

[12] **Hosmer D.W., Lemeshow S. (2000),** ***Applied Logistic Regression***. *John Wiley & Sons I*nc.: New York;

[13] **Huberty C.J., Olejnik S. (2006),** ***Applied MANOVA and Discriminant Analysis***. *John Wiley & Sons* Inc.: New York;

[14] **Kaplan R, Urwitz G. (1979),** ***Statistical Models of Bond Ratings: A Methodological Inquiry***. *Journal of Business*, 52, 231-261;

[15] **Ong M.K. (2002),** ***Credit ratings. Methodologies, Rationale and Default Risk***. Risk Books: London;

[16] **Pinches G., Mingo K.A. (1973),** ***A Multivariate Analysis of Industrial Bond Ratings***. *The Journal of Finance*, 28, 1 – 18;

[17] **Rencher A.C. (2002),** ***Methods of Multivariate Analysis***. *John Wiley & Sons I*nc.: New York;

[18] **Rerolle J, Rimaud C. (2009),** ***Does Independent Credit Research Add Value?*** *Alternative Intelligence Quotient* 2009, 32, 31 – 37;

[19] **Surkan, A.J., Singleton J.C. (1990),** ***Neural Networks for Bond Rating Improved by Multiple Hidden Layers***. *Proceedings of the International Joint Conference on Neural Networks, 2, San Diego, California*, 157-162;

[20] **Tabachnik B.G., Fidell L.S. (2007),** ***Using Multivariate Statistics***. *Pearson Education*, Inc.: Boston;

[21] **Waagepetersen R. (2010),** ***A Statistical Modeling Approach to Building an Expert Credit Rating System*****.** *The Journal of Credit Risk*, 6, 81 – 94;

[22] **Wingler T.R., Watts J.M. (1980),** ***An Analysis of Bond Rating Changes in the Electric Utility Industry***. *The Financial review*, 15, 1 – 89.