FINANCIAL ENGINEERING LABORATORY

Technical University of Crete

Combining Accounting Data and a Structural Model for Predicting Credit Ratings: Empirical Evidence from European Listed Firms

Michael Doumpos
Dimitrios Niklis
Constantin Zopounidis
Kostas Andriosopoulos

Working Paper 2014.03

May 2014

FINANCIAL ENGINEERING LABORATORY

School of Production Engineering & Management
Technical University of Crete

List of Working Papers

Editorial Committee

2010.01 Modelling banking sector stability with multicriteria approaches
C. Gaganis, F. Pasiouras, M. Doumpos, C. Zopounidis

2010.02 Bank productivity change and off-balance-sheet activities across different levels of economic development
A. Lozano-Vivas, F. Pasiouras

2010.03 Developing multicriteria decision aid models for the prediction of share repurchases
D. Andriosopoulos, C. Gaganis, F. Pasiouras, C. Zopounidis

2010.04 Developing an employee evaluation management system: The case of a healthcare organization
E. Grigoroudis, C. Zopounidis

2010.05 Analysis of domestic and cross-border mega-M&As of European commercial banks
M. Nnadi, S. Tanna

2010.06 Corporate culture and the tournament hypothesis
N. Ozkan, O. Talavera, A. Zalewska

2011.01 Mutual funds performance appraisal using a multicriteria decision making approach
V. Babalos, N. Philippas, M. Doumpos, C. Zopounidis

2011.02 Assessing financial distress where bankruptcy is not an option: An alternative approach for local municipalities
S. Cohen, M. Doumpos, E. Neophytou, C. Zopounidis

2012.01 Multicriteria decision aid models for the prediction of securities class actions: Evidence from the banking sector
V. Balla, C. Gaganis, F. Pasiouras, C. Zopounidis

2012.02 Service quality evaluation in the tourism industry: A SWOT analysis approach
M. Tsitsiloni, E. Grigoroudis, C. Zopounidis

2012.03 Forecasting the prices of credit default swaps of Greece by a neuro-fuzzy technique
G.S. Atsalakis, K.I. Tsakalaki, C. Zopounidis

2012.04 Sensitivity of consumer confidence to stock markets' meltdowns
E. Ferrer, J. Salaber, A. Zalewska

2012.05 Efficiency and performance evaluation of European cooperative banks
M. Doumpos, C. Zopounidis

2012.06 Rating mutual funds through an integrated DEA-based multicriteria performance model: Design and information content
V. Babalos, M. Doumpos, N. Philippas, C. Zopounidis

2012.07 Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines
D. Niklis, M. Doumpos, C. Zopounidis

2013.01 Firm performance and ownership by financial institutions: The case of Cyprus
K. Chatzimichael, C. Gaganis, F. Pasiouras, P. Theodossiou, C. Zopounidis

2013.02 Evaluation of the efficiency of Greek hospitals: A non-parametric framework and managerial implications
G. Fragkiadakis, M. Doumpos, C. Zopounidis, C. Germain

2013.03 Bank diversification and overall financial strength: International evidence around the crisis
M. Doumpos, C. Gaganis, F. Pasiouras

2014.01 The robustness concern in preference disaggregation approaches for decision aiding: An overview
M. Doumpos, C. Zopounidis

2014.02 An integrated approach for energy efficiency analysis in European Union countries
G. Makridou, K. Andriosopoulos, M. Doumpos, C. Zopounidis

2014.03 Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from European listed firms
M. Doumpos, D. Niklis, C. Zopounidis, K. Andriosopoulos

Combining Accounting Data and a Structural Model for Predicting Credit Ratings: Empirical Evidence from European Listed Firms

Michael Doumpos[a,*], Dimitrios Niklis[a], Constantin Zopounidis[a,b], Kostas Andriosopoulos[c]

[a] *Technical University of Crete, School of Production Engineering and Management, Financial Engineering Laboratory, University Campus, 73100 Chania, Greece*
[b] *Audencia Group, Nantes School of Management, France*
[c] *ESCP Europe Business School, Research Centre for Energy Management, UK*

Abstract

Ratings issued by credit rating agencies (CRAs) play an important role in the global financial environment. Among other issues, past studies have explored the potential for predicting these ratings using a variety of explanatory factors and modeling approaches. This paper describes a multicriteria classification approach that combines accounting data with a structural default prediction model in order to obtain improved predictions and test the incremental information that a structural model provides in this context. Empirical results are presented for a panel data set of European listed firms during the period 2002–2012. The analysis indicates that a distance to default measure obtained from a structural model adds significant information compared to popular financial ratios. Nevertheless, its power is considerably weakened when market capitalization is also considered. The robustness of the results is examined over time and under different specifications of the rating categories.

Keywords: Credit ratings, Rating agencies, Black-Scholes-Merton model, Multiple criteria decision making

JEL classification: C44, G24, G13

*Corresponding author. Phone: +30 28210 37318, Fax: +30 28210 69410

Email addresses: `mdoumpos@dpem.tuc.gr` (Michael Doumpos), `dniklis@isc.tuc.gr` (Dimitrios Niklis), `kostas@dpem.tuc.gr` (Constantin Zopounidis), `kandriosopoulos@escpeurope.eu` (Kostas Andriosopoulos)

1. Introduction

Credit risk modeling plays a crucial role in financial risk management, in areas such as banking, corporate finance, and investments. Credit risk management has evolved rapidly over the past decades. Altman and Saunders (1997) list some factors that have contributed to the increasing importance of credit risk management, including: (i) the worldwide increase in the number of bankruptcies, (ii) the trend towards disintermediation by the highest quality and largest borrowers, (iii) the increased competition among credit institutions, (iv) the declining value of real assets and collateral in many markets, and (v) the growth of new financial instruments with inherent default risk exposure, such as credit derivatives. The recent global financial crisis confirms that these factors are still valid today.

Credit ratings are important ingredients of the credit risk management process, and they are widely used for estimating default probabilities, supporting credit granting decisions, pricing loans, and managing loan portfolios. Credit ratings are either obtained through models developed internally by financial institutions (Treacy and Carey, 2000) or are provided externally by credit rating agencies (CRAs). The latter, despite the criticisms on their scope and accuracy (e.g., Frost, 2007; Pagano and Volpin, 2010; Tichy et al., 2011), are widely used by investors, financial institutions, and regulators, and they have been extensively studied in academic research (for a recent overview, see Jeon and Lovo, 2013).

In this context, models that explain and replicate the ratings issued by CRAs can be useful in various ways, as they can facilitate the understanding of the factors that drive CRAs' evaluations, provide investors and regulators with early-warning signals and information for important rating changes, and support the credit risk assessment process for firms not covered by the CRAs. Following this line of research, Huang et al. (2004) compared two data mining techniques (neural networks and support vector machines) using two data sets from Taiwan and USA for predicting credit ratings using purely financial ratios. Pasiouras et al. (2006) analyzed the bank credit ratings issued by Fitch for a cross-country sample in relation to a wide range of explanatory factors covering the regulatory and supervisory framework and bank-specific characteristics. Hwang et al. (2010) employed an ordered nonlinear semiparametric probit model to predict the ratings issued by Standard and Poor's (S&P) for a static sample of listed US companies, using financial and market variables as well as industry effects. In a similar framework, Mizen and Tsoukas (2012) found that using a nonlinear probit model with state dependencies improves the prediction results, whereas Hwang (2013) further found time-varying effects to be important. On the other hand, Lu et al. (2012) analyzed the information in news articles, combined with financial and market variables, to predict changes in the S&P ratings for firms in the USA.

Previous studies, such as the ones noted above, have used a number of explanatory factors to analyze and predict credit ratings, including firm-specific data (usually in the form of financial ratios) as well as market variables. Some studies have also considered default risk estimates from structural models (Hwang et al., 2010; Hwang, 2013; Lu et al., 2012), which are based on the contingent claims approach introduced by Black and Scholes (1973) and Merton (1974). Nevertheless, no systematic analysis has been conducted in the context of credit ratings prediction with respect to the additional information that the estimates of structural models provide compared to accounting-based data, nor have the benefits that can be obtained from the combination of these two different approaches been examined. On the other hand, several studies have explored this issue for models that aim to predict bankruptcy and default. For instance, Hillegeist et al. (2004) found the estimates obtained from a structural market-based model to be significantly superior for predicting bankruptcies when compared to popular accounting-based models. Bharath and Shumway (2008) introduced a very simple structural model that they concluded was useful for predicting defaults when combined with other default predictor variables. Agarwal and Taffler (2008), on the other hand, used a data set of UK bankruptcies and concluded that a structural model does not add much information when compared to accounting-based models. Campbell et al. (2008) also reached a similar conclusion for a sample of bankruptcies in the USA, while Das et al. (2009) noted that rather than viewing accounting and market information as substitutes, they should be combined in order to predict default more accurately.

Some evidence on the information content provided by the estimates of structural models in the context of credit ratings prediction can be found in the works of Du and Suo (2007) and Chen and Hill (2013). Du and Suo (2007) used a sample of corporate ratings by S&P for US firms over the period 1985–2002 and tested the power of a structural model compared to the variables upon which the model is based, but no analysis undertaken for comparison to popular accounting-based data that also affect corporate credit ratings. In a different context, Chen and Hill (2013) studied the relationship between different measures of default risk and stock returns. Their comparison between default models and the ratings of Moody's and S&P for a sample of UK firms indicated that the accounting-based z-score model had a stronger correlation to the ratings of the agencies than two structural models did. This led the authors to note that "this suggests a relatively high reliance on accounting ratios in the default risk assessments of the rating agencies."

Based on this discussion of the relevant literature, this study provides a further contribution through the analysis of the usefulness and information content that the estimates of a structural model provide for credit rating prediction, when combined with accounting-based data. To this end, a non-parametric modeling approach is employed based on a multicriteria decision making (MCDM) technique. MCDM is well suited to the ordinal nature of credit ratings and the features of credit scorecards, while taking into account the nonlinearities observed in previous studies (Hwang et al., 2010; Mizen and Tsoukas, 2012) through an easy to comprehend additive modeling form. The robustness of the results is analyzed through the implementation of an approach that uses walk-forward model development and testing, together with the consideration of two alternative ways for modeling credit ratings (the first based on a multi-grading scheme and the second on a dichotomic investment vs. speculative rating setting). Furthermore, in contrast to previous studies that have heavily relied on US and UK data, we present empirical results for a sample of European companies from different countries over the period 2002–2012. During that period, and particularly after the outbreak of the European sovereign debt crisis, the role of CRAs has received much attention among authorities, regulators, and governments in Europe. Thus, it is particularly interesting to examine how the findings of studies conducted in other regions and time-periods translate into this context. The results indicate that the output of a market model provides significant additional information when compared to traditional financial data. However, its significance is significantly reduced when market capitalization is also considered.

The rest of the paper is organized as follows. Section 2 discusses the market model used in the analysis and the multicriteria approach employed for constructing the credit rating classification models. Section 3 is devoted to the description of the data set and the variables, whereas section 4 presents the empirical framework and analyzes of the obtained results. Finally, section 5 concludes the paper and discusses some future research directions.

2. Methodology

2.1. Market Model

The works of Black and Scholes (1973) and Merton (1974) led to the development of the research on structural models for credit risk modeling. In this framework (henceforth referred to as BSM), a firm is assumed to have a simple debt structure, consisting of a single liability with face value L maturing at time T. The firm defaults on its debt at maturity, if its assets' market value is lower than L. In this context, the firms' market value of equity (E) is modeled as a call option on the underlying assets (A), whose value is given by the Black-Scholes option pricing formula:

$$E = A\mathcal{N}(d_1) - Le^{-rT}\mathcal{N}(d_1 - \sigma\sqrt{T}) \tag{1}$$

where r is the risk-free rate, σ is the volatility of the asset returns, $\mathcal{N}(\cdot)$ represents the cumulative normal distribution function, and

$$d_1 = \frac{\ln(A/L) + (r + 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

Furthermore, under the Merton's assumption that equity is a function of assets and time, the following equation is derived from Itô's lemma (Hull, 2011):

$$\sigma_E = \frac{A}{E}\sigma\mathcal{N}(d_1) \tag{2}$$

Solving equations (1) and (2) simultaneously or with iterative procedures (Hillegeist et al., 2004; Vassalou and Xing, 2004) leads to an estimate of the market value of assets (A) and the volatility of the assets' return (σ). Then, a distance to default measure can be defined as the number of standard deviations that the firm is away from default (i.e., how much $\ln(A/L)$ should deviate from its mean in order for default to occur; Vassalou and Xing, 2004):

$$DD = \frac{\ln(A/L) + (\mu - 0.5\sigma^2)T}{\sigma\sqrt{T}} \tag{3}$$

where μ is the expected return on assets, which can be estimated from the annual changes in A obtained from the solution of equations (1) and (2). In the context of the basic BSM model, several variants have been introduced in the literature (see Agarwal and Taffler, 2008 for a comparative analysis).

2.2. Multicriteria analysis approach

In this study, the development of models to explain and predict credit ratings is based on a non-parametric multicriteria decision making (MCDM) approach. MCDM has evolved into a major discipline in operations research involved with decision problems under multiple criteria, and has been extensively used in various areas of financial risk management (Zopounidis and Doumpos, 2013), including credit scoring and rating (Doumpos and Pasiouras, 2005; Doumpos and Zopounidis, 2011). In this context, we introduce and employ a variant of the UTADIS multicriteria classification method (Doumpos and Zopounidis, 2002) in order to cope with the multi-class nature of credit ratings. We use an evaluation (scoring) model expressed in the form of an additive value function, which is widely used by financial institutions for credit scoring and rating (Krahnen and Weber, 2001):

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} w_k v_k(x_{ik}) \tag{4}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})$ is the data vector for firm i on K independent variables (evaluation criteria), w_k is the (non-negative) trade-off constant for criterion k (the trade-off are normalized to sum up to one), and $v_k(\cdot)$ is the corresponding marginal value function scaled in [0, 1]. The criteria trade-offs act as proxies for the relative importance of the independent attributes, whereas the marginal value functions decompose the overall evaluation result (i.e., credit score) into partial scores at the criteria level. The marginal value functions are non-decreasing for profit-related criteria and non-increasing for cost-related criteria and have a functional-free form (piece-wise linear) inferred directly through the model fitting process. This enables the additive model to capture the nonlinear (monotone) relationships between the independent attributes and the ratings of the firms. Nevertheless, in contrast to other popular nonlinear data mining algorithms (e.g., neural networks), the additive form of the model makes it easy to comprehend as it adopts the structure of a simple credit scorecard.

On the basis of its global value (i.e., credit score) as defined by (4), a firm i is classified into risk grade R_ℓ if and only if $t_\ell < V(\mathbf{x}_i) < t_{\ell-1}$, where $1 > t_1 > t_2 > \cdots t_{N-1} > 0$ are thresholds that distinguish a set of N ordered rating classes $R_1 \succ R_2 \succ \cdots \succ R_N$ (class R_1 is the low risk grade and R_N is the high risk one). Given a training sample consisting of m_ℓ observations from each rating class R_ℓ, the additive model can be developed by the solution of the following optimization problem:

$$\begin{aligned}
\min \quad & \sum_{\ell=1}^{N} \frac{1}{m_\ell} \sum_{i\in R_\ell} \sum_{h=1}^{N} (y_{ih}^{+} + y_{ih}^{-}) && && (5)\\
\text{s.t.} \quad & V(\mathbf{x}_i) + y_{i\ell}^{+} \geq t_\ell + \delta && \forall\, i \in \{R_1, \ldots, R_\ell\},\ \ell = 1, \ldots, N-1 && (6)\\
& V(\mathbf{x}_i) - y_{i\ell}^{-} \leq t_\ell - \delta && \forall\, i \in \{R_1, \ldots, R_\ell\},\ \ell = 2, \ldots, N && (7)\\
& t_\ell - t_{\ell+1} \geq \varepsilon && \ell = 1, \ldots, N-2 && (8)\\
& w_1 + w_2 + \ldots w_K = 1 && && (9)\\
& v_k(x_{ik}) - v_k(x_{jk}) \geq 0 && \forall\, x_{ik} \geq x_{jk} && (10)\\
& 0 \leq v_k(x_{ik}) \leq 1,\ w_k,\ y_{i\ell}^{+},\ y_{i\ell}^{-} \geq 0 && \forall\, i,\ k,\ \ell && (11)
\end{aligned}$$

The objective of this optimization formulation is to fit an additive scoring model to the data, so that the total weighted classification error (downgrade and upgrade errors) is minimized. The classification errors in the objective function are weighted by the number of training observations from each rating class, thereby avoiding the construction of a scoring model that is biased towards larger classes. Constraint (6) defines the downgrade errors $y_{i\ell}^{+}, y_{i,\ell+1}^{+}, \ldots, y_{i,N-1}^{+}$ for a firm i from rating class R_ℓ ($\ell = 1, \ldots, N-1$) as the violations of the lower bounds of classes $R_\ell, \ldots, R_{N-1}$. Similarly, constraint (7) defines the upgrade errors $y_{i2}^{-}, \ldots, y_{i\ell}^{-}$ for a firm i from rating class R_ℓ ($\ell = 2, \ldots, N$), as the violations of the upper bounds of the classes $R_2, \ldots, R_\ell$. In these constraints, δ is a user-defined small positive constant used to avoid arbitrary results that arise when the credit score of a firm is equal to one of the class limits. Constraint (8) ensures that class limits are non-increasing (with ε being a small positive constant), whereas (9) ensures that the attributes' trade-offs sum up to one. Finally, constraint (10) imposes the monotonicity of the marginal value functions (i.e., assuming that all attributes are expressed in maximization form).

Using a piece-wise linear modeling scheme for the marginal value functions, this optimization model can be expressed in linear programming form (for the details see Doumpos and Zopounidis, 2002), thereby allowing the fitting of model (4) to large scale data sets, which are common in credit scoring and rating.

3. Data and variables

The empirical analysis is based on a panel data set consisting of 1,325 firm-year observations involving European listed companies over the period 2002–2012. The sample covers eight different countries and five business sectors, as illustrated in Table 1.

[Table 1 about here.]

Financial data for the firms in the sample were collected from the Osiris database, whereas Bloomberg was used to get the firm's ratings from S&P. Due to the sparsity of the data set with respect to the number observations from each rating grade in the S&P scale, the observations were re-grouped. Two schemes are considered for this purpose. In the first one, the sample firms are grouped into five major risk classes, as follows: (1) class R_1 consisting of the lowest risk cases with ratings in the range from AA- up to AAA, (2) class R_2 consisting of cases with ratings A-, A, and A+, (3) classes R_3 with cases from low investment-level grades (BBB-, BBB, BBB+), (4) class R_4 with speculative ratings (BB-, BB, BB+), and (5) the high risk category R_5 that includes highly speculative ratings (i.e., D up to B+). Alternatively, we also consider a two-group setting distinguishing between speculative (D to BB+) and investment grades (BBB- to AAA). The percentage of observations in each rating group under these schemes is shown in Table 2.

[Table 2 about here.]

For every observation in the sample for year t, the S&P long-term rating is recorded at the end of June, while annual financial data are taken from the end of year $t-1$. The financial data involve four financial ratios, namely return on assets (ROA, profit before taxes/total assets), interest coverage (earnings before interest and taxes/interest expenses, EBIT/IE), solvency (equity/total assets, EQ/TA), and the long-term debt leverage ratio (equity/long term debt, EQ/LTD). ROA is the primary indicator used to measure corporate profitability. Interest coverage assesses the firms' ability to cover their debt obligations through their operating profits. The solvency ratio analyzes the capital adequacy of the firms, whereas the long-term debt leverage ratio takes into consideration the long-term debt burden of the firms in relation to their equity. In addition to these financial ratios, we also take into account the size of the firms, as measured by the logarithm of their market capitalization (CAP), and a country risk indicator (CRI) defined as a binary variable that takes the value of one for countries rated by S&P as AAA (in a given year) and zero otherwise.

The DD measure from the BSM model is also employed, estimated from the daily stock prices of the firms in year $t-1$. The risk-free rate for countries in the Eurozone is taken from the 3-month Euribor rate, whereas for UK and Switzerland, the 3-month treasury-bill and LIBOR rates are employed, respectively. The face value of debt is defined using current liabilities plus half the long-term debt. Finally, following similar simplifications to the ones of Bharath and Shumway (2008) and Li and Miu (2010), the annual equity returns (bounded below by the risk-free rate) and the corresponding volatilities over year $t-1$ are used as proxies of the expected return on assets and its volatility, respectively.

Table 3 presents the averages (over all years) of the financial variables and the distance to default (DD) indicator for each rating group in the sample. All variables have a clear monotone (increasing) relationship with the ratings level. Highly rated firms (e.g., rating group R_1) are more profitable, have higher interest coverage, are better capitalized and leveraged in terms of the long-term debt, and have higher market capitalization. The distance to default is also considerable larger for highly rated firms than for low-rated ones.

The differences between the rating groups are statistically significant at the 1% level for all variables under the Kruskal-Wallis non-parametric test. The country indicator was also found significant at the 5% level according to the χ^2 test ($p = 0.045$).

[Table 3 about here.]

The relationship between the selected variables and the rating of the firms is also verified with the Kendall's τ rank correlation coefficient, as illustrated in Table 4. The results are reported both for the five rating categories defined above as well as for the full rating scale of S&P. For comparison purposes, we also include the correlations for the logarithm of total assets (TA) as an alternative size indicator to market capitalization. Results for two alternative procedures for estimating the distance to default measure are also reported, including the iterative procedure of Vassalou and Xing (2004) as well as the model of Bharath and Shumway (2008). These are denoted as DD-VX and DD-BS, respectively. All correlations are found significant at the 1% level. Market capitalization has the strongest correlation with the ratings. Total assets are also strongly associated with the ratings, but the correlation is considerably weaker compared to capitalization, thus confirming that capitalization incorporates additional information when compared to accounting-based measures of a firm's size. Among the three estimates of distance to default, the one obtained with the procedure adopted in this study has the highest correlation with the ratings. Finally, it is also worth noting the marginal differences of correlations between the 5-grade scheme and the full rating scale of of S&P. These slight differences indicate that no significant loss of information occurs due to the adopted simplification of the rating scale.

[Table 4 about here.]

4. Results

4.1. Empirical setting

The robustness of the results is tested in a realistic setting by developing and validating a series of models through a walk-forward approach. In particular, the data for the period

2002–2005 are used first for model fitting, whereas the subsequent period 2006–2012 serves as the holdout sample. In a second run, the training data are extended up to 2006, and the holdout data span the period 2007–2012. The same process is repeated up to the case where the training data cover the period 2002–2010. Thus, six training and test runs are performed, which enable the analysis of the stability and the dynamics of the results over time. Henceforth, these walk-forward runs will be referred to as F05 (model fitting on data up to 2005) up to F10 (model fitting on data up to 2010).

Each run of the walk-forward approach develops four models. The first model (henceforth referred to as M1) is based solely on the four financial ratios and the country indicator. The second model (M2) additionally considers the distance to default. A third model (M3) extends M1 by adding capitalization, whereas the most comprehensive model (M4) combines all independent attributes. The consideration of these four models provides a decomposition of the results covering both the information content of the market model compared to accounting ratios, as well as its explanatory power as opposed to the size of the firms, which was found to be the strongest predictor.

In the section below, we analyze the results obtained using the 5-point rating classification described in the previous section. As a robustness test, the two-class investment-speculative scheme will be discussed in separate section.

4.2. Overall results

Table 5 summarizes the trade-offs of all variables in the models, averaged over all the six tests of the walk-forward approach. As a measure of the variability of the results over the six tests, the coefficients of variation are also reported (in parentheses). Under the two models that do not consider the capitalization of the firms (i.e., models M1 and M2), ROA and interest coverage (EBIT/IE) are the strongest variables, followed by solvency (EQ/TA), whereas the country risk rating indicator appears to be practically irrelevant in both models. The distance to default measure (DD) has a strong effect in model M2, with an average trade-off constant of 23.39%. The introduction of capitalization in models M3 and M4 has a significant impact on the relative importance of the attributes. In particular, both models show capitalization as the dominant factor, followed by ROA and interest coverage. The relative importance of the distance to default indicator in considerably weaker in model M4 than in model M2, which ignores capitalization. This result is consistent with the finding on bankruptcy prediction of Agarwal and Taffler (2008), who used a data set of UK bankruptcies and concluded that market-based models do not carry much information over simpler market variables, in particular market capitalization.

The coefficients of variation of the variables' trade-offs are generally low in all models. This indicates that the conclusions drawn above are robust over time (i.e., under the six tests conducted through the walk-forward approach). Details on the evolution of the variables' relative importance over the six walk-forward tests are illustrated in Figure 1 for all variables except for the country indicator, which as explained above makes only a minor contribution in the models. The relative importance of ROA and interest coverage clearly follows a slightly decreasing trend over time. On the other hand, the trade-off of long-term leverage steadily increases. The same is also observed for capitalization, whose relative importance appears higher in the walk-forward tests that incorporate data for 2008–2010. The trade-off of the distance to default follows an increasing trend under model M2, but the introduction of capitalization in model M4 maintains it at an almost steady level around 10%. The same behavior is also evident for the solvency ratio.

[Table 5 about here.]

[Figure 1 about here.]

Table 6 presents the results of the comparison of the classification accuracies of all models in the holdout samples. The rows in the two panels correspond to the six walk-forward tests, while the columns correspond to the years in the holdout samples. For each year in the holdout samples, panel A presents the differences between the classification accuracy of model M2 as opposed to model M1, whereas panel B presents the same information for the comparison of M4 to M3. The two last columns of the table present the overall classification accuracies of the models for the full holdout panel data.

[Table 6 about here.]

Both models M1 and M2 clearly perform poorly as their classification accuracies are below 50%. The accuracies for model M1, which only considers the four financial ratios and the country indicator, range between 37–40%. The introduction of DD into model M2 improves the results (over M1) by 4.5% on average (according to the McNemar's test, the differences between the two models are statistically significant at the 5% level except for the walk-forward test F07 which is based on the data up to 2007). The comparisons of the two models by year indicate that M2 performs almost consistently better than M1 in all cases and it is significantly better in 2011 and 2012. The introduction of capitalization in models M3 and M4 considerably improves the results. Model M3 outperforms M1 by more than 18% on average and M4 performs better than M2 by more than 15%. When compared to model M3, which combines the financial ratios, the country indicators, and capitalization, the introduction of DD in model M4 provides a slight improvement of 1.2% on average, but the differences are not significant at the 5% level according to the McNemar test.

Table 7 presents the overall average classification matrix (from all years) for model M4, as well as the mean absolute error (MAE) for each rating category. MAE indicates the mean absolute notch difference between the classifications of the model and the actual ratings of the sample observations. All misclassifications are evidently restricted to two notches. In fact, 94% of the errors involve one-notch misclassifications. The downgrades from the investment grades (R_1–R_3) to speculative classes are 18% of the total number of observations in the investment classes, whereas the upgrade rate (from speculative to investment grades) is 23% of the total number of observations in the speculative categories R_4 and R_5.

[Table 7 about here.]

The classification results are further tested using the area under the receiver operating characteristic curve (AUROC). The AUROC is usually employed to measure the discriminating power of two-class credit scoring and rating models; however, it can be extended to multi-class cases, such as the one considered in this study. To this end, the generalization proposed by Hand and Till (2001) is employed, which indicates the probability that a firm from any rating category R_ℓ ($\ell = 1, \ldots, N-1$) has a higher evaluation score (credit score) compared to any other firm with a worse rating (i.e., from any of the classes $R_{\ell+1}$, ..., R_N). Figure 2 illustrates the average AUROC for all models. The results are in accordance with the observations made above regarding the classification accuracies. In particular, model M2 consistently outperforms M1 (by an average of 2.3%), whereas the two models that take capitalization into account (M3 and M4) are almost indistinguishable, as their differences range between 0.04% and 0.3%.

[Figure 2 about here.]

4.3. Speculative vs investment grades

We tested the robustness of the results presented above by also considering a binary classification scheme, based on discriminating between speculative and investment grades. Table 8 summarizes the trade-offs of the variables in each of the four models. The results confirm the findings discussed previously (e.g., Table 5). In particular, among the financial ratios ROA, interest coverage, and solvency have the highest relative importance in the models. The country indicator indicator once again has marginal relevance. The distance to default indicator has a relative contribution in model M2 equal to 28.24%, which is quite similar to its trade-off (23.39%) in same model with the five-class scheme. Capitalization is again the dominant factor in models M3 and M4, with its trade-off being consistently higher than 40% and exhibiting low variability over the six walk-forward tests (as indicated by the coefficient of variation, which is equal to 0.03). As with the observation made above for model M2, the relative importance of distance to default in model M4 under this two-class setting is almost identical to the multi-grade specification discussed in the previous section (i.e., 10.94% in the two-class scheme vs 10.7% in the five-grade one).

[Table 8 about here.]

Details on the classification results (accuracy rate and AUROC) are shown in Figure 3. As seen with the multi-class setting, the introduction in model M2 improves both the classification accuracy (by 1.77% on average) and AUROC (by 1.6% on average), with the improvements becoming larger in the walk-forward tests based on the most recent data. On the other hand, the introduction of DD provides no noticeable benefit when capitalization is already included in the model. Under the most comprehensive model M4, the overall (average) accuracy is 77%, with the downgrade error rate being 23.5% and the update error being 21.8%.

[Figure 3 about here.]

5. Conclusions and future perspectives

The analysis of the ratings issued by CRAs has received much attention in the financial literature, due to their significance in the context of credit risk management and their widespread use by investors, policy makers, and managers. In this study, we sought to explain and predict the credit ratings issued by S&P on the basis of financial and market data, using a cross-country panel data set from Europe over the period 2002–2012. The BSM structural model was employed to obtain distance to default estimates and their information content in predicting credit ratings was compared to models based on financial and market capitalization data. For the analysis, we employed a multicriteria decision making technique, based on a linear programming formulation for fitting additive credit scoring models to the data. The results demonstrate that even though the BSM model significantly enhances the predictions based solely on financial ratios, its information power is considerably weaker when market capitalization is considered. Furthermore, the relative importance of market capitalization and long-term leverage has increased after the outbreak of the crisis, whereas the relative importance of profitability ratios decreased. The developed multicriteria models showed good and robust behavior, with most classification errors restricted to one-notch divergences from the actual ratings.

These empirical results could be extended in a number of directions. Among others, these include: (a) the combination of financial data, structural models, and credit ratings to obtain improved default probability estimates, (b) the consideration of additional predictor attributes including macroeconomic factors, data from the CDS markets, and variables related to corporate governance, and (c) the extension of the analysis to non-listed companies.

References

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. Journal of Banking and Finance 32 (8), 1541–1551.

Altman, E. I., Saunders, A., 1997. Credit risk measurement: Developments over the last 20 years. Journal of Banking and Finance 21 (11-12), 1721–1742.

Bharath, S. T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. Review of Financial Studies 21 (3), 1339–1369.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81 (3), 637–659.

Campbell, J. Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. The Journal of Finance 63 (6), 2899–2939.

Chen, J., Hill, P., 2013. The impact of diverse measures of default risk on UK stock returns. Journal of Banking and Finance.

Das, S. R., Hanouna, P., Sarin, A., 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. Journal of Banking and Finance 33 (4), 719–730.

Doumpos, M., Pasiouras, F., 2005. Developing and testing models for replicating credit ratings: A multicriteria approach. Computational Economics 25 (4), 327–341.

Doumpos, M., Zopounidis, C., 2002. Multicriteria Decision Aid Classification Methods. Kluwer Academic Publishers, Dordrecht.

Doumpos, M., Zopounidis, C., 2011. A multicriteria ouranking modeling approach for credit rating. Decision Sciences 42 (3), 721–742.

Du, Y., Suo, W., 2007. Assessing credit quality from the equity market: can a structural approach forecast credit ratings? Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration 24 (3), 212–228.

Frost, C. A., 2007. Credit rating agencies in capital markets: A review of research evidence on selected criticisms of the agencies. Journal of Accounting, Auditing and Finance 22 (3), 469–492.

Hand, D. J., Till, R. J., 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. Machine Learning 45 (2), 171–186.

Hillegeist, S. A., Keating, E. K., Cram, D. P., Lundstedt, K. G., 2004. Assessing the probability of bankruptcy. Review of Accounting Studies 9 (1), 5–34.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems 37 (4), 543–558.

Hull, J., 2011. Options, Futures, and Other Derivatives, 8th Edition. Prentice Hall, New Jersey.

Hwang, R.-C., 2013. Forecasting credit ratings with the varying-coefficient model. Quantitative Finance, 1–19.

Hwang, R.-C., Chung, H., Chu, C., 2010. Predicting issuer credit ratings using a semiparametric method. Journal of Empirical Finance 17 (1), 120–137.

Jeon, D.-S., Lovo, S., 2013. Credit rating industry: A helicopter tour of stylized facts and recent theories. International Journal of Industrial Organization.

Krahnen, J. P., Weber, M., 2001. Generally accepted rating principles: A primer. Journal of Banking and Finance 25 (1), 3–23.

Li, M.-Y. L., Miu, P., 2010. A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. Journal of Empirical Finance 17 (4), 818–833.

Lu, H.-M., Tsai, F.-T., Chen, H., Hung, M.-W., Li, S.-H., 2012. Credit rating change modeling using news and financial ratios. ACM Transactions on Management Information Systems 3 (3), 14:1–14:30.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance 29 (2), 449–470.

Mizen, P., Tsoukas, S., 2012. Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model. International Journal of Forecasting 28 (1), 273–287.

Pagano, M., Volpin, P., 2010. Credit ratings failures and policy options. Economic Policy 25 (62), 401–431.

Pasiouras, F., Gaganis, C., Zopounidis, C., 2006. The impact of bank regulations, supervision, market structure, and bank characteristics on individual bank ratings: A cross-country analysis. Review of Quantitative Finance and Accounting 27 (4), 403–438.

Tichy, G., Lannoo, K., Gwilym, O., Alsakka, R., Masciandaro, D., Paudyn, B., 2011. Credit rating agencies: Part of the solution or part of the problem? Intereconomics 46 (5), 232–262.

Treacy, W. F., Carey, M., 2000. Credit risk rating systems at large US banks. Journal of Banking and Finance 24 (1-2), 167–201.

Vassalou, M., Xing, Y., 2004. Default risk in equity returns. The Journal of Finance 59 (2), 831–868.

Zopounidis, C., Doumpos, M., 2013. Multicriteria decision systems for financial problems. TOP 21 (2), 241–261.

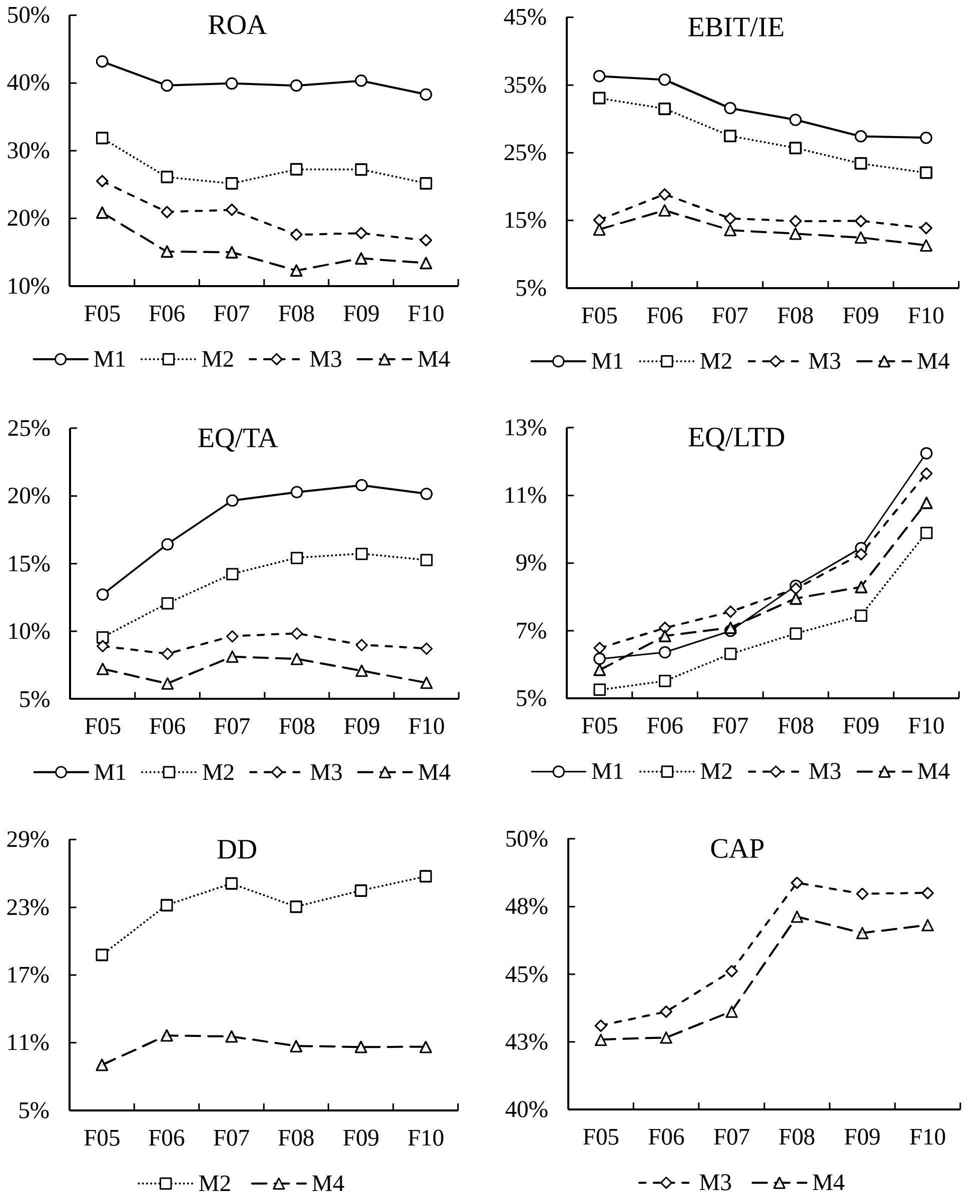

Figure 1: The trade-offs of the variables over time

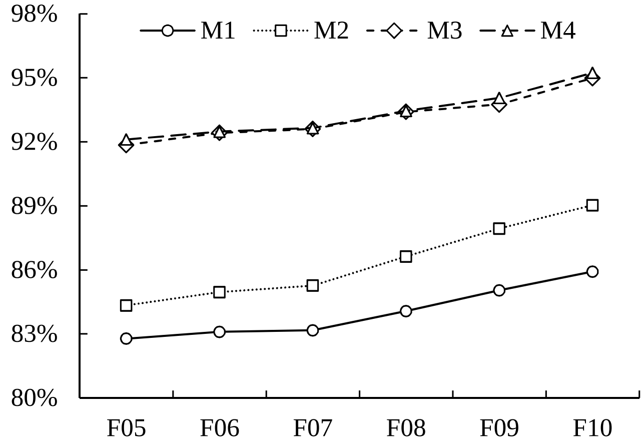

Figure 2: AUROC of the models, averaged over all years

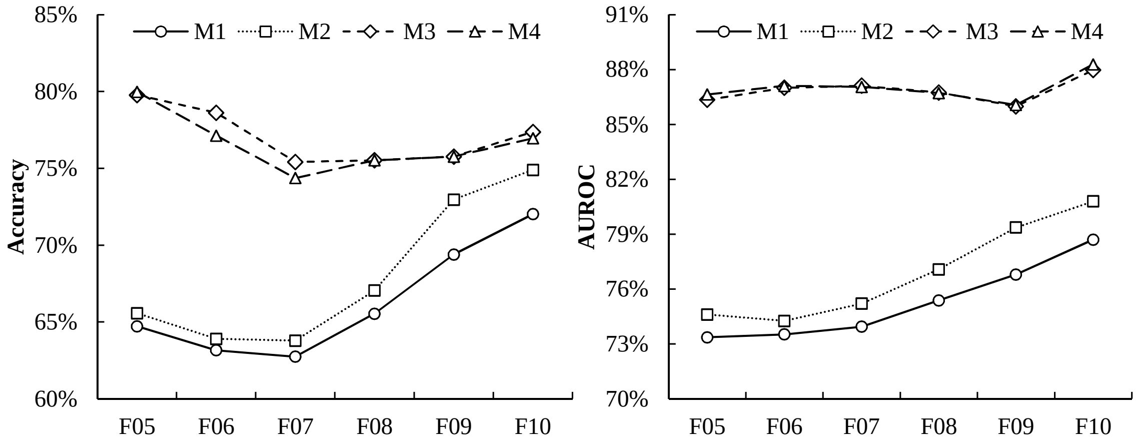

Figure 3: Classification accuracy and AUROC for the two-class models (averaged over all years)

Table 1: Sample composition (number of observations) by year, country, and business sector

Year		Country		Sector	
2002	38	Germany	308	Manufacturing	853
2003	102	France	303	Information & communication	220
2004	115	UK	298	Wholesale & retail trade	130
2005	126	Switzerland	135	Transportation & storage	90
2006	135	Netherlands	130	Construction	32
2007	138	Italy	88		
2008	140	Spain	33		
2009	139	Belgium	30		
2010	149				
2011	151				
2012	92				

Table 2: Percentage of sample observations in each risk category

	R_1	R_2	R_3	R_4	R_5	Investment	Speculative
2002	5.3	31.6	39.5	18.4	5.3	76.3	23.7
2003	5.9	30.4	43.1	12.7	7.8	79.4	20.6
2004	7.8	31.3	39.1	13.0	8.7	78.3	21.7
2005	7.1	28.6	39.7	16.7	7.9	75.4	24.6
2006	7.4	24.4	43.7	20.0	4.4	75.6	24.4
2007	7.2	23.9	45.7	19.6	3.6	76.8	23.2
2008	5.7	24.3	42.1	22.9	5.0	72.1	27.9
2009	4.3	25.9	41.0	20.1	8.6	71.2	28.8
2010	4.0	24.8	41.6	22.1	7.4	70.5	29.5
2011	4.0	23.2	43.7	23.2	6.0	70.9	29.1
2012	5.4	27.2	45.7	15.2	6.5	78.3	21.7
Overall	5.8	26.3	42.4	19.0	6.5	74.5	25.5

Table 3: Averages of independent variables by rating group

	R_1	R_2	R_3	R_4	R_5	Invest.	Specul.
ROA	13.69	8.02	6.58	4.26	−2.79	7.64	2.47
EBIT/IE	18.33	8.42	5.54	3.12	0.54	7.56	2.46
EQ/TA	46.86	34.74	30.48	27.82	21.33	1.23	0.85
EQ/LTD	2.29	1.27	1.06	0.87	0.78	33.26	26.17
CAP	18.01	16.75	15.74	14.65	13.12	16.27	14.26
DD	8.96	5.94	4.17	2.89	1.31	5.17	2.49

Table 4: Kendall's τ rank correlations with the credit ratings

	5 risk grades	Full rating scale
ROA	-0.335	-0.327
EBIT/IE	-0.391	-0.372
EQ/TA	-0.199	-0.181
EQ/LTD	-0.202	-0.186
CAP	-0.565	-0.570
TA	-0.404	-0.396
DD	-0.382	-0.375
DD-VX	-0.361	-0.352
DD-BS	-0.370	-0.362

Table 5: Trade-offs (in %) of the independent variables (averages over all years with coefficients of variation in parentheses)

	ROA	EBIT/IE	EQ/TA	EQ/LTD	CI	DD	CAP
M1	40.17 (0.04)	31.37 (0.13)	18.34 (0.17)	8.26 (0.28)	1.86 (0.09)	–	–
M2	27.13 (0.09)	27.20 (0.16)	13.71 (0.18)	6.89 (0.24)	1.67 (0.07)	23.39 (0.11)	–
M3	19.98 (0.16)	15.49 (0.11)	9.08 (0.06)	8.38 (0.22)	1.03 (0.08)	–	46.03 (0.05)
M4	15.12 (0.20)	13.43 (0.13)	7.13 (0.12)	7.80 (0.22)	0.93 (0.13)	10.70 (0.09)	44.89 (0.05)

Table 6: Comparison of classification accuracies

	Panel A: M2 - M1 differences							Overall accur.	
	2006	2007	2008	2009	2010	2011	2012	M1	M2
F05	−2.22	2.90	−0.71	2.88	4.03	4.64	11.96	38.77	41.74
F06		6.52	−2.14	2.16	3.36	5.30	14.13	37.70	42.03
F07			−2.14	0.72	0.67	4.64	9.78	37.11	39.34
F08				0.00	3.36	3.97	14.13	39.17	43.69
F09					4.70	5.30	10.87	40.05	46.43
F10						5.96	8.70	39.92	46.91
	Panel B: M4 - M3 differences							M3	M4
F05	0.74	5.07	−0.71	−1.44	1.34	2.65	2.17	54.03	55.40
F06		1.45	−0.71	−0.72	1.34	3.97	−2.17	57.97	58.71
F07			0.00	−2.16	0.67	4.64	0.00	56.18	56.93
F08				−0.72	2.68	2.65	−1.09	56.87	58.00
F09					2.68	3.31	−2.17	58.42	60.20
F10						0.66	2.17	60.91	62.14

Table 7: Average classification matrix and mean absolute error (MAE) for model M4

		Estimated rating					
		R_1	R_2	R_3	R_4	R_5	MAE
Actual rating	R_1	90.23%	6.32%	3.45%	0.00%	0.00%	0.13
	R_2	17.53%	48.09%	31.35%	3.03%	0.00%	0.55
	R_3	1.49%	12.60%	57.82%	26.94%	1.16%	0.45
	R_4	0.00%	1.60%	27.67%	60.43%	10.29%	0.41
	R_5	0.00%	0.00%	2.17%	35.65%	62.17%	0.40

Table 8: Average trade-offs (in %) of the independent variables (coefficients of variation in parentheses) under the two-class setting

	ROA	EBIT/IE	EQ/TA	EQ/LTD	CI	DD	CAP
M1	36.38 (0.15)	31.85 (0.16)	20.51 (0.36)	9.47 (0.23)	1.78 (0.09)	–	–
M2	23.03 (0.23)	25.26 (0.28)	14.91 (0.36)	6.79 (0.22)	1.76 (0.21)	28.24 (0.25)	–
M3	14.78 (0.27)	18.24 (0.13)	11.94 (0.44)	7.32 (0.10)	1.48 (0.44)	–	46.24 (0.03)
M4	11.26 (0.22)	15.50 (0.18)	10.34 (0.46)	6.41 (0.12)	1.31 (0.44)	10.94 (0.19)	44.24 (0.03)