Available online at www.sciencedirect.com
ScienceDirect
Journal of Banking & Finance 32 (2008) 434–446
Journal of BANKING & FINANCE Applied Section
www.elsevier.com/locate/jbf

Improving performance of corporate rating prediction models by reducing financial ratio heterogeneity ☆

Martin Niemann *, Jan Hendrik Schmidt, Max Neukirchen

German Office of McKinsey & Company, Inc., Sophienstrasse 26, 80333 Munich, Germany

Received 31 May 2005; accepted 15 May 2007
Available online 15 September 2007

Abstract

We introduce a new approach to improve the performance of rating prediction models for multinational corporations. In this segment, the low number of defaults poses a challenge, as it prevents rating models to be constructed for individual industry sectors or regions. We show that reducing group-level heterogeneity in financial ratios results in a rating prediction model with better performance than both unadjusted models and models adjusted by including industry dummies or other simpler procedures. Our approach fills a gap in cases where a limited dataset does not permit the construction of separate models for individual industries or regions.

JEL classification: G33; C65

Keywords: Rating models; Default prediction models; Financial ratios; Industry effects; Heterogeneity

1. Introduction

Rating models have become ubiquitous in banking in the decades since Horrigan (1966) and Altman (1968) published the first major rating and default prediction studies. Statistical models based on financial statements have increasingly supported and complemented the decisions of banks' credit analysts. The Basel II capital accord has further stimulated their advancement, as it requires banks to use model-driven estimates of the probability of default (PD)[1] as a basis for risk-based capital allocation.

In the context of Basel II implementation, the credit customer segment of multinational corporations poses a special challenge to the construction of rating models since it lacks current default data: large corporations experience too few defaults to provide sufficient data to construct models differentiating directly between defaults and non-defaults. Banks follow two common alternative paths to address this problem. Many use Merton-type models that estimate PDs based on companies' stock volatilities and capital structures. However, outside the US, a considerable share of large companies is still unlisted and thus, especially European banks have to rely on financial statement data to rate credit exposures. For this approach, emulating ratings of international agencies such as Standard and Poor's (S&P) is a feasible option. It alleviates the lack of default information by introducing 22 classes of credit quality for modeling. Still, the problem of data availability remains, given that the S&P-rated universe contains only roughly 2000 non-financial corporations.[2] While this population – as a whole – is sufficient to obtain statistically significant regression models, it is again too small to construct specific

☆ This paper was reviewed and accepted during the tenure of the past Editorial Board in which professor Giorgio Szego acted as Managing Editor of the Journal of Banking and Finance.

* Corresponding author. Tel.: +49 89 5594 8214; fax: +49 89 5594 8215.
E-mail address: Martin_Niemann@mckinsey.com (M. Niemann).

[1] For the purpose of this paper, PD refers to a one-year probability of default, in concordance with paragraph 285 of the Basel II capital accord (see Bank for International Settlements, 2006).

[2] Based on our sample of S&P ratings from 2003.

0378-4266/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2007.05.015

models for each industry sector or region. We therefore need to integrate data from several industries and countries, which leads to a high degree of sectoral and regional heterogeneity in the sample: Every subgroup's financials exhibit a potentially different relationship to ratings, leading to suboptimal model performance. This is the *financial ratio heterogeneity* we seek to address with this paper.

For banks, the quality of rating systems is a key determinant of profitability for their credit business. Given the large loan and bond portfolios held by financial institutions, small improvements in model performance translate into substantial economic benefits: banks profit from lower loan losses, more accurate risk-adjusted pricing, and a more efficient utilization of regulatory and economic capital.

In the search for better model performance, prior rating studies have mostly focused on assessing the predictive power of different classification algorithms. This has led to a multitude of available model types for predicting PDs, given a sufficiently large dataset of historic defaults. However, the challenges of smaller, heterogeneous datasets call for approaches beyond the choice of classification algorithm. Techniques to enhance the significance and informational content of predictive factors offer further potential.

This paper presents a new approach to increase the performance of rating prediction models for multinational corporations by reducing group-level heterogeneity in financial ratios. We start with a brief survey of past rating and default prediction studies in Section 2. Thereby we focus on methods to enhance the significance of predictors and resulting models, rather than evaluating the performance of different model types. In Section 3, we introduce a new technique to measure financial ratio heterogeneity caused by sectoral and regional subgroups in the sample, and to perform effective adjustments to reduce its impact on predictor variables. In Section 4, we construct a set of rating models to compare the performance of our approach with prior methods reducing heterogeneity. We present our conclusions in Section 5.

2. Overview of previous rating studies

When developing a rating model, several key levers determine the attainable performance of the model (see also Liu, 2002). Collecting high quality data is the obvious precondition for a working model. After that, three main levers for rating model performance remain. The first and most intuitive is the *classification algorithm* used for model development. Consequently, most default prediction studies focused on testing the discriminative power of different classification techniques. The second performance lever is *factor definition*, i.e. the definition and analytical adjustment of financial ratios for their use as predictor variables. Finally, *statistical factor transformation* changes the statistical attributes of predictor variables to make them more suitable for the chosen classification technique and enhance their statistical significance. The next three sections briefly survey past rating studies and analyze where methodological innovations occurred for each of the three performance levers.

2.1. Choice of classification algorithm

For solving the question of credit quality, the choice of the ideal *classification algorithm* has attracted by far the most attention from scholars. In a comprehensive study, Aziz and Dar (2006) gather and compare the predictive accuracy of models developed in 89 default prediction studies. They divide the universe of classification algorithms into three broad categories: statistical, artificial intelligence expert systems (AIES), and theoretic models. They find that the majority of default risk studies use statistical techniques such as multinomial discriminant analysis (MDA), followed by logit. Among AIES models, neural networks and recursive partitioning algorithms are most commonly used. The most prominent theoretic model type is the option-based approach developed by Merton (1974) based on Black and Scholes' (1973) theory of option pricing. Aziz and Dar do not find significant performance differences between classification algorithms; still, AIES models reach a marginally better average performance. Balcaen and Ooghe (2004) confirm the largely uniform performance, but their conclusion contradicts Aziz and Dar: "Although the alternative [AIES] methods are computationally more complex and more sophisticated than the classical cross-sectional statistical methods, it is not clear whether they produce better performing corporate failure prediction models".

In sum, given today's state of the art in classification algorithms, the choice of model type provides little opportunity for further improvements in the performance of rating systems. This decision is mostly dependent on the available data and the classification task. Consequently, methods to extract rating-relevant information from the dataset, and enhance the statistical significance of predictors, provide the remaining potential for performance increases.

2.2. Factor definition

The second performance lever, *factor definition*, includes all conversions of raw data into statistically meaningful predictors, based on economic reasoning and justifiable through economic theory. The primary application is financial ratio analysis. While most bankruptcy and rating literature has relied on combining standard "accounting book ratios", several studies have developed promising methods to extract more rating-relevant information from financial data. Belkaoui (1980) finds that using *economic categorization* of predictors increases model performance over previous models. Choosing the best ratio from each category, e.g. debt coverage ratios, tends to increase the amount of uncorrelated information in the model, improving its performance. The utility of *multi-year transformations* such as averages, trends, and volatilities in the development of rating models has not yet been the focus of a published rating study, although some scholars have employed them successfully. Fisher (1959) uses nine-year earnings volatility as an indicator of business risk, and in most models by Perry et al. (1984), a five-year average debt-to-equity ratio outperforms its one-year version.

Further to the *definition* of predictive factors, *analytical adjustments* offer potential to increase their predictive power. Removing one-time effects or reconciling between accounting standards are common examples. In addition, several rating studies use adjustments to manage *heterogeneity* of financial ratios, i.e. different statistical relationships between predictive factors and PD due to regional or industry subgroups in the sample.[3] Three common approaches to managing this type of heterogeneity are *model segmentation*, *dummy variables*, and *financial ratio adjustments*.

The first approach is the *segmentation* of a rating model into several submodels for industry sectors and geographic regions. Perry et al. (1984) find that the combined results of several sector-specific models outperform a pooled model. This approach requires, however, a sufficiently large sample to obtain valid regression models for subgroups. The second approach to reduce heterogeneity is to introduce *dummy variables*. Chava and Jarrow (2004) include industry dummies for both slope and intercept parameters in their logistic bankruptcy prediction model. Although most coefficients of industry dummy variables are significant, their inclusion only provides "a modest improvement" in classification accuracy. The third method is to perform *financial ratio adjustments* for industry and country effects. Horrigan (1966) adjusts for industry effects by dividing the difference between a company's ratio and the industry average by the industry average, thus obtaining a standardized relative measure across industries. Comparably, Moody's (2001a) perform country adjustments by subtracting the country average from companies' ratios and normalizing distributions with the country standard deviation.

2.3. Statistical factor transformation

Statistical factor transformation changes the statistical properties of predictor variables to suit the requirements of the chosen classification algorithm, e.g. normality, and increase model performance. Kane and Richardson (1998) point out that "accounting data do not meet the specifications for most statistical models". Restrictive requirements of classification algorithms such as MDA motivated the first studies focusing on methods of statistical data transformation. These and other studies developed four main approaches to deal with non-normality and enhance the statistical significance of predictors: *mathematical transformations* to normalize distributions, *outlier removal*, *rank transformation*, and *univariate submodels*.

First, *mathematical transformations* such as square root and log-transformations are found by Deakin (1976) to accomplish approximate normality in some cases. Watson (1990) improves this approach by using a family of modified power transformations suggested by Box and Cox (1964):

$$x'(\lambda) = \frac{(x^{\lambda} - 1)}{\lambda} \quad \text{for } \lambda \neq 0,\ x > 0 \quad \text{and}$$
$$x'(\lambda) = \ln(x) \quad \text{for } \lambda = 0,\ x > 0. \tag{1}$$

It attains a high variety of functional forms, depending on the value of the transformation factor λ, whereby x represents a financial ratio's values. In most cases, this approach is able to generate symmetrical, approximate-normal distributions (Watson, 1990). However, Frecka and Hopwood (1983) find that "the inclusion of outliers can produce a [...] potentially severe impact on the parameter estimates [of statistical models]", which cannot be fully compensated by mathematical transformations. Therefore, techniques such as *outlier removal* through "winsorizing", i.e. replacing extreme values by substitute values (Copeland and Ingram, 1982, in Watson, 1990), and other trimming procedures help to bring distributions sufficiently close to normality for they can be used in statistical models (Frecka and Hopwood, 1983). The third approach switches to non-parametric statistics. Kane and Richardson (1998) find that *rank transformation* is a powerful alternative when dealing with outliers and non-normality. While this procedure eliminates the information about absolute distances between data points, companies' relative ranks tend to stay constant over time. Statistical models using ranked values are more robust against cyclical effects, such as recession, compared to models using unranked data. *Univariate submodels* constitute the fourth technique to enhance statistical stability. Moody's models the relationship between individual factors and default, fitting a logit function to the data that smoothes the raw data and caps the extreme values. With its asymptotic symmetrical form, logistic regression is suitable to achieve the truncation and normalization of input ratios. Multivariate models based on univariate submodels are more robust, because outliers have a smaller impact on parameter estimates (Moody's Investors Service, 2001b).

In sum, scholarly attention has largely focused on the first performance lever, the choice of *classification algorithm*. However, at today's stage of advancement, this decision is often more a matter of practicality than a performance driver. Regarding *factor definition* and *statistical factor transformation*, past studies found better-performing predictors and proved the benefits of economic categorization and model segmentation. Another line of studies focused on ensuring that the statistical assumptions of classification algorithms be met, which led to several valuable methods to improve the distributional properties

[3] This definition of the term *heterogeneity* differs from the one used in the context of hazard/survival models. Thereby the influence of *unobserved* heterogeneity (or frailty) on default probabilities is modeled using frailty variables. Chava et al. (2006) provide a recent example of this methodology. Our paper focuses on reducing *observed* heterogeneity due to sectoral or regional subgroups in the sample.

of predictors. While methods to reduce sectoral and regional effects have been shown to improve model accuracy, no study has yet focused on addressing the effects of financial ratio heterogeneity.

3. A methodology for measuring and managing financial ratio heterogeneity

Having identified an opportunity for improvement in model performance, our focus is on the reduction of *financial ratio heterogeneity* as a further performance lever. We develop an indicator of heterogeneity that provides a basis to make effective model segmentation decisions and to perform financial ratio adjustments.

3.1. Definition of financial ratio heterogeneity

Financial ratio heterogeneity, for the purpose of this paper, stems from diverse statistical relationships between corporate financial data and companies' ratings. It arises when financial statement structures differ between subgroups in the sample, because of common distinguishing factors. For instance, *industry sectors* can differ substantially in terms of balance-sheet structure, e.g. fixed asset intensity, and profitability standards. At the *regional* level, different interest rate environments, tax regimes, wage levels, and access to capital markets are prominent drivers of typical profitability levels and capital structures. These affect corporate ratings accordingly. For instance, S&P imposes caps on the ratings of otherwise financially sound companies, if they face high industry or country risk (Standard & Poor's, 2003).

The main problem with a heterogeneous sample is that the same rating coincides with substantially different values of financial ratios across different subgroups, and vice versa. For instance, 'AAA'/'AA'-rated Consumer Discretionary companies have a median 3-year-average profit margin of 5%, whereas 'AAA'/'AA' rated Healthcare companies have a median margin of 18%. Conversely, if a Healthcare company had a profit margin of 5%, its expected rating would be 'BBB'. It is obvious that this dispersion reduces the statistical significance of predictors for use in regression models.

3.2. An indicator of financial ratio heterogeneity

For this type of heterogeneity, one intuitive indicator is the *overlap* of two subgroups' distributions for a given level of risk. If a financial ratio is distributed identically in two sectors, the overlap of their distributions will be 100%, and if the sample consisted only of these two groups, it would be perfectly homogeneous for that ratio. Conversely, for very dissimilar groups, the overlap of financial ratio distributions approaches zero.

Consider the distributions of a financial ratio variable X_i for two subgroups – as defined by a categorical variable, such as industry sector or country – and a given level of credit risk,[4] i.e. a rating grade. The overlap measure can be based, for example, on the hypothesis test for the difference between the means of two samples drawn randomly from a population: Its confidence level correlates strongly with the distribution overlap. Since this heterogeneity test assumes normality, we need to obtain at least approximate-normal, symmetric distributions. The studies reviewed in Section 2.3 provide a variety of statistical transformation methods for this purpose. One viable solution is the family of modified power transformations proposed by Box and Cox. To prepare financial ratio distributions for the procedure, we first winsorize financial ratio distributions by replacing all values lower than the 0.5% percentile and higher than the 99.5% percentile with the respective values of the two boundaries. We then apply the Box–Cox transformation from Eq. (1) to obtain approximate-normal financial ratio distributions, where $x'_{ij}(\lambda)$ is the normalized value for financial ratio i of company j, and is the Box–Cox transformation factor for each financial ratio:

$$x'_{ij}(\lambda) = \frac{(x_{ij}^{\lambda_i} - 1)}{\lambda_i} \quad \text{for } \lambda_i \neq 0,\ x_{ij} > 0 \quad \text{and}$$
$$x'_{ij} = \ln(x_{ij}) \quad \text{for } \lambda_i = 0,\ x_{ij} > 0. \tag{2}$$

In line with the use of proxy distributions for the subsequent calculation of heterogeneity scores, a ratio's optimal λ_i is the transformation factor that causes a ratio's median $X_{i0.5}$ to lie in the middle of the quantiles $X_{i0.1}$ and $X_{i0.9}$.[5] The resulting normalized variables X'_i form the basis for the mean difference test.

We approximate the difference of the two subgroups' (indicated by suffixes 1 and 2) sample means using their medians as substitutes, which are approximately equal to the means after performing the Box–Cox transformation. The sign of the difference is irrelevant for the measurement of overlap, hence we set

$$d_{\overline{X}'} = -\left|\overline{X}'_1 - \overline{X}'_2\right|. \tag{3}$$

Our null-hypothesis for the difference test is $E(X'_1) = E(X'_2)$. The difference $d_{\overline{X}'}$ is an occurrence of the random variable $D_{\overline{X}'}$, which is the distribution of the differences of randomly drawn sample-means from a population, and follows approximately a normal distribution with the following parameters:

$$\mu_{D\bar{x}'} = \overline{X}'_1 - \overline{X}'_2 = 0, \tag{4}$$

and

$$\sigma_{D\bar{x}'} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \tag{5}$$

[4] For the construction of our homogenized model in Section 4, we use the ten S&P industry sectors (excluding financial institutions), leaving nine industry groups for model development.

[5] We use the value range between 10% and 90% quantiles, because measured standard deviations would otherwise be very high due to remaining extreme values in the sample. More elaborate methods such as the χ^2-test would not lead to substantial improvements for our goal of achieving approximate-normal, symmetric distributions as an input for our heterogeneity indicator.

where $\sigma^2_{D\bar{x}'}$ is the pooled variance of the two sample means and n_1 and n_2 are the respective subgroups' sample sizes. The variances σ^2_1 and σ^2_2 of the two samples are approximated by using the values of their 10%- and 90%-quantiles,

$$\sigma^2 = \left(X'_{0.9} - X'_{0.1}\right)^2 \tag{6}$$

for both subgroups.[6] Inserting $d_{\overline{X}'}$ in the cumulative normal distribution function Φ gives

$$\Phi\left(\frac{d_{\overline{X}'}}{\sigma_{D\bar{x}'}}\right) = \frac{\alpha}{2}, \quad \text{whereby } 0 < \alpha < 1. \tag{7}$$

Defining the *heterogeneity score* as $H = 1 - \alpha$ delivers an easy-to-interpret indicator, bounded by 0% and 100%: the higher the heterogeneity score, the greater the difference between the two subgroups.

Using the nine S&P sectors (see Appendix 2) in our development sample as an example for the use of this indicator, we get a total of 36 unique industry pairs, whereby we consolidate the 22 S&P rating classes into five broad categories[7] to achieve adequate sample sizes for similar levels of credit risk. With the resulting matrix of 180 heterogeneity measurements for each predictive factor, we obtain a detailed profile on how industry groups affect the statistical relationship between ratings and financial data.

3.3. Using heterogeneity scores to increase prediction accuracy

To use the heterogeneity score for model improvement, we propose two aggregate heterogeneity measures: *financial ratio heterogeneity* and *relative group heterogeneity*.

Financial ratio heterogeneity is the weighted average heterogeneity score for one financial ratio over all subgroup combinations across different rating categories.[8] This measure shows the extent to which a financial ratio is affected by sectoral or regional heterogeneity. Reducing financial ratio heterogeneity allows a closer fit of regression functions to the underlying relationship of corporate financials and ratings. Our hypothesis is that this also increases the performance of multivariate models, which we test in Section 4. We propose the following procedure: we homogenize the Box–Cox transformed financial ratios X'_i by minimizing their heterogeneity scores. For each group g among the total number of groups G, financial ratio distributions X'_{ig} are adjusted by setting a company j's ratio value

$$x''_{igj} = x'_{igj} + h_{ig} \quad \text{for } 1 \leqslant g \leqslant (G-1), \tag{8}$$

where h_{ig} is the group homogenization factor for ratio i and x'_{igj} are the Box–Cox transformed financial ratio values. For G subgroups, only $G-1$ factors are needed, because the coefficient for the last group would be redundant. By minimizing the ratio's heterogeneity score with an iterative procedure, we reduce the group-level heterogeneity for the respective financial ratio.

Relative group heterogeneity is the average heterogeneity score of one subgroup in relation to all other groups in the sample across different rating classes. This measure indicates the extent to which a sector or region differs from others for the respective financial ratio. By aggregating this measure for several financial ratios from different economic categories, this analysis provides a guideline for effective model segmentation decisions. Fig. 1 gives the resulting heterogeneity matrix for nine sector groups, averaged for nine financial ratios.

Utilities exhibit the highest financial ratio heterogeneity compared to the rest of the sample. Therefore, the best sectoral model segmentation would be separate models for utilities and "non-utilities". Since Perry et al. (1984) have already proven the benefits of segmentation, we will not construct group-specific models here.

Still, we should look at the regional level to determine which dimension – regional or sectoral – is most relevant for our dataset. Due to the low number of rated issuers outside the U.S., it is not feasible to measure heterogeneity for individual countries. The 70% share of U.S./Canadian companies in the sample allows only a limited improvement from group-level adjustments for the remaining 30% (see Appendix 2). We therefore believe that the sectoral dimension provides the more instructive example for the application of the proposed procedure to reduce financial ratio heterogeneity.

4. Rating models

This section describes the construction of four rating prediction models to compare different approaches to manage heterogeneity. We test a model with homogenized financial ratios against one that uses sector dummy variables, a model with adjusted sector medians, and a basic "one-size-fits-all" model without any adjustments. Drawing from the most promising methods summarized in our review, we apply the following process for rating model construction: After defining predictive factors across 11 economic categories, we calculate two versions of factors with reduced sectoral heterogeneity. The first one is obtained by applying the homogenization procedure developed in Section 3.3 of this paper. The second one is a simplified version of the procedure that adjusts financial ratio distributions so that their medians are equal. In this way, we can determine the value of the added complexity of het-

[6] Of course, defining one standard deviation as encompassing 80% of a distribution's values is not accurate. However, as all measured heterogeneity scores increase by consistent amounts, subsequent relative comparisons are not affected. Moreover, this procedure limits the influence of remaining outliers.

[7] The rating categories were defined as AAA to A+, A to BBB+, BBB to BBB−, BB+ to BB−, and B+ to B−, reflecting major groups of rated industrials.

[8] Individual heterogeneity score observations are weighted by the sample size of the primary subgroup for each pair to arrive at an average heterogeneity score for a subgroup within a rating category. Then, these rating-specific results are averaged across different rating categories, weighted by their sample size.

Industry Sectors	Consumer Staples	Industrials	Materials	Consumer Discretionary	Information Technology	Telecommunication Services	Health Care	Energy	Utilities	**Average (ex. self)**
Consumer Staples	0%	30%	36%	40%	51%	54%	57%	63%	68%	50%
Industrials	30%	0%	33%	45%	53%	53%	57%	69%	68%	51%
Materials	36%	33%	0%	51%	51%	50%	55%	67%	68%	51%
Consumer Discretionary	40%	45%	51%	0%	51%	56%	50%	63%	74%	54%
Information Technology	51%	53%	51%	51%	0%	59%	39%	57%	69%	54%
Telecommunication Services	54%	53%	50%	56%	59%	0%	64%	48%	55%	55%
Health Care	57%	57%	55%	50%	39%	64%	0%	56%	70%	56%
Energy	63%	69%	67%	63%	57%	48%	56%	0%	64%	61%
Utilities	68%	68%	68%	74%	69%	55%	70%	64%	0%	67%
Average (ex. self)	50%	51%	51%	54%	54%	55%	56%	61%	67%	**55%**

Legend:
- relatively low heterogeneity (< 40%)
- average heterogeneity (40 - 60%)
- relatively high heterogeneity (> 60%)

Notes: Sectoral heterogeneity scores for financial statements from the period 1998-2002 and 2003 S&P-ratings, aggregated to relative group heterogeneity*, and averaged for nine financial ratios. The ratios considered include the EBITDA margin, EBITDA/debt service, short-term debt intensity, leverage (3-year average), net debt/EBITDA, sales, gross margin (5-year volatility), net investment coverage, and capital turnover.* Multi-year transformations *of financial ratios use data starting from 2000 (3-years) and 1998 (5-years) respectively.*

Fig. 1. Sectoral heterogeneity matrix based on nine financial ratios.

erogeneity scores. Subsequently, we estimate logistic univariate submodels for each predictive factor and its homogenized versions, which we use as inputs for the construction of multivariate linear regression models.

4.1. Data

Our data sample consists of 2048 S&P-rated companies with public financial data from the Compustat/GlobalVantage databases. We predict 2003 S&P ratings using annual financial statements from 1998 to 2002. Choosing a random holdout sample of 403 companies (20%) leaves a sample of 1645 companies for model development.

4.2. Factor definition

To extract a maximum of rating-relevant information from available financial data, we define 11 *economic categories* of financial ratios that describe distinct features of a corporation's financials (see Appendix 3). We do not use stock market variables since we aim to predict agency ratings based on financial data. This approach is supported e.g. by results from Löffler (2004), who shows that "it is not evident that ratings are generally inferior to market-based rules" for managing credit portfolios.

Where appropriate, we define several *multi-year transformations* of factors, which enhance their informational content. *Averages* (2, 3, and 5 years) filter out systematic effects such as business cycles but also idiosyncratic factors such as one-off items, reducing distortions of financial ratio distributions. A five-year average would include financial data from 1998 to 2002; a two-year average would consist of the years 2001 and 2002. The same applies to the other multi-year derivations. *Trends* (3 and 5 years) are defined as the average absolute change in a factor's values, in order to avoid unwanted denominator effects.[9] The trend components in sales, profits, leverage and coverage relationships measure improvement and deterioration in a company's business, market position, profitability, and financial risk, respectively. Thus, they serve as proxies partially covering the qualitative business risk factors which rating agencies analyze. The same applies to *volatility* (5 years) of sales, margins, and other ratios. It is defined as the average deviation from a linear trend, which avoids correlations of volatilities with trend strength. Finally, pattern analyses such as the *ever-negative* transformation (5 years) permit asking specific "questions" about a company's track record, for instance, whether EBITDA has ever been negative recently.

[9] A trend ratio based on a very small number in the denominator would not return meaningful figures.

4.3. Classification algorithm and statistical factor transformation

Since our focus is on showing the effect of reducing heterogeneity, we will discuss only briefly the chosen *classification algorithm*. Several methods would be appropriate for rating prediction. Since ratings are ordered categories of credit quality, we could use *ordinal regression* by McCullagh (1980). It estimates a logit probability function for each class of the dependent variable. As this algorithm requires ordinal inputs, financial ratios need to be transformed into quantile classes. While this offers the higher stability of rank transformed variables, model predictions can exhibit undue jumps between rating periods when a predictive factor's value moves over a quantile boundary. An alternative to ordinal regression is based on the approach used in Moody's RiskCalc models (see Section 2.3). It uses the rating grades' corresponding historic PDs as dependent variable for regression. This allows taking into account the actual difference in credit risk in addition to the ordinal ranking of rating classes. Since the PDs of S&P rating classes are approximately exponential, we use their logarithms. Each rating grade is replaced with a *PD-score* based on its historical PD:

$$S(\mathrm{PD}) = \ln\left(\frac{1}{\mathrm{PD}} - 1\right). \tag{9}$$

To enhance the stability of predictor variables, we estimate logistic *univariate submodels*. The technique offers several advantages. The logistic transformation caps extreme values that would otherwise have an undue impact on a company's overall rating. It results in easily interpretable relationships between financial ratios and ratings. Moreover, it allows taking into account the absolute differences in risk of different rating classes. In addition, since univariate logistic models already take into account non-linearities, we can use linear regression for multivariate analysis. Finally, the output of such a regression function would be a prediction of a company's log PD, which can be easily transformed back to its PD and the corresponding S&P rating grade.

4.4. Multivariate analysis

The high number of candidate-variables to derive rating models poses a challenge to which *stepwise regression* offers a shortcut (Efroymson, 1969). It starts by entering the variable with the lowest probability of the *F*-statistic, i.e. the highest significance of the correlation to ratings, into the model. Each step adds the variable with the lowest *F*-value as long as it complies with a 5% significance level as the entry criterion. Variables already in the equation are removed if their significance level exceeds 10%. The process stops when no additional variables comply with the entry or exit criteria. Stepwise regression quickly generates models containing the most significant factors. However, it is sensitive to inter-correlation among groups of variables, resulting in a model that is subject to overfit and that has counterintuitive coefficients. One cannot be sure whether stepwise regression delivers the optimal result (Montgomery and Peck, 1992), and counterintuitive rating models will certainly perform poorly out-of-sample (Hayden, 2003). As the second drawback of this automated approach, it eliminates variables with a less significant relationship to ratings right from the beginning: these predictors could still add valuable information to the model at a later stage.

Consequently, we only use stepwise regression to find a "core model", which combines a set of four to five predictors with the highest significance, as a basis for further modeling. Manual additions of selected factors from economic categories not yet represented in the model complement this approach. For the construction of the homogenized rating model, we also test both the normal and the homogenized versions of a predictor, in order to use the predictor with the higher in-sample significance.

4.5. Multivariate regression results

Tables 1–4 show the parameterization of the four rating models with the highest in-sample performance for each model type. The indicated weights are the product of the range of a ratio's log PD-scores and its regression coefficient. This best represents a factor's actual impact on predicted ratings. Please refer to Appendix 1 for detailed parameterization of multivariate regression and univariate submodels.

The modeling results have several implications. Across the four models, the economic core categories are *size*, *Return on Sales*, *Leverage*, *Debt Service Coverage*, and to a lesser extent, *Business Risk*. While other economic categories such as *Return on Capital* also contain powerful predictors, they do not provide sufficient uncorrelated information beyond that in *Return on Sales*. Common "early warning" measures such as the Quick Ratio or Days Sales Outstanding, i.e. the *Liquidity* category, are not relevant for a long-term corporate rating, in line with prior findings (e.g. Altman and Rijken, 2004). As expected, *multi-year transformations* generate valuable predictors. All models include a three-year average profit margin, which is more significant than its one-year version. The models partly reflect some qualitative factors which rating agencies evaluate. Aspects such as market position are emulated by *size* variables (log. of Equity), while the *volatility* and *ever-negative* transformations of various profitability measures partly cover the business risk assessment.

Comparing the industry dummy and the homogenized model to the "one-size-fits-all" model reveals the significance of sectoral heterogeneity. As expected, based on

Table 1
Parameterization of the "one-size-fits-all" model

Predictor variable	Predictive factor base formula	Economic category	Weight %[a]	Significance[b]
Total assets	Total assets	Size	23.6	<.001
Total equity	Total stockholder's equity	Size/static leverage	19.9	<.001
Profit margin, 3-year average	Net income/sales	Return on sales	19.2	<.001
Total debt 1/EBITDA	(Long-term debt + debt, current portion + notes payable)/(EBIT + depreciation)	Dynamic leverage	13.8	<.001
Total debt 2/total capital	(Long-term debt + debt, current portion + notes payable + preferred stock)/(long-term debt + debt, current portion + notes payable + total stockholder's equity)	Static leverage	7.0	.029
EBIT/interest	EBIT/interest expense	Debt service coverage	6.1	<.001
EBIT, 5-year volatility	EBIT	Business risk	6.0	<.001
Adj. operating cash flow, 5-year ever-negative	Net operating cash flow – CAPEX	Business risk	4.4	<.001

Notes: Model parameterization based on development sample ($N = 1645$) with financials from the period 2002 and 2003 S&P ratings (converted into log PD-scores). Multi-year transformations of financial ratios use data starting from 2000 (3-years) and 1998 (5-years), respectively. Financial statement captions are used as defined in the Compustat/GlobalVantage databases.
[a] Range of predicted log PD scores multiplied by regression coefficient.
[b] As indicated by *t*-test.

Table 2
Parameterization of the industry dummy model

Predictor variable	Predictive factor base formula	Economic category	Weight %[a]	Significance[b]
Total assets	Total assets	Size	17.2	<.001
Profit margin, 3-year average	Net income/sales	Return on sales	13.1	<.001
Total equity	Total stockholder's equity	Size/static leverage	12.8	<.001
EBITDA/interest	(EBIT + depreciation)/interest expense	Debt service coverage	10.4	<.001
Dummy utilities	If Sector = "Utilities" THEN 1 ELSE 0	Sector	9.7	<.001
Total debt 2/total capital	(Long-term debt + debt, current portion + notes payable + preferred stock)/(long-term debt + debt, current portion + notes payable + total stockholder's equity)	Static leverage	8.6	<.001
Total debt 1/Adj. EBITDA	(Long-term debt + debt, current portion + notes payable)/(EBIT + depreciation – CAPEX)	Dynamic leverage	7.1	<.001
Dummy telecom	If Sector = "Telecommunication Services" THEN 1 ELSE 0	Sector	6.3	<.001
Dummy consumer staples	If Sector = "Consumer Staples" THEN 1 ELSE 0	Sector	4.5	<.001
EBIT, 5-year volatility	EBIT	Business risk	3.6	.002
Adj. operating cash flow, 5-year ever-negative	Net operating cash flow – CAPEX	Business risk	3.5	<.001
Dummy industrials	If Sector = "Industrials" THEN 1 ELSE 0	Sector	3.0	<.001

Notes: Model parameterization based on development sample ($N = 1645$) with financials from the period 2002 and 2003 S&P ratings (converted into log PD-scores). Multi-year transformations of financial ratios use data starting from 2000 (3-years) and 1998 (5-years), respectively. Financial statement captions are used as defined in the Compustat/GlobalVantage databases.
[a] Range of predicted log PD scores multiplied by regression coefficient.
[b] As indicated by *t*-test.

the analysis of *relative group heterogeneity* (see Section 3.3), the dummy for Utilities is most significant. In the homogenized model, the sectoral adjustment factor for Utilities is generally either the highest or the lowest value, indicating the high dissimilarity of the sector's financials compared to other industries (see Appendix 4).

Table 3
Parameterization of the homogenized model

Predictor variable	Predictive factor base formula	Economic category	Weight %[a]	Significance[b]
Total equity (h)	Total stockholder's equity	Size/static leverage	30.0	<.001
EBITDA/interest (h)	(EBIT + depreciation)/interest expense	Debt service coverage	15.3	<.001
Total assets	Total assets	Size	14.7	<.001
Profit margin, 3-year average	Net income/sales	Return on sales	14.3	<.001
Total debt 1/EBITDA (h)	(Long-term debt + debt, current portion + notes payable)/(EBIT + depreciation)	Dynamic leverage	13.3	<.001
Operating margin, 5-year volatility (h)	EBIT/sales	Business risk	8.0	<.001
Adj. operating cash flow, 5-year ever-negative	Net operating cash flow – CAPEX	Business risk	4.5	<.001

Notes: (h) homogenized version of financial ratio based on minimizing a financial ratio's heterogeneity score with respect to industry sectors.
Model parameterization based on development sample ($N = 1645$) with financials from the period 2002 and 2003 S&P ratings (converted into log PD-scores). Multi-year transformations of financial ratios use data starting from 2000 (3-years) and 1998 (5-years), respectively. Financial statement captions are used as defined in the Compustat/GlobalVantage databases.
[a] Range of predicted log PD scores multiplied by regression coefficient.
[b] As indicated by *t*-test.

Table 4
Parameterization of the median-shift model

Predictor variable	Predictive factor base formula	Economic category	Weight %[a]	Significance[b]
Total assets	Total assets	Size	26.6	<.001
Total equity (m)	Total stockholder's equity	Size/static leverage	17.0	.007
Profit margin, 3-year average	Net income/sales	Return on sales	16.9	<.001
Total debt 2/total capital	(Long-term debt + debt, current portion + notes payable + preferred stock)/(long-term debt + debt, current portion + notes payable + total stockholder's equity)	Static leverage	12.5	<.001
EBITDA/interest (m)	(EBIT + depreciation)/interest expense	Debt service coverage	9.7	<.001
Operating margin, 5-year volatility (m)	EBIT/sales	Business risk	9.3	<.001
Total debt 1/EBITDA (m)	(Long-term debt + debt, current portion + notes payable)/(EBIT + depreciation)	Dynamic leverage	4.5	<.001
Adj. operating cash flow, 5-year ever-negative	Net operating cash flow – CAPEX	Business risk	3.5	<.001

Notes: (m) homogenized version of financial ratio based on shifting financial ratio values so that the respective sectors' median values are equal.
Model parameterization based on development sample ($N = 1645$) with financials from the period 2002 and 2003 S&P ratings (converted into log PD-scores). Multi-year transformations of financial ratios use data starting from 2000 (3-years) and 1998 (5-years), respectively. Financial statement captions are used as defined in the Compustat/GlobalVantage databases.
[a] Range of predicted log PD scores multiplied by regression coefficient.
[b] As indicated by *t*-test.

4.6. Performance evaluation

We evaluate the four rating models in terms of correlation coefficients and *hitrates*, using out-of-sample measures for the comparison of model performance. The ROC curve and corresponding accuracy ratios, which are typically used in default prediction studies, are less appropriate for the prediction of agency ratings for several reasons. The ROC curve measures a model's ability to concentrate observed defaults among the lowest model scores in the sample. While this is an intuitive measure to assess a model's capability to differentiate between two states – default

Legend: **best performance**

	Basic Model		Dummy Model		Homogenized Model		Median-Shift Model	
	Training	Holdout	Training	Holdout	Training	Holdout	Training	Holdout
Pearson Correlation*	0.784	0.784	0.806	0.796	**0.808**	**0.817**	0.797	0.803
Kendall's tau-b*	0.608	0.603	0.634	0.623	**0.641**	**0.647**	0.620	0.625
Hitrate +/- 0**	0.209	0.183	0.209	0.159	0.209	**0.201**	**0.214**	0.163
Hitrate +/- 1 (cum.)**	0.544	0.519	0.567	0.520	**0.589**	**0.542**	0.562	0.524
Hitrate +/- 2 (cum.)**	0.783	0.748	0.802	0.754	**0.811**	0.771	0.787	**0.774**

* Significant at the 0.01 level

** Additional bootstrap validation confirms better out-of-sample hitrates of Homogenized Model compared to Median-Shift Model, with difference between means of +0.007 (+/- 0), +.026 (+/- 1), and +.019 (+/- 2), respectively. Training and Holdout samples were resampled N=1645 times at random.

Notes: Performance measures for four rating models for training (N=1645) and holdout (N = 403) samples, based on financial data from the period 1998-2002 and 2003 S&P ratings (converted into log PD-scores). Both correlation coefficients indicate predictive accuracy with regard to PD-scores. Hitrates +/- X indicate the percentage of correctly predicted S&P ratings within a margin of X notches.

Fig. 2. Evaluation of rating model performance.

and non-default – the construction of a ROC curve would be much more complex for 22 rating classes. Even if we did so, it would still not give an indication as to how closely predicted ratings correspond to actual ratings, due to its ordinal nature.

We therefore use the Pearson correlation coefficient to measure fit in absolute terms (based on log PD-scores), and Kendall's tau-b in order to measure fit in ordinal terms. We define *hitrate* $\pm X$ as the share of model predictions being within X rating notches from the actual S&P rating. The homogenized model performs best: for both correlation coefficients, it outperforms the "one-size-fits-all" model by 3–4% and the dummy model by about 2%. The same applies to hitrates, where the homogenized model performs 2–4% better across all hitrate definitions. In contrast, the dummy model only achieves slight performance increases over the "one-size-fits-all" model in terms of correlation, and almost no improvement in classification rates. Compared to the median-shift model, the homogenized model outperforms in 4 out of 5 performance measures. Fig. 2 summarizes the performance evaluation.

All measures uniformly show that the added computational complexity of the proposed homogenization procedure pays off in form of increased model performance.

5. Conclusions and implications for further research

The contribution of this study is twofold. First, we evaluate previous rating studies and classified innovations according to three main performance levers: the *classification algorithm*, *factor definition*, and *statistical factor transformation*. This overview of techniques to increase model performance provides a toolset for rating modelers in both scientific and practical contexts. We find that most previous studies focus on finding the ideal *classification algorithm*, while tending to neglect the other two performance levers. Second, we show that reducing sectoral *financial ratio heterogeneity* with the heterogeneity score method increases performance of rating models over a "one-size-fits-all" model, as well as models that include industry dummies or those adjusted by simply shifting sector medians.

This new approach to managing financial ratio heterogeneity fills a gap in cases where the sample size and structure do not permit the construction of separate models for individual subgroups in the sample. It adds to the available techniques of *factor definition*, providing more significant predictors in heterogeneous datasets with limited default information.

The results have several implications for further research. While we tested the independent contribution of reducing sectoral heterogeneity in financial ratios, a further performance increase could be achieved by also applying the approach to regional heterogeneity, given a suitable sample structure. A different mathematical specification of the heterogeneity adjustment – e.g. using slope adjustment factors in addition to the proposed intercept adjustments – might also improve model accuracy. Furthermore, although the benefit of multi-year derivations of financial ratios becomes intuitively clear (in the fact that they replace their one-year, less significant counterparts in our final models), measuring their independent contribution would be another appealing area of research. Finally, it is safe to assume that by combining the approaches of this study with alternative methods of statistical data transformation (e.g. rank transformation) and different classification techniques, such as ordinal regression, one could achieve further performance increases.

Acknowledgements

We thank the McKinsey & Company Risk Management Practice for its support and Professor Günter Bamberg, University of Augsburg, Germany, and a referee, for helpful comments.

Appendix 1. Detailed rating model parameterization

Variables	Unstandardized coefficients				Parameterization for logit transformation			
	β	Std. error	t	Sig.	$y1$	$y2$	$y3$	$y4$
"One-size-fits-all" model								
(Constant)	−6.397	0.446	−14.33	1.37E−43				
Total assets	0.576	0.118	4.90	1.09E−06	2.933	1.954	2.900	2.959
Total debt to total capital	0.220	0.100	2.19	2.89E−02	−2.286	4.419	7.870	0.837
Log. equity	0.476	0.113	4.20	2.79E−05	2.997	2.385	1.101	6.129
Profit margin 1, 3-year average	0.546	0.046	11.83	7.73E−31	2.513	2.331	38.952	−0.006
Adj. operating cashflow, 5-year ever-negative	0.286	0.055	5.15	2.92E−07	−1.098	4.534	1000.000	0.500
EBIT, 5-year volatility	0.228	0.060	3.78	1.64E−04	−36.265	38.993	13.816	−0.210
EBIT/interest	0.182	0.053	3.46	5.54E−04	2.401	2.505	1.924	1.723
Total debt/EBITDA	0.335	0.058	5.82	7.34E−09	−2.952	5.450	0.510	3.825
Dummy model								
(Constant)	−6.789	0.433	−15.68	3.33E−51				
Total assets	0.546	0.114	4.80	1.72E−06	2.933	1.954	2.900	2.959
Total debt to total capital	0.349	0.095	3.67	2.53E−04	−2.286	4.419	7.870	0.837
EBITDA/interest	0.362	0.046	7.92	4.85E−15	2.678	2.190	1.485	2.717
Log. equity	0.398	0.109	3.64	2.86E−04	2.997	2.385	1.101	6.129
Profit margin 1, 3-year average	0.486	0.044	11.02	3.89E−27	2.513	2.331	38.952	−0.006
Total debt/Adj. EBITDA	0.274	0.051	5.35	1.00E−07	−29.040	32.098	0.213	−11.298
Adj. operating cashflow, 5-year ever-negative	0.295	0.055	5.34	1.11E−07	−1.098	4.534	1000.000	0.500
EBIT, 5-year volatility	0.180	0.058	3.09	2.02E−03	−36.265	38.993	13.816	−0.210
Dummy industrials	0.275	0.075	3.68	2.39E−04				
Dummy utilities	0.901	0.087	10.40	1.79E−24				
Dummy telecom	0.589	0.142	4.14	3.71E−05				
Dummy consumer staples	0.420	0.099	4.24	2.41E−05				
Homogenized model, (h) indicates homogenized predictors								
(Constant)	−5.128	0.290	−17.65	4.12E−63				
EBITDA/interest (h)	0.345	0.052	6.69	3.19E−11	3.025	2.192	0.270	−27.314
Adj. operating cashflow, 5-year ever-negative	0.278	0.052	5.36	9.97E−08	−1.098	4.534	1000.000	0.500
Profit margin 1, 3-year average	0.388	0.044	8.73	7.31E−18	2.513	2.331	38.952	−0.006
Operating margin, 5-year volatility (h)	0.219	0.054	4.09	4.48E−05	−2.498	4.721	0.792	−4.068
Total debt/EBITDA (h)	0.343	0.055	6.28	4.41E−10	−2.636	5.053	1.689	−4.451
Log. equity (h)	0.585	0.073	7.96	3.56E−15	3.499	2.066	7.831	−0.550
Total assets	0.342	0.084	4.06	5.10E−05	2.933	1.954	2.900	2.959
Median-shift model, (m) indicates adjusted predictors								
(Constant)	−7.022	0.369	−19.02	5.83E−72				
EBITDA/interest (m)	0.342	0.043	7.95	3.82E−15	2.395	2.517	1.832	2.673
Total debt/Adj. EBITDA (m)	0.191	0.046	4.12	4.04E−05	−1.994	5.066	0.694	4.277
Operating margin, 5-year volatility (m)	0.266	0.062	4.32	1.66E−05	−41.065	43.533	8.359	−0.371
Total assets	0.765	0.076	10.10	3.37E−23	2.933	1.954	2.900	2.959
Profit margin 1, 3-year average	0.566	0.043	13.29	4.96E−38	2.513	2.331	38.952	−0.006
Total debt to total capital (m)	0.409	0.075	5.46	5.78E−08	−2.574	4.691	6.010	0.778
Adj. operating cashflow, 5-year ever-negative	0.265	0.055	4.81	1.65E−06	−1.098	4.534	1000.000	0.500
Log. equity (m)	0.206	0.077	2.70	7.11E−03	13.242	−1.754	0.156	7.820

Parameters of logit transformation: The *logit* function is defined as a function with 4 degrees of freedom y_j:

$$\text{PD} - \text{Score} = Y(x_i) = \frac{y_1}{1 + e^{-y_3(x_i - y_4)}} + y_2,$$

where x_i is a company's financial ratio.

Appendix 2. Shares of industry sectors and countries in the sample

	Count	%
Sector		
Consumer discretionary	363	22.1
Industrials	279	17.0
Materials	199	12.1
Utilities	198	12.0
Information technology	147	8.9
Consumer staples	136	8.3
Health care	123	7.5
Energy	121	7.4
Telecommunication services	79	4.8
Country		
United States of America	1094	66.5
Japan	154	9.4
United Kingdom	57	3.5
Canada	49	3.0
Australia	34	2.1
France	28	1.7
Germany	21	1.3
Netherlands	19	1.2
Mexico	14	0.9
Italy	11	0.7
Sweden	11	0.7
Bermuda	10	0.6
Chile	10	0.6
Spain	10	0.6
Switzerland	10	0.6
New Zealand	8	0.5
Brazil	7	0.4
Cayman Islands	7	0.4
Taiwan	7	0.4
Finland	6	0.4
Indonesia	6	0.4
Other	72	4.4

Appendix 3. Definition of economic categories

Economic category	Economic rationale
Return (on sales)	Accrual and cash flow return measures on sales assess the overall health of a company's operations. This category comprises ratios such as EBITDA/sales, net profit margin, and pre-tax-pre-interest return on sales
Return (on capital)	Return-on-investment ratios encompass the various types of returns on different "slices" of a company's asset or capital base, such as ROA and ROE
Leverage (static)	Static leverage ratios indicate the overall level of financial risk inherent in a company's capital structure in terms of debt vs. equity, for instance, total liabilities/equity. Static debt coverage ratios are simply reverse versions of leverage ratios, for example, book value coverage of total debt
Leverage (dynamic)	Dynamic leverage ratios analyze a company's debt burden in the context of its earnings and cash flows. Ratios such as debt/EBITDA can also be interpreted as debt repayment periods
Debt coverage (dynamic)	Earnings coverage of debt service charges indicates a company's ability to service its debt and repay principal, for instance, EBITDA/(interest + current debt)
Liquidity	This contains the small group of ratios which measure the various intensities of short-term liabilities and working capital against assets, debt, or sales
Investment coverage	Investment coverage ratios such as equity/fixed assets and depreciation/CAPEX indicate the sustainability of a company's asset base. Static and dynamic ratios were combined in one group due to the low number of ratios
Size	Asset size indicates asset protection for a company's debt. Size in terms of sales is a proxy for its market position
Business trends	Trends of sales and earnings contain information about the development and outlook of a company and its industry
Business risk	Volatilities of sales and earnings figures indicate riskiness of a sector and also capture management's ability to steer the company through the business cycle
Differentiating factors	Intensities such as fixed assets/total assets are not significantly correlated to ratings, but can act as a differentiator between sectors in a model. The same applies to productivity and turnover ratios

Appendix 4. Sectoral adjustment factors for homogenized predictors

Sector	EBITDA/interest	Log. equity	Operating margin 5-year volatility	Total debt/EBITDA
Consumer discretionary	0.00	0.00	0.00	0.00
Consumer staples	−0.29	0.05	0.29	−0.67
Energy	0.64	0.00	−2.48	−0.30
Health care	−4.49	−0.02	0.05	0.65
Industrials	1.95	0.02	−0.07	−0.50
Information technology	−7.13	−0.09	−1.48	0.31
Materials	1.29	−0.01	−0.57	−0.62
Telecommunication services	2.63	−0.10	−1.20	−0.45
Utilities	8.94	0.06	−1.24	−1.35

Notes: Sectoral adjustment factors h_g as calculated with Eq. (8) in the development sample ($N = 1645$).

References

Altman, E., 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. The Journal of Finance 23, 589–609.

Altman, E., Rijken, H., 2004. How rating agencies achieve rating stability. Journal of Banking and Finance 28, 2679–2714.

Aziz, M., Dar, H., 2006. Predicting corporate bankruptcy: Where we stand? Corporate Governance 6, 18–33.

Balcaen, S., Ooghe, H., 2004. Alternative methodologies in studies on business failure: Do they produce better results than the classical statistical methods? Vlerik Leuven Gent Working Paper Series 2004/16, Ghent University, Belgium.

Bank for International Settlements, 2006. International Convergence of Capital Measurement and Capital Standards, Basel Committee on Banking Supervision, Basel, Switzerland. Available from: <http://www.bis.org/publ/bcbs128b.pdf> (accessed 11.24.06).

Belkaoui, A., 1980. Industrial bond ratings: A new look. Financial Management 9, 44–51.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) 26, 211–252.

Chava, S., Jarrow, R., 2004. Bankruptcy prediction with industry effects. Review of Finance 8, 537–569.

Chava, S., Stefanescu, C., Turnbull, S., 2006. Modeling expected loss. Working Paper, Texas A&M University.

Deakin, E., 1976. Distributions of financial accounting ratios: Some empirical evidence. Accounting Review 51, 90–97.

Efroymson, M., 1969. Multiple regression analysis. In: Rolston, A., Wilf, H. (Eds.), Mathematical Models for Digital Computers. Wiley, New York.

Fisher, L., 1959. Determinants of risk premiums on corporate bonds. The Journal of Political Economy 67, 217–237.

Frecka, T., Hopwood, W., 1983. The effects of outliers on the cross-sectional distributional properties of financial ratios. The Accounting Review 57, 115–128.

Hayden, E., 2003. Are credit scoring models sensitive with respect to default definitions? Evidence from the Austrian Market. In: EFMA 2003 Helsinki Meetings.

Horrigan, J., 1966. The determination of long-term credit standing with financial ratios. Journal of Accounting Research 4 (Suppl.), 44–62.

Kane, G., Richardson, F., 1998. Rank transformations and the prediction of corporate failure. Contemporary Accounting Research 15, 145–166.

Liu, Y., 2002. A framework of data mining application process for credit scoring. University of Göttingen, Germany, Institute of Business Computer Science.

Löffler, G., 2004. Ratings versus market-based measures of default risk in portfolio governance. Journal of Banking and Finance 28, 2715–2746.

McCullagh, P., 1980. Regression models for ordinal data. Journal of the Royal Statistical Society, Series B (Methodological) 42, 109–142.

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

Montgomery, D., Peck, E., 1992. Introduction to Linear Regression Analysis. John Wiley & Sons, New York.

Moody's Investors Service – Global Credit Research, 2001a. RiskCalc™ Public – Europe – Rating Methodology. Moody's Investors Service, New York.

Moody's Investors Service – Global Credit Research, 2001b. RiskCalc™ for Private Companies – The German Model – Rating Methodology. Moody's Investors Service, New York.

Perry, L., Henderson Jr., G., Cronan, T., 1984. Multivariate analysis of corporate bond ratings and industry classification. Journal of Financial Research 7, 27–36.

Standard & Poor's, 2003. Corporate Ratings Criteria. The McGraw-Hill Companies, New York.

Watson, C., 1990. Multivariate distributional properties, outliers, and transformation of financial ratios. The Accounting Review 65, 682–695.