Classification and Rating of Firms in the Presence of Financial and Non-financial Information

Thomas Mählmann

Chair of Banking, University of Cologne
Albertus-Magnus-Platz
50923 Koeln
Germany
Tel: (0049)+221-4702628
Fax: (0049)+221-4702305
maehlmann@wiso.uni-koeln.de

Version: February 2004

Abstract

Under the proposed Basel Capital Accord, internal credit ratings are expected to gain in importance because of their potential role in determining the adequacy of regulatory capital. A standard method of deriving an internal (sub) rating consists of estimating a credit scoring function and assigning rating grades based on credit scores. However, the usual purpose of credit scores is to classify firms into two groups: good and bad. Using a model for the joint distribution of financial and non-financial risk factors, we study the effect of common data sample characteristics on the accuracy of 2-group classification and rating of firms on the basis of their credit scores as derived from two well known scoring techniques: linear discriminant analysis and logistic regression. Our main focus is on bias in scoring function estimation and it is shown that such a bias has a strong negative impact on the accuracy of ratings, but not on that of classification. Therefore this study suggests that a Basel II conforming rating system should be based on a scoring technique that at least leads to consistent (i.e. asymptotically unbiased) estimates of scoring function coefficients for a wide class of risk factor distributions.

Key words: credit ratings, credit scores, bias
JEL classification: C13, G21, G28

1 Introduction

The proposed new regulation of the supervisory standards for banks' credit risk capital requirements (Basel II) concentrates the interest of banking supervision and commercial banks on internal ratings. Within the „Internal Ratings Based (IRB) Approach" of Basel II, Banks are allowed to use their own internal ratings for the calculation of capital requirements, provided that they fulfill a number of so-called minimum requirements. One central requirement is that ratings are based not only on traditional quantitative, financial factors (e.g. accounting and industry sector data) but also on qualitative, non-financial information (e.g. market position, management quality etc.).[1] This qualification is supported by empirical work that confirms the additional information content of „soft facts" (see Grunert/Norden/Weber (2004) and Becchetti/Sierra (2003)). In addition, studies that analyze the structure of internal rating systems in use show that banks already rely regularly on qualitative information for the assignment of ratings (see Basel Committee on Banking Supervision (2000) and Treacy/Carey (2000)). As a result, Krahnen/Weber (2001) view the consideration of qualitative factors as part of „Generally Accepted Rating Principles".

Having emphasized the importance of both types of information, the issue becomes, how to combine all the different risk factors into an overall rating, i.e. searching for the optimal weighting scheme. For this task we can, in principle, differentiate between two approaches. A partially statistical and expert judgement-based approach is to combine financial factors into a scoring function via credit scoring techniques and to convert the metric scores to ordinally scaled rating grades (e.g. AAA, AA, A etc.). The financial rating is then adjusted to reflect non-financial data by a credit analyst, based on his subjective opinion (see Crouhy/Galai/Mark (2001)). A second, purely statistical, approach is to combine financial and non-financial factors into one scoring function (see Becchetti/Sierra (2003)). Both procedures involve the transformation of scores to ratings. Well known scoring techniques, such as linear discriminant analysis (Altman (1968)), logistic regression (Martin (1977)) and neural networks (Altman/Marco/Varetto (1994)), were formerly applied to support credit granting decisions (2-group classifications); that is, to separate firms into default and non-default groups. Altman/Narayanan (2002) compare scoring functions, based mainly on financial information, developed in several countries for 2-group classifications. Even though it is now common practice to assign (financial) ratings based on credit scores (see Bardos (1998) or Frerichs/Wahrenburg (2003)), it is not clear whether credit scores are appropriate for this new task.

We study the effects of several factors on the accuracy of the 2-group classification of, and the assignment of ratings to, firms on the basis of their credit scores as derived from linear discriminant analysis and logistic regression. A necessary condition for credit scores being applicable to the assignment of ratings is given and it is shown that this condition is not, in general, met by one popular scoring technique: linear discriminant analysis. Evidence is provided

[1] In what follows, the terms quantitative and financial, and qualitative and non-financial, respectively, are perfect substitutes.

for a systematic bias of ratings, deduced from discriminant credit scores, even in large data samples. The same is not true for logistic regression based rating assignments. We offer advice for the design of data samples used in developing scoring functions.

Our discussion is focused on the more general case of combining financial and non-financial information into an overall rating, but the main results are also applicable to the special case of using only financial data for credit scoring. Incorporating qualitative factors into a scoring function requires the transformation of nominal or ordinal scaled data to a metric scale. A standard method of doing this is to use a dummy (0/1) variable for each different category of the qualitative factor, except for a reference category. The dummy takes the value 1 only if the realization of the qualitative factor equals the specified category. Obviously, dummy variables are not normally distributed and so they violate the central assumption of classical discriminant analysis.[2] Therefore Blochwitz/Eigermann (2000) proposed a different method for handling qualitative data in discriminant analysis, the Fisher-Lancaster (FL) scaling procedure, which tries to achieve scaled qualitative data that follows approximately a normal distribution (see Elpelt/Hartung (1986) for further details). As Blochwitz/Eigermann (2000) do not consider dummy encoding of their data, results on the relative merits of FL scaling are absent to date. We try to resolve this shortcoming.

The paper proceeds as follows. Section 2 describes the linear discriminant analysis and the logistic regression approach for estimating credit scoring function coefficients and shows two common methods for transforming scores to ratings. A hypothesis concerning the effect of biased coefficient estimation on classification as well as rating assignment accuracy of credit scores is derived. Whereas section 2 considers the general case without specifying the type of the scoring function risk factors, section 3 analyzes the properties of coefficient estimates for the case of combining financial and non-financial data. The Model of section 3 is used to test our hypothesis via Monte Carlo simulations. Section 4 provides the Monte Carlo setup. Section 5 states the simulation results concerning the accuracy of classification and rating of firms on the basis of their credit scores. The paper concludes in section 6.

2 Credit scoring and rating assignment

It is assumed that in total p relevant risk factors (e.g. accounting ratios, industry sector information, management quality etc.), included in the random vector $X' = (X_1, X_2, \ldots, X_p)$, are given and common knowledge. The determination of the scoring function $S(x) = \alpha + x'\beta$ reduces therefore to the problem of estimating the coefficients α and $\beta' = (\beta_1, \ldots, \beta_p)$. The question of identifying appropriate risk factors is excluded in this work. In the following section we show, without specifying the type of the variables in X, how these coefficients

[2] Note that using independent dummy variables does not violate the assumption of logistic regression, which is that the conditional distribution of the dependent variable has logistic form.

are estimated by two prominent credit scoring techniques: linear discriminant analysis (LDA) and logistic regression.

The standard case of 2-group discriminant analysis can be described as follows:

Let G be the set of all firms, and G_1 and G_0 be, respectively, the subsets of defaulted and of non-defaulted firms with $G = G_1 \cup G_0$ and $G_1 \cap G_0 = \varnothing$. Let π_1 and π_0 be, respectively, the *a priori* group probabilities. Let $f_1(x)$ and $f_0(x)$ be, respectively, the group density functions. Suppose that Y denotes a dichotomous grouping variable given by

$$y_i = \begin{cases} 1, & if\ firm\ i \in G_1, \\ 0, & if\ firm\ i \in G_0. \end{cases} \tag{1}$$

Each firm should now be classified into G_1 or G_0 depending on its specific value of X. In general, this could be done in the following way: the sample space $\mathbb{R}^p$ of X is separated into two disjunct subsets $\mathbb{R}^p_1$ and $\mathbb{R}^p_0$ and it is agreed upon the rule

$$X \in \mathbb{R}^p_Y \iff the\ firm\ is\ classified\ to\ G_Y \ (Y = 0, 1). \tag{2}$$

The probabilities of misclassification are given by:

$$P(0|1) = P(X \in \mathbb{R}^p_0 |\ firm \in G_1) = \int_{\mathbb{R}^p_0} f_1(x)dx, \tag{3}$$

$$P(1|0) = P(X \in \mathbb{R}^p_1 |\ firm \in G_0) = \int_{\mathbb{R}^p_1} f_0(x)dx. \tag{4}$$

The subsets $\mathbb{R}^p_1$ and $\mathbb{R}^p_0 = \mathbb{R}^p \backslash \mathbb{R}^p_1$ are chosen in such a way that the error rate e (i.e. the overall probability of misclassification) is minimized.[3]

$$\begin{aligned} e &= P(firm\ is\ misclassified) \\ &= P(1|0) \cdot \pi_0 + P(0|1) \cdot \pi_1 \\ &= \pi_1 + \int_{\mathbb{R}^p_1} (\pi_0 f_0(x) - \pi_1 f_1(x))dx. \end{aligned} \tag{5}$$

Obviously the set $\mathbb{R}^p_1$ should contain all elements of X for which the difference $\pi_0 f_0(x) - \pi_1 f_1(x)$ is negative:

$$\mathbb{R}^p_1 \equiv \{x \in \mathbb{R}^p | \pi_0 f_0(x) - \pi_1 f_1(x) \leq 0\}. \tag{6}$$

This results in the following classification rule:

[3] The following expositions are equally valid for the case of minimizing the expected costs of misclassification, except that π_1 and π_0 must be replaced by $\pi_1 C(0|1)$ and $\pi_0 C(1|0)$, respectively, where $C(0|1)$ and $C(1|0)$ denote the costs of both error types.

$$\text{the firm is classified to } G_1 \iff \frac{f_1(x)}{f_0(x)} \geq \frac{\pi_0}{\pi_1}. \tag{7}$$

The expression $\frac{f_1(x)}{f_0(x)}$ is also known as the likelihood ratio at point x. To render the above rule applicable, we assume that $f_1(x)$ and $f_0(x)$ denote multivariate normal density functions with mean vector μ_Y, $Y = 0, 1$, and common covariance matrix $\Sigma_1 = \Sigma_0 = \Sigma$. For the log likelihood ratio results:

$$\log \frac{f_1(x)}{f_0(x)} = D(x) + \log \frac{\pi_0}{\pi_1}, \tag{8}$$

with

$$\begin{aligned} D(x) &\equiv \alpha_{LDA} + \beta'_{LDA} x, \\ \beta_{LDA} &\equiv (\mu_1 - \mu_0)' \Sigma^{-1}, \\ \alpha_{LDA} &\equiv \log \frac{\pi_1}{\pi_0} - \frac{1}{2} \left(\mu_1' \Sigma^{-1} \mu_1 - \mu_0' \Sigma^{-1} \mu_0 \right). \end{aligned} \tag{9}$$

$D(x)$ gives the classical linear discriminant function of Fisher (1936) and Welch (1939). The following classification rule applies:

$$\text{the firm is classified to } G_1 \iff D(x) \geq 0. \tag{10}$$

In practical situations the unknown parameters π_0, π_1, μ_0, μ_1, Σ will be replaced by their maximum likelihood estimates $\widehat{\pi}_0$, $\widehat{\pi}_1$, $\overline{x}_0$, $\overline{x}_1$, $\widehat{\Sigma}$ and we get $\widehat{\alpha}_{LDA} = \log \frac{\widehat{\pi}_1}{\widehat{\pi}_0} - \frac{1}{2} \left(\overline{x}_1 - \overline{x}_0 \right)' \widehat{\Sigma}^{-1} (\overline{x}_1 + \overline{x}_0) \right)$ as estimator for the constant α_{LDA} and $\widehat{\beta}_{LDA} = (\overline{x}_1 - \overline{x}_0)' \, \widehat{\Sigma}^{-1}$ as estimator for the coefficient vector β_{LDA}.

Applying Bayes' formula yields a relationship between the *a posteriori* default probability $P(y = 1|x)$ and the likelihood ratio $\frac{f_1(x)}{f_0(x)}$:

$$P(y = 1|x) = \frac{\pi_1 f_1(x)/\pi_0 f_0(x)}{1 + \pi_1 f_1(x)/\pi_0 f_0(x)} = \frac{\exp(S(x))}{1 + \exp(S(x))}, \tag{11}$$

with $S(x) = \log(\pi_1 f_1(x)/\pi_0 f_0(x)) = \log(\pi_1/\pi_0) + \log(f_1(x)/f_0(x))$. We see that when the vector of risk factors X follows a multivariate normal distribution in both groups, $D(x)$ equals the *a posteriori* log odds ratio for group 1 versus group 0 having observed x, i.e.:[4]

$$P(y = 1|x) = \frac{\exp(D(x))}{1 + \exp(D(x))} \tag{12}$$

$$\Leftrightarrow \log \left(\frac{P(y = 1|x)}{P(y = 0|x)} \right) = D(x). \tag{13}$$

[4] Expression (13) shows that an allocation based on *a posteriori* probabilities, i.e. the firm is allocated to G_1 if $P(y = 1|x) - P(y = 0|x) \geq 0$, is identical to classification rule (10).

The LDA-approach thus offers the possibility of estimating the coefficients α and $\beta' = (\beta_1, \ldots, \beta_p)$ in a linear logit model of the form:

$$P(y = 1|x) = \frac{\exp(\alpha + x'\beta)}{1 + \exp(\alpha + x'\beta)} = \Lambda(\alpha + x'\beta). \tag{14}$$

with $\Lambda(z) \equiv \exp(z)/(1 + exp(z))$ denoting the cumulative distribution of a standard-logistic distributed random variable Z.

Normally, the determination of the logit model parameters using logistic regression is based on a conditional maximum likelihood approach. Under the assumption that the conditional distribution of Y, given X, has logistic form, the following conditional likelihood function can be set up:

$$\begin{aligned} &L_C(\alpha, \beta; y_1, \ldots, y_n, x_1, \ldots, x_n) \\ &= \prod_{i=1}^{n} \Lambda(\alpha + x_i'\beta)^{y_i} (1 - \Lambda(\alpha + x_i'\beta))^{1-y_i} \ . \end{aligned} \tag{15}$$

Maximizing this function with respect to α and β yields the well known conditional maximum likelihood (CML)-estimators $\widehat{\alpha}_{CML}$ and $\widehat{\beta}_{CML}$ for the logit model. As a result, there are now two possibilities available for estimating the scoring function $S(x)$ or the *a posteriori* default probabilities $P(y = 1|x) = \Lambda(\alpha + x'\beta)$ derived from $S(x)$: the CML- and the LDA-approaches.

Regarding the transformation of credit scores into rating grades, we can differentiate between the following methods:[5]

1. A simple approach (named R1 in what follows) attempts to divide the whole range of credit scores into a small number (equal to the number of rating grades) of subintervals. Different methods for determining the subinterval boundaries have been proposed. Baetge/Huss/Niehaus (1988) build symmetrical intervals around the value $\widehat{S}_{LDA}(x) = 0$, while Blochwitz/Eigermann (1999) use the default probabilities $\widehat{P}_{LDA}(y = 1|x)$ belonging to $\widehat{S}_{LDA}(x)$ as an orientation for subdividing $\widehat{S}_{LDA}(x)$.[6] Carey/Hrycay (2001) construct intervals for $\widehat{P}_{CML}(y = 1|x)$ by considering default rates for Moody's rating classes. In addition to these essentially expert judgement-based techniques it is also possible to derive interval boundaries in a purely statistical fashion. To achieve this, all firms of a given sample are ordered by the value of their credit score and then intervals with the same number of firms are delimited. Examples for such a procedure using the discriminant scores $\widehat{S}_{LDA}(x)$ can be found in Taffler (1984), Micha (1984) and Bardos (1998). Frerichs/Wahrenburg (2003) derive statistical interval boundaries for the logistic regression scores $\widehat{S}_{CML}(x)$.

[5] Because *a posteriori* probabilities result from a monotonic transformation of scores, it does not matter whether rating grade assignments are based on $S(x)$ or $P(y = 1|x)$.

[6] Lincoln (1984) and Izan (1984) utilize $\widehat{P}_{LDA}(y = 1|x)$ as an continuous index of default risk and go without defining rating classes.

2. A second, more advanced, statistical approach (named R2 here) goes one step further (see, for example, Fritz/Popken/Wagner (2002) or Höse/ Huschens (2003)). First, in analogy to R1, all firms of a sample are ordered by their credit scores and grouped into even-sized (e.g. 50) buckets. For each bucket the default rate (i.e. the proportion of defaulted firms, which is an unbiased estimate of the default probability) for a given default horizont, is determined. Using common parametric or nonparametric regression analysis it is possible to infer the functional form of the relationship PD= $f(S(x))$ between credit scores ($S(x)$) and estimated default probabilities (PD). Often the PDs are calibrated in order to reflect, not the sample default rate, but a projected (long run) population default rate. By comparing PDs with historical default rates for Moody's or Standard & Poor's rating classes, one can map score intervals to well known rating grades.

We can now make some general statements about the properties of the CML- and LDA-estimators. If the assumption of multivariate normal distributed risk factors in both groups are fulfilled, $\widehat{\alpha}_{LDA}$ and $\widehat{\beta}_{LDA}$ are unconditional maximum likelihood estimators. Otherwise, they are not consistent. The CML-estimators equal unconditional maximum likelihood estimators if the marginal distribution of X contains no information about α and β. Because this is not true for normal distributed risk factors, the LDA-estimators are in this case not only consistent, but also asymptotically efficient, relative to the CML-estimators (Efron (1975); McLachlan/Byth (1979)). On the other hand, the maximum likelihood approach is more robust, because the logistic form of the conditional distribution of Y, given X, holds for a wide class of distributions of (Y, X) (see Day/Kerridge (1967)).

We can conclude that if the vector X of relevant risk factors is not multivariate normal distributed the estimators $\widehat{\alpha}_{LDA}$ and $\widehat{\beta}_{LDA}$ are asymptotically biased, i.e. $\lim_{n \longrightarrow \infty} E_n(\widehat{\alpha}_{LDA}) \neq \alpha$ and $\lim_{n \longrightarrow \infty} E_n(\widehat{\beta}_{LDA}) \neq \beta$. In the following we try to identify the consequences of biased estimation of scoring function coefficients for 2-group classifications and assignments of ordinal rating grades based on credit scores.[7] Consider Figure 1. The range of scores is given by $[S_{\min}, S_{\max}]$. Biased coefficient estimation results in a biased score for an arbitrary firm i, where bias is defined as the difference $S_i - E_n(\widehat{S}_i)$ between the unknown true score S_i and the mean $E_n(\widehat{S}_i)$ of the estimated score, given sample size n. A 2-group classification of firm i in this example with S^* as cutoff score should, on average, bring about the same result, regardless of whether S_i or $\widehat{S}_i$ is used. Note that the same is not true if the range $[S_{\min}, S_{\max}]$ is divided into, for example, ten equally-sized intervals (rating classes). The mean estimated rating grade (D) deviates from the true grade (B). We summarize these considerations in the subsequent hypothesis:

H_1: Biased estimates of scoring function coefficients have a negative impact

[7] 2-group classifications form the binary special case of ordinal rating systems with only one class each for defaulted and non-defaulted firms.

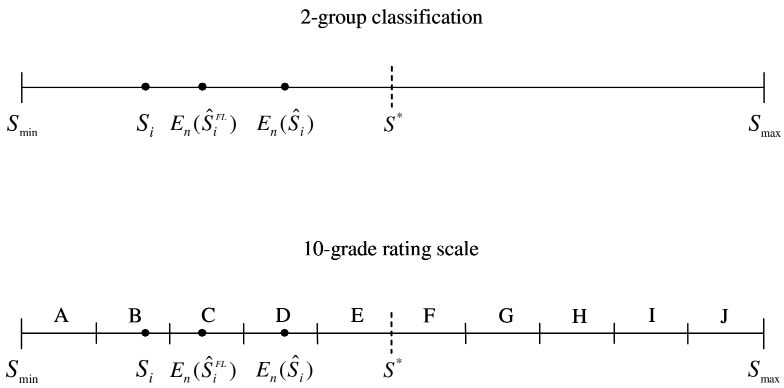

Figure 1: 2-group classification and rating assignment

on the accuracy of rating assignments based on credit scores. 2-group classifications should not be affected.

Since qualitative risk factors that have, in general, only a few different categories can hardly be seen as realizations of normal random variables, the combined processing of quantitative and qualitative data in a scoring function estimated under the LDA-approach is subject to H_1. Against this background, the Fisher-Lancaster (FL) scaling might be more beneficial than using dummy variables. According to the FL method the scale values $k^l_{X_j}$, $l = 1, \ldots, q_j$, for the q_j different categories of a qualitative variable X_j, $j \in \{1, 2, \ldots, p\}$, are identified in such a way as to maximize their discriminatory power between the categories of Y. In addition, the resulting scaled variables $K_Y \in \{k^0_Y, k^1_Y\}$ and $K_{X_j} \in \{k^1_{X_j}, \ldots, k^{q_j}_{X_j}\}$ should be adapted, as well as is possible, to a bivariate standard normal distribution. Both objectives correspond to maximizing the first empirical canonical correlation between K_Y and K_{X_j}. If the FL procedure leads to a better adaption of scaled qualitative variables to a normal distribution and this leads to less biased coefficient estimates for the LDA-approach (for example, $E_n(\widehat{S}_i)$ changes to $E_n(\widehat{S}^{FL}_i)$), we should see a better rating assignment accuracy of LDA discriminant scores, given that H_1 is true. We formulate hypothesis H_2:

H_2: Assume H_1 is true. Then the accuracy of rating assignments based on LDA scoring functions rises if FL scaling, rather than dummy variables, is used to process qualitative, non-financial information. 2-group classifications should not be affected.

If H_1 cannot be rejected and (14) holds, then in large samples of data that do not follow a normal distribution we should see a higher rating assignment accuracy for a scoring function based on the CML-approach than for one based on the LDA-approach. Since we know from the literature (see King/Zeng (2001)) that the CML-estimates of α and β are biased in small samples of rare events data, we do not expect a great difference between the two approaches in small samples with few defaults. If H_1 and H_2 cannot be rejected, then FL scaling improves ratings based on the LDA-approach, but this does not mean that the accuracy of the consistent CML-approach is reached.[8]

To test hypothesis H_1 and H_2 we need a model for the distribution of mixed quantitative and qualitative variables. In the following sector one such model, first introduced by Olkin/Tate (1961) and given the name "location model" by Afifi/Elashoff (1969), is described and the properties of the CML- and LDA-estimators in this special case are analyzed.

3 A location model for mixtures of financial and non-financial risk factors

Suppose $X^{(1)}$ denotes the vector of p_1 quantitative variables (e.g. accounting or industry ratios) and $X^{(2)}$ denotes the vector of $p_2 = p - p_1$ qualitative variables (e.g. market position, accounting behavior etc.), i.e $X = (X^{(1)\prime}, X^{(2)\prime})^{\prime}$.[9] Let risk factor j, $j = 1, \ldots, p_2$, in $X^{(2)}$ have q_j different categories. If we arrange the qualitative variables in a p_2-dimensional table, we get a table with $m = \sum_{j=1}^{p_2} q_j$ uniquely defined multinomial cells, where each cell stands for a particular realization of the p_2 qualitative risk factors in the subvector $X^{(2)}$. For example, each cell might represent a different combination of market position and accounting behavior. Concerning the quantitative information, it is assumed that $X^{(1)}$ has a multivariate normal distribution with mean vector $\mu_{Yc}^{(1)}$ in cell c and group Y ($c = 1, \ldots, m$; $Y = 0, 1$) and a common covariance matrix Σ in all cells of both groups, i.e

$$X^{(1)}|X^{(2)} \sim N(\mu_{Yc}^{(1)}, \Sigma) \quad Y = 0, 1; \, c = 1, \ldots, \, m. \tag{16}$$

Having varying mean vectors for different cells makes it possible to model the correlation between financial and non-financial information. To get an expression for the joint distribution $f_Y(x)$ of $X^{(1)}$ and $X^{(2)}$ the distribution of $X^{(2)}$ must be specified. Therefore, let $V^{\prime} = (V_1, \ldots, V_{m-1})$ be an $(m-1)$-dimensional random vector containing dummy variables. Here, each dummy represents the indicator function for the first $m-1$ cells corresponding to one of the first $m-1$ possible distinct patterns that $X^{(2)}$ may form.[10] The vector $\theta_Y = (\theta_{Y1}, \ldots, \theta_{Ym-1})$ gives the appropriate cell probabilities in group Y.

[8] The consistency property of the CML-estimates depends on assuming the logistic form for the conditional distribution of Y. Violation of this supposition leads to a misspecified likelihood function and, in general, to inconsistent CML-estimates.

[9] The main results of this sector are based on Hosmer/Hosmer/Fisher (1983).

[10] Using a dummy variable for each cell, excepting one reference cell, is equivalent to the more standard procedure of employing dummies for the main and interaction effects of the qualitative variables separately. This second approach might have the advantage of not mixing main and interaction effects together.

Using $\theta_{Y.} \equiv \sum_{c=1}^{m-1} \theta_{Yc}$ the multinomial distribution of V in group Y may be expressed in terms of the natural parameterization for the exponential family by

$$f_Y(v) = \exp\left\{\sum_{c=1}^{m-1} v_c \ln\left(\frac{\theta_{Yc}}{1-\theta_{Y.}}\right) + \ln(1-\theta_{Y.})\right\}. \quad (17)$$

Writing the mean vector of $X^{(1)}$ as $\mu_{Yc}^{(1)} = \mu_{Ym}^{(1)} + \Delta_Y v$ where Δ_Y is a $(p_1 \times m-1)$ matrix, we get

$$f_Y(x) = f_Y(x^{(1)}, x^{(2)}) = f_Y(x^{(1)}, v) = \phi(\mu_{Ym}^{(1)} + \Delta_Y v, \Sigma) f_Y(v), \quad (18)$$

where $\phi(\cdot)$ denotes the normal density. Limiting our discussion to the case of equal interactions between the quantitative and qualitative variables in both groups, i.e. interactions that do not affect the response Y, we can set $\Delta_1 = \Delta_0 = \Delta$.[11] Under this assumption the log likelihood ratio is linear in X and we obtain for the true coefficients of our scoring function $S(x)$

$$S(x) = \alpha + \beta_v^{'} v + \beta_{x^{(1)}}^{'} x^{(1)}, \quad (19)$$

with

$$\alpha = \log\frac{\pi_1}{\pi_0} + \log\left(\frac{1-\theta_{1.}}{1-\theta_{0.}}\right) - \frac{1}{2}\left(\mu_{1m}^{(1)} - \mu_{0m}^{(1)}\right)^{'} \Sigma^{-1}\left(\mu_{1m}^{(1)} + \mu_{0m}^{(1)}\right), \quad (20)$$

$$\beta_v = \log\left(\frac{\theta_1(1-\theta_{0.})}{\theta_0(1-\theta_{1.})}\right) - \left(\mu_{1m}^{(1)} - \mu_{0m}^{(1)}\right)^{'} \Sigma^{-1}\Delta, \quad (21)$$

$$\beta_{x^{(1)}} = \Sigma^{-1}\left(\mu_{1m}^{(1)} - \mu_{0m}^{(1)}\right). \quad (22)$$

The first term in (21) denotes the $(m-1)$-dimensional vector of the log odds ratios $\log\left(\frac{\theta_{1c}(1-\theta_{0.})}{\theta_{0c}(1-\theta_{1.})}\right)$, $c = 1, \ldots, m-1$. The vector $\beta_{x^{(1)}}$ contains the coefficients for the p_1 quantitative variables and β_v contains the coefficients for the $m-1$ dummy variables that represent the qualitative information. The CML estimates of α and $\beta = (\beta_v^{'}, \beta_{x^{(1)}}^{'})^{'}$ are consistent because (14) holds. For the estimated coefficients using the LDA-approach we have the general expressions (see section 2)

$$\widehat{\alpha}_{LDA} = \log\frac{\widehat{\pi}_1}{\widehat{\pi}_0} - \frac{1}{2}\left(\overline{x}_1 - \overline{x}_0)^{'} \widehat{\Sigma}^{-1}(\overline{x}_1 + \overline{x}_0\right) \quad (23)$$

$$\widehat{\beta}_{LDA} = (\overline{x}_1 - \overline{x}_0)^{'} \widehat{\Sigma}^{-1}. \quad (24)$$

[11] The background for this assumption is provided by the observation that in most empirical scoring functions interaction terms are also missing.

The unconditional mean vector μ_Y and covariance matrix V_Y of $X = (X^{(1)'}, X^{(2)'})' = (X^{(1)'}, V')'$ in group Y are equal to

$$\mu_Y = ((\mu^{(1)}_{Ym} + \Delta\theta_Y)', \theta'_Y)' \tag{25}$$

$$V_Y = \begin{bmatrix} \Sigma_{Yx^{(1)}x^{(1)}} & \Sigma_{Yx^{(1)}v} \\ \Sigma'_{Yx^{(1)}v} & \Sigma_{Yvv} \end{bmatrix}, \tag{26}$$

with $\Sigma_{Yx^{(1)}x^{(1)}} = \Sigma + \Delta\Sigma_{Yvv}\Delta'$, $\Sigma_{Yx^{(1)}v} = \Delta\Sigma_{Yvv}$, and $\Sigma_{Yvv} = Diag(\theta_Y) - \theta_Y\theta'_Y$. If follows that as $n \to \infty$, $\overline{x}_Y$ and $\widehat{\Sigma}$ converge in probability to μ_Y and $V = \pi_1 V_1 + \pi_0 V_0$, respectively. On substitution of these limiting values in (23) and (24), the asymptotic form of $E_n(\widehat{S}_{LDA}(x)) = E_n(\widehat{D}(x))$ can be expressed as

$$\lim_{n\to\infty} E_n(\widehat{D}(x)) = (\alpha + b_\alpha) + (\beta_v + b_v)'v + \beta'_{x^{(1)}} x^{(1)}, \tag{27}$$

with

$$b_\alpha = -\log\left(\frac{1-\theta_{1.}}{1-\theta_{0.}}\right) - \frac{1}{2}(\theta_1 + \theta_0)'(\pi_1\Sigma_{1vv} + \pi_0\Sigma_{0vv})^{-1}(\theta_1 - \theta_0), \tag{28}$$

$$b_v = -\log\left(\frac{\theta_1(1-\theta_{0.})}{\theta_0(1-\theta_{1.})}\right) + (\theta_1 - \theta_0)'(\pi_1\Sigma_{1vv} + \pi_0\Sigma_{0vv})^{-1}. \tag{29}$$

It can be seen that the use of the normal-based LDA-estimates $\widehat{\alpha}_{LDA}$ and $\widehat{\beta}_{LDA}$ provides consistent estimation of $\beta'_{x^{(1)}}$, which contains the scoring function coefficients of the quantitative risk factors. However, the constant term α and the coefficients β_v of the qualitative risk factors are estimated inconsistently, their asymptotic biases being given by b_α and b_v, respectively. For the asymptotic bias of the *a posteriori* default probability $\widehat{P}_{LDA}(y = 1|x)$ results

$$b_P = \lim_{n\to\infty} E_n\left[P(y=1|x;\alpha,\beta) - \widehat{P}_{LDA}(y=1|x;\widehat{\alpha}_{LDA},\widehat{\beta}_{LDA})\right] \tag{30}$$

$$= \frac{1-\exp(g(v))}{1+\exp(g(v))+\exp(S(x))\exp(g(v))+\exp(-S(x))}, \tag{31}$$

with $g(v) = b_\alpha + b_v v$ and $S(x)$ given in (19). We can now use Monte Carlo simulations based on the developed location model (18) to test hypotheses H_1 and H_2. Furthermore, we can analyze in this way the influence of other factors (proportion of defaulted firms π_1, correlation structure between quantitative and qualitative data etc.) on the accuracy of 2-group classification and rating of firms based on their credit scores.

4 Monte Carlo setup

The range for the free parameters in the simulations is chosen in accordance with Krzanowski (1975) and Schmitz/Habbema/Hermans (1985), who also applied special variants of a location model for the generation of mixtures of continuous and binary variables.

a) Specification of the qualitative risk factors

Only the cases of $p_2 = 1$ and $p_2 = 2$ qualitative factors, each with three different categories ($q_j = 3$) are considered. For $p_2 = 1$ the number of different cells is $m = 3$ and for $p_2 = 2$ we have $m = 9$. The distribution of $X^{(2)}$ is determined through a specification of the cell probability vectors θ_Y. Using the vector of odds ratios $OR' = (OR_1, \dots, OR_{m-1})$, where

$$OR_c = \frac{\theta_{1c}(1-\theta_{0.})}{\theta_{0c}(1-\theta_{1.})}, \quad c = 1, \dots, m-1, \tag{32}$$

the relationship between the elements of θ_1 and θ_0 can be expressed as

$$\theta_{1c} = \frac{h_c}{1 + \sum_{c=1}^{m-1} h_c}, \tag{33}$$

where $h_c = (\theta_{0c} OR_c)/(1 - \sum_{c=1}^{m-1} \theta_{0c})$. Selecting values for OR and θ_0 (θ_1) fixes therefore the elements of θ_1 (θ_0). In analogy to Hosmer/Hosmer/Fisher (1983) we pick the numbers 0.3, 0.5, and 0.7 for $\theta_{0.}$, where each element of θ_0 is given by $\theta_{0c} = (\theta_{0.}/(m-1))$, $c = 1, \dots, m-1$. Hence the situation of equal probabilities for the first $m-1$ cells in group G_0 is examined. The values for the odds ratios are chosen as follows

$$\begin{array}{ll} OR = (4,\ 5) & if\ p_2 = 1, \\ OR = (2,\ 2.1,\ 2.5,\ 3,\ 3.2,\ 4,\ 4.5,\ 8) & if\ p_2 = 2. \end{array} \tag{34}$$

Odds ratios reflect not only the strength, but also the direction, of the relationship between the categories of the qualitative variables and the response variable Y. If $X^{(1)}$ and $X^{(2)}$ are independent, a value of $OR_c = 1$ means that cell c, i.e. the corresponding combination of qualitative risk factor categories, has no impact on group membership and thus $\beta_{v_c} = 0$. $OR_c > 1$ reflects a positive ($\beta_{v_c} > 0$) and $OR_c < 1$ a negative ($\beta_{v_c} < 0$) influence of c on Y. Therefore in the case $p_2 = 2$ we have both combinations of categories that are weakly positive related to Y and combinations that stand in a strong positive relationship to Y.

b) Specification of the quantitative risk factors

The analysis is limited to $p_1 = 3$ and $p_1 = 5$ quantitative factors. We have to find appropriate choices for the parameters $\mu_{Ym}^{(1)}$, Σ and Δ. Since a suitable transformation of $X^{(1)}$ can always be found such that the new variables are conditionally independent with unit variance, we set $\Sigma = I$. Without loss of generality, we can set

$$\mu_{Yc}^{(1)} = r_{Yc} T \quad , \ c = 1, \dots, m, \tag{35}$$

where T denotes a p_1-dimensional vector containing 1s. The correlation structure between the quantitative and qualitative data is governed by the choice for r_{Yc}. We consider two situations:

1. Financial and non-financial variables are uncorrelated. This leads to identical mean vectors in each cell, i.e. $\mu_{Yc}^{(1)} = \mu_Y^{(1)}$ (or $r_{Yc} = r_Y$), $c = 1, \ldots,$ m, and all elements of Δ are zero. In order to study the effect of distance between the two groups, for $r_1 = 0$ the values of r_0 are varied as follows, writing $d_c^2 = (\mu_{1c}^{(1)} - \mu_{0c}^{(1)})' \Sigma^{-1} (\mu_{1c}^{(1)} - \mu_{0c}^{(1)}) = d^2$ for the Mahalanobis squared distance between G_0 and G_1, conditional on the realization of the non-financial variables falling in multinomial cell c.

r_1	r_0	d^2 if $p_1 = 3$	d^2 if $p_1 = 5$
0	0.5	0.75	1.25
0	1	3	5

Table 1: Group distance

2. Financial and non-financial variables are correlated. We choose for r_{1c}, according to Krzanowski (1975), the expectation of the cth order statistic from a sample of size m taken from a standard normal distribution. These three ($p_2 = 1$) and nine ($p_2 = 2$) expectations are (see Pearson/Hartley (1972), table 9):

$$\begin{array}{lr} \pm 0.84628 \;\; 0.0 & if\ m = 3, \\ \pm 1.48501 \quad \pm 0.93230 \quad \pm 0.57197 \quad \pm 0.27453 \;\; 0.0 & if\ m = 9. \end{array}$$

To exclude interaction effects between $X^{(1)}$ and $X^{(2)}$ on Y, the difference $\mu_{1c}^{(1)} - \mu_{0c}^{(1)}$ (or $r_{1c} - r_{0c}$) must be equal in each cell. We take $r_{0c} = r_{1c} + 1$, $c = 1, \ldots, m$, and restrict therefore the Mahalanobis distance to $d^2 = 3$ for $p_1 = 3$ and to $d^2 = 5$ for $p_1 = 5$. The matrix Δ is given by $\mu_{Yc}^{(1)} = \mu_{Ym}^{(1)} + \Delta v$.

Following the structure of the data samples used by Moody's for the development of different country-specific versions of RiskCalc$^{\text{TM}}$ (see Table 2 , all tables and figures are in the Appendix) the sample size is set to $n = 1,000$ and $n = 10,000$ firms, with three different default proportions $\pi_1 \in \{0.01, 0.05, 0.1\}$ in each case.[12] We define two different scoring model types: (i) a simple model A with only three financial ($p_1 = 3$) and one non-financial variables ($p_2 = 1$), and (ii) a more complex model B with five financial ($p_1 = 5$) and two non-financial factors ($p_2 = 2$).[13] For both models and sample sizes 27 different parameter situations are analyzed by varying π_1, $\theta_{0.}$, Δ and d^2 (see Table 3). Three coherent

[12] Even if it is taken into account that the numbers in Table 2 also include validation samples, a development sample size of $n = 10,000$ seems not to be unrealistically large. This gives rise to the conjecture that the consequences of inconsistent coefficient estimation, documented in this paper, might even be more pronounced for some real data sets.

[13] The number of independent variables presumed here seems appropriate for the task of default forecasting. For example, in the empirical study of Shumway (2001), only five risk factors (two accounting ratios and three market-driven variables) are sufficient for a highly accurate classification. In addition, the German central bank (see Deutsche Bundesbank (1999)) uses scoring functions that have a structure similar to scoring model A.

situations can always be summarized in the form of a triple, i.e. the situations 1 to 3 form the first triple (1, 2, 3), the situations 4 to 6 the second (4, 5, 6) etc. The first two elements of each triple describe situations that have uncorrelated quantitative and qualitative risk factors, i.e. Δ equals a zero matrix, but with a rising Mahalanobis distance d^2 between both groups. The third situation of each triple is distinguished from the second by considering correlation, i.e. all elements of Δ are different from zero. The first three triples employ an *a priori* default probability (group proportion) of $\pi_1 = 0.01$, the next three triples (situations 10 to 18) use $\pi_1 = 0.05$ and the last three utilize $\pi_1 = 0.1$. The three triples that have an identical value of π_1 can be differentiated by their value for the vector of cell probabilities, i.e. by $\theta_{0.} = 0.1$, 0.5 or 0.7. Each situation is replicated $S = 1,000$ times.

The true coefficients for our simple and complex scoring functions with cell dummies are given by (19). We estimate these coefficients for the simulated data sets using the CML-(i.e. logistic regression) and the LDA-(i.e. linear discriminant analysis) approaches. It is known from the theoretical discussion above that all CML-estimates are asymptotically unbiased, whereas the LDA-estimates of the dummy coefficients are asymptotically biased. Scaling the qualitative factors using the FL method and estimating the coefficients for the scaled variables using linear discriminant analysis is called the FL/LDA-approach. We can test hypothesis H_1 by comparing the accuracy of 2-group classifications and ratings of firms based on the CML- and LDA-approach. Comparison of the LDA- and the FL/LDA-approach yields a test of H_2.[14]

Now we define our measures of classification and rating assignment accuracy.

- 2-group classification accuracy[15]
  The derivation of the optimal Bayes 2-group classification rule for our location model is straightforward if we take into account that the quantitative variables, given a particular realization of the qualitative variables (i.e. given a particular multinomial cell), are multivariate normal distributed with a common covariance matrix. Hence, the assumptions of LDA are satisfied in each cell and so the optimal rule leads to a different linear discriminant function for each of the m cells, with cutoff points determined by the *a priori* group probabilities and the discrete component of the model. The theoretical error rate e_{LM} based on this optimal rule is given by (see Krzanowski (1975))

$$e_{LM} = \pi_1 P(0|1) + \pi_0 P(1|0), \tag{36}$$

[14] Through the choice for the odds ratio values in the case $p_2 = 2$, all main and interaction effects of the two qualitative variables have a positive impact on the response Y. In order not to treat FL/LDA unfairly, an interaction term between the scaled variables is also considered here.

[15] In the following we assume equal costs ($C(0|1) = C(1|0)$) of misclassification.

where

$$P(0|1) = \sum_{c=1}^{m} \theta_{1c} \Phi \left(\frac{\ln\left(\frac{\pi_0 \theta_{0c}}{\pi_1 \theta_{1c}}\right) - \frac{1}{2} d_c^2}{d_c} \right), \tag{37}$$

$$P(1|0) = \sum_{c=1}^{m} \theta_{0c} \Phi \left(\frac{\ln\left(\frac{\pi_1 \theta_{1c}}{\pi_0 \theta_{0c}}\right) - \frac{1}{2} d_c^2}{d_c} \right), \tag{38}$$

$\theta_{Ym} = 1 - \theta_{Y.}$ and $\Phi(x)$ is the cumulative standard normal distribution function. Under the assumption of equal Mahalanobis distance ($d_c^2 = d^2$ for $c = 1, \ldots, m$) the coefficients of the m cell specific discriminant functions are identical and only the cutoff points differ.

In linear discriminant analysis, classification decisions are reached according to rule (10) on the basis of a single discriminant function $D(x)$. If all parameters of the location model are known, the optimal LDA classification rule is: allocate the firm to group G_1 if

$$D(x) \geq 0 \Longleftrightarrow (\alpha + b_\alpha) + (\beta_v + b_v)' v + \beta'_{x^{(1)}} x^{(1)} \geq 0 \tag{39}$$

and otherwise to G_0. Note that $D(x)$ in (39), given the realization of the qualitative variables falls in cell c, is normally distributed in group Y with mean $(\alpha + b_\alpha) + (\beta_{v_c} + b_{v_c}) + \beta'_{x^{(1)}} \mu_{Yc}^{(1)}$ and variance $\beta'_{x^{(1)}} \Sigma \beta_{x^{(1)}}$. Now it can easily be shown that the best achievable theoretical error rate e_{LDA} for the LDA-approach is given by

$$e_{LDA} = \pi_1 P(0|1) + \pi_0 P(1|0), \tag{40}$$

with

$$P(0|1) = \sum_{c=1}^{m} \theta_{1c} \Phi \left(\frac{\ln\left(\frac{\pi_0 \theta_{0c}}{\pi_1 \theta_{1c}}\right) + t_c - \frac{1}{2} d_c^2}{d_c} \right), \tag{41}$$

$$P(1|0) = \sum_{c=1}^{m} \theta_{0c} \Phi \left(\frac{\ln\left(\frac{\pi_1 \theta_{1c}}{\pi_0 \theta_{0c}}\right) - t_c - \frac{1}{2} d_c^2}{d_c} \right), \tag{42}$$

$$\begin{aligned} t_c = {} & \frac{1}{2} (\theta_1 + \theta_0)' (\pi_1 \Sigma_{1vv} + \pi_0 \Sigma_{0vv})^{-1} (\theta_1 - \theta_0) \\ & - (\theta_1 - \theta_0)' (\pi_1 \Sigma_{1vv} + \pi_0 \Sigma_{0vv})^{-1} - \ln \frac{\theta_{0c}}{\theta_{1c}}. \end{aligned} \tag{43}$$

Since rule (10) is not, in a Bayes sense, optimal for the location model we have $e_{LDA} \geq e_{LM}$.[16]

[16] Differences between the two classification rules (and therefore between e_{LDA} and e_{LM}) rest only on differences among the cell cutoff points, expressed by t_c.

Group allocations using standard logistic regression (CML-approach) are based on *a posteriori* probabilities, i.e. a firm is classified to group G_1 if

$$P(y=1|x) = \Lambda(S(x)) \geq 0.5$$
$$\Longleftrightarrow S(x) \geq 0 \Longleftrightarrow \alpha + \beta'_v v + \beta'_{x^{(1)}} x^{(1)} \geq 0. \quad (44)$$

and otherwise to G_0. Note that $S(x)$ in (44), given the realization of the qualitative variables falls in cell c, is normally distributed in group Y with mean $\alpha + \beta_{v_c} + \beta'_{x^{(1)}} \mu^{(1)}_{Yc}$ and variance $\beta'_{x^{(1)}} \Sigma \beta_{x^{(1)}}$. It can easily be seen that the best achievable theoretical error rate e_{CML} for the CML-approach equals the optimal error rate, i.e. $e_{CML} = e_{LM}$. As the calculation of theoretical error rates requires that the model parameters are known, the identification of the actual error rate ec for an estimated classification rule (i.e. scoring function), given a specific data sample, is of greater practical interest. Of the numerous methods available for determining this conditional error rate (see for example McLachlan (1992), Ch. 10), the 0.632-estimator $\widehat{ec}_{(0.632)}$ of Efron (1983) seems to be best with respect to the mean-squared error (MSE) criterion. The 0.632-estimator corrects the optimistic bias of the apparent error rate $\widehat{ec}_{app}$ using a bootstrap approach, i.e. drawing with replacement of n firms from the original sample to construct B new independent samples, where the reason for the remarkably low MSE of $\widehat{ec}_{(0.632)}$ is the lack of negative correlation between the estimated bias and $\widehat{ec}_{app} - ec$.[17] Taking the average of $\widehat{ec}_{(0.632)}$ for a number S of independent, equally-sized samples gives an estimate of the expected (unconditional) actual error rate eu.[18] We define $\widehat{eu}_{(0.632)} \equiv 1/1,000 \sum_{s=1}^{1,000} \widehat{ec}_{(0.632);s}$ as 0.632-estimator of the unconditional error rate eu. The difference $dif = \widehat{eu}_{(0.632)} - e_{LM}$ between the estimated unconditional error rate $\widehat{eu}_{(0.632)}$ and the optimal theoretical error rate e_{LM} is our measure of the accuracy of 2-group classifications.

- Accuracy of rating assignments
  We construct rating systems according to both methods (R1 and R2) described above. Within the scope of the R1 approach we form ten rating grades of equal size (i.e. deciles) for a given sample, using both the true and the (three) estimated credit scoring functions. For each simulated firm i, $i = 1, \ldots, n$, the true rating class rc_i using (19) is compared to the estimated class $\widehat{rc}_i$ based on the CML-, the LDA- or the FL/LDA-approach. The average proportion of corresponding assignments for each approach is taken as measure of the accuracy of rating assignments under the R1 rating method, i.e.

[17] By analogy with Efron (1983) the number of bootstrap replications is set here to $B = 200$.

[18] While the conditional error rate ec gives the accuracy of an particular estimated classification rule, the unconditional error rate eu describes the accuracy of a particular approach, such as CML or LDA, for constructing classification rules. Hence, the intended comparison of procedures in this paper is based on the unconditional error rate.

$$\text{Rating quality (R1)} = \frac{1}{1,000} \sum_{s=1}^{1,000} \sum_{i=1}^{n} Q_{R1}\left(rc_{is}, \widehat{rc}_{is}\right)/n \ , \qquad (45)$$

where rc_i, $\widehat{rc}_i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $Q_{R1}\left(rc_{is}, \widehat{rc}_{is}\right) = 1$ if $rc_{is} = \widehat{rc}_{is}$ and $Q_{R1}\left(rc_{is}, \widehat{rc}_{is}\right) = 0$ else.

The functional form of the relationship among credit scores and default rates is ascertained under the R2 rating method employing exponential regression, always calibrating the estimated default probabilities (PD's) to an average default rate of 2%.[19] In a next step the estimated PDs could be mapped into rating grades and the grade assignments of the different approaches could be compared with grades that make use of true PDs. True default probabilities are derived as described, but utilizing the true scoring function coefficients given in (19). Since the economic interest lies not primarily in differences in rating assignments but in the resulting capital requirements, we compare capital requirements calculated using estimated and true PDs for each firm.[20] The average proportion of firms with an absolute difference of equal to or less than 0.5% between estimated ($\widehat{cr}$) and true (cr) capital requirements is taken as a criterion for rating accuracy under the R2 rating method, i.e.

$$\text{Rating quality (R2)} = \frac{1}{1,000} \sum_{s=1}^{1,000} \sum_{i=1}^{n} Q_{R2}\left(cr_{is}, \widehat{cr}_{is}\right)/n \ , \qquad (46)$$

where $Q_{R2}\left(cr_{is}, \widehat{cr}_{is}\right) = 1$ if $\left|cr_{is} - \widehat{cr}_{is}\right| \leq 0.005$ and $Q_{R2}\left(cr_{is}, \widehat{cr}_{is}\right) = 0$ else.

5 Results

In this section we present our simulation results, based on (18), concerning classification and rating assignment accuracy of the three analyzed approaches: logistic regression (CML) and discriminant analysis (LDA) with dummy variables and discriminant analysis with scaled qualitative variables (FL/LDA).

a) Accuracy of 2-group classifications

Table 4 gives the values of the theoretical error rates e_{LM} and e_{LDA} for the 27 parameter situations and both scoring model types (model A contains three financial and one non-financial variables and model B includes five financial

[19] The rating agency Moody's employed an average one-year default rate of 1.6% for the calibration of the German version of their RiskCalc™ model (see Escott/Glormann/Kocagil (2001), p. 7).

[20] Capital requirements are computed according to the latest Basel II proposal for firms with annual sales greater than or equal to €50 million (see Basel Committee on Banking Supervision (2003), § 241-242). Furthermore, we set the loss given default to 45%, the maturity to 2.5 years and we explicitly take into account that Basel II imposes a minimum rating class default probability of 0.03%.

and two non-financial variables). It can be seen that error rates grow with the proportion π_1 of defaulted firms and fall if the group distance is increased (d^2 rises from 0.75 to 3 in model A and from 1.25 to 5 in model B by moving from the first to the second situation of each triple). Correlation among quantitative and qualitative variables has no impact on theoretical error rates (compare the second and third situation of each triple) and increasing the frequency of the first $m-1$ cells (i.e. $\theta_{0.}$) weakly increases the probability of misclassification.[21] For model A in particular, a high degree of correspondence between the optimal (e_{LM}) and the best with the LDA-approach achievable error rate (e_{LDA}) can be noted, which was to be expected.

To assess the effect of inconsistent scoring function estimation on the accuracy of 2-group classification, consider Figure 2 (model A) and Figure 3 (model B), which show for the CML- and the LDA-approach the course of our accuracy measure $dif = \widehat{eu}_{(0.632)} - e_{LM}$ relative to 54 distinct parameter situations (27 for each sample size). Concerning $n = 1,000$, differences in the value of dif for both procedures are clearly visible. For example, only the maximum likelihood CML-approach shows a pronounced positive correlation and distance effect if $\pi_1 = 0.01$ (default proportion 1%). In contrast to this, the LDA-approach demonstrates greater fluctuations in dif for varying $\theta_{0.}$. When $n = 10,000$, all of these differences disappear almost completely and the estimated unconditional error rates of both procedures approach their theoretical limits, e_{LM} and e_{LDA}, respectively.[22] This agrees with the results of Amemiya/Powell (1983) and Press/Wilson (1978) with respect to the robustness of LDA classifications. Increasing sample size does not lead to a decision in favor of the consistent CML-approach if the accuracy of classification is used as a selection criterion, i.e. asymptotically biased coefficient estimates should not be of practical concern for 2-group classifications.

In order to test the appropriateness of FL scaling, Figure 4 (model A) and Figure 5 (model B) compare the classification accuracy of LDA with that of the FL/LDA-approach. Obviously, without considering the correlation between $X^{(1)}$ and $X^{(2)}$, the expected probability of misclassification is almost identical for both procedures. While correlation influences LDA only slightly (compare the value for dif among the second and third situation of each triple), the FL/LDA-approach is strongly adversely affected.[23] Note that the resulting differences in classification accuracy for LDA and FL/LDA are independent of

[21] In model B, $\theta_{0.}$ sometimes also has a slightly negative effect on the error rate e_{LDA}.

[22] Note that the following convergence results hold:

a) for LDA: $\lim_{n\to\infty} eu_{LDA} = e_{LDA} > e_{LM}$,

b) for CML: $\lim_{n\to\infty} eu_{CML} = e_{CML} = e_{LM}$.

Thus dif approaches zero for CML and moves towards a small but positive limiting value ($e_{LDA} - e_{LM} > 0$) for LDA.

[23] For example, in situation 21 ($\pi_1 = 0.1$, $\theta_{0.} = 0.3$, $d^2 = 3$ and **correlation**) of model A and $n = 10,000$ we get the following numbers : $\widehat{eu}_{(0.632)} = 7.5073\%$ for LDA, $\widehat{eu}_{(0.632)} = 9.1462\%$ for FL/LDA and $e_{LM} = 7.4497\%$. This gives the differences $dif = 0.0576\%$ for LDA and $dif = 1.6965\%$ for FL/LDA. In contrast, in situation 20 ($\pi_1 = 0.1$, $\theta_{0.} = 0.3$, $d^2 = 3$ and **no correlation**) of model A and $n = 10,000$ we get the differences $dif = 0.0260\%$ for LDA and $dif = 0.0233\%$ for FL/LDA.

sample size and model type. By using FL scaling it is consequently not possible to reproduce the interaction structure among the quantitative and qualitative risk factors.[24] Employing dummy variables, on the other hand, allows a higher flexibility in modelling interactions, which nevertheless is connected with an increasing risk of overfitting, especially in small samples.

b) Accuracy of Rating assignments[25]

Figure 6 (model A) and Figure 7 (model B) show the course of rating accuracy under the R1 rating method for all three approaches (CML, LDA and FL/LDA) relative to 27 distinct parameter situations for each sample size. Remember that rating accuracy, given in (45), is defined here as the average proportion of firms for which the true and estimated rating grades correspond. It can be seen that, in general, rating accuracy rises with group distance d^2 and the proportion of defaulted firms π_1, which is 1%, 5% or 10%. The evident increase when 50 defaulted firms are present (i.e. $n = 1,000$, $\pi_1 = 0.05$) indicates that it is primarily not the proportion but the number of defaulted firms that matters, especially if the starting number is small. This supposition is confirmed by comparing the assignment accuracy in situations that lead to the same number of 100 defaulted firms despite different proportions (see situations 19 to 27 for $n = 1,000$ and 1 to 9 for $n = 10,000$). The course of rating accuracy for these both groups of situations is almost identical. Extending the sample size through additional non-defaulted firms might, therefore, have only a small influence on the accuracy of the approaches. Increasing the frequency of the first $m - 1$ cells has a negative impact on assignment accuracy for all three procedures (see the three triples with the same π_1). The reason is that with growing θ_0. the proportion of firms, for which the realization of the qualitative variables falls in one of the first $m - 1$ cells, rises and thus the accuracy of coefficient estimation for the cell dummies is of greater importance. For linear discriminant analysis employing dummy variables (LDA-approach) no correlation effect can be detected (see the horizontal course of rating accuracy among the second and third situation of a triple). This is also true for the maximum likelihood CML-approach, except in cases in which the number of defaulted firms is small (see situations 3, 6, 9, and additional 12, 15, 18 for model B). In these situations, defaulted firms are not always represented in each cell (i.e. no defaulted firm in the sample has, for example, a particular market position) and as a consequence the coefficients of the corresponding dummies cannot be determined by the CML-approach and the dummies are missing in the estimated scoring function. Because qualitative

[24] Note that by assuming $\Delta_1 = \Delta_0 = \Delta$ the interaction structure between $X^{(1)}$ and $X^{(2)}$ in both groups is identical and so interaction terms are not significant. However, this statement is not true for interactions among the qualitative variables and therefore the additional term $K_{X_1^{(2)}} \cdot K_{X_2^{(2)}}$ is employed for the FL/LDA-approach in model B.

[25] The rating accuracy of all three approaches is overestimated, due to the use of in-sample data for its determination. In contrast, the estimation of classification accuracy is based on a cross-validation/bootstrap approach. In the case of using out-of-sample data, any advantage of CML and LDA relative to FL/LDA might slightly be reduced in model B for $n = 1,000$ and $\pi_1 = 0.01$, because of possible overfitting of CML and LDA.

variables are of greater importance just when there are differences between them in the distribution of the quantitative variables, i.e. when correlation is present, missing dummy variables are especially problematic in this case.[26]

As was to be expected, rating accuracy falls for all procedures and situations with the number of coefficients to be estimated (compare model A with model B).

Figure 8 (model A) and Figure 9 (model B) display the course of rating accuracy for the three approaches under the R2 rating method. Rating accuracy is defined here (see (46)) as the average proportion of firms for which the difference between true and estimated capital requirements is small ($\leq 0.5\%$). The effects of correlation among both types of variables and increasing group distance d^2, established under the R1 rating method, also apply to the R2 method, i.e. d^2 has a positive bearing throughout and correlation influences, namely negatively, only the rating accuracy of the FL/LDA-approach.[27] Differences exist with respect to π_1 and $\theta_{0.}$. The negative effect of $\theta_{0.}$ is almost absent and rating accuracy is slightly diminished by the number of defaulted firms, particularly if $n = 1,000$.

The consequences of asymptotically biased coefficient estimation on rating accuracy become evident if the differences between the CML- and LDA-approaches for both scoring model types and rating methods are analyzed separately for each sample size. While the assignment accuracy of both approaches is almost equal for $n = 1,000$, the accuracy of the consistent CML procedure is always greater than that for the inconsistent LDA procedure when $n = 10,000$. The differences between the two procedures thus increase with sample size.[28] Because both statements of H_1 are true for the scoring functions A and B and rating methods R1 and R2, H_1 cannot be rejected based on evidence from a location model. Biased coefficent estimation hence has a negative bearing on rating accuracy and almost no effect on classification accuracy.

The introduction of correlation has a strong negative influence on discriminant analysis when using scaled qualitative variables (FL/LDA-approach). While the rating accuracy of LDA and FL/LDA differs little without correlation, the accuracy of FL/LDA partly drops about 40% and more when allowing for correlation (third situation of each triple). This reflects the results of analyzing the classification accuracy (see Figures 4 and 5). The FL scaling does therefore not lead to an improvement in rating or classification accuracy for the LDA-approach compared to using dummy variables. H_2 can be rejected for the

[26] In the presence of correlation the term $\left(\mu_{1m}^{(1)} - \mu_{0m}^{(1)}\right)' \Sigma^{-1}\Delta$ must be considered in addition to the odds ratio vector for the identification of the true coefficients, which can largely affect their values. For example, in model B in situations with $\theta_{0.} = 0.3$ we get for β_v without correlation $\beta_v = (0.69, 0.74, 0.92, 1.10, 1.16, 1.39, 1.50, 2.08)'$ and with correlation $\beta_v = (15.54, 12.83, 11.20, 9.90, 8.59, 7.44, 6.07, 4.84)'$.

[27] The negative correlation effect for the maximum likelihood CML-approach in situations with few defaults (i.e. 10 or 50) is preserved in model B.

[28] Consider, for example, situation 25 ($\pi_1 = 0.1$, $\theta_{0.} = 0.7$, $d^2 = 1.25$ and no correlation) for model B under rating method R1 (see Figure 7). The rating accuracy measure (45) is 48.92% (CML) and 46.74% (LDA) for $n = 1,000$ and 79.67% (CML) and 65.93% (LDA) for $n = 10,000$. Hence the difference rises from 2.18% to 13.74%.

considered cases.

6 Concluding remarks

The central results of this paper are:

- We have shown that consistent (i.e. asymptotically unbiased) estimation of credit scoring function coefficients is a necessary condition for obtaining rating grade assignments of high accuracy for the now more commonly available large samples of defaulted and non-defaulted firms. Because linear discriminant analysis estimators are not consistent if the independent variables are not normally distributed, this technique should not be used to derive credit scores for rating grade assignments.[29]

- 2-group classifications are not greatly affected by biased coefficient estimates. The difference between logistic regression and linear discriminant analysis is not large enough to favor one technique over the other if the task is classification of firms into only a few groups.

- Qualitative information can be incorporated into a scoring function using dummy variables. This should be unproblematic, even in small samples, as long as sample firms from both (default and non-default) groups are available for every category of the qualitative variables.

- There is no advantage to scaling qualitative factors with the FL method instead of using dummy variables in linear discriminant analysis. If there are (non-significant) interaction effects between the quantitative and the qualitative data, FL scaling of variables even results in a very low level of classification and rating assignment accuracy.

- Besides bias, the number of defaulted sample firms ($\pi_1 n$) and the distance between the two groups of firms (d^2) seem to be the most important factors in determining the accuracy of rating grade assignments. The sign of the effect of $\pi_1 n$ depends on the way in which ratings are derived from credit scores.

The default risk model of the German central bank (see Deutsche Bundesbank (1999)) offers a practical application of FL scaling. One part of this model covers three scoring functions that are specific to industry sectors. These functions are estimated with the FL/LDA-approach and equal in structure model

[29] As part of the minimum requirements for the IRB approach that apply to statistical models for assigning borrower ratings, the Basel Committee is demanding that „the model must be accurate on average across the range of borrowers to which the bank is exposed and there must be no known material biases" (Basel Committee on Banking Supervision (2003), § 379). Since this paper documents a material bias for LDA based ratings, the LDA technique seems not to conform to Basel II.

A, except that one function uses five instead of three quantitative variables (accounting ratios). Accounting behavior, subdivided into the categories „progressive", „neutral" and „conservative", is employed as a qualitative, non-financial risk factor. The satisfactory classification accuracy of these three functions, proved in benchmark studies (see Blochwitz/Liebig/Nyberg (2000)), indicates a low correlation between the considered accounting ratios and accounting behavior.[30] As shown above (see Figure 4), in such cases LDA and FL/LDA differ only slightly, but the negative consequences of biased coefficient estimation are prevalent (see Figures 6 and 8). The presented work justifies the assertion that the rating accuracy of the CML-approach cannot be attained by the Bundesbank FL/LDA-approach, especially when large development samples are employed.

This paper can be extended in several ways. First, while our focus is on asymptotic bias, it might also be of interest to study the effect of methods to correct credit scores for finite sample bias. In view of the above results and because finite sample bias is amplified by rare events, like defaults, it seems reasonable to assume a distinct impact of such correction methods on rating accuracy for at least some parameter situations. Second, a different model for the distribution of mixtures of financial and non-financial data could be considered. For instance, the properties of the LDA-estimates can be analyzed in a model that does not require conditional normality for the quantitative variables. Finally, the rating accuracy of other common scoring techniques, such as neural networks[31], can be compared to logistic regression or discriminant analysis.

[30] This conjecture makes sense because by taking advantage of accounting latitude, existing differences between firms should just be balanced. In contrast to this, interactions between a qualitative variable indicating the industry sector of a firm and particular accounting ratios are entirely expected. Interestingly, the Bundesbank does not utilize such an industry sector variable, scaled with the FL technique. Instead, a scoring function is estimated for each distinct sector.

[31] Note that our results also have consequences for so called perceptrons, which constitute a special class of neural networks. It is well known (see Ripley (1994), p. 414) that a logistic perceptron (a perceptron without a hidden layer and employing the logistic function $\Lambda(z)$ as activation function in the output unit) minimizing the Kullback-Leibler distance by using the Newton-Raphson method equals ordinary logistic regression. Therefore, all statements concerning the CML-approach also apply to a logistic perceptron, specified as described.

7 References

Afifi, A. A./Elashoff, R. M. (1969): Multivariate Two-Sample Tests with Dichotomous and Continuous Variables. 1. The Location Model. *Annals of Mathematical Statistics*, 40, 290-298.

Altman, E. I. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23, 589-609.

Altman, E. I./Narayanan, P. (2002): Business Failure Classification Models: An International Survey. In: Bankruptcy, Credit Risk, and High Yield Junk Bonds, ed. by E. I. Altman, Oxford 2002, 81-130.

Altman, E. I./Marco, G./Varetto, F. (1994): Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking & Finance*, 18, 505-529.

Amemiya, T./Powell, J. L. (1983): A Comparison of the Logit Model and Normal Discriminant Analysis when the Independent Variables are Binary. In: Studies in Econometrics, Time Series and Multivariate Statistics, ed. by S. Karlin, T. Amemiya and L. A. Goodman, New York 1983, 3-30.

Baetge, J./Huss, M./Niehaus, H.-J. (1988): The Use of Statistical Analysis to Identify the Financial Strength of Corporations in Germany. *Studies in Banking and Finance*, 7, 183-196.

Bardos, M. (1998): Detecting the Risk of Company Failure at the Banque de France. *Journal of Banking & Finance*, 22, 1405-1419.

Basel Committee on Banking Supervision (2000): Range of Practice in Banks' Internal Rating Systems, Discussion Paper, Basel January 2000.

Basel Committee on Banking Supervision (2003): The New Basel Capital Accord, Consultative Document, Basel April 2003.

Becchetti, L./Sierra, J. (2003): Bankruptcy Risk and Productive Efficiency in Manufacturing Firms. *Journal of Banking & Finance*, 27, 2099-2120.

Blochwitz, S./Eigermann, J. (1999): Effiziente Kreditrisikobeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen. In: Handbuch Kreditrisikomodelle und Kreditderivate, hrsg. v. R. Eller, W. Gruber und M. Reif, Stuttgart 1999, 4-22.

Blochwitz, S./Eigermann, J. (2000): Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen. *Zeitschrift für betriebswirtschaftliche Forschung*, 52, 58-73.

Blochwitz, S./Liebig, T./Nyberg, M. (2000): Benchmarking Deutsche Bundesbank's Default Risk Model, the KMV^{TM} Private Firm Model and Common Financial Ratios for German Corporations, Working Paper, Moody's/ KMV, New York November 2000.

Carey, M./Hrycay, M. (2001): Parameterizing Credit Risk Models with Rating Data. *Journal of Banking & Finance*, 25, 197-270.

Crouhy, M./Galai, D./Mark, R. (2001): Prototype Risk Rating System. *Journal of Banking & Finance*, 25, 47-95.

Day, N. E./Kerridge, D. F. (1967): A General Maximum Likelihood Discriminant. *Biometrics*, 23, 313-324.

Deutsche Bundesbank (1999): The Bundesbank's Method of Assessing the Creditworthiness of Business Enterprises. *Deutsche Bundesbank Monthly Report*, January 1999, 51-63.

Efron, B. (1975): The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis. *Journal of the American Statistical Association*, 70, 892-898.

Efron, B. (1983): Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association*, 78, 316-331.

Elpelt, B./Hartung, J. (1986): On Scaling of Qualitative Data. In: Classification as a Tool of Research, ed. by W. Gaul and M. Schader, Amsterdam 1986, 113-120.

Escott, P./Glormann, F./Kocagil, A. E. (2001): Moody's RiskCalc^{TM} for Private Companies: The German Model, Working Paper, Moody's Investors Service, New York November 2001.

Fisher, R. A. (1936): The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, 179-188.

Frerichs, H./Wahrenburg, M. (2003): Evaluating Internal Credit Rating Systems Depending on Bank Size, Working Paper, Johann Wolfgang Goethe University, Frankfurt am Main September 2003.

Fritz, S. G./Popken, L./Wagner, C. (2002): Scoring and Validating Techniques for Credit Risk Rating Systems. In: Credit Ratings - Methodologies, Rationale and Default Risk, ed. by M. K. Ong, London 2002, 189-212.

Grunert, J./Norden, L./Weber, M. (2004): The Role of Non-financial Factors in Internal Credit Ratings. *Journal of Banking & Finance*, forthcoming.

Höse, S./Huschens, S. (2003): From Credit Scores to Stable Default Probabilities: A Model Based Approach. In: Exploratory Data Analysis in Empirical Research, ed. by M. Schwaiger and O. Opitz, Berlin 2003, 454-462.

Hosmer, T./Hosmer, D./Fisher, L. (1983): A Comparison of the Maximum Likelihood and Discriminant Function Estimators of the Coefficients of the Logistic Regression Model for Mixed Continuous and Discrete Variables. *Communications in Statistics - Simulation and Computation*, 12, 23-43.

Izan, H. Y. (1984): Corporate Distress in Australia. *Journal of Banking & Finance*, 8, 303-320.

King, G./Zeng, L. (2001): Logistic Regression in Rare Events Data. *Political Analysis*, 9, 137–163.

Krahnen, J. P./Weber, M. (2001): Generally Accepted Rating Principles: A Primer. *Journal of Banking & Finance*, 25, 3-23.

Krzanowski, W. J. (1975): Discrimination and Classification using both Binary and Continuous Variables. *Journal of the American Statistical Association*, 70, 782-790.

Lincoln, M. (1984): An Empirical Study of the Usefulness of Accounting Ratios to Describe Levels of Insolvency Risk. *Journal of Banking & Finance*, 8, 321-340.

Martin, D. (1977): Early Warning of Bank Failure - A Logit Regression Approach. *Journal of Banking & Finance*, 1, 249-276.

McLachlan, G. J. (1992): Discriminant Analysis and Statistical Pattern Recognition, New York 1992.

McLachlan, G. J./Byth, K. (1979): Expected Error Rates for Logistic Regression versus Normal Discriminant Analysis. *Biometrical Journal*, 21, 47-56.

Micha, B. (1984): Analysis of Business Failures in France. *Journal of Banking & Finance*, 8, 281-291.

Olkin, I./Tate, R. F. (1961): Multivariate Correlation Models with Mixed Discrete and Continuous Variables. *Annals of Mathematical Statistics*, 32, 448-465.

Pearson, E. S./Hartley, H. O. (1972): Biometrika Tables for Statisticians - Volume II, Cambridge 1972.

Press, J. S. /Wilson, S. (1978): Choosing between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association*, 73, 699-705.

Ripley, B. D. (1994): Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society*, 56, 409-456.

Schmitz, P. I. M./Habbema, J. D. F./Hermans, J. (1985): A Simulation Study of the Performance of Five Discriminant Analysis Methods for Mixtures of Continuous and Binary Variables. *Journal of Statistical Computation and Simulation*, 23, 69-95.

Shumway, T. (2001): Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74, 101-124.

Taffler, R. J. (1984): Empirical Models for the Monitoring of UK Corporations. *Journal of Banking & Finance*, 8, 199-227.

Treacy, W. F./Carey, M. (2000): Credit Risk Rating Systems at Large US Banks. *Journal of Banking & Finance*, 24, 167-201.

Welch, B. L. (1939): Note on Discriminant Functions. *Biometrika*, 31, 218-220.

8 Appendix

RiskCalc^{TM} model for	n	π_1
Australia	29,636	8.4998%
Austria	19,524	11.0428%
Belgium	102,594	6.4897%
Canada	8,115	6.1738%
France	253,268	9.9614%
Germany$^{\#}$	4,866	9.9671%
Italy	52,327	1.8308%
Japan	41,557	2.7504%
Korea	32,228	10.2302%
Mexico	3,797	46.4051%
The Netherlands	19,327	2.2559%
Nordic Region	178,436	5.8873%
Portugal	18,137	2.2937%
Singapore	4,441	14.2310%
Spain	140,790	1.6088%
UK	64,531	7.3190%
USA	33,964	4.1014%

Table 2: Number of all firms (n) and proportion of defaulted firms (π_1) based on development and validation samples for particular, country-specific versions of Moody's RiskCalc^{TM} default model (source: Moody's Investors Service)

$^{\#}$ only development sample

situation	triple	default proportion π_1			cell probability $\theta_{0.}$			corre-lation		distance d^2	
		0.01	0.05	0.1	0.1	0.5	0.7	no	yes	0.75	3
1		X			X			X		X	
2	1	X			X			X			X
3		X			X				X		X
4		X				X		X		X	
5	2	X				X		X			X
6		X				X			X		X
7		X					X	X		X	
8	3	X					X	X			X
9		X					X		X		X
10			X		X			X		X	
11	4		X		X			X			X
12			X		X				X		X
13			X			X		X		X	
14	5		X			X		X			X
15			X			X			X		X
16			X				X	X		X	
17	6		X				X	X			X
18			X				X		X		X
19				X	X			X		X	
20	7			X	X			X			X
21				X	X				X		X
22				X		X		X		X	
23	8			X		X		X			X
24				X		X			X		X
25				X			X	X		X	
26	9			X			X	X			X
27				X			X		X		X

Table 3: Parameter constellation for the 27 distinct simulation situations (instead of 0.75 and 3 we have the distance values 1.25 and 5 in model B)

situation	model A		model B	
	e_{LM}	e_{LDA}	e_{LM}	e_{LDA}
1	1.0000%	1.0000%	0.9992%	1.0530%
2	0.9700%	0.9712%	0.8659%	0.9246%
3	0.9700%	0.9712%	0.8659%	0.9246%
4	1.0000%	1.0000%	0.9995%	1.0056%
5	0.9748%	0.9750%	0.8692%	0.8891%
6	0.9748%	0.9750%	0.8692%	0.8891%
7	1.0000%	1.0000%	0.9997%	1.0006%
8	0.9792%	0.9794%	0.8761%	0.8839%
9	0.9792%	0.9794%	0.8761%	0.8839%
10	4.9829%	4.9870%	4.8698%	5.2139%
11	4.2457%	4.2563%	3.3640%	3.4993%
12	4.2457%	4.2563%	3.3640%	3.4993%
13	4.9917%	4.9919%	4.8932%	5.0066%
14	4.3057%	4.3088%	3.3772%	3.4410%
15	4.3057%	4.3088%	3.3772%	3.4410%
16	4.9958%	4.9959%	4.9165%	4.9528%
17	4.3773%	4.3823%	3.4227%	3.4546%
18	4.3773%	4.3823%	3.4227%	3.4546%
19	9.7561%	9.7768%	9.2323%	9.5783%
20	7.4497%	7.4680%	5.6038%	5.7298%
21	7.4497%	7.4680%	5.6038%	5.7298%
22	9.8500%	9.8519%	9.3130%	9.4847%
23	7.5818%	7.5880%	5.6211%	5.6908%
24	7.5818%	7.5880%	5.6211%	5.6908%
25	9.9085%	9.9098%	9.4192%	9.4944%
26	7.7632%	7.7759%	5.7077%	5.7480%
27	7.7632%	7.7759%	5.7077%	5.7480%

Table 4: Theoretical error rates

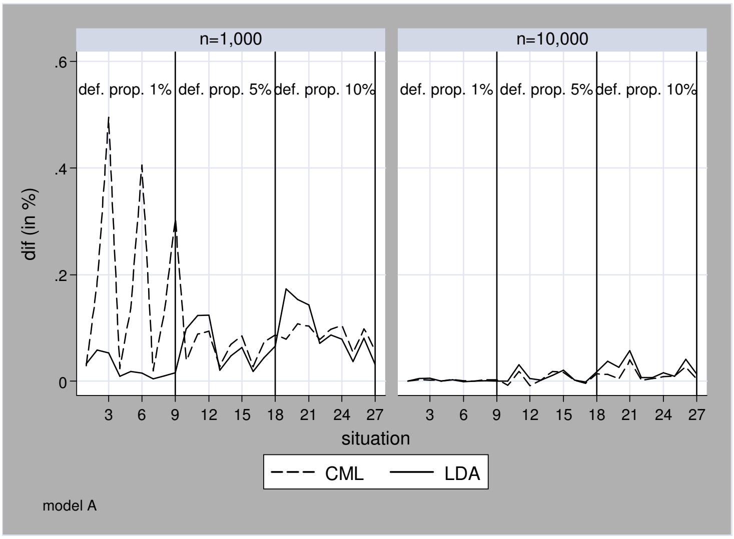

Figure 2: Course of the difference (dif) between the estimated unconditional error rate $\widehat{eu}_{(0.632)}$ and the optimal error rate e_{LM} for the procedures CML and LDA in model A ($p_1 = 3$, $p_2 = 1$)

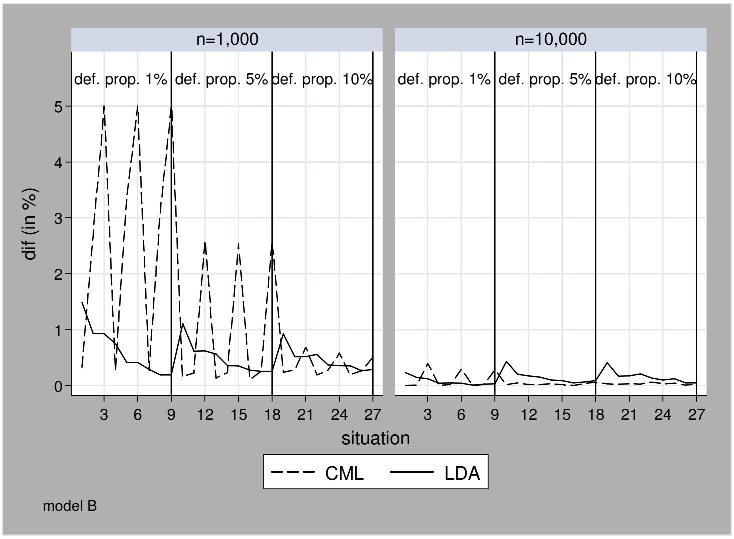

Figure 3: Course of the difference (dif) between the estimated unconditional error rate $\widehat{eu}_{(0.632)}$ and the optimal error rate e_{LM} for the procedures CML and LDA in model B ($p_1 = 5$, $p_2 = 2$); for the purpose of presentation the differences for CML in situations 3, 6 and 9 are restricted to 5%, the real values are: 12.5687%, 16.2035% and 15.6539%.

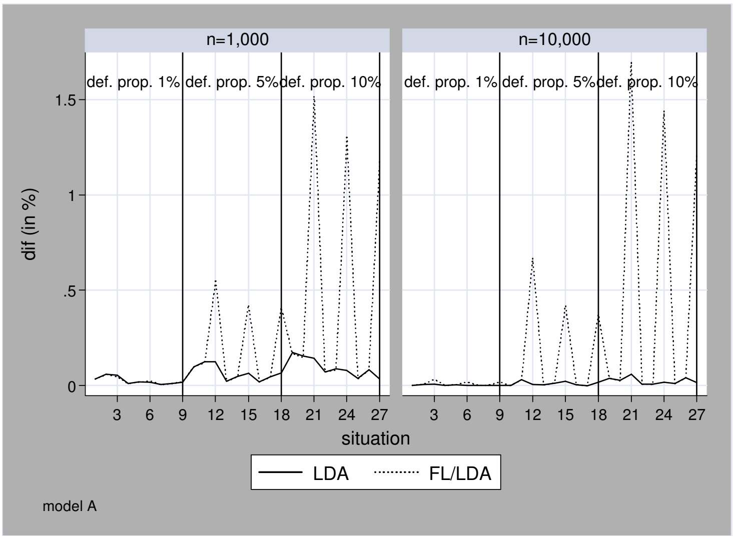

Figure 4: Course of the difference (dif) between the estimated unconditional error rate $\widehat{eu}_{(0.632)}$ and the optimal error rate e_{LM} for the procedures LDA and FL/LDA in model A ($p_1 = 3$, $p_2 = 1$)

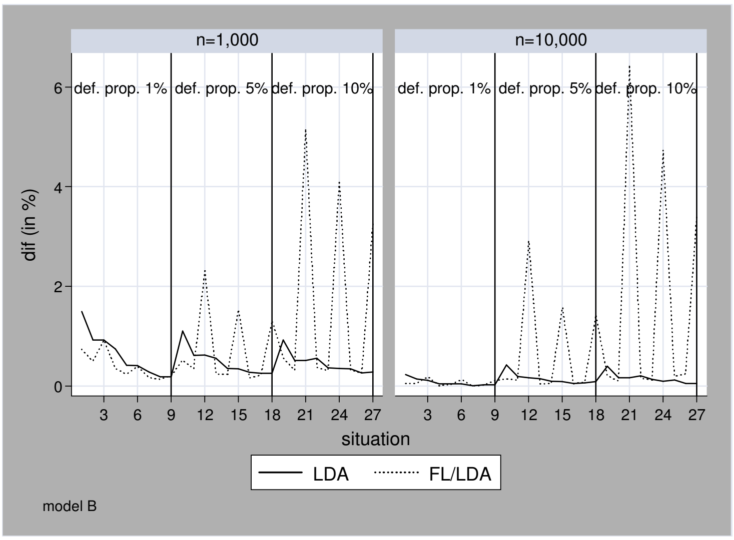

Figure 5: Course of the difference (dif) between the estimated unconditional error rate $\widehat{eu}_{(0.632)}$ and the optimal error rate e_{LM} for the procedures LDA and FL/LDA in model B ($p_1 = 5$, $p_2 = 2$)

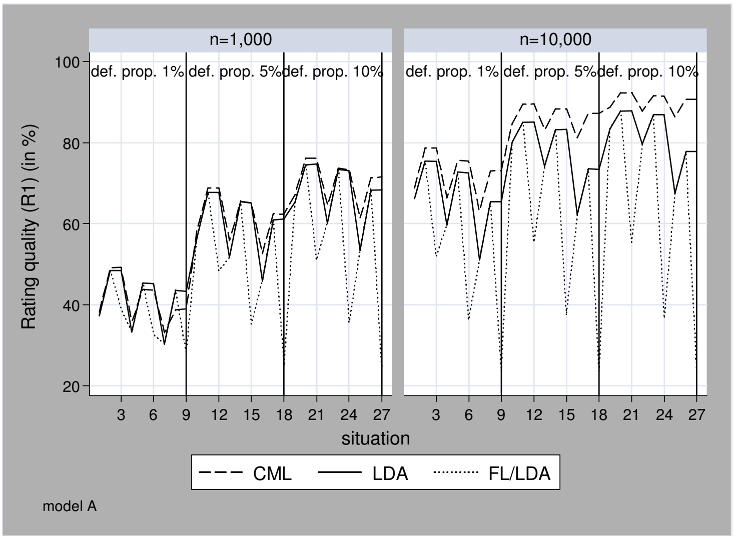

Figure 6: Course of the average proportion of firms with correspondence between true and estimated rating assignments for procedures CML, LDA and FL/LDA in model A ($p_1 = 3$, $p_2 = 1$)

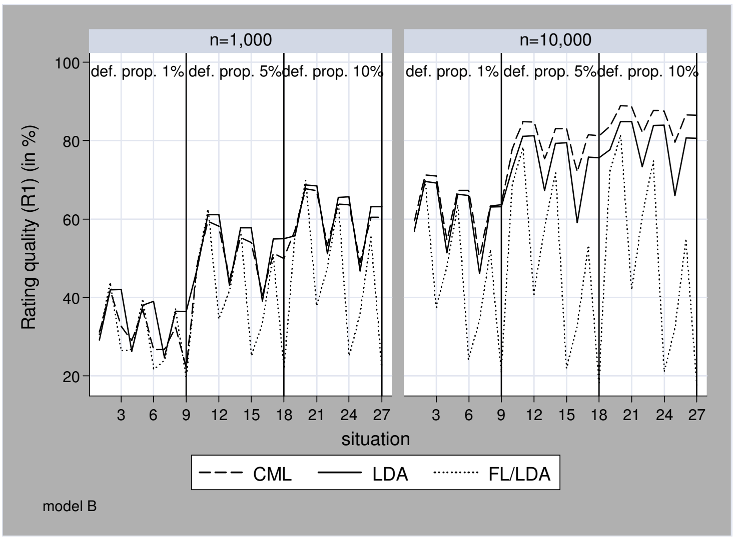

Figure 7: Course of the average proportion of firms with correspondence between true and estimated rating assignments for procedures CML, LDA and FL/LDA in model B ($p_1 = 5$, $p_2 = 2$)

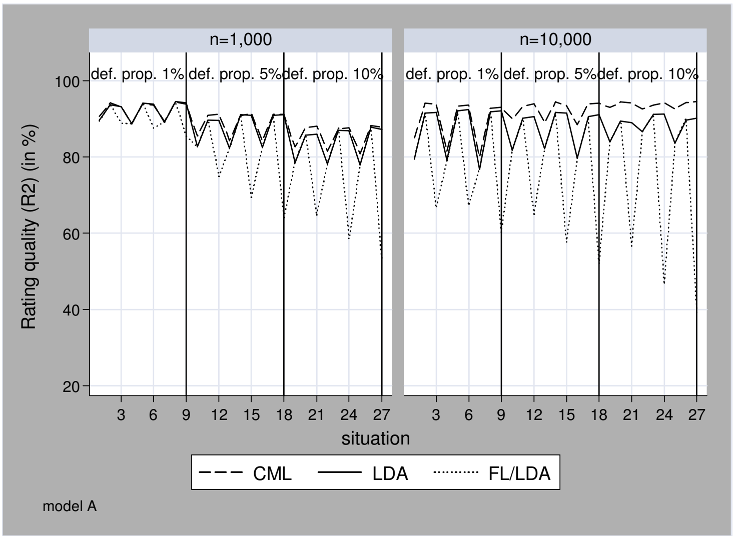

Figure 8: Course of the average proportion of firms with a small (≤ 0.005) absolute difference among true and estimated capital requirements for procedures CML, LDA and FL/LDA in model A ($p_1 = 3$, $p_2 = 1$)

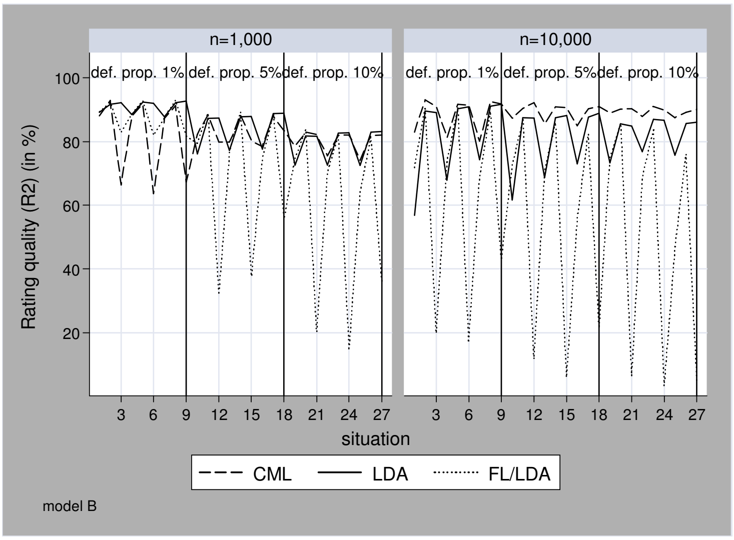

Figure 9: Course of the average proportion of firms with a small (≤ 0.005) absolute difference among true and estimated capital requirements for procedures CML, LDA and FL/LDA in model B ($p_1 = 5$, $p_2 = 2$)