JOHANN WOLFGANG GOETHE-UNIVERSITÄT
FRANKFURT AM MAIN

FACHBEREICH WIRTSCHAFTSWISSENSCHAFTEN

Hergen Frerichs / Mark Wahrenburg

Evaluating internal credit rating systems depending on bank size

No. 115
September 2003

WORKING PAPER SERIES: FINANCE & ACCOUNTING

Hergen Frerichs[a,*], Mark Wahrenburg[a]

Evaluating internal credit rating systems depending on bank size

No.115

September 2003

ISSN 1434-3401

[a] Chair of Banking & Finance, University of Frankfurt (Main), P.O. Box 11 19 32, 60054 Frankfurt (Main), Germany.

* Corresponding author: Tel.: ++49-69-79828959, facsimile: ++49-69-79822143. *E-mail addresses*: frerichs@wiwi.uni-frankfurt.de, wahrenburg@wiwi.uni-frankfurt.de

This paper is a result of a research cooperation with Deutsche Bundesbank. We wish to thank Richard Cantor, Ian Cooper, Lorne Switzer, an anonymous referee, the credit risk research group of Deutsche Bundesbank, participants at the 6th conference of the Swiss Society for Financial Markets Research, the 2003 Global Finance Conference, the 2003 Annual Meeting of the European Financial Management Association, and seminar participants at Frankfurt and Münster for helpful comments.

Abstract

Under a new Basel capital accord, bank regulators might use quantitative measures when evaluating the eligibility of internal credit rating systems for the internal ratings based approach. Based on data from Deutsche Bundesbank and using a simulation approach, we find that it is possible to identify strongly inferior rating systems out-of-time based on statistics that measure either the quality of ranking borrowers from good to bad, or the quality of individual default probability forecasts. Banks do not significantly improve system quality if they use credit scores instead of ratings, or logistic regression default probability estimates instead of historical data. Banks that are not able to discriminate between high- and low-risk borrowers increase their average capital requirements due to the concavity of the capital requirements function.

Key words: credit risk, credit ratings, bank regulation, Basel II

JEL classification: G2, G21, G28, C52

0. Introduction

Under the proposal of a new Basel capital accord (Basel Committee, 2003), banks will be allowed to use their own default probability estimates for regulatory capital calculation. Bank regulators will have to decide whether a bank's internal credit rating system meets certain minimum requirements specified in the proposal. We implement quantitative measures to evaluate system quality. While these measures will not be sufficient to decide on a system's admittance to the internal ratings based approach, they may serve as valuable and objective quality indicators. Bank regulators might carry out an analysis like ours to define threshold values in order to separate high-quality systems from low-quality ones.

Based on Deutsche Bundesbank annual accounts and default data in the time period 1994-99 of, on average, 24,000 medium-sized and large companies per year, we investigate whether it is possible to identify low-quality internal credit rating systems based on quantitative measures. In the course of the analysis, we treat two further research questions: First, should banks abandon their current practice of aggregating credit scores into rating classes, and of using historical default rates to estimate default probabilities instead of rating model derived estimates? Second, how does a rating system's quality influence capital requirements?

Our empirical approach proceeds as follows: We start by defining rating systems of different quality. We then choose a quantitative measure that summarizes system quality, and empirically simulate its distribution for the various systems defined. Finally, we compare the statistics' distributions for high-quality systems with those for low-quality ones. If the distributions sufficiently differ, the quantitative measure performs well in discriminating between both kinds of systems. We perform our analysis for different bank sizes and different portfolio default rates.

We evaluate the performance of three statistical measures that have been widely used in other fields of science (medicine, psychology, meteorology), but also in economics[1]: the area-under-curve (AUC), the Brier score, and the grouped Brier score.

The AUC evaluates the ranking of borrowers from good to bad, while the Brier score evaluates individual default probability forecasts, and the grouped Brier score average rating class default probability forecasts. Each of the three measures is associated with a different economic loss function. As evaluations are sensitive to the loss function chosen[2], we need to analyze whether a given statistical measure makes sense economically. While the AUC dominates the current discussion, we find that the Brier score and the grouped Brier score are more closely related to the evaluator's objectives. Regulatory capital requirements directly depend on default probability estimates, and therefore the error in default probability estimates ought to be measured in order to evaluate an internal credit rating system.

The main result of our study is that it is possible to identify strongly inferior internal credit rating systems out-of-time based on both the AUC and the Brier score. For example, a rating system which is based on one randomly drawn financial ratio is identified as inferior with a high probability even if there are only three out-of-time defaults. The grouped Brier score does not work well in our study as it is exclusively concerned with precision and not with discriminatory power. Because average rating class default probabilities are relatively precisely measured even by the low-quality systems we define, the grouped Brier score is not able to discriminate between good and bad systems.

The identification frequency of inferior systems depends positively on the number of out-of-time defaults meaning that it is similar if either a large portfolio with a low portfolio default rate or a small portfolio with a high portfolio default rate is taken.

The identification frequency decreases considerably if more sophisticated systems are examined. For example, a system that uses the financial ratios of the Altman Z"-score, which are selected based on U.S. data, is only recognized as being inferior

---

[1] In economics, the AUC is used in (Sobehart et al., 2000), (Blochwitz et al., 2000), (Engelmann et al., 2003). The Brier score is used in (Diebold and Rudebusch, 1989), (Winkler, 1994), (Lopez, 1999, 2001). The grouped Brier score is used in (Diebold and Rudebusch, 1989).

[2] (Lopez, 2001)

with a high probability for large banks (51 or more out-of-time defaults). As none of the financial ratios used for the Altman Z”-score is chosen in a stepwise selection procedure based on the Deutsche Bundesbank database, this result is not satisfactory and reflects the limitations of our approach.

Our results are derived for two different evaluation approaches. In the first approach, regulators simply set critical thresholds on the value of a quantitative measure. Performing worse than these thresholds indicates a system’s underperformance. In the second approach, regulators set critical thresholds on the p-values of tests of equality of a bank’s own system’s quality measures and the respective measures of a benchmark model calibrated by regulators. This approach is more complex than the first one. It performs worse in rejecting the system, which is based on one randomly drawn financial ratio, and better in rejecting the system based on the Altman Z”-score financial variables.

With respect to our additional research questions, we find that banks do not significantly improve their systems’ quality if they do not aggregate credit scores into rating classes, or if they use rating model derived default probability estimates instead of estimates based on internal default histories. These results accord with the current behaviour of many banks.

Banks might have an incentive to use logistic regression default probability estimates instead of historical rating class default rates, if capital requirements can be lowered in this way. For one of the rating systems we define this is actually the case. A systematic underestimation of low default rates along with an overestimation of high default rates in connection with a concave capital requirements function, leads to lower average capital requirements. For all other systems capital requirements based on logistic regression default probability estimates are higher on average. These systematic differences also exist with probit regression and linear discriminant analysis, but can be overcome with a non-parametric approach.

Concerning capital requirements, rating systems that are not able to discriminate between high and low risks are punished by higher capital requirements. This is also due to the concavity of the Basel II capital requirements function.

As far as we know, none of our research questions has so far been treated in the credit risk research literature. Therefore, we believe that our insights represent a contribution to the small body of papers on the validation of credit risk models.

Related literature can be classified into four areas: first, classical papers on credit scoring; second, theoretical papers on the refinement of statistical tools for analyzing statistical default models; third, empirical papers applying refined assessment criteria; and fourth, papers discussing rating system evaluation procedures:

1. In the classical literature starting with (Altman, 1968) simple error rates are used for model validation. A review of this big body of literature along with a discussion of basic problems is given by (Rosenberg and Gleit, 1994).

2. Statistical tools that go beyond simple error rates are theoretically discussed in (DeLong et al., 1988) and (Swets, 1996), as representatives of studies in medical and psychological research, (Wilkie, 1992), (Hand, 1994, 1997), (Hand and Henley, 1997), (Sobehart et al., 2000), and (Engelmann, Hayden and Tasche, 2003).

3. Recent empirical papers applying refined assessment criteria to credit risk include (Sobehart et al., 2000), (Blochwitz et al., 2000), and (Carey and Hrycay, 2001). (Sobehart et al., 2000) uses accuracy and entropy ratios to measure the accuracy of six different scoring and rating models, based on balance sheet and market information from public companies. (Blochwitz et al., 2000) use gini coefficients to compare Deutsche Bundesbank's credit scoring system (discriminant analysis plus expert system) with the KMV private firm model using data from Deutsche Bundesbank. (Carey and Hrycay, 2001) empirically examine properties of mapping- and scoring-model methods, which they use to estimate average default probabilities by rating grade, and present evidence of potential problems of bias, instability, and gaming. (Grunert, Norden and Weber, 2003) find that qualitative factors significantly increase the performance of internal credit rating systems based on a small sample of internal credit ratings given by four German banks in the time period 1992-1996 using Brier scores among other measures.

4. (RMA Capital Working Group, 2000) and (Carey, 2001) propose an alternative evaluation procedure for credit rating systems. In the peer group approach, banks are asked to provide ratings for a given sample of borrowers. Banks whose

ratings differ significantly from ratings given by other banks are regarded as outliers and are further investigated. (Tabakis and Vinci, 2002) propose an intermediate approach: First, a set of publicly available financial variables is used to obtain a core rating, and then, additional information from peer group ratings is combined in a variance-minimizing way to obtain a benchmark rating.

The remainder of the paper is organized as follows. Section 1 describes the data. In section 2, we define six kinds of internal credit rating systems differing in quality. Section 3 describes the quantitative measures we use. Section 4 presents the simulation set-up and simulation results and section 5 concludes.

1. Data

Deutsche Bundesbank's annual accounts database is the most comprehensive collection of annual accounts of German non-financial companies. Nevertheless, due to its rediscount business origin, it is somewhat biased towards large public limited West German manufacturing companies, and thus not entirely representative of the German economy.[3]

An important characteristic of the database is that companies usually submit annual accounts based on tax law to Deutsche Bundesbank. While annual accounts based on commercial law generally have to be finished within three to six months after the end of the financial year, the compilation period for those based on tax law is generally up to one year. This characteristic causes problems for default prediction. If a company defaulted, e.g. in 1998, then it is quite probable that there are no annual accounts of 1997 in the database, and there might not even be annual accounts of 1996.

Default is defined as the formal initiation of insolvency proceedings. Default information is retrieved from public sources and incorporated into the database as soon as it becomes known. This default definition is narrower than the Basel II definition. There, default is additionally triggered if the obligor is unlikely to pay its debt obligations in full, without recourse by the bank to actions such as realising

[3] (Deutsche Bundesbank, 1998)

security, or if the obligor is past due more than 90 days on any material credit.[4] As a consequence, default rates in the Deutsche Bundesbank database are relatively low.

Our dataset contains annual accounts and default data of medium-sized and large German companies in the time period 1990-2000. Companies have to satisfy at least two of the following criteria:

1. total assets larger than 3,438 million Euros,
2. revenues larger than 6,875 million Euros,
3. a yearly average of more than 50 employees.[5]

Small companies are excluded from the dataset because they have more means of distorting annual accounts data than medium-sized and large companies. By focusing on medium-sized and large companies, we also reduce the problem that small business lending relies more on qualitative information, which we do not have.[6]

Companies that have once satisfied the size criterion are retained in later years even if the size criterion is no longer satisfied. This reflects the lending behaviour of banks that cannot easily get rid off a customer after granting a credit even if she no longer belongs to the target group, which is here defined by size. The proportion of annual accounts based on tax law equals 88%.

The classification of annual accounts as solvent and insolvent, respectively, is based on a simplified view of a bank granting a loan to a company. If a company asks for a loan in $t=0$ (t being spaced in monthly intervals), the bank tries to predict the company's default by $t=12$ based on the company's latest annual accounts (which are typically based on tax law). Since we define the compilation period for annual accounts based on tax law to equal one year, the time difference d between default and the latest annual accounts available lies in the interval $d \in [1,24]$. For example, if

---

[4] (Basel Committee, 2003), § 414

[5] §267 HGB (Handelsgesetzbuch / German Corporate Law) defining the size criterion for public limited companies.

[6] (Berger et al., 2003) provide empirical evidence for this point. Note that the minimum total assets in our sample is larger than the 75%-quantile of the sample of Berger et al. Furthermore, the primary empirical result refers to companies without financial statements, which do not exist in our sample.

the financial year ends on 31 December 2000, the annual accounts are compiled and submitted on 10 January 2001, the company is granted credit on 15 January 2001, and defaults on 20 January 2001, then $d = 1$. If the company is granted a credit on 15 December 2001, compiles and submits its annual accounts for the financial year ending in December 2000 on 20 December 2001, and defaults on 19 December 2002, then $d = 24$.

1,446 annual accounts are classified as insolvent because they lie within the interval $d \in [1,24]$. Since we do not have default information after January 2001, we are not able to classify annual accounts after 31 January 1999 as solvent or insolvent, which leads to a reduction in data available for default prediction.

For the present study, we use data from 1994-1999. As banks and regulators are primarily interested in the future performance of internal credit rating systems, we divide the sample into a 1994-1998 training sample and a 1999 validation sample. The training sample time period of five years complies with the Basel II minimum historical time period.[7] While in retail credit a maximum of three years of data is commonly used to derive credit scoring functions,[8] the lack of default events in corporate credit forces especially small banks to use more years of data if available.

The training sample consists of 98,910 observations of 29,607 companies with an average default rate of 0.58%. The validation sample consists of 18,671 observations (= companies) with an average default rate of 0.74%.

Our credit scoring is based on a set of forty-eight financial variables, which have been found to be good default indicators in the German credit risk literature, and one variable from (Altman, 1968) (Table 1). Forty-one financial variables are taken from (Niehaus, 1987), three ratios from (Hüls, 1995), and four ratios from (Deutsche Bundesbank, 1999).

Data input errors are handled by winsorizing financial ratios at the 0.5%- and the 99.5%-quantile. Missing values are conservatively set to the 0.5% (99.5%)-quantile if low (high) values of the financial variable indicate high default risk.

---

[7] (Basel Committee, 2003), § 425

[8] (Lewis, 1994), p. 35

2. Defining internal credit rating systems of different quality

In our empirical approach, we define six kinds of qualitatively different internal credit rating systems (ordered in presumed ascending quality): the trivial system, the optimized Altman system, the Z-score system, the stepwise system, the benchmark variables system, and the pooled system. An overview over the main characteristics of the different system types is given in Table 2.

For all systems we assume that banks use just one credit scoring function for all borrowers. This may be a strong assumption as annual accounts of, for example, manufacturing and trading companies differ structurally. For this reason, (Deutsche Bundesbank, 1999) derives three different scoring functions for manufacturing, trading, and other sectors. On the other hand, in the well-known ZETA model, (Altman et al., 1977) argue that there are financial variables that are good default predictors and that behave similarly for manufacturers and trading companies such that both can be analyzed on an equal basis. The Z''-score of (Altman, 1993) is based on the same reasoning. As our goal is not to derive a perfect credit scoring system for the Deutsche Bundesbank dataset, we believe that our assumption will not affect results in an adverse way.

We also assume that banks determine credit scores and default probabilities exclusively based on annual accounts information. They do not add any qualitative information. The inclusion of qualitative factors is not a prerequisite for admittance to the internal-ratings based approach of Basel II.[9] Yet, many banks base their internal credit ratings on some qualitative components like management quality.[10] Since we do not have any additional qualitative information in our dataset, we might underestimate system performance, particularly for small banks. Given a credit quality threshold derived on data without taking into account qualitative information,

---

[9] (Basel Committee, 2003), § 379

[10] (Basel Committee, 2000) surveyed large international banks. Most banks assign ratings using considerable judgmental elements. The relative importance of qualitative versus quantitative factors ranged from very minor to more than 60%. (Günther and Grüning, 2000) report that 72 of 146 surveyed German banks use qualitative criteria for default prediction. 38 of 49 banks state that the quality of default prediction has been improved by the inclusion of qualitative factors.

small banks would violate the threshold more often and would consequently be put under additional investigation more frequently. Considering that it is more difficult to evaluate the quality of systems that largely rely on qualitative information, this consequence might even be desirable.

The six systems we define are designed to reflect different approaches to system calibration. We expect some of these approaches to result in inferior system quality, while others may not differ significantly from a benchmark model. The systems differ in their use of different information sources in the process of system calibration. Information can be

- internal to the bank,
- external to the bank, but internal to the economy (represented by the Deutsche Bundesbank database), or
- external to both bank and economy.

In the stepwise system, banks completely rely on their own data. In the trivial and the optimized Altman system financial variables are determined external to bank and economy, while weights are derived on the bank's data. The benchmark variables system uses information that is external to the bank, but internal to the economy to select financial variables, and information internal to the bank to derive weights. The Z-score and the pooled system are calibrated without any reference to a bank's own data base, but relying on information internal to the economy.

2.1 The trivial system

The trivial system represents the bottom end of our quality scale. Banks simply draw one financial variable by chance and use logistic regression to derive credit scores and default probabilities from this single variable. There is no doubt that this system's quality is inadequate such that it should always be detected as being of inferior quality.

2.2 The optimized Altman and the Z-score system

The Altman Z-score as well as its claim to be applicable in a broad range of applications is widely known. Yet, as accounting variables are differently defined in Germany than in the U.S., it does not seem reasonable to apply the specific credit scoring function without modifications. The optimized Altman and the Z-score system represent two ways banks might use to apply the Z-score to German data.

The optimized Altman system is defined by taking the financial ratios of Altman's Z''-score, and deriving the financial ratios' optimal weights by a logistic regression based on a bank's own dataset.

The Altman Z''-score is a modified version of the (Altman, 1968) Z-score, and is designed to account better for private companies and industry effects.[11] It consists of four financial variables: working capital / total assets (V22), retained earnings / total assets (V49), earnings before interest and taxes / total assets (V46), and book value of equity / book value of total liabilities (V25).

As will be seen in the description of the pooled system, none of these variables works particularly well with the Deutsche Bundesbank database. Therefore, we expect the quality of this system to be rather low, although better than that of the trivial system.

The Z-score system can be seen as the research output of a credit risk researcher who applies the design of the Altman Z-score to German data. The result is a specific logistic credit scoring function, which banks may opt to use. The system is based on information on the 39 largest defaulters from the Deutsche Bundesbank database and 39 randomly drawn non-defaulters, both satisfying the size criterion of revenues larger than fifty million Euros. Financial variables are chosen by a logistic stepwise selection procedure (with a significance level of 5%). The credit scoring function contains six financial variables (V8, V10, V28, V30, V34, and V42) covering revenues, profitability, equity, debt, and short-term debt.

Banks using this system need to correct the average predicted default probability of 50% to correspond to the bank's actual in-sample default rate. Banks perform this prior correction by replacing the intercept $\hat{b}_o$ of the logistic regression function by the consistent corrected estimate

$$\hat{b}_0 - \ln\left[\left(\frac{1-t}{t}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right], \qquad (1)$$

[11] (Altman, 1993), p. 204f

where t is the fraction of defaults in the bank's sample, and $\bar{y} = 0.5$ is the fraction of defaults in the sample which was used to derive the credit scoring function.[12] This correction will usually not be completely successful (because datasets differ). Therefore, banks scale all predicted default probabilities linearly such that average predicted default probabilities coincide with average actual default rates.[13]

It is not clear a priori whether the Z-score system performs better than the optimized Altman system. The advantage that the Z-score system selects financial variables based on the Deutsche Bundesbank database may be offset by the large company bias and the small sample size.

2.3 The stepwise system

In the stepwise system banks rely completely on their own data. They select from a set of financial variables by a logistic stepwise selection procedure, and then apply logistic regression to derive credit scores and default probabilities.

In choosing a significant level for the stepwise selection procedure, we take into account the lack of independence between firm-year observations. We apply a conservative procedure proposed by (Shumway, 2001) and multiply the value of the partial *F* test statistic, which is necessary to obtain a confidence level of 90%, by the average number of firm-years per company in each training sample. We use these new values of the partial F test as thresholds to decide on the significance of a variable.

The average number of financial ratios chosen by the stepwise selection procedures varies from about one for banks with three out-of-time defaults to about six for banks with 102 out-of-time defaults. We expect the stepwise system to be particularly favourable for large banks, while small banks might not be able to achieve a high quality because of their small databases.

---

[12] (Manski and Lerman, 1977)

[13] This is similar to the approach in (Falkenstein et al., 2000), p. 15.

2.4 The benchmark variables and the pooled system

The benchmark variables system represents the best rating system available for banks in our study. The story for this system is that bank regulators publish a set of financial ratios which they believe work well in measuring credit risk without revealing the specific form of the credit scoring function. This is done, for example, by (Deutsche Bundesbank, 1999).

The benchmark variables system consists of six financial ratios (V1, V16, V34, V39, V42, V43). Two of these ratios are also part of the Z-score system (V34, V42). The financial ratios were chosen based on the pooled system to be described next. V8 was dropped because it is structurally similar to V34.

The pooled system is only available to bank regulators that have access to a database of annual accounts and defaults covering the whole economy (like in Germany or France). In order to assess internal credit rating systems, regulators need to have an intuition about a good system's performance. Is a system with an area under curve of 80% extraordinarily good or only average? Regulators may use their own database to develop this intuition by calibrating some sort of a benchmark model. To be clear, this benchmark model is not meant to be the perfect credit scoring model for the economy in question. In this case every bank would be well advised to use it. The only purpose of the benchmark model is to help regulators in their task of evaluating system quality.

In our study, the pooled system serves as the benchmark model. The pooled system is derived on the complete Deutsche Bundesbank training sample using the procedure defined for the stepwise system. It consists of seven variables (V1, V8, V16, V34, V39, V42, and V43). The danger of overfitting is low due to the large dataset and the strict entry criterion for the stepwise selection procedure.

3. Quantitative measures of system quality

For each kind of internal credit rating system, we simulate the distribution of quantitative measures summarizing the system's quality. We consider three different measures: the area-under-curve (AUC) of the receiver operating characteristic (ROC) curve, the Brier score, and the grouped Brier score.

3.1 Area-under-curve (AUC)

The AUC measures the quality of ranking borrowers from high to low default risk. If low credit scores are defined to indicate high default probabilities, then all borrowers that actually defaulted in a learning sample should be assigned a relatively low credit score, and those that did not default a relatively high credit score. The AUC is only concerned with ranking, and does not assess the accuracy of default probability estimates.

Under what circumstances is such a measure useful? The ranking of borrowers is sufficient for credit risk management if banks are not able to charge different credit risk premiums for different customers in the market.[14] In this case, banks maximize their risk-adjusted returns by not granting credit to customers with negative expected returns which is equivalent to defining a minimum credit score. Yet, this line of thought does not lead us to the AUC as a measure of system quality, but to the concept of minimized expected error costs.[15] The AUC measures the quality of the complete ranking and not only of one threshold. Only if the threshold is difficult to define in practice, the AUC may be a sensible measure.

If banks are able to charge customer-specific risk premiums in the market, then the quality of the ranking of borrowers can serve as an approximation to the quality of default probability estimates. At least, borrowers with a low credit score ought to have a higher default probability than those with a high credit score. Yet, in this case the Brier score discussed in the next section is more appropriate.

Formally, the AUC is derived from the ROC curve. The ROC curve is obtained by sorting credit scores from low to high, and plotting the empirical distribution function (EDF) of scores of non-defaulting companies on the x-axis, and the EDF of scores of defaulting companies on the y-axis. If low scores are defined to indicate a high default probability, then x-values represent the error rate that a solvent company is classified as insolvent (type-II error) and y-values represent one minus the error rate that an insolvent company is classified as solvent (type-I error). Thus, the ROC curve is a complete representation of type-I and type-II errors. The area under the ROC

[14] For example, this is the case in consumer credit as described in (Jacobson and Roszbach, 2003).

[15] As developed for linear discriminant analysis by (Joy and Tollefson, 1975).

curve (AUC) is a summarizing accuracy measure. It is equivalent to the two independent sample Mann-Whitney non-parametric test statistic $\hat{q}$, which estimates the probability that the score of a randomly chosen defaulted company from the sample of defaulted companies is (correctly) lower than the score of a randomly chosen solvent company from the sample of solvent companies:

$$\hat{q} = \frac{1}{N_D N_{ND}} \sum_{(D,ND)} u_{D,ND} \ , \tag{2}$$

where N_D, and N_{ND} is the number of defaulters and non-defaulters, respectively, and

$$u_{D,ND} = \begin{cases} 1 & if \quad s_D < s_{ND} \\ 1/2 & if \quad s_D = s_{ND} \\ 0 & if \quad s_D > s_{ND} \end{cases} \tag{3}$$

with s_D, s_{ND} being the score of a defaulter and a non-defaulter, respectively. The sum in (2) is taken over all pairs of defaulters and non-defaulters (D, ND) in the sample.[16]

The AUC ranges from 0% to 100%. A perfect AUC value of 100% is attained if exactly those borrowers defaulting in the future receive the lowest credit scores. A value of below 50% would mean that the system performs worse than a system which randomly allocates credit scores to borrowers.

(DeLong et al., 1988) provide for an asymptotically valid test of the hypothesis that the AUC values of two different systems calculated on the same dataset are equal.

3.2 Brier score

The Brier score B is not only concerned with the ranking of borrowers, but also with the accuracy of default probability estimates. It is defined as

$$B = \frac{\sum_{i=1}^{n} (\hat{p}_i - I_i)^2}{n} \ , \tag{4}$$

where $\hat{p}_i$ is a system's default probability estimate for borrower i, $i = 1,...,n$, and I_i is the indicator variable of default (1 if default, zero otherwise).

---

[16] (DeLong et al., 1988). The notation is taken from (Engelmann et al., 2003).

The Brier score relies on a quadratic loss function, which is often used in economics. Other scoring rules are available. An important property of the Brier score is that it is a strictly proper scoring rule, meaning that banks minimize their expected score by reporting their probability estimates honestly.[17]

The Brier score ranges from zero (defaulters are attached a default probability of 100% percent and non-defaulters one of 0%) to some maximum value (defaulters are attached a default probability of zero percent and non-defaulters one of 100%). A system with an AUC of 100% does not necessarily have a Brier score of zero, as default probabilities for defaulters will mostly be below 100%, and those for non-defaulters above zero. Vice versa, a system with a Brier score of zero will also have an AUC of 100% showing that the Brier score evaluates ranking accuracy plus the accuracy of default probability estimates.

(Bloch, 1990) provides for a test of the hypothesis that the Brier scores of two different internal credit rating systems calculated on the same dataset are equal.

As indicated in the previous section, the Brier score may be more appropriate in settings in which banks are able to charge customer-specific credit risk premiums. As credit markets develop into this direction, and as it is one goal of the Basel II reform that banks price their loans according to the borrower's risk, the Brier score seems to dominate the AUC.

One disadvantage of the Brier score is that it strongly depends on the overall default rate level in a given sample. Therefore, the Brier score of a low-risk bank cannot directly be compared with that of a high-risk bank. So-called skill scores are proposed in the literature to make scores comparable that result from systems with different event probabilities. A system's Brier score is compared with the score of an unsophisticated system, e.g. by measuring the percentage improvement over the unsophisticated system or simply the difference in scores. Unfortunately, these modified scoring rules are not strictly proper and / or change with linear transformations.[18]

[17] Cf. (Winkler, 1994).

[18] Cf. (Winkler, 1994)

(Winkler, 1994) proposes an asymmetric scoring rule to standardize scores in a way such that the property of being strictly proper is retained. We implemented the application of this scoring rule to the Brier score. The unsophisticated system is represented by a system that uses the average in-sample default rate as individual default probability estimate for each borrower. We find that the standardization works, but the power is lower than that of the unmodified Brier score.

The main problem of applying the Winkler-proposal to credit risk is the difficulty to motivate the asymmetrical treatment of default probability forecasts. With respect to defaulters, a system is increasingly punished the lower its default probability forecasts are; a treatment which is intuitive. Yet, if default probability forecasts drop below the forecast of the trivial system, the punishment becomes a lot more severe (compare Winkler, 1994, Figure 2). The opposite holds for non-defaulters. The punishment increases with the default probability forecast, but it increases much slower if the default probability forecast exceeds the forecast of the trivial system. Because this asymmetrical treatment lacks a motivation, we stick to the original Brier score. We take its dependence on overall default rates into account by simulating thresholds depending on bank size and portfolio default rate (see Table 5, Panel B).

3.3 Grouped Brier score

Under Basel II, banks will have to construct at least eight rating classes. Most banks will estimate an average default probability for each rating class (usually derived from historical rating class default rates), and use these default probabilities to calculate capital requirements. Bank regulators will therefore be interested in the precision of these rating class-specific default probability estimates.

The grouped Brier score gB measures exactly this precision. It is defined as

$$gB = \frac{\sum_{g=1}^{G} (\hat{p}_g - \overline{p}_g)^2}{G}, \tag{5}$$

where $\hat{p}_g$ is the mean default probability estimate for borrowers of rating class $g = 1,...,G$, and $\overline{p}_g$ is the actual default rate of borrowers in rating class g.

The grouped Brier score is positioned between the Brier score, which measures system quality if there are as many borrowers as rating classes, and a system with just one rating class, in which case we simply measure the difference between the average default rate and the average default probability estimate.

As such the grouped Brier score seems to represent the most adequate measure for our purposes. Yet, the following example shows that there is an important caveat.

Consider two rating systems with two rating classes each. In the first system defaulters are evenly distributed across the two rating classes, while in the second all defaulters are classified into one of the two classes. Both systems are assumed to be perfect at predicting rating class default probabilities. Thus, both systems will have a zero grouped Brier score. But the first system is obviously not much in line with the spirit of Basel II which calls for risk differentiation, while the second is.

Confirming our reasoning, we find in our simulations that the grouped Brier score does not succeed at all in identifying inferior systems. As all of our systems are equally good at the task of predicting average rating class default rates, we do not treat the grouped Brier score any further.

4. Simulation set-up and results

4.1 Simulation set-up

Our economy consists of banks of four different sizes and of three different levels of portfolio default rates. We use in-sample bank sizes of 1,875, 3,750, 7,500, and 15,000 observations in a bank's 1994-1998 training sample. As each balance sheet in the training sample is counted as a separate observation, the number of companies per sample is less than the number of observations. There are on average 3.34 balance sheets per company such that the number of training sample companies ranges from 561 to 4,490.

As levels of portfolio default rates, we use 0.85%, 1.7%, and 3.4%. The value of 1.7% is taken from (Carey, 1998) as being a representative default rate for commercial loan portfolios of large U.S. banks.[19] Unfortunately, we do not have data on representative default rates of German credit portfolios.

---

[19] Take the portfolio structure for commercial loan portfolios of large U.S. banks in (Carey, 1998), p. 1380 and multiply it with default probabilities given in Table III, Panel B, second column.

The 1999 out-of-time sample of each bank consists of all companies that are part of the training sample and that stay customers in 1999. In addition, we simulate new business by randomly drawing new customers such that the bank's portfolio size and portfolio default rate stays constant. In doing this, we construct the situation that the out-of-time sample reflects an average year.[20]

By defining portfolio size and portfolio default rate, we also lock in the number of defaults in a credit portfolio. The number of defaults is the single most important variable influencing system quality. A relatively high default rate does not result in a good system, if the portfolio is relatively small. As well, a large portfolio does not result in a good system, if the default rate is relatively low.

We define six different classes of number of defaults (in-sample: 16, 32, 64, 128, 255, out-of-time: 3, 6, 13, 26, 51, 102; cf. Table 3). Later on, we will see that it is possible to state most results depending on the number of defaults instead of bank size / portfolio default rate combinations.

For each company, banks have access to the complete annual accounts history as it is available in the Deutsche Bundesbank database. For each bank size / portfolio default rate combination, we randomly draw 1,000 credit portfolio compositions representing different banks.

This procedure is based on two assumptions: first, banks do not specialize in regions or industries; second, banks are not able to collect additional annual accounts data to improve their credit scoring performance.

The first assumption will not be true especially for small banks that are often regionally focused. As a consequence, we may underestimate the credit scoring performance of small banks relative to large banks that are more diversified. The second assumption also concerns small banks that only have few defaults in their training samples such that they obtain a credit scoring with low predictive power.

[20] Standard credit scoring models do not address the question of systematic risk factors influencing the level of default rates. We do not treat the problem of systematic risk in this paper, as under Basel II systematic risk is addressed by introducing an asset value model with a relatively high asset correlation.

While we use the first assumption to simplify our analysis, the second assumption can be motivated to some extent by high costs of gathering additional annual accounts and default information especially for private companies.

4.2 Scores + logistic default probabilities vs. ratings + historical default rates

Today, most banks use credit scoring systems to classify borrowers into a set of rating classes defined by a range of credit scores.[21] By aggregating borrowers into rating classes, additional information inherent in credit scores is lost.

In addition, banks often do not transform credit scores into individual default probabilities parametrically, but rather apply historical rating class default rates to all borrowers in a rating class.[22]

We examine the question, whether banks are able to improve system quality if they use credit scores instead of ratings to rank borrowers, and if they use logistic regression estimates instead of average historical rating class default rates to determine individual default probabilities.

For this purpose, we conduct $?^2$-tests of the null hypothesis that the difference between the mean values of our quality measures is equal comparing the two cases described above:

$$\frac{(m_1 - m_2)^2}{s_1^2 + s_2^2 - 2s_{12}} \sim c^2(1) \qquad (6)$$

Table 4 summarizes the p-values of these tests depending on system type, bank size, and portfolio default rate.

In Panel A, the AUC derived from credit scores is compared with the AUC derived from credit ratings. Only for some medium-sized and large banks with medium or

---

[21] For simplicity, we define a set of eight rating classes of equal size. This is the Basel II minimum number, (Basel Committee, 2003), § 366.

[22] Historical rating class default rates can either be obtained from a bank's internal default history or from external rating agency data. The latter source is attractive because of long time series, yet it is often not applicable to non-U.S. banks because of a strong U.S. bias in the data.

high default rates, the p-value is below or equal to 10%. In these few cases the use of credit scores would significantly increase system quality.

In Panel B, the Brier score based on logistic default probabilities is compared with the one based on average historical rating class default rates. Again, there are some cases in which the null hypothesis is rejected, but in these cases the difference is positive favouring the Brier score based on historical default rates.

Based on this evidence, banks do not have an incentive to change the methods currently used. Results in the following sections will consequently be based on current behaviour.

4.3 Identification of inferior internal credit rating systems

We propose two different procedures to identify inferior internal credit rating systems. In the first procedure, systems are classified as inferior if their AUC or Brier score is worse than a given threshold. Banks simply submit their statistics to the bank regulator, who evaluates the system. The regulator does not have to publish neither the threshold nor any information about the way thresholds are derived.

The second procedure is more complex. The AUC or Brier score needs to be calculated for the bank's own system and the regulator's benchmark system, both based on the bank's credit portfolio. Bank regulators set a lower threshold on the p-value of the test of equality of the two statistics such that all banks whose system performs worse than the benchmark system and whose p-value falls below the threshold are classified as inferior systems.

This procedure is more difficult to put into practice. Either the regulator has to distribute her benchmark system to banks, which might give the impression that regulators arrogate to have the best system, or banks have to submit a large amount of data to regulators. Additional costs are generated, if regulators need financial ratios not produced by a bank's system by default.

Therefore, the first procedure will be preferred if it is sufficiently powerful. The main task to be solved for both procedures is the derivation of threshold values, which is explained along with simulation results in the next two sections.

4.3.1 Using critical thresholds on AUC values and Brier scores

To identify inferior internal credit rating systems, supposedly inferior systems need to be compared with a predefined benchmark system. We use the pooled system as the benchmark system as it is calibrated on the largest information set.[23]

Critical thresholds are derived as quantiles of the quality measure's distribution for the pooled system. For the AUC, we use the 10%-quantile such that lower values indicate inferiority. For the Brier score, we use the 90%-quantile such that higher values indicate inferiority.

Simulations are carried out for all bank size / portfolio default rate combinations specified in Table 3. The analysis of simulation results reveals that the probability to identify an inferior system predominantly depends on the number of out-of-time defaults and not on the specific bank size / portfolio default rate combination.

For the AUC, it is even possible to express thresholds only depending on the number of defaults, while for the Brier score, thresholds are expressed depending on the bank size / portfolio default rate combination (cf. discussion in section 3.2).

Table 5 presents out-of-time simulation results. For the AUC, the trivial system is always identified as an inferior system with a probability of at least 50%. For the optimized Altman system, this holds only for large banks and for medium-sized banks if default rates are high. The Z-score-system is identified as inferior with a high probability if the bank is large and the default rate is high.

For the stepwise system, the identification rate decreases from 31% for small banks with low default rates to 4% for large banks with high default rates. This confirms our intuition that the stepwise system is not adequate for small banks, while for large banks it is superior to the pooled system. As out-of-time samples primarily consist of the same borrowers as training samples, large banks using the stepwise system benefit relative to using the pooled system.

[23] If regulators are not able to calibrate the pooled system because they lack data, they might use the benchmark variables or the stepwise system instead. The power to identify inferior systems will decrease using these systems.

For the Brier score, results are similar. Identification frequencies are lower for small banks and medium-sized banks with low default rates using the trivial system and higher for large banks and medium-sized banks with high default rates using the Z-score system.

4.3.2 Using critical thresholds on p-values

In Table 6, simulation results are shown for the procedure that inferior systems are identified based on a lower p-value threshold.

Each bank performs a test of the null hypothesis that the AUC or Brier score of its own system is equal to the respective values of the regulator's system. Regulators set the critical p-value such that at least half of the banks using the optimized Altman system are identified as using an inferior system. Statistically, this means that the threshold equals the median of the p-value distribution resulting from comparing the optimized Altman with the pooled system conditioned on the fact that the optimized Altman system performs worse than the pooled system.

Results in Table 6 for the trivial system look similar to results presented in Table 5. The trivial system is always identified as inferior with a high probability. For small banks and for medium-sized banks with low or medium default rates the power is even higher, if the test is based on p-values rather than on the AUC or Brier score itself. Yet, it must be taken into account that in these cases the benchmark variables system is also identified as inferior with a high probability. Therefore, the discriminatory power of the test is not very high for these bank types.

Is the procedure based on p-values better than the procedure based on critical AUC and Brier score thresholds?

To answer this question, we set the critical p-values such that the power to identify the optimized Altman system as inferior is equal to the power shown in Table 5.

Results given in Table 7 are mixed. Identification frequencies are lower for the trivial system and higher for the Z-score system. Overall, the performance of the p-value approach and the critical AUC / Brier score approach seem to be similar. Differences of some percentage points may be due to simulation noise.

4.4 Capital requirements depending on system quality

We now address the question how the quality of internal credit rating systems influences Basel II capital requirements. First, we will look at absolute capital requirements depending on system quality. Then, we will pick up our question whether it is preferable for banks to derive default probability estimates based on historical rating class default rates instead of using logistic default probability estimates. If capital requirements are lower using logistic regression estimates, then banks might have a reason to adjust their models.

In Table 8, capital requirements (C) are calculated based on the latest Basel II proposal:[24]

$$C = 0.45N\left((1-R)^{-0.5}N^{-1}(PD)+\left(\frac{R}{1-R}\right)^{0.5}N^{-1}(0.999)\right)\left(\frac{1}{1-1.5(0.08451-0.05898\log(PD))^2}\right), \tag{7}$$

$$R = 0.12\frac{1-\exp(-50PD)}{1-\exp(-50)}+0.24\left(1-\frac{1-\exp(-50PD)}{1-\exp(-50)}\right)-0.04\left(1-\frac{S-5}{45}\right)$$

where R is the asset correlation, PD is the default probability estimate, and S equals sales in million Euros. We have to take into account the firm-size adjustment for small- and medium-sized entities as the median firm size in the Deutsche Bundesbank database equals 17 million Euros. The loss given default is set to 45%, and the maturity to 2.5 years.

Panel A shows that in-sample capital requirements roughly increase with decreasing system quality.[25] They are lowest for the benchmark variables system, and highest for the trivial system. Inferior systems, i.e. systems that are not able to discriminate well between high- and low-risk borrowers, are actually punished by higher capital

---

[24] (Basel Committee, 2003), § 241-242

[25] We deliberately show in-sample results here, because all systems are perfectly calibrated in-sample. Average in-sample default probabilities equal average in-sample default rates. Yet, the same behaviour can also be seen out-of-time.

requirements. The reason for this behaviour is due to the concavity of the capital requirements function.

For example, consider a portfolio of borrowers, each of which having sales of 17 million Euros, and an average default probability of 1.7%. If the rating system is not at all able to differentiate between risks, then it effectively consists of only one rating class yielding a capital requirement of 8.1%. If it is able to differentiate between risks such that there is one rating class with an average default probability of 0.17% containing 90% of the borrowers, and one with a default probability of 15.47% containing 10% of the borrowers, then the capital requirement equals only 4.7%.

The trivial system almost performs as bad as if it effectively produces only one rating class, while the benchmark variables system is not quite as good as the second system in the example.

The observation that capital requirements based on historical default rates increase with bank size is primarily caused by the fact that there often are no defaults in the high-quality rating classes of small banks. As the curvature of the capital requirement function is strongest for very low default probabilities, capital requirements are comparatively low for small banks even if the minimum default probability of 0.03% is imposed.

In Panel B, we report median differences in capital requirements if default probabilities are estimated from logistic regression instead of historical rating class default rates.

Except for the Z-score system all differences are positive independent of bank size and portfolio default rate. Median differences range from 0.1% to 0.7%. For the Z-score system, differences are negative, ranging from -0.9% to -1.5%. The Z-score system probably behaves differently because it is the only system for which we carry out a large prior correction (from 50% to 0.85%-3.4%). These calculations take into account the minimum default probability of 0.03% for the calculation of capital requirements. Otherwise differences would even be higher.

The reason for these differences lies in the fact that the capital requirement is a concave function of the default probability, and that the logistic regression produces systematic estimation errors depending on the level of the default rate. For the Z-score system (all other systems) it underestimates (overestimates) default rates if they are relatively low while overestimating (underestimating) default rates if they are relatively high. Due to the concavity of the capital requirement function, small differences between predicted and historical rating class default rates lead to relatively large differences in capital requirements if historical rating class default rates are low, and large differences between predicted and historical rating class default rates lead to relatively small differences in capital requirements if historical rating class default rates are high. On average, we obtain the differences in capital requirements reported in Panel B of Table 8.

Thus, only banks using the Z-score system benefit from using logistic regression default probabilities instead of historical rating class default rates. Banks using any other system have to have more capital. As in most of our analyses the Z-score system is not identified as an inferior system, there actually is a chance for banks to reduce their capital requirements without a significant loss in system quality. Yet, there are other considerations to be taken into account. For example, a bank using the Z-score system might not be able to charge the relatively high credit spreads to risky borrowers. Losing these borrowers will lead to a situation in which the system is not calibrated any more, reducing the system's quality and increasing the probability of being detected as an inferior system.

Results in Panel B depend on the particular function which logistic regression uses to transform credit scores into default probabilities. Our simulations show that this bias also exists with probit regression and parametric linear discriminant analysis. It could be removed by using a non-parametric approach.[26]

5. Conclusion

Based on a large database of Deutsche Bundesbank, we examined quantitative measures summarizing the quality of internal credit rating systems. Our main result is that both the AUC and the Brier score are valuable statistics in identifying low-quality systems. Which statistic should bank regulators choose? While the AUC dominates the current discussion, we believe that the Brier score measures more closely those errors that are important for capital regulation. Capital requirements are based on default probability estimates and not on the ranking of borrowers. Therefore, capital

[26] For example, cf. (Hausman et al., 1998), p. 250ff

requirements are correct only if default probability estimates are correct. As it is current practice that default probabilities are estimated for rating classes and not on an individual base, we also considered the grouped Brier score as an evaluation measure. The problem with the grouped Brier score is that it is only concerned with precision. It does not at all evaluate a system's ability to discriminate between high- and low-risk borrowers.

Other results of our study are that banks do not significantly improve system quality if they do not aggregate credit scores into rating classes, or if they use logistic regression to estimate default probabilities instead of historical rating class default rates. If banks are not able to discriminate between high- and low-risk borrowers, they increase their average capital requirements due to the concavity of the capital requirements function. The use of parametric methods to derive default probabilities from credit scores might lead to an over- or underestimation of capital requirements relative to using historical rating class default rates.

An interesting question for future research is how the power of identifying low-quality internal credit rating systems will develop as more data becomes available over time. In this paper, we used a training sample covering five years which is the minimum amount of data allowed under Basel II after the transition period. As our data covers the years 1990-2000, we are actually able to investigate this issue by starting with a training sample from 1990 to 1995, and then increasing the sample by additional years. Increasing training samples might lead to an improved average system performance, while the pooling of validation samples might also lead to improved validation results.

References

Altman, Edward I. “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, Journal of Finance 23 (September 1968), 568-609.

Altman, Edward I. Corporate financial distress and bankruptcy, Second edition. New York: John Wiley & Sons, Inc. 1993.

Altman, Edward I., Haldeman, Robert G., Narayanan, P. “ZETA analysis: a new model to identify bankruptcy risk of corporations”, Journal of Banking & Finance 1 (1977), 29-54.

Basel Committee on Banking Supervision. “The new Basel capital accord”, Consultative document (April 2003), Basel.

Basel Committee on Banking Supervision. “Range of practice in banks’ internal ratings systems”, (January 2000), Basel.

Berger, Allen N., Miller, Nathan H., Petersen, Mitchell A., et al. “Does function follow organizational form? Evidence from the lending practices of large and small banks!”, Working paper (March 2003), Board of Governors of the Federal Reserve System and NBER.

Bloch, Daniel A. “Evaluating predictions of events with binary outcomes: an appraisal of the Brier score and some of its close relatives”, Technical Report No. 135 (May 1990), Stanford University, Division of Biostatistics.

Blochwitz, Stefan, Liebig, Thilo, Nyberg, Mikael. “Benchmarking Deutsche Bundesbank’s default risk model, the KMV Private Firm Model and common financial ratios for German corporations”, Working paper Deutsche Bundesbank, KMV (November 2000).

Carey, Mark. “Some evidence on the consistency of banks’ internal credit ratings”, Working paper (April 2001), Federal Reserve Board.

Carey, Mark. “Credit risk in private debt portfolios”, Journal of Finance 53, 4 (August 1998), 1363-1387.

Carey, Mark, Hrycay, Mark. “Parameterizing credit risk models with rating data”, Journal of Banking & Finance 25 (2001), 197-270.

DeLong, Elizabeth R., DeLong, David M., Clarke-Pearson, Daniel L. “Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach”, Biometrics 44 (September 1988), 837-845.

Deutsche Bundesbank. “The methodological basis of the Deutsche Bundesbank’s corporate balance sheet statistics”, Deutsche Bundesbank Monthly Report (October 1998).

Deutsche Bundesbank. “The Bundesbank’s method of assessing the creditworthiness of business enterprises”, Deutsche Bundesbank Monthly Report (January 1999).

Diebold, Francis X., Rudebusch, Glenn D. “Scoring the leading indicators”, Journal of Business 62 (July 1989), 369-391.

Engelmann, Bernd, Hayden, Evelyn, Tasche, Dirk. “Testing rating accuracy”, Risk (January 2003), 82-86.

Falkenstein, Eric, Boral, Andrew, Kocagil, Ahmet E. “RiskCalc for private companies II: More results and the Australian model”, Moody’s Investors Service, Global Credit Research, (December 2000).

Grunert, Jens, Norden, Lars, Weber, Martin. “The role of non-financial factors in internal credit ratings”, Working paper, University of Mannheim, Center for Economic Policy Research, London.

Günther, Thomas, Grüning, Michael. „Einsatz von Insolvenzprognoseverfahren bei der Kreditwürdigkeitsprüfung im Firmenkundenbereich“, Die Betriebswirtschaft 1 (2000), 39-59 (Use of processes for default prediction for corporate borrowers, in German with English abstract).

Hand, D.J. “Assessing classification rules”, Journal of Applied Statistics 21, 3 (1994), 3-16.

Hand, D.J., Henley, W.E. “Statistical classification methods in consumer credit scoring: a review”, Journal of the Royal Statistical Society A, 160, Part 3 (1997), 523-541.

Hand, D.J. Construction and assessment of classification rules, Chichester: John Wiley & Sons. 1997.

Hausman, J.A., Abrevaya, Jason, Scott-Morton, F.M. “Misclassification of the dependent variable in a discrete-response setting”, Journal of Econometrics 87 (1998), 239-269.

Hüls, Dagmar. Früherkennung insolvenzgefährdeter Unternehmen. Düsseldorf: IDW-Verlag, 1995 (Early identification of companies with high default risk, in German).

Jacobson, Tor, Roszbach, Kasper “Bank lending policy, credit scoring and value-at-risk”, Journal of Banking & Finance 27 (2003), 615-633.

Joy, O.Maurice, Tollefson, John O. “On the financial applications of discriminant analysis”, Journal of Financial and Quantitative Analysis 10 (December 1975), 723-739.

Lewis, Edward M. An introduction to credit scoring. San Rafael, CA: The Athena Press.

Lopez, Jose A. “Evaluating the predictive accuracy of volatility models”, Journal of Forecasting 20 (June 1999), 87-109.

Lopez, Jose A. “Regulatory evaluation of value-at-risk models”, Journal of Risk 1 (June 1999), 37-64.

Manski, C.F., Lerman, S.R. “The estimation of choice probabilities from choice based samples”, Econometrica, 45 (1977), 1977-1988.

Niehaus, Hans-J. Früherkennung von Unternehmenskrisen. Düsseldorf: IDW-Verlag. 1987 (Early identification of company crises, in German).

RMA Capital Working Group. “EDF estimation: a “test-deck“ exercise”, The RMA Journal (November 2000), 54-61.

Rosenberg, Eric, Gleit, Alan. “Quantitative methods in credit management: a survey”, Operations Research 42 (August 1994), 589-613.

Shumway, Tyler. “Forecasting bankruptcy more accurately: a simple hazard model”, Journal of Business 74 (2001), 101-124.

Sobehart, Jorge R., Keenan, Sean C., Stein, Roger M. “Benchmarking quantitative default risk models: a validation methodology”, Moody’s Investors Service, Global Credit Research, (March 2000).

Swets, John A. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers, Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Tabakis, Evangelos, Vinci, Anna. “Analysing and combining multiple credit assessments of financial institutions”, Working paper No. 123 (February 2003), European Central Bank, Frankfurt.

Wilkie, A.D. “Measures for comparing scoring systems”; In: L.C. Thomas, J.N. Crook, D.B. Edelman, ed., Credit scoring and credit control. Oxford: Clarendon Press. 1992, pp. 123-138.

Winkler, Robert L. „Evaluating probabilities: asymmetric scoring rules”, Management Science 40 (November 1994), 1395-1405.

Table 1: Financial ratios used as independent variables in credit scoring

Financial variables are taken from Niehaus (1987), Hüls (1995), Deutsche Bundesbank (1999), and Altman (1968) (cf. footnote to table). The column 'Hypothesis' indicates whether the value of the financial variable is expected to be generally lower or higher, respectively, for insolvent (I) observations than for solvent (S) observations.

Variable	Ratio	Hypothesis
V1	operating profit (before taxes) / revenues	I < S
V2	EBITDA (excl. extraordinary items) / revenues	I < S
V3	earnings before financial expenses / total assets	I < S
V4	operating profit (before taxes and financial expenses) / total assets	I < S
V5	EBITDA (excl. extraordinary items) / total assets	I < S
V6	(EBITDA (excl. extraordinary items) + financial expenses) / total assets	I < S
V7	EBITDA (incl. extraordinary items) / total assets	I < S
V8	(revenues – expenses for raw materials and supplies – amortization of fixed assets – other operating expenses) / total assets	I < S
V9	EBITDA (incl. extraordinary items) / revenues	I < S
V10	EBITDA (excl. extraordinary items) / total debt	I < S
V11	EBITDA (incl. extraordinary items) / total debt	I < S
V12	EBITDA (excl. extraordinary items) / (total debt – cash)	I < S
V13	EBITDA (incl. extraordinary items) / (total debt – cash)	I < S
V14	EBITDA (excl. extraordinary items) / (total debt – cash – securities – trade receivables)	I < S
V15	EBITDA (incl. extraordinary items) / (total debt – cash – securities – trade receivables)	I < S
V16	EBITDA (excl. extraordinary items) / short-term debt	I < S
V17	EBITDA (incl. extraordinary items) / short-term debt	I < S
V18	(short-term debt * 360) / revenues	I > S
V19	(trade payables + liabilities from accepted bills) * 360 / revenues	I > S
V20	(cash + securities + trade receivables) / short-term debt	I < S
V21	working assets / short-term debt	I < S
V22	(working assets – short-term debt) / total assets	I < S
V23	(working assets – short-term debt) / revenues	I < S
V24	(cash + securities + trade receivables – short-term debt) / (operating expenses – amortization of fixed assets)	I < S
V25	adjusted equity capital / total assets	I < S
V26	(equity capital + total earnings) / total assets	I < S
V27	adjusted equity capital / total debt	I < S
V28	(equity capital + total earnings) / total debt	I < S
V29	short-term debt / total assets	I > S

Table 1: Financial ratios used as independent variables in credit scoring (continued)[27]

Variable	Ratio	Hypothesis
V30	short-term bank debt / total debt	I > S
V31	(adjusted equity capital + pension provisions + long-term debt) / long-term assets	I < S
V32	adjusted equity capital / (total assets – cash – properties)	I < S
V33	adjusted equity capital / (fixed assets – properties)	I < S
V34	revenues / total assets	I < S
V35	(debt from accepted bills + trade payables) * 12 / expenses for raw materials and supplies	I > S
V36	trade receivables * 12 / revenues	I > S
V37	finished goods * 12 / revenues	I > S
V38	raw materials and supplies * 12 / expenses for raw materials and supplies	I > S
V39	amortization / (fixed assets + reductions of fixed assets + amortization)	I < S
V40	investments / (fixed assets + reductions of fixed assets + amortization – investments)	I < S
V41	investments / amortization	I < S
V42	(adjusted equity capital + provisions/2) / total assets	I < S
V43	(trade payables + debt from accepted bills + bank debt) / (total debt – received advance payments)	I > S
V44	(trade receivables + inventories) / revenues	I > S
V45	(adjusted equity capital + pension provisions) / total assets	I < S
V46	earnings before taxes on income and interest paid / total assets	I < S
V47	earnings before taxes on income / adjusted equity capital	I < S
V48	net interest result / revenues	I < S
V49	retained earnings / total assets	I < S

[27] Variables V1-V41 are taken from (Niehaus, 1987, p. 75-76). The variable 21 of (Niehaus, 1987) is not sufficiently defined so that we do not use it. V42-44 are from (Hüls, 1995), p. 241, Table 22. (V42 = K_122, V43 = K_68A, V44 = K_85) The variable K_08EP cannot be calculated because we do not have data on the change in pension provisions, but V5 is very similar. The variable K_35 = V19, and the variable K_79 = V34. V45-48 are from (Deutsche Bundesbank, 1999), p. 55. (V45 = Equity/pension provision ratio, V46 = Return on total capital employed, V47 = Return on equity, V48 = Net interest rate) The capital recovery rate cannot be calculated because it is not sufficiently defined. The equity ratio equals V26. V49 is taken from (Altman, 1968).

Table 2: Overview of rating system types

Name	Description
Trivial	Bank randomly draws one financial variable from set of 49 variables, and derives optimal logistic credit scoring function based on its own data.
Optimized Altman	Bank takes financial variables of Altman's Z''-score calibrated on US data, and derives optimal logistic credit scoring function based on its own data.
Z-score	Bank applies logistic credit scoring function derived on a sample of the 39 largest defaulters in the Deutsche Bundesbank database (revenues > 50 million Euros) and 39 randomly drawn non-defaulters of the same size. No reference to bank's own data.
Stepwise	Bank selects financial variables by logistic stepwise selection procedure, and derives optimal logistic credit scoring function based on its own data.
Benchmark variables	Bank uses a set of six financial variables that work well for the complete dataset, and derives optimal logistic credit scoring function based on its own data.
Pooled	Logistic credit scoring function derived on the complete learning sample. Serves as benchmark function to evaluate all other systems.

Table 3: Overview of number of defaults resulting from bank size / portfolio default rate combinations

	Portfolio default rate		
Bank size (# observations)	0.85% (low)	1.70% (medium)	3.40% (high)
	In-sample		
1,875 (small)	16	32	64
3,750 (medium I)	32	64	128
7,500 (medium II)	64	128	255
15,000 (large)	128	255	512
	Out-of-time		
375 (small)	3	6	13
750 (medium I)	6	13	26
1,500 (medium II)	13	26	51
3,000 (large)	26	51	102

Table 4: P-Values of ?²-tests (Scores + logistic default probabilities vs. ratings + historical default rates) (in %)

The table shows p-values of ?2-tests of the null hypothesis that the difference between two means equals zero. The (+)-sign indicates that the difference is positive. In Panel A, the AUC derived from credit scores is compared with the one derived from credit ratings. In Panel B, the Brier score based on individual logistic default probabilities is compared with the one based on average in-sample rating class default rates. Results are shown for five different systems (Trivial, Optimized Altman, Z-score, Stepwise, Benchmark variables), three levels of portfolio default rates (Low= 0.85%, Med= 1.7%, High= 3.4%), and four bank sizes (Small: 375 out-of-time observations, Med I: 750, Med II: 1,500, Large: 3,000), and are based on 1,000 out-of-time simulations.

	Trivial			Optimized Altman			Z-score			Stepwise			Benchmark variables		
Bank size	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High
Panel A															
Small	84	87	83	79	77	64	78	68	53	86	71	52	62	51	30
Med I	83	79	76	69	63	66	65	48	40	73	47	26	50	28	16
Med II	80	72	68	68	62	56	46	27	20	45	20	9 (+)	32	16	3 (+)
Large	68	62	52	55	59	40	25	10 (+)	1 (+)	17	5 (+)	1 (+)	14	4 (+)	0 (+)
Panel B															
Small	92	98	98	96	93	88	68	63	74	82	89	91	91	98	76
Med I	98	91	81	99	86	74	49	51	60	85	90	90	100	81	63
Med II	92	73	61	82	72	55	24	34	43	89	91	77	92	72	42
Large	77	69	55	76	52	21	8 (+)	9 (+)	11	89	80	40	80	58	15

Table 5: Relative frequencies of identifying inferior systems using critical AUC and Brier score thresholds (in %)

The table shows the relative frequency that a quality measure of a given system performs worse than a threshold value. In Panel A, the quality measure is the AUC, and lower thresholds (given in %) are derived as 10%-quantiles of the AUC distribution for the pooled system depending on the number of in-sample defaults. In Panel B, the quality measure is the Brier score, and upper thresholds (given in %) are derived as 90%-quantiles of the Brier score distribution for the pooled system depending on portfolio default rate and portfolio size. Results are shown for five different systems (Trivial, Optimized Altman, Z-score, Stepwise, Benchmark variables), and are based on 1,000 out-of-time simulations for each of twelve portfolio size / portfolio default rate combinations (see Table 3).

# out-of-time defaults	Trivial	Optimized Altman	Z-score	Stepwise	Benchmark variables
Panel A	Thresholds: 68-74-76-79-80-82				
3	50	28	14	31	13
6	67	33	17	30	13
13	79	38	19	23	11
26	90	51	26	11	11
51	97	75	44	6	10
102	100	99	82	4	12
Panel B	Thresholds: Low: 0.848-0.842-0.837-0.833, Medium: 1.664-1.646-1.634-1.626, High: 3.18-3.14-3.11-3.09				
3	26	18	13	24	14
6	47	22	14	23	14
13	77	33	22	26	14
26	93	54	40	17	14
51	99	85	72	12	15
102	100	100	100	9	23

Table 6: Relative frequencies of identifying inferior systems using critical p-values I (in %)

Panel A and B show the relative frequency that a p-value is smaller than a threshold value. The p-value results from a test that a quality measure for the bank's own system is equal to the one for the pooled system based on a bank's own dataset. Threshold values are set depending on bank size such that 50% of those banks using the Optimized Altman-system are identified as using inferior systems. In Panel A, the quality measure is the AUC. In Panel B, it is the Brier score. Results are shown for five different systems (Trivial, Optimized Altman, Z-score, Stepwise, Benchmark variables), and are based on 1,000 out-of-time simulations for each of twelve portfolio size / portfolio default rate combinations (see Table 3).

# out-of-time defaults	Trivial	Optimized Altman	Z-score	Stepwise	Benchmark variables
Panel A	Thresholds: 43-35-32-24-11-2.5				
3	66	50	47	55	49
6	72	50	40	48	39
13	81	50	41	44	27
26	90	50	43	19	17
51	92	50	41	3	5
102	97	50	29	0	0
Panel B	Thresholds: 30-23-16-10-3-2.5				
3	59	50	43	49	36
6	66	50	43	47	29
13	75	50	43	40	24
26	86	50	47	20	15
51	90	50	46	3	6
102	95	50	34	0	0

Table 7: Relative frequencies of identifying inferior systems using critical p-values II (in %)

Panel A and B show the relative frequency that a p-value is smaller than a threshold value. The p-value results from a test that a quality measure for the bank's own system is equal to the one for the pooled system based on a bank's own dataset. Threshold values are set depending on bank size such that the power for the optimized Altman system equals the one reported in Table 5. In Panel A, the quality measure is the AUC. In Panel B, it is the Brier score. Results are shown for five different systems (Trivial, Optimized Altman, Z-score, Stepwise, Benchmark variables), and are based on 1,000 out-of-time simulations for each of twelve portfolio size / portfolio default rate combinations (see Table 3).

# out-of-time defaults	Trivial	Optimized Altman	Z-score	Stepwise	Benchmark variables
Panel A	Thresholds: 17-17-19-24-30-32				
3	46	28	24	31	19
6	59	33	26	33	21
13	74	38	30	33	18
26	90	51	44	20	17
51	96	75	67	8	14
102	100	99	96	3	10
Panel B	Thresholds: 9-8-8-11-11-6				
3	27	18	13	19	12
6	38	22	15	20	10
13	61	33	27	25	12
26	88	54	50	23	17
51	97	78	76	12	19
102	100	99	98	2	14

Table 8: Capital requirements depending on system quality (in %)

Panel A shows capital requirements depending on system type, bank size, and portfolio default rate. Panel B shows the median difference between capital requirements using individual logistic regression default probability estimates and historical rating class default rates. Results are based on in-sample data, which means that average portfolio default rates are equal to average predicted default probabilities for all models shown. Results are shown for five different systems (Trivial, Optimized Altman, Z-score, Stepwise, Benchmark variables), three levels of portfolio default rates (Low= 0.85%, Med= 1.7%, High= 3.4%), and four bank sizes (Small: 375 out-of-time observations, Med I: 750, Med II: 1,500, Large: 3,000), and are based on 1,000 in-sample simulations.

	Trivial			Optimized Altman			Z-score			Stepwise			Benchmark variables		
	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High
Panel A															
Small	5.2	7.3	10.0	4.7	6.6	9.4	4.4	6.1	8.6	4.5	6.3	8.7	4.1	5.8	8.4
Med I	5.5	7.6	10.1	5.0	6.9	9.6	4.6	6.2	8.8	4.8	6.3	8.6	4.3	6.0	8.5
Med II	5.7	7.6	10.2	5.1	7.0	9.7	4.7	6.3	8.8	4.7	6.1	8.6	4.4	6.1	8.6
Large	5.7	7.6	10.2	5.2	7.1	9.7	4.7	6.3	8.9	4.5	6.1	8.6	4.5	6.1	8.6
Panel B															
Small	0.7	0.6	0.2	0.6	0.7	0.4	-0.9	-1.1	-1.3	0.7	0.7	0.6	0.5	0.6	0.5
Med I	0.4	0.3	0.1	0.4	0.4	0.2	-1.1	-1.3	-1.4	0.5	0.5	0.5	0.4	0.5	0.4
Med II	0.3	0.2	0.1	0.3	0.3	0.1	-1.2	-1.3	-1.5	0.4	0.5	0.4	0.3	0.4	0.3
Large	0.2	0.2	0.1	0.3	0.3	0.1	-1.2	-1.4	-1.5	0.4	0.4	0.3	0.3	0.4	0.3

Working Paper Series: Finance & Accounting

No.114: **Raimond Maurer/ Frank Reiner/ Ralph Rogalla**, Risk and Return of Open-End Real-Estate Funds: The German Case, September 2003

No.113: **Patrick Behr/ André Güttler/ Thomas Kiehlborn**, Der deutsche Hypothekenbankenmarkt: Ergebnisse einer empirischen Untersuchung, September 2003

No.112: **Reinhard H. Schmidt/ Andreas Hackethal/ Valentin Marinov**, Die Bankenmärkte Russlands und Bulgariens, July 2003

No.111: **Reinhard H. Schmidt/ Marcel Tyrell,** What constitutes a financial system in general and the German financial system in particular?, July 2003

No.110: **Falko Fecht**, On the Stability of Different Financial Systems, June 2003

No.109: **Raimond Maurer/ Shohreh Valiani**, Hedging the Exchange Rate Risk in International Portfolio Diversification: Currency Forwards versus Currency Options, June 2003

No.108: **Raimond Maurer/ Frank Reiner/ Steffen Sebastian**, Financial Characteristics of International Real Estate Returns: Evidence from the UK, US, and Germany, May 2003

No.107: **Anne d'Arcy/ Michiyo Mori/ Christine Rossbach**, The impact of valuation rules for intangible assets in Japanese and Germanaccounts of listed companies, April 2003

No.106: **Andreas Hackethal**, German banks – a declining industry?, March 2003

No.105: **Ingo E. Tschach**, The long term impact of microfinance on income, wages and the sectoral distribution of economic activity, April 2003

No.104: **Reinhard H. Schmidt/ Marco Weiß**, Shareholder vs. Stakeholder: Ökonomische Fragestellungen, January 2003

No.103: **Ingo E. Tschach**, The Theoretical Derivation of Credit Market Segmentation as the Result of a Free Market Process, March 2003

No.102: **Samuel Lee/ Nina Moisa/ Marco Weiss**, Open Source as a Signalling Device – An Economic Analysis, March 2003

No.101: **Christian Gaber**, Bewertung von Fertigerzeugnissen zu Voll- oder Teilkosten? Ansatz von Forderungen zum Nennwert oder Barwert? Eine agencytheoretische Analyse zur zielkongruenten Performancemessung, December 2002

No.100: **Oliver Ruß/ Günther Gebhardt,** Erklärungsfaktoren für den Einsatz von Währungsderivaten bei deutschen Unternehmen – eine empirische Logit-Analyse, August 2002

No.99: **Christian Gaber**, Gewinnglättung und Steuerung dezentraler Investitionsentscheidungen bei sich gegenseitig ausschließenden Investitionsprojekten, September 2002

No.98: **Volker Laux**, On the Value of Influence Activities for Capital Budgeting, September 2002

No.97: **Gunter Löffler**, Avoiding the rating bounce: Why rating agencies are slow to react to new information, June 2002

No.96: **Andreas A. Jobst**, Collateralized Loan Obligations (CLOs) – A Primer, December 2002

No.95: **Günther Gebhardt/ Rolf Reichardt/ Carsten Wittenbrink,** Accounting for Financial Instruments in the Banking Industry, November 2002

No.94: **Ulf Herold/ Raimond Maurer,** Portfolio choice and estimation risk – A comparison of Bayesian approaches to resampled efficiency, June 2002

No.93: **Olivia S. Mitchell/ David McCarthy,** Annuities for an Ageing World, June 2002

No.92: **Ulf Herold/ Raimond Maurer,** How much foreign stocks? Classical versus Bayesian approaches to asset allocation, June 2002

No.91: **Gunter Löffler/ Patrick F. Panther/ Erik Theissen,** Who Knows What When? – The Information Content of Pre-IPO Market Prices, June 2002

No.90: **Reinhard Hujer/ Sandra Vuletic/ Stefan Kokot,** The Markov switching ACD model, April 2002

No.89: **Markus C. Arnold/ Robert M. Gillenkirch,** Stock Options as Incentive Contracts and Dividend Policy, April 2002

No.88: **Anne d'Arcy/ Sonja Grabensberger,** The Quality of Neuer Markt Quarterly Reports - an Empirical Investigation, January 2002

No.87A: **Reinhard H. Schmidt /Ingo Tschach,** Microfinance as a Nexus of Incentives, May 2001

No.87: **Reinhard H. Schmidt/ Ingo Tschach,** Microfinance als ein Geflecht von Anreizproblemen, Dezember 2001 (erscheint in den Schriften des Vereins für Sozialpolitik, 2002)

No.86: **Ralf Elsas/ Yvonne Löffler,** Equity Carve-Outs and Corporate Control in Germany, December 2001

No.85: **Günther Gebhardt/ Stefan Heiden/ Holger Daske,** Determinants of Capital Market Reactions to Seasoned Equity Offers by German Corporations, December 2001

No.84: **Hergen Frerichs/ Gunter Löffler,** Evaluating credit risk models: A critique and a proposal, October 2001 (erschienen in: Journal of Risk, 5, 4, Summer 2003, 1-23)

No. 83: **Ivica Dus/ Raimond Maurer,** Integrated Asset Liability Modelling for Property Casuality Insurance: A Portfolio Theoretical Approach, October 2001 (erscheint in Handbuch Asset-Liability Management, hrsg. von M. Rudolph u.a.)

No.82: **Raimond Maurer/ Frank Reiner,** International Asset Allocation with Real Estate Securities in a Shortfall-Risk Framework: The Viewpoint of German and US Investors, September 2001

No.81: **Helmut Laux,** Das Unterinvestitionsproblem beim EVA-Bonussystem, August 2001

No.80: **Helmut Laux,** Bedingungen der Anreizkompatibilität, Fundierung von Unternehmenszielen und Anreize für deren Umsetzung, July 2001

No. 79: **Franklin Allen/ Douglas Gale**, Banking and Markets, July 2001

No.78: **Joachim Grammig/ Michael Melvin/ Christian Schlag,** Price Discovery in International Equity Trading, July 2001 (erscheint in: Journal of Empirical Finance unter dem Titel “Internationally Cross-Listed Stock Prices During Overlapping Trading Hours: Price Discovery and Exchange Rate Effects”)

No.77: **Joachim Grammig/ Reinhard Hujer/ Stefan Kokot,** Tackling Boundary Effects in Nonparametric Estimation of Intra-Day Liquidity Measures, July 2001

No.76: **Angelika Esser/ Christian Schlag ,** A Note on Forward and Backward Partial Differential Equations for Derivative Contracts with Forwards as Underlyings, June 2001 (erschienen in “Foreign Exchange Risk”, Hakala, J.; Wystup, U. (eds), 2002, 115-124)

No.75: **Reinhard H. Schmidt/ Marcel Tyrell/ Andreas Hackethal,** The Convergence of Financial Systems in Europe, May 2001 (erschienen in: German Financial Markets and Institutions: Selected Studies, Special Issue 1-02 of Schmalenbach Business Review (2002), S. 7-53)

No.74: **Ulf Herold**, Structural positions and risk budgeting - Quantifying the impact of structural postions and deriving implications for active portfolio management, May 2001

No.73: **Jens Wüstemann,** Mängel bei der Abschlußprüfung: Tatsachenberichte und Analyse aus betriebswirtschaftlicher Sicht, April 2001 (erschienen in: „Der Wirtschaftsprüfer als Element der Corporate Governance“, Zentrum für Europäisches Wirtschaftsrecht, Bonn 2001, S. 25-60)

No.72: **Reinhard H. Schmidt**, The Future of Banking in Europe, March 2001 (erschienen in: Financial Markets and Portfolio Management, Vol. 15 (2001), S. 429-449)

No.71: **Michael H. Grote/ Britta Klagge**, Wie global sind Japans Banken? Die Veränderung institutioneller Bedingungen und ihre Auswirkungen auf die internationale Präsenz japanischer Kreditinstitute, April 2001

No.70: **Stefan Feinendegen/ Eric Nowak**, Publizitätspflichten börsennotierter Aktiengesellschaften im Spannungsfeld zwischen Regelberichterstattung und Ad-hoc-Publizität - Überlegungen zu einer gesetzeskonformen und kapitalmarktorientierten Umsetzung, März 2001 (erscheint in: Die Betriebswirtschaft)

No.69: **Martin F. Grace/ Robert W. Klein/ Paul R. Kleindorfer**, The Demand for Homeowners Insurance with Bundled Catastrophe Coverages, March 2001

No.68: **Raimond Maurer/ Martin Pitzer/ Steffen Sebastian**, Konstruktion transaktionsbasierter Immobilienindizes: Theoretische Grundlagen und empirische Umsetzung für den Wohnungsmarkt in Paris, Februar 2001

No.67: **Gyöngyi Bugár/ Raimond Maurer**, International Equity Portfolios and Currency Hedging: The Viewpoint of German and Hungarian Investors, February 2001 (erscheint in. ASTIN-Bulletin)

No.66: **Rainer Brosch**, Portfolio-aspects in real options management, February 2001

No.65a: **Marcel Tyrell/ Reinhard H. Schmidt**, Pension Systems and Financial Systems in Europe:A Comparison from the Point of View of Complementarity, July 2001 (erschienen in ifo-Studien, Vol. 47, 2001, S. 469-503)

No.65: **Marcel Tyrell/ Reinhard H. Schmidt**, Pensions- und Finanzsysteme in Europa: Ein Vergleich unter dem Gesichtspunkt der Komplementarität, Februar 2001 (erschienen in gekürzter Fassung in: „Private Versicherung und Soziale Sicherung", Festschrift zum 60. Geburtstag von Prof. Dr. Dr. h.c. Roland Eisen, hrsg. von H.-C. Mager, H. Schäfer, K. Schrüfer, Metropolis: Marburg),

No.64: **Jutta Dönges/ Frank Heinemann**, Competition for Order Flow as a Coordination Game, January 2001

No.63: **Eric Nowak/ Alexandra Gropp**, Ist der Ablauf der Lock-up-Frist bei Neuemissionen ein kursrelevantes Ereignis, Dezember 2000 (erschienen in Zeitschrift für betriebswirtschaftliche Forschung, Februar 2002)

No.62: **Ulrich Kaiser/ Andrea Szczesny**, Einfache ökonometrische Verfahren für die Kreditrisikomessung: Verweildauermodelle, Dezember 2000

No.61: **Ulrich Kaiser/ ndrea Szczesny**, Einfache ökonometrische Verfahren für die Kreditrisikomessung: Logit- und Probit-Modelle, Dezember 2000

No.60: **Andreas Hackethal**, How Unique Are US Banks? - The Role of Banks in Five Major Financial Systems, , December 2000 (erschienen in: Zeitschrift für Nationalökonomie und Statistik, Vol. 221, S. 592-619)

No.59: **Rolf Elgeti/ Raimond Maurer**, Zur Quantifizierung der Risikoprämien deutscher Versicherungsaktien im Kontext eines Multifaktorenmodells, Oktober 2000 (erschienen in: Zeitschrift für die gesamte Versicherungswissenschaft 4/2000, S. 577- 603.)

No.58: **Harald A. Benink/ Reinhard H. Schmidt**, Towards a Regulatory Agenda for Banking in Europe, September 2000 (erschienen in: Research in Financial Services-Bank Crises: Causes, Analysis and Prevention, Vol.12, JAI Press-Elsevier Science, hrsg. von George G. Kaufman, 2000)

No.57: **Thomas G. Stephan/ Raimond Maurer/ Martin Dürr**, A Multiple Factor Model for European Stocks, September 2000

No.56: **Martin Nell/ Andreas Richter**, Catastrophe Index-Linked Securities and Reinsurance as Substituties, August 2000

No.55: Four short papers on Development Finance, August 2000

Reinhard H. Schmidt, Entwicklungsfinanzierung; (erschienen in: Handwörterbuch des Bank- und Finanzwesens, 3. Aufl., hrsg. von Wolfgang Gerke und Manfred Steiner, Stuttgart: Schäffer-Poeschel, 2001)

Reinhard H. Schmidt, Banking Regulation contra Microfinance; (erschienen in: Savings and Development, Vol. 24 (2000) , S.111-121.)

Ingo Tschach, The Impact of Inflation on Long-Term Housing Loans;

Eva Terberger-Stoy/ Marcel Tyrell/ Joseph E. Stiglitz (erschienen in: Entwicklung und Zusammenarbeit, 41. Jahrgang (2000), S. 46-49)

No.54: **Raimond Maurer/ Thomas G. Stephan**, Vermögensanlagevorschriften für deutsche Versicherungsunternehmen: Status Quo und finanzwirtschaftliche Bewertungen, Juli 2000 (erschienen in: Handbuch Spezialfonds (hrsg. von J.M. Kleeberg und C. Schlenger), Bad Soden 2000, S. 143-176.)

No.53: **Joachim Grammig/ Reinhard Hujer/Stefan Kokot**, Bias-free Nonparametric Estimation of Intra-Day Trade Activity Measures, June 2000

No.52: **Raimond Maurer/ Steffen Sebastian/ Thomas G. Stephan**, Immobilienindizes im Portfolio-Management, Mai 2000 (erscheint in Deutscher Aktuarverein (Hrsg.): Investmentmodelle für das Asset-Liability-Modelling von Versicherungs-unternehmen, 2002)

No.51: **Raimond Maurer/ Steffen Sebastian,** Inflation Risk Analysis of European Real Estate Securities, Mai 2000 (erscheint in: Journal of Real Estate Research, 2002)

No.50: **Andreas Hackethal/ Reinhard H. Schmidt,** Finanzsysteme und Komplementarität, April 2000 (erschienen in: Kredit und Kapital, Beiheft 15 "Neue finanzielle Arrangements: Märkte im Umbruch", 2000, S. 53-102)

No.49: **Mark Wahrenburg/ Susanne Niethen,** Vergleichende Analyse alternativer Kreditrisikomodelle, April 2000 (erschienen in: Kredit und Kapital, Heft 2, 2000)

No.48: **Christian Leuz,** IAS versus US GAAP: A "New Market" Based Comparsion, January 2000 (erschienen in: Journal of Accounting Research, 41 (2003), S. 445-472)

No.47: **Ralf Elsas/ Mahmoud El-Shaer/ Erik Theissen,** Beta and Returns Revisited – Evidence from the German Stock Market, December 1999 (erschienen in: Journal of International Financial Markets, Institutions and Money 13 (2003), 1-18)

No.46: **Michael H. Grote/ Sofia Harrschar-Ehrnborg/ Vivien Lo,** Technologies and Proximities: Frankfurt´s New Role in the European Financial Centre System, December 1999

No.45: **Reinhard H. Schmidt/ Adalbert Winkler,** Building Financial Institutions in Developing Countries, November 1999 (erschienen in: "Journal für Entwicklungspolitik", XVI/3, 2000, S. 329-346)

No.44: **Konstantin Korolev/ Kai D. Leifert/ Heinrich Rommelfanger,** Arbitragetheorie bei vagen Erwartungen der Marktteilnehmer, November 1999

No.43: **Reinhard H. Schmidt/ Stefanie Grohs,** Angleichung der Unternehmensverfassung in Europa –Ein Forschungsprogramm, November 1999 (erschienen in: Systembildung und Systemlücken in Kerngebieten des Europäischen Privatrechts, hrsg. von Stefan Grundmann, Tübingen: Mohr Siebeck, 2000, S. 146-188)

No.42: **Konstantin Kovolev/ Kai D. Leifert/ Heinrich Rommelfanger,** Optionspreistheorie bei vagen Daten, Oktober 1999

No.41: **Christian Leuz/ Robert E. Verrecchia,** The Economic Consequences of Increased Disclosure, June 2000 (erschienen in: Journal of Accounting Research 38 (Supplement 2000), 91-124)

No.40: **Christian Leuz,** The Development of Voluntary Cash Flow Statements in Germany and the Influence of International Reporting Standards, July 1999 (erschienen in: Schmalenbach Business Review, Vol. 52 (2) (April 2000), S. 182-207)

No.39: **Ulrike Stefani,** Quasirenten, Prüferwechsel und rationale Adressaten, Juni 1999

No.38: **Michael Belledin/ Christian Schlag,** An Empirical Comparison of Alternative Stochastic Volatility Models, June 1999

No.37: **Jens Wüstemann,** Internationale Rechnungslegungsnormen und neue Institutionenökonomik, Mai 1999

No.36: **Robert Gillenkirch/ Matthias M. Schabel,** Die Bedeutung der Periodenerfolgsrechnung für die Investitionssteuerung – Der Fall ungleicher Zeitpräferenzen, April 1999 (die überarbeitete Fassung "Investitionssteuerung, Motivation und Periodenerfolgsrechnung bei ungleichen Zeitpräferenzen" erscheint voraussichtlich 2001 in der ZfbF)

No.35: **Reinhard H. Schmidt**, Differences between Financial Systems in Europe: Consequences for EMU, April 1999 (erschienen in "The Monetary Transmission Mechanism: Recent Developments and Lessous for Europe", hrsg. v. Deutsche Bundesbank, Houndsmill (UK), 2001, S. 208-240)

No.34: **Theodor Baums/ Erik Theissen,** Banken, bankeigene Kapitalanlagegesellschaften und Aktienemissionen, März 1999 (erschienen in: Zeitschrift für Bankrecht und Bankwirtschaft, 11 (1999), Heft 3, S. 125-134)

No.33: **Andreas Hackethal/ Reinhard H. Schmidt,** Financing Patterns: Measurement Concepts and Empirical Results, May 2000

No.32: **Michael H. Haid/ Eric Nowak,** Executive compensation and the susceptibility of firms to hostile takeovers – An empirical investigation of the U.S. oil industry, March 1999

No.31: **Reinhard H. Schmidt/ Jens Maßmann,** Drei Mißverständnisse zum Thema "Shareholder Value", Februar 1999 (erschienen in Kumar/ Osterloh/ Schreyögg (Hrsg.):Unternehmensethik und die Transformation des Wettbewerbs, Festschrift für Professor Dr. Dr. h. c. Horst Steinmann zum 65. Geburtstag,1999, Schäffer-Poeschel Verlag Stuttgart, S.125-157)

No.30: **Eberhard Feess/ Michael Schieble,** Credit Scoring and Incentives for Loan Officers in a Principal Agent Model, January 1999

No.29: **Joachim Grammig/ Dirk Schiereck/ Erik Theissen,** Informationsbasierter Aktien-handel über IBIS, Januar 1999 (erschienen in: Zeitschrift für betriebswirtschaftlicher Forschung 52 (2000), 619-642)

No.28: **Ralf Ewert/ Eberhard Feess/ Martin Nell,** Auditor Liability Rules under Imperfect Information and Costly Litigation – The Welfare Increasing Effect of Liability Insurance, January 1999 (erschienen in: European Accounting Review)

No.27: **Reinhard H. Schmidt/ Gerald Spindler,** Path Dependence, Corporate Governance and Complementarity, March 2000 (erschienen in: International Finance, Vol. 5 (2002), No. 4, S. 311-333; erscheint in: Jeffrey Gordon & Mark Roe, eds.: Convergence and Rersistence of Corporate Governance Systems, University of Chicago Press, 2001)

No.26: **Thorsten Freihube/ Carl-Heinrich Kehr/ Jan P. Krahnen/ Erik Theissen,** Was leisten Kursmakler? Eine empirische Untersuchung am Beispiel der Frankfurter Wertpapierbörse, Dezember 1998 (erschienen in: Kredit und Kapital 32(1999), Heft3, S. 426-460)

No. 25: **Jens Maßmann/ Reinhard H. Schmidt,** Recht, internationale Unternehmensstrategien und Standortwettbewerb, December 1998 (erschienen in: Jahrbuch für Neue Politische Ökonomie, Band 18, hrsg. von K.-E. Schenk u.a., Tübingen 2000, S. 169-204)

No. 24: **Eberhard Feess/ Martin Nell,** The Manager and the Auditor in a Double Moral Hazard Setting: Efficiency through Contingent Fees and Insurance Contracts, December 1998

No. 23: **Carl-Heinrich Kehr/ Jan P. Krahnen/ Erik Theissen,** The Anatomy of a Call Market: Evidence from Germany, December 1998 (erschienen in: Journal of Financial Intermediation 10 (2001), S. 249-270)

No. 22: **Christian K. Muus,** Non-voting shares in France: An empirical analysis of the voting premium, December 1998

No. 21: **Christian Leuz,** Voluntary Disclosure of Cash Flow Statements and Segment Data in Germany, September 1998 (erscheint in: The Economics and Politics of Accounting: International Essays, C. Leuz, D. Pfaff and A. Hopwood (eds), Oxford University Press)

No. 20: **Anne D`Arcy,** The Degree of Determination of National Accounting Systems – An Empirical Investigation, September 1998

No. 19: **Helmut Laux,** Marktwertmaximierung und CAPM im Ein- und Mehrperioden-Fall, September 1998 (erschienen in Unternehmensführung, Ethik und Umwelt, Festschrift zum 65. Geburtstag von Hartmut Kreikebaum, hrsg. von Gerd-Rainer Wagner, Wiesbaden 1999, S. 226-251)

No. 18: **Joachim Grammig/ Reinhard Hujer/ Stefan Kokot/ Kai-Oliver Maurer**, Ökonometrische Modellierung von Transaktionsintensitäten auf Finanzmärkten; Eine Anwendung von Autoregressive Conditional Duration Modellen auf die IPO der Deutschen Telekom, August 1998

No. 17: **Hanne Böckem**, An Investigation into the Capital Market Reaction on Accounting Standards Enforcement, July 1998

No. 16: **Erik Theissen,** Der Neue Markt: Eine Bestandsaufnahme, April 1998 (erschienen in: Zeitschrift für Wirtschafts- und Sozialwissenschaften, Heft 4/98, S. 623-652)

No. 15: **Jan Pieter Krahnen,** Finanzierungstheorie: Ein selektiver Überblick, April 1998 (englische Fassung erschienen in "Gutenberg Centennial", hrsg. von Horst Albach, Berlin, 2000)

No. 14: **Erik Theissen,** Liquiditätsmessung auf experimentellen Aktienmärkten, April 1998 (erschienen in: Kredit und Kapital, 32(1999), Heft 2, S. 225-264)

No. 13: **Reinhard H. Schmidt,** Erich Gutenberg und die Theorie der Unternehmung, February 1998 (englische Fassung erschienen in "Theory of the Firm", hrsg. von Horst Albach u.a., Berlin 2000, S. 3-39)

No. 12: **Adalbert Winkler,** Financial Development, Economic Growth and Corporate Governance, February 1998 (erschienen in: Asian Financial Markets, hrsg. von Lukas Menkhoff/Beate Reszat, Baden-Baden 1998, S. 15-44)

No. 11: **Andreas R. Hackethal/ Marcel Tyrell,** Complementarity and Financial Systems – A Theoretical Approach, December 1998

No. 10: **Reinhard H. Schmidt/ Andreas Hackethal/ Marcel Tyrell,** Disintermediation and the Role of Banks in Europe: An International Comparison, January 1998 (erschienen in: Journal of Financial Intermediation, Vol. 8, 1999, S.37-67)

No. 9: **Stefan Heiden/ Günther Gebhardt/ Irmelin Burkhardt,** Einflußfaktoren für Kurs-reaktionen auf die Ankündigung von Kapitalerhöhungen deutscher Aktiengesellschaften, December 1997

No. 8: **Martin Nell,** Garantien als Signale für die Produktqualität?, November 1997 (erscheint in: Zeitschrift für betriebswirtschaftliche Forschung)

No. 7: **Robert M. Gillenkirch,** Anreizwirkungen und Bewertung von Erfolgsbeteiligungen im Portefeuillemanagement, November 1997 (erschienen in: ZfB, Sonderheft Finanzmanagement 1999)

No. 6: **Reinhard H. Schmidt/ C.-P. Zeitinger,** Critical Issues in Microbusiness Finance and the Role of Donors, October 1997 (erschienen in: Strategic Issues in Microfinance, ed. by Kimenyi/Wieland/Von Pischke, Averbury, UK, 1998, S. 27-51)

No. 5: **Erik Theissen/ Mario Greifzu,** Performance deutscher Rentenfonds, September 1997 (erschienen in: Zeitschrift für betriebswirtschaftliche Forschung, 50. Jg., 1998, S. 436-461)

No. 4: **Jan Pieter Krahnen/ Martin Weber,** Marketmaking in the Laboratory: Does Competition Matter?, September 1997

No. 3: **Reinhard H. Schmidt,** Corporate Governance: The Role of Other Constituen-cies, July 1997 (erschienen in: Pezard, Alice;Thiveaud, Jean-Marie (Hrsg.): Corporate Governance: Cross Border Experience, Paris, 1997, S. 61-74)

No. 2: **Ralf Ewert/ Christian Ernst,** Strategic Management Accounting, Coordination and Long-term Cost Structure, July 1997 (erschienen unter dem Titel "Target Costing, Coordination and Strategic Cost Management" in Euopean Accounting Review, Vol.8, No.1 (1999), S. 23-49)

No. 1: **Jan P. Krahnen/ Christian Rieck/ Erik Theissen,** Insider Trading and Portfolio Structure in Experimental Asset Markets with a Long Lived Asset, July 1997 (erschienen in European Journal of Finance, Vol. 5, Nr. 1, March 1999, S. 29-50)

Kontaktadresse für Bestellungen:

Professor Dr. Reinhard H. Schmidt
Wilhelm Merton Professur für
Internationales Bank- und Finanzwesen
Mertonstr. 17
Postfach 11 19 32 / HPF66
D-60054 Frankfurt/Main

Tel.: +49-69-798-28269
Fax: +49-69-798-28272
e-mail: rschmidt@wiwi.uni-frankfurt.de
http://www.finance.uni-frankfurt.de/schmidt/WPs/wp/wpliste.html

Mit freundlicher Unterstützung der Unternehmen der
Sparkassen-Finanzgruppe Hessen-Thüringen.