

Are Credit Scoring Models Sensitive With Respect to Default Definitions? Evidence from the Austrian Market

Evelyn Hayden

University of Vienna
Department of Business Administration
Chair of Banking and Finance
Brünnerstrasse 72
A-1210 Vienna
Austria
Tel.: +43 (0) 1 - 42 77 - 38 076
Fax: +43 (0) 1 - 42 77 - 38 074

E-Mail: Evelyn.Hayden@univie.ac.at

April 2003

This article is based on the chapters two to five of my dissertation. I thank Engelbert Dockner, Sylvia Frühwirth-Schnatter, David Meyer, Otto Randl, Michaela Schaffhauser-Linzatti and Josef Zechner for their helpful comments as well as participants of the research seminar at the University of Vienna, of the European Financial Management Association Meetings 2001, and of the Austrian Working Group on Banking and Finance 2001. Besides, I gratefully acknowledge financial support from the Austrian National Bank (ÖNB) under the Jubiläumsfond grant number 8652 and the contribution of three Austrian commercial banks, the Austrian Institute of Small Business Research, and the Austrian National Bank for providing the necessary data for this analysis.

Are Credit Scoring Models Sensitive With Respect to Default Definitions? Evidence from the Austrian Market

April 2003

Abstract: In this paper models of default prediction conditional on financial statements of Austrian firms are presented. Apart from giving a discussion on the suggested 65 variables the issue of potential problems in developing rating models is raised and possible solutions are reviewed. A unique data set on credit risk analysis for the Austrian market is constructed and used to derive rating models for three different default definitions, i.e. bankruptcy, restructuring, and delay-in-payment. The models are compared to examine whether the models developed on the tighter default criteria, that are closer to the definition proposed by Basel II, do better in predicting these credit loss events than the model estimated on the traditional and more easily observable default criterion bankruptcy. Several traditional methods to compare rating models are used, but also a rigorous statistical test is discussed and applied. All results lead to the same conclusion that not much prediction power is lost if the bankruptcy model is used to predict the credit loss events of rescheduling and delay-in-payment instead of the alternative models specifically derived for these default definitions. In the light of Basel II this is an interesting result. It implies that traditional credit rating models developed by banks by exclusively relying on bankruptcy as default criterion are not automatically outdated but can be equally powerful in predicting the comprising credit loss events provided in the new Basel capital accord as models estimated on these default criteria.

JEL Classification: G33, C35, C52

I. Introduction

In January 2001 the Basel Committee on Banking Supervision released the second version of its proposal for a new capital adequacy framework. In this release the Committee announced that an internal ratings-based approach could form the basis for setting capital charges for banks with respect to credit risk in the near future. Besides, the Basle Committee on Banking Supervision (2001a) defined default as any credit loss event associated with any obligation of the obligor, including distressed restructuring involving the forgiveness or postponement of principal, interest, or fees and delay in payment of the obligor of more than 90 days. According to the current proposal for the new capital accord banks will have to use this tight definition of default for estimating internal rating-based models. However, historically credit risk models were typically developed using the default criterion bankruptcy, as this information was relatively easily observable. Now an important question is whether ‘old’ rating models that use only bankruptcy as default criterion are therefore outdated, or whether they can compete with models derived for the tighter Basel II default definitions in predicting those more complex default events. Stated differently: is the structure and the performance of credit scoring models sensitive to the default definitions that were used to derive them? Should the answer be no, then banks would not have to re-calibrate their rating models but could stick to their traditional ones by just adjusting the default probability upwards to reflect the fact that the Basel II default events occur more frequently than bankruptcies. This knowledge would be especially valuable for small banks, as - due to their limited number of clients - they typically face severe problems when trying to collect enough data for being able to statistically reliably update their current rating models within a reasonable time period.

Up to the authors knowledge the present work is the first to try to answer this question. To do so, credit risk rating models based on balance sheet information of Austrian firms using the default definitions of bankruptcy, loan restructuring and 90 days past due are estimated and compared. Besides, apart from giving a discussion on the suggested 65 variables the issue of potential problems in developing rating models is raised and possible solutions are reviewed. Several traditional methods to compare rating models like the Accuracy Ratio popularized by Moody’s¹ are presented, but also a rigorous statistical test based on Receiver Operating Characteristic Curves as described in Engelmann, Hayden, and Tasche (2003) is discussed and applied.

The data necessary for this analysis was provided by three major Austrian commercial banks, the Austrian National Bank and the Austrian Institute of Small Business Research. By combining these data pools a unique data set on credit risk analysis for the Austrian market of more than 100.000 balance sheet observations was constructed.

The remainder of this work is composed as follows: In Section II the model selection is described, while Section III depicts the data and Section IV details the applied methodology. The results of the analysis are discussed in Section V. Finally, Section VI concludes.

¹See for example Sobehart, Keenan, and Stein (2000a).

II. Model Selection

As already mentioned in the introduction it is the aim of this study to develop rating systems based on varying default definitions to test whether these models show differences concerning their default prediction power. To do so, the first step is to decide on the following five questions: which parameters shall be estimated; which input variables are used; which type of model shall be estimated; how is default defined; and which time horizon is chosen? In this section these questions will be answered for the work at hand.

II.1. Parameter Selection

When banks try to predict credit risk, they actually are interested to predict the potential loss that they might incur. So the credit quality of a borrower does not only depend on the default probability, the most popular credit risk parameter, but also on the exposure-at-default, the outstanding and unsecured credit amount at the event of default, and the loss-given-default, which usually is defined as a percentage of the exposure-at-default. However, historically most studies concentrated on the prediction of the default probability. Besides, also Basel II differentiates between the Foundation and the Advanced IRB Approach, where for the Foundation Approach banks only have to estimate default probabilities. Due to these reasons and data unavailability for the exposure-at-default and the loss-given-default, the current study will focus on rating models based on default probabilities, too.

II.2. Choice of Input Variables

Essentially, there are three main possible model input categories: accounting variables, market-based variables such as market equity value and so-called soft facts such as the firm's competitive position or management skills. Historically banks used to rely on the expertise of credit advisors who looked at a combination of accounting and qualitative variables to come up with an assessment of the client firm's credit risk, but especially larger banks switched to quantitative models during the last decades.

One of the first researchers who tried to formalize the dependence between accounting variables and credit quality was Edward I. Altman (1968) who developed the famous Z-Score model and showed that for a rather small sample of observations financially distressed firms can be separated from the non-failed firms in the year before the declaration of bankruptcy at an in-sample accuracy rate of better than 90% using linear discriminant analysis. Later on more sophisticated models using linear regressions, logit or probit models and lately neural networks were estimated to improve the out-of-sample accuracy rate and to come up with true default probabilities (see f. ex. Lo (1986) and Altman, Agarwal, and Varetto (1994)). Yet all the studies mentioned above have in common that they only look at accounting variables. In contrast to this in the year 1993 KMV

published a model where market variables were used to calculate the credit risk of traded firms. As KMV's studies assert, this model based on the option pricing approach originally proposed by Merton (1974) does generally better in predicting corporate distress than accounting-based models. Besides, they came up with the idea to separate stock corporations of one sector and region and to regress their default probabilities derived from the market-value based model on accounting variables and then use those results to estimate the credit risk of similar but small, non-traded companies (see Nyberg, Sellers, and Zhang (2001)).

Due to those facts at first sight one might deduce that one should always use market-value based models when developing rating models. However, there are some countries where almost no traded companies exist. For example, according to the Austrian Federal Economic Chamber in the year 2000 stock corporations accounted for only about 0.5% of all Austrian companies. Furthermore, as Sobehart, Keenan, and Stein (2000a) point out in one of Moody's studies, the relationship between financial variables and default risk varies substantially between large public and usually much smaller private firms, implying that default models based on traded firm data and applied to private firms will likely misrepresent actual credit risk. Therefore it might be preferable to rely exclusively on the credit quality information contained in accounting variables when fitting a rating model to such markets. Besides, one could also consider to include soft facts into the model building process. However, for the study at hand, due to the inherent subjectivity of candidate variables and data unavailability, soft facts were excluded from the model, too, leaving accounting variables as the main input to the analysis.

II.3. Model-Type Selection

In principle, three main model categories exist:

- Judgements of experts (credit advisors)
- Statistical models ²
 - Linear discriminant analysis
 - Linear regressions
 - Logit and Probit models
 - Neural networks
- Theoretical models (option pricing approach)

However, as already evident from the arguments in Section II.2, the choice of the model-type and the selection of the input variables have to be adapted to each other. The option pricing model, for example, can only be used if market-based data is available, which for the majority of Austrian companies is not the case. Therefore this model is not appropriate. Excluding the informal, rather

²For a comprehensive review of the literature on the various statistical methods that have been used to construct default prediction models see for example Dimitras, Zanakas, and Zopoundis (1996).

subjective expert-judgements from the model-type list, only statistical models are left. Within this group of models, on the one hand logit and probit models, that generally lead to similar estimation results, and on the other hand neural networks are the state of the art. This study focuses on logit models mainly because of two reasons. Firstly, although there is some evidence in the literature that artificial neural networks are able to outperform probit or logit regressions in achieving higher prediction accuracy ratios, as for example in Charitou and Charalambous (1996), there are also studies like the one of Barniv, Agarwal, and Leach (1997) finding that differences in performance between those two classes of models are either non-existing or marginal. Secondly, the chosen approach allows to check easily whether the empirical dependence between the potential input variables and default risk is economically meaningful, as will be demonstrated in Section IV.

II.4. Default Definition

Historically credit risk models were developed using the default criterion bankruptcy, as this information was relatively easily observable. But of course banks also incur losses before the event of bankruptcy, for example when they move payments back in time without compensation in hopes that at a later point in time the troubled borrower will be able to repay his debts. Therefore the Basle Committee on Banking Supervision (2001a) defined the following reference definition of default:

A default is considered to have occurred with regard to a particular obligor when one or more of the following events has taken place:

- it is determined that the obligor is unlikely to pay its debt obligations (principal, interest, or fees) in full;
- a credit loss event associated with any obligation of the obligor, such as a charge-off, specific provision, or distressed restructuring involving the forgiveness or postponement of principal, interest, or fees;
- the obligor is past due more than 90 days on any credit obligation; or
- the obligor has filed for bankruptcy or similar protection from creditors.

According to the current proposal for the New Capital Accord banks will have to use the above regulatory reference definition of default in estimating internal rating-based models. However, up to the authors knowledge until now there does not exist a single study testing whether traditional, bankruptcy-based rating models are indeed inferior to models derived for the tighter Basel II default definitions in predicting those more complex default events. Stated differently: is the structure and the performance of credit scoring models sensitive to the default definitions that were used to derive them? Should the answer be no, then banks would not have to re-calibrate their rating models but could stick to their traditional ones by just adjusting the default probability upwards to reflect the fact that the Basel II default events occur more frequently than bankruptcies. This knowledge would be especially valuable for small banks, as - due to their limited number of

clients - they typically face severe problems when trying to collect enough data for being able to statistically reliably update their current rating models within a reasonable time period. To answer this question (at least concerning accounting input), rating models using the default definitions of bankruptcy, loan restructuring and 90 days past due are estimated and compared.

II.5. Time Horizon

As the Basle Committee on Banking Supervision (1999a) illustrates for most banks it is common habit to use a credit risk modeling horizon of one year. The reason for this approach is that one year is considered to reflect best the typical interval over which

- a) new capital could be raised;
- b) loss mitigation action could be taken to eliminate risk from the portfolio;
- c) new obligor information can be revealed;
- d) default data may be published;
- e) internal budgeting, capital planning and accounting statements are prepared; and
- f) credits are normally reviewed for renewal.

But also longer time horizons could be of interest, especially when decisions about the allocation of new loans have to be made. To derive default probabilities for such longer time horizons, say 5 years, two methods are possible: firstly, one could calculate the 5-year default probability from the estimated one-year value, however, this calculated value might be misleading as the relationship between default probabilities and accounting variables could be changing when altering the time horizon. Secondly, a new model for the longer horizon might be estimated, but usually here data unavailability imposes severe restrictions. As displayed in Section III and Appendix A, about two thirds of the largest data set used for this study and almost all observations of the two smaller data sets are lost when default should be estimated based on accounting statements prepared 5 years before the event of default - therefore this study sticks to the convention of adopting a one-year time horizon, the method currently proposed by the Basle Committee on Banking Supervision (2001b).

III. The Data Set

As illustrated in Section II, in this study accounting variables are the main input to the credit quality rating model building process based on logistic regressions. The necessary data for the statistical analysis was supplied by three major commercial Austrian banks, the Austrian National Bank and the Austrian Institute for Small Business Research. The original data set consisted of about 230.000 firm-year observations spanning the time period 1975 to 2000. However, due to obvious mistakes in the balance sheets and gain and loss accounts, such as assets being different from liabilities or negative sales, the data set had to be reduced to 199.000 observations. Besides, certain firm types were excluded, i.e. all public firms including large international corporations, as they do not represent the typical Austrian company, and rather small single owner firms with a turnover of less than 5m ATS (about 0.36m EUR), whose credit quality often depends as much on the finances of a key individual as on the firm itself. After also eliminating financial statements covering a period of less than twelve months and checking for observations that were twice or more often in the data set almost 160.000 firm-years were left. Finally those observations were dropped, where the default information was missing or dubious. By using varying default definitions, three different data sets were constructed. The biggest data set defines the default event as the bankruptcy of the borrower within one year after the preparation of the balance sheet and consists of over 1.000 defaults and 123.000 firm-year observations spanning the time period 1987 to 1999. The second data set, which is less than half as large as the first one, uses the first event of loan restructuring (for example forgiveness or postponement of principal, interest, or fees without compensation) or bankruptcy as default criterion, while the third one includes almost 17.000 firm-year observations with about 1.600 defaults and uses 90 days past due as well as restructuring and bankruptcy as default event. The different data sets are summarized in Table 1.

Table 1
Data set characteristics using different default definitions

This table displays the number of observed balance sheets, distinct firms and defaults as well as the covered time period for three data sets that were built according to the default definition of bankruptcy, rescheduling, and delay in payment (arising within one year after the reference point-in-time of the accounting statement). The finer the default criterion is, the higher is the number of observed defaults, but the lower is the number of total firm-year observations as some banks only record bankruptcy as default criterion.

default definition	bankruptcy	restructuring	90 days past due
firm-years	124,479	48,115	16,797
companies	35,703	14,602	6,062
defaults	1,024	1,459	1,604
time-period	1987-1999	1992-1999	1992-1999

Each observation consists of the balance sheet and the gain and loss account of a particular firm for a particular year, the firm's legal form, the sector in which it is operating according to the ÖNACE-classification ³, and the information whether default occurred within one year after the accounting statement was prepared.

The composition of the data for the largest data set (bankruptcy) is illustrated in Table 2 as well as in Figure 1 to Figure 4. The corresponding graphs for the other two data sets, that depict similar patterns as the figures for the bankruptcy data, are shown in Appendix A.

Table 2
Number of observations and defaults per year for the bankruptcy data set

This table shows the total number of the observed balance sheets and defaults per year. The last column displays the yearly default frequency according to the bankruptcy data set, that varies substantially due to the varying data contribution of different banks.

year	observations	in %	defaults	in %	default ratio in %
1987	2,235	1.80	1	0.10	0.04
1988	2,184	1.75	9	0.88	0.41
1989	2,055	1.65	8	0.78	0.39
1990	2,084	1.67	14	1.37	0.67
1991	2,406	1.93	20	1.95	0.83
1992	7,789	6.26	31	3.03	0.40
1993	9,894	7.95	32	3.13	0.32
1994	12,697	10.20	49	4.79	0.39
1995	16,814	13.51	103	10.06	0.61
1996	19,096	15.34	156	15.23	0.82
1997	19,837	15.94	208	20.31	1.05
1998	17,745	14.26	249	24.32	1.40
1999	9,643	7.75	144	14.06	1.49
Total	124,479	100.00	1,024	100.00	0.82

In Table 2 the number of observations and defaults per years is depicted. It is noticeable that the ratio of defaults to total observations is rather volatile. It varies much more than could be explained purely by macro-economic changes. The reason for this pattern lies in the composition of the data set. Not all banks were able to deliver data for the whole period of 1987 to 1999, and while some banks were reluctant to make all their observations of good clientele available but delivered all their defaults, others did not record their defaults for the entire period. The consequence is that macro-economic influences can not be studied with this data set. Besides, it is important to guarantee that the accounting schemes of the involved banks are (made) comparable, because one can not easily control for the influence of different banks as - due to the above mentioned circumstances - they delivered data with rather in-homogeneous default frequencies. Therefore only

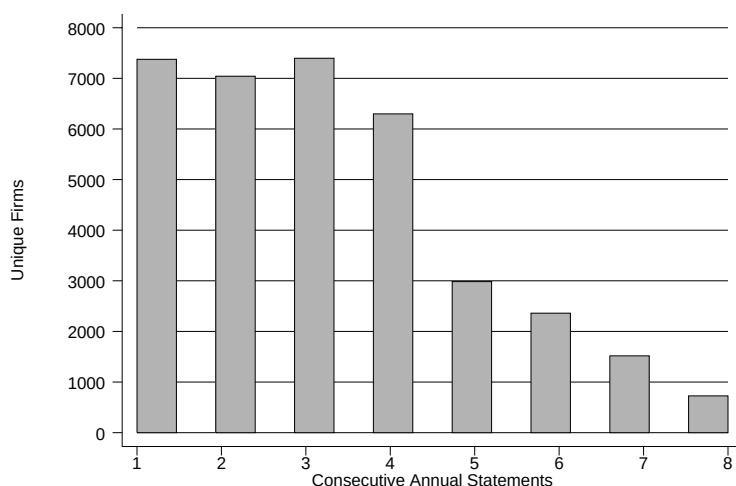
³The ÖNACE-classification is the Austrian version of the NACE-classification of the European Union, the "nomenclature générale des activités économiques dans les communautés européennes".

major positions of the balance sheets and gain and loss accounts could be used. The comparability of those items was proven when they formed the basis for the search of observations that were reported by more than one bank and several thousands of those double counts could be excluded from the data set.

Figure 1 groups the companies according to the number of consecutive financial statement observations that are available for them. For about 7,000 firms only one balance sheet belongs to the bankruptcy data set, while for the rest two to eight observations exist. These multiple observations will be important for the evaluation of the extent to which trends in financial ratios help predict defaults.

Figure 1. Obligor Counts by Number of Observed Yearly Observations

This figure shows the number of borrowers that have either one or multiple financial statement observations for different lengths of time. Multiple observations are important for the evaluation of the extent to which trends in financial ratios help predict defaults.



In contrast to the former graphs Figures 2 to 4 are divided into a development and a validation sample. The best way to test whether an estimated rating model does a good job in predicting default is to apply it to a data set that was not used to develop the model. In this work the estimation sample includes all observations for the time period 1987 to 1997, while the test sample covers the last two years. In that way the default prediction accuracy rate of the derived model can be tested on an out-of-sample, out-of-time and - as depicted in the next three graphs - slightly out-of-universe data set that contains about 40% of total defaults.

Figure 2. Distribution of Financial Statements by Legal Form

This figure displays the distribution of the legal form. The test sample differs slightly from the estimation sample as its percentage of limited liability companies is a few percentages higher.

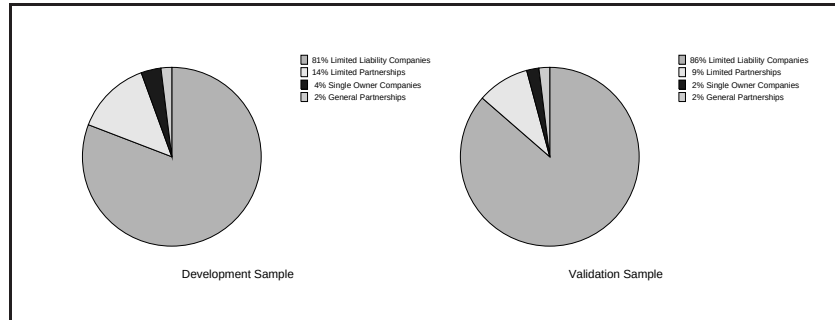


Figure 3. Distribution of Financial Statements by Sales Class

This graph shows the distribution of the accounting statements grouped according to sales classes for the observations in the estimation and the test sample. Differences between those two samples according to this criterion are only marginal.

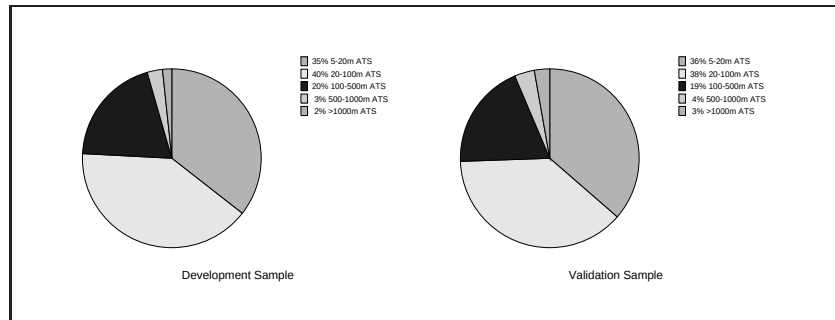
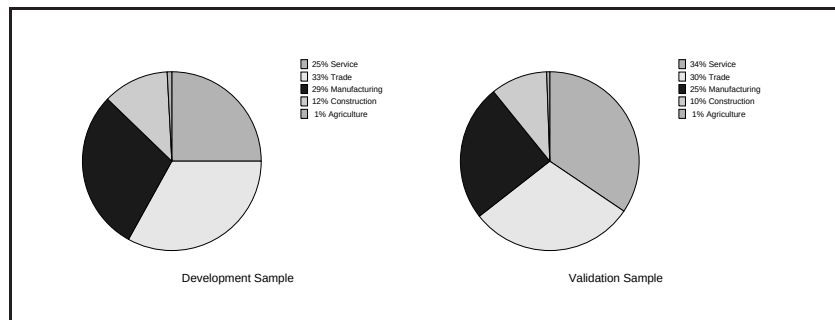


Figure 4. Distribution of Financial Statements by Industry Segments

This figure shows that the distribution of firms by industry differs between the development and the validation sample as there are more service companies in the test sample. This provides a further element of out-of-universe testing.



IV. Methodology

For reasons described in Section II, the credit risk rating models for Austrian companies shall be developed by estimating a logit regression and using accounting variables as the main input to it. The exact methodology, consisting of the selection of candidate variables, the testing of the linearity assumption inherent in the logit model, the estimation of univariate regressions, the construction of the final models and their validation will be explained in the following section.

IV.1. Selection of Candidate Variables

To derive a credit quality models, in a first step candidate variables for the final model have to be selected. As there is a huge number of possible candidate ratios and according to Chen and Shimerda (1981) in the literature out of more than 100 financial items almost 50% were found useful in at least one empirical study, the selection strategy described below was chosen.

In a first step all potential candidate variables that could be derived from the available data set are defined and calculated. Already at that early stage some variables cited in the literature had to be dropped, either because of data unavailability or because of interpretation problems. An example for the first reason of exclusion is the productivity ratio “Net Sales / Number of Employees” mentioned in Crouhy, Galai, and Mark (2001), as in the current data set the number of employees for a particular firm is not available. Interpretation problems would arise if for example the profitability ratio “Net Income / Equity” was considered, as - in contrast to most Anglo-American studies of large public firms - the equity of the observed companies sometimes is negative. Usually one would expect that the higher the return on equity, the lower the default probability is. However, if equity can be negative, a firm with a highly negative net income and a small negative equity value would generate a huge positive return-on-equity-ratio and would therefore wrongly obtain a prediction of low default probability. To eliminate those problems all accounting ratios were excluded from the analysis where the variable in the denominator could be negative.

Then, in a second step the accounting ratios were classified according to the ten categories leverage, debt coverage, liquidity, activity, productivity, turnover, profitability, firm size, growth rates and leverage development, which represent the most obvious and most cited credit risk factors. Table 3 lists all ratios that were chosen for further examination according to this scheme.

Leverage

The credit risk factor group leverage contains ten accounting ratios. Those measuring the debt proportion of the assets of the firm should have a positive relationship with default, those measuring the equity ratio a negative one. In the literature leverage ratios are usually calculated by just using the respective items of the balance sheet, however, Baetge and Jerschensky (1996) and Khandani, Lozano, and Carty (2001) suggested to adjust the equity ratio in the following way to counter creative accounting practices:

Table 3
Promising Accounting Ratios

In this table all accounting ratios that are examined in this study are listed and grouped according to ten popular credit risk factors. Besides, in the fourth column the expected dependence between accounting ratio and default probability is depicted, where + symbolizes that an increase in the ratio leads to an increase in the default probability and - symbolizes a decrease in the default probability given an increase in the ratio. Finally, column five lists some current studies in which the respective accounting ratios are used, too.

	Accounting Ratio	Credit Risk Factor	Hypothesis	Literature
1	Liabilities / Assets	Leverage	+	a, c, d, e, f
2	Equity / Assets	Leverage	-	a, c
3	Equity / Assets*	Leverage	-	b
4	Liabilities / Tangible Assets	Leverage	+	a, d
5	Long term Liabilities / Assets	Leverage	+	d, e
6	Bank Debt / Assets	Leverage	+	a, c
7	Bank Debt / Assets*	Leverage	+	b
8	Bank Debt/(Assets - Bank Debt)	Leverage	+	a
9	Bank Debt / (Assets - Bank Debt)*	Leverage	+	b
10	Bankdebt / Liabilities	Leverage	+	d
11	EBIT / Interest Expenses	Debt Coverage	-	a, f
12	Cash Flow / (Liab.-Advances)*	Debt Coverage	-	b
13	Current Assets / Current Liabilities	Liquidity	-	a, c, d, e, f
14	Current Assets / Liabilities	Liquidity	-	d
15	Working Capital / Assets	Liquidity	-	a, b, d, e
16	Current Liabilities / Assets	Liquidity	+	d
17	Current Assets / Assets	Liquidity	-	d, e
18	Cash / Assets	Liquidity	-	a, d, e
19	Working Capital / Net Sales	Liquidity	-/+	d, e
20	Cash / Net Sales	Liquidity	-/+	d
21	Current Assets / Net Sales	Liquidity	-/+	d
22	Quick Assets / Net Sales	Liquidity	-/+	d
23	Short Term Bank Debt / Bank Debt	Liquidity	+	a
24	Cash / Current Liabilities	Liquidity	-	d, e
25	Working Capital / Current Liabilities	Liquidity	-	d
26	Quick Ratio	Liquidity	-	a, d, e, f
27	Inventory / Operating Income	Activity	+	c
28	Inventory / Net Sales	Activity	+	d, e
29	Inventory / Material Costs	Activity	+	a, c
30	Accounts Receivable / Net Sales	Activity	+	a, e
31	Accounts Receivable / Operating Income	Activity	+	c
32	Accounts Receivable / Liabilities	Activity	-	d
33	Accounts Receivable / Liabilities*	Activity	-	

a...Falkenstein, Boral, and Carty (2000)

b...Khandani, Lozano, and Carty (2001)

c...Lettmayr (2001)

d...Chen and Shimerda (1981)

e...Kahya and Theodossiou (1999)

f...Crouhy, Galai, and Mark (2001)

CPI...Consumer Price Index 1986

*...assets, equity and liabilities adjusted for intangible assets and cash

Table 3 continued
Promising Accounting Ratios

In this table all accounting ratios that are examined in this study are listed and grouped according to ten popular credit risk factors. Besides, in the fourth column the expected dependence between accounting ratio and default probability is depicted, where + symbolizes that an increase in the ratio leads to an increase in the default probability and - symbolizes a decrease in the default probability given an increase in the ratio. Finally, column five lists some current studies in which the respective accounting ratios are used, too.

	Accounting Ratio	Credit Risk Factor	Hypothesis	Literature
34	Accounts Receivable / Material Costs	Activity	+	a
35	Accounts Payable / Material Costs	Activity	+	a
36	Accounts Payable / Net Sales	Activity	+	b
37	Accounts Receivable / Inventory	Activity	+	d
38	Personnel Costs / Net Sales	Productivity	+	b
39	Operating Income / Personnel Costs	Productivity	-	c
40	(Net Sales-Material Costs)/Personnel Costs	Productivity	-	c
41	Material Costs / Operating Income	Productivity	+	c, f
42	Net Sales / Assets	Turnover	-	a, d, e
43	Net Sales / Assets*	Turnover	-	b
44	Operating Income / Assets	Turnover	-	c, e
45	EBIT / Assets	Profitability	-	a, d
46	EBIT / Assets*	Profitability	-	b
47	EBIT / Net Sales	Profitability	-	d, f
48	(EBIT+Interest Income)/Operating Income	Profitability	-	c
49	(EBIT + Interest Income) / Assets	Profitability	-	c
50	Ordinary Business Income / Assets	Profitability	-	c
51	Ordinary Business Income / Assets*	Profitability	-	b
52	(Ord.Bus.Income+Interest+Depr.) / Assets*	Profitability	-	b
53	Ord. Business Income / Operating Income	Profitability	-	a, c
54	Net Income / Assets	Profitability	-	a, d, e, f
55	Net Income / Assets*	Profitability	-	b
56	Net Income / Net Sales	Profitability	-	d, f
57	Net Income / Operating Income	Profitability	-	c
58	Retained Earnings / Assets	Profitability	-	a, d, e
59	Assets / CPI	Size	-	a, e
60	Assets / CPI*	Size	-	b
61	Net Sales / CPI	Size	-	a, e
62	Net Sales / Last Net Sales	Growth Rates	-/+	a, b
63	Operating Income / Last Op. Income	Growth Rates	-/+	a
64	(Liab./Assets) / (Last Liab./Assets)	Leverage Change	+	a
65	(Bankdebt/Assets)/(Last Bankdebt/Assets)	Leverage Change	+	a

a...Falkenstein, Boral, and Carty (2000) b...Khandani, Lozano, and Carty (2001)

c...Lettmayr (2001)

d...Chen and Shimerda (1981)

e...Kahya and Theodossiou (1999)

f...Crouhy, Galai, and Mark (2001)

CPI...Consumer Price Index 1986

*...assets, equity and liabilities adjusted for intangible assets and cash

- Subtract intangible assets from equity and assets as the value of these assets generally is considerable lower than the accounting value in the case of default;
- Subtract cash and equivalents from assets (and debt) as one course of action for a firm wishing to improve its reported liquidity is to raise a short-term loan at the end of the accounting period and hold it in cash.

Therefore also such adjusted accounting ratios are considered in the study at hand and are marked with a star in Table 3 whenever either assets, equity, debt or several of those items are adjusted for a certain accounting ratio.

Debt Coverage

Debt coverage either measures the earnings before interest and taxes to interest expenses or the cash flow to liabilities ratio. Here liabilities were adjusted by subtracting advances from customers in order to account for industry specificities (e.g. construction), where advances traditionally play an important role in financing.

Liquidity

Liquidity is a common variable in most credit decisions and can be measured by a huge variety of accounting ratios. The most popular ratio is the current ratio, calculated as current assets divided by current liabilities. In general the hypothesis is that the higher liquidity, i.e. the higher cash and other liquid positions or the lower short-term liabilities, the lower is the probability of default. However, for the four liquidity ratios that are scaled by sales instead of assets or liabilities, another effect has to be taken into account. As discussed below, the larger the turnover of a firm the lower is its default probability, implying that the smaller the reciprocal of turnover the more creditworthy a company is. Therefore one has the effect that - as for example a large “Working Capital / Net Sales” ratio can be caused by good liquidity or by small sales - the overall influence of an increase in these ratios on the default probability is unclear. Nevertheless those ratios were often used in older studies, and as they were found to be useful for the credit risk analysis in Tamari (1966), Deakin (1972) and Edmister (1972) they were also selected for further examination in the work at hand.

Activity Ratios

Activity ratios are accounting ratios that reflect some aspects of the firm that have less straightforward relations to credit risk than other variables, but that nevertheless capture important information. Most of the ratios considered in this study either display the ability of the firm’s customers to pay their bills, measured by accounts receivable, or they evaluate the company’s own payment habit in looking at accounts payable. For example a firm that suffers from liquidity problems would have higher accounts payable than a healthy one. Therefore the default probability should increase with these ratios. The only exception is “Accounts Receivable / Liabilities”, as here an increasing ratio means that a larger fraction of the firms own debt can be repaid by outstanding claims. For activity ratios that use inventory in the numerator again a positive relationship to

the default probability is expected, as a growing inventory reveals higher storage costs as well as non-liquidity.

Productivity

Here the costs for generating the company's sales are measured by looking at the two big cost categories personal and material expenses. The higher the costs, the worse the firm is off.

Turnover

As for example illustrated in Coenenberg (1993), asset turnover reflects the efficiency with which the available capital is used. According to Lettmayr (2001) a high "Sales / Assets" ratio is a prerequisite to obtain high returns with relatively low investment and has a positive effect on the liquidity of the firm, therefore reducing the default probability.

Profitability

Profitability can be expressed in a variety of accounting ratios that either measure profit relative to assets or relative to sales. As higher profitability should raise a firm's equity value and also implies a longer way of revenues to fall or costs to rise before losses incur, a company's creditworthiness is positively related to its profitability.

Size

According to Falkenstein, Boral, and Carty (2000) sales or total assets are almost indistinguishable as reflections of size risk. Both items are divided by the consumer price index to correct for inflation. Usually smaller firms are less diversified and have less depth in management, which implies greater susceptibility to idiosyncratic shocks. Therefore larger companies should default less frequently than smaller firms.

Growth Rates

As Khandani, Lozano, and Carty (2001) point out, the relationship between the rate at which companies grow and the rate at which they default is not as simple as that between other ratios and default. The reason is that while it is generally better for a firm to grow than to shrink, companies that grow very quickly often find themselves unable to meet the management challenges presented by such growth - especially within smaller firms. Furthermore, this quick growth is unlikely to be financed out of profits, resulting in a possible build up of debt and the associated risks. Therefore one should expect that the relationship between the growth ratios and default is non-monotone, what will be examined in detail lateron.

Change of Leverage

Lenders are often more interested in where the firm is going than where it has been. For that purpose, trends are often analyzed. The most important trend variables are probably the change in profits and the change in liabilities. However, former studies such as the one of Falkenstein,

Boral, and Carty (2000) find that ratio levels in general do better in discriminating between good and defaulting firms than their corresponding growth ratios. Nevertheless, the impact of a change in liabilities shall be examined in this work. However, profit growth ratios will not be explored as they suffer from the problem of possible negative values in the denominator discussed above.

IV.2. Test of Linearity Assumption

After having selected the candidate accounting ratios, the next step is to check whether the underlying assumptions of the logit model apply to the data. The logit model can be written as

$$Prob(Default) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \quad (1)$$

This implies a linear relationship between the Log Odd and the input accounting ratios.

$$LogOdd = \ln\left(\frac{\Pi}{1 - \Pi}\right) = \alpha + \beta x \quad (2)$$

To test for this linearity assumption, the variables are divided into about 50 groups that all contain the same number of observations, and within each group the historical default rate respectively the empirical log odd is calculated. Finally a linear regression of the log odd on the mean values of the variable intervals is estimated.

What I find is that for most accounting ratios the linearity assumption is indeed valid. As an example the relationship between the variable “Current Liabilities / Total Assets” and the empirical log odd for the bankruptcy criterion as well as the estimated linear regression is depicted in Figure 5. The fit of the regression is as high as 82.02%.

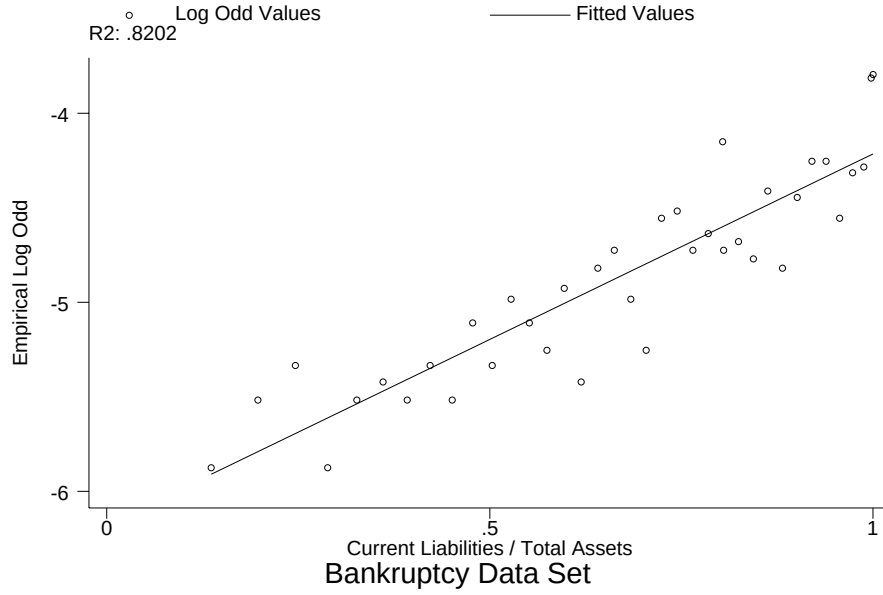
However, for some accounting ratios the functional dependence between the log odd and the variable is nonlinear. In most of these cases the relationship is still monotone, as for example for “Bank Debt / (Assets-Bank Debt)” depicted in Figure 6. Therefore there is no need to adjust these ratios at that stage of the model building process, as one will get significant coefficients in univariate logit regressions, the next step for identifying the most influential variables, anyway.

But there are also two accounting ratios, i.e. “Sales Growth” and “Operating Income Growth”, that show non-monotone behavior just as was expected. The easiest way would be to exclude those two variables from further analysis, however, other studies claim that sales growth would be a very helpful ratio in predicting default. So to be able to investigate whether this is true for Austria, the two variables have to be linearized before logit regressions can be estimated. This is done in the following way: the points obtained from dividing the ratios into groups and plotting them against empirical log odds are smoothed by an adapted version of a filter proposed in Hodrick and Prescott (1997) to reduce noise. The formula for the Hodrick-Prescott filter was intended for the examination of the growth component of time series and looks like

$$Min_{(g_i)} \sum_i (y_i - g_i)^2 + \lambda \sum_i [(g_i - g_{i-1}) - (g_{i-1} - g_{i-2})]^2 \quad (3)$$

Figure 5. Linearity Test for the “Current Liabilities/Total Assets” Ratio (Bankruptcy Data Set)

This figure shows the relationship between the variable “Current Liabilities / Total Assets” and the empirical log odd for the bankruptcy criterion, which is derived by dividing the accounting ratio into about 50 groups and calculating the historical default rate respectively the empirical log odd within each group. Finally a linear regression of the log odd on the mean values of the variable intervals is estimated and depicted, too. One can see that for the “Current Liabilities / Total Assets” ratio the linearity assumption is valid.



But as in our application the observed intervals in the input variable are not stable as it is the case for time, the filter has to be adapted to

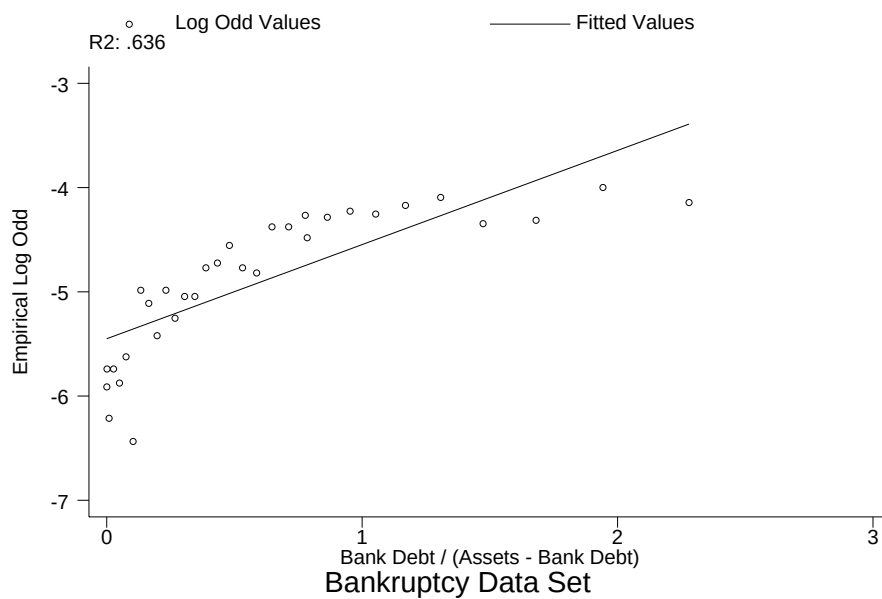
$$\text{Min}_{(g_i)} \sum_i (y_i - g_i)^2 + \lambda \sum_i \left[\left(\frac{g_i - g_{i-1}}{x_i - x_{i-1}} \right) - \left(\frac{g_{i-1} - g_{i-2}}{x_{i-1} - x_{i-2}} \right) \right]^2 \quad (4)$$

where y is the empirical log odd, x is the corresponding value of the accounting ratio, λ is the smoothing parameter that was set to 0.005 and g defines the log odd after smoothing. So this filter minimizes the squared difference between the original and the filtered log odds subject to a smoothness constraint on the smoothed log odd values. The larger the value of λ , the smoother the result is, as the variability in growth of the filtered log odds is penalized more severely. This implies that if λ approached infinity a least squares linear regression would be fitted to the data.

Figure 7 shows the resulting relationship between the ratio “Sales Growth” and the log odd for the bankruptcy data set. Now the accounting ratios are transformed to log odds according to these smoothed relationships and in any further analysis the transformed log odd values replace the original ratios as input variables.

Figure 6. Linearity Test for the “Bank Debt / (Assets-Bank Debt)” Ratio (Bankruptcy Data Set)

This figure shows the relationship between the variable “Bank Debt / (Assets-Bank Debt)” and the empirical log odd for the bankruptcy criterion, which is derived by dividing the accounting ratio into about 50 groups and calculating the empirical log odd within each group. Then a linear regression of the log odd on the mean values of the variable intervals is estimated and depicted, too. One can see that for the “Bank Debt / (Assets-Bank Debt)” ratio the linearity assumption is not valid, but nevertheless the graph displays a monotone relationship between the variable and the default probability.



Besides, this test for the appropriateness of the linearity assumption also allows for a first check whether the univariate dependence between the considered accounting ratios and the default probability is as expected. As can be seen in Table 4 in Section V, all variables behave in an economically meaningful way. Also the suspicion that for the four liquidity ratios that are scaled by sales the overall influence of an increase in these ratios on the default probability is unclear is verified. Dependent on which of the two conflicting effects is larger, two of those variables show a positive empirical relationship to default and the other two a negative one.

Another important result already derived at this early stage of the model building process is the fact that the functional dependence between log odd and input variable is the same for all three default definitions for all examined variables. So if the relationship between log odd and accounting variable is linear for the default criterion bankruptcy, it is also linear for the criteria loan restructuring and 90 days past due. This can be interpreted as a first hint that perhaps models that were developed by using a certain default definition also do well when used to predict default based on other default criteria.

Figure 7. Smoothed Relationship between “Sales Growth” and the Empirical Log Odd

This figure shows the smoothed relationship between the variable “Sales Growth” and the log odd for the bankruptcy data set. In any further analysis the transformed log odd values are used as input variable instead of the corresponding accounting ratio.

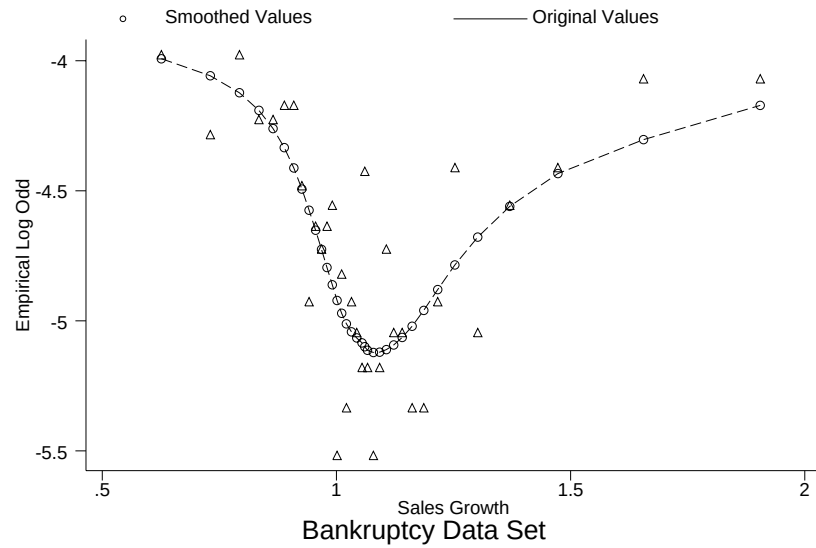
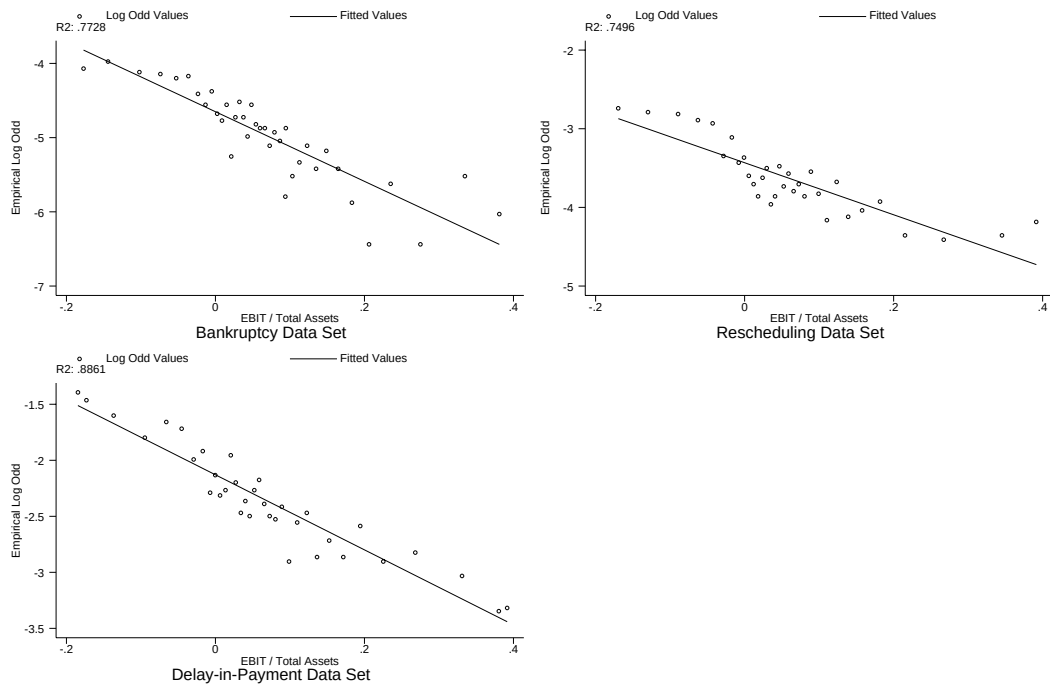


Figure 8. Functional Dependence between “EBIT / Total Assets” and the Default Probability

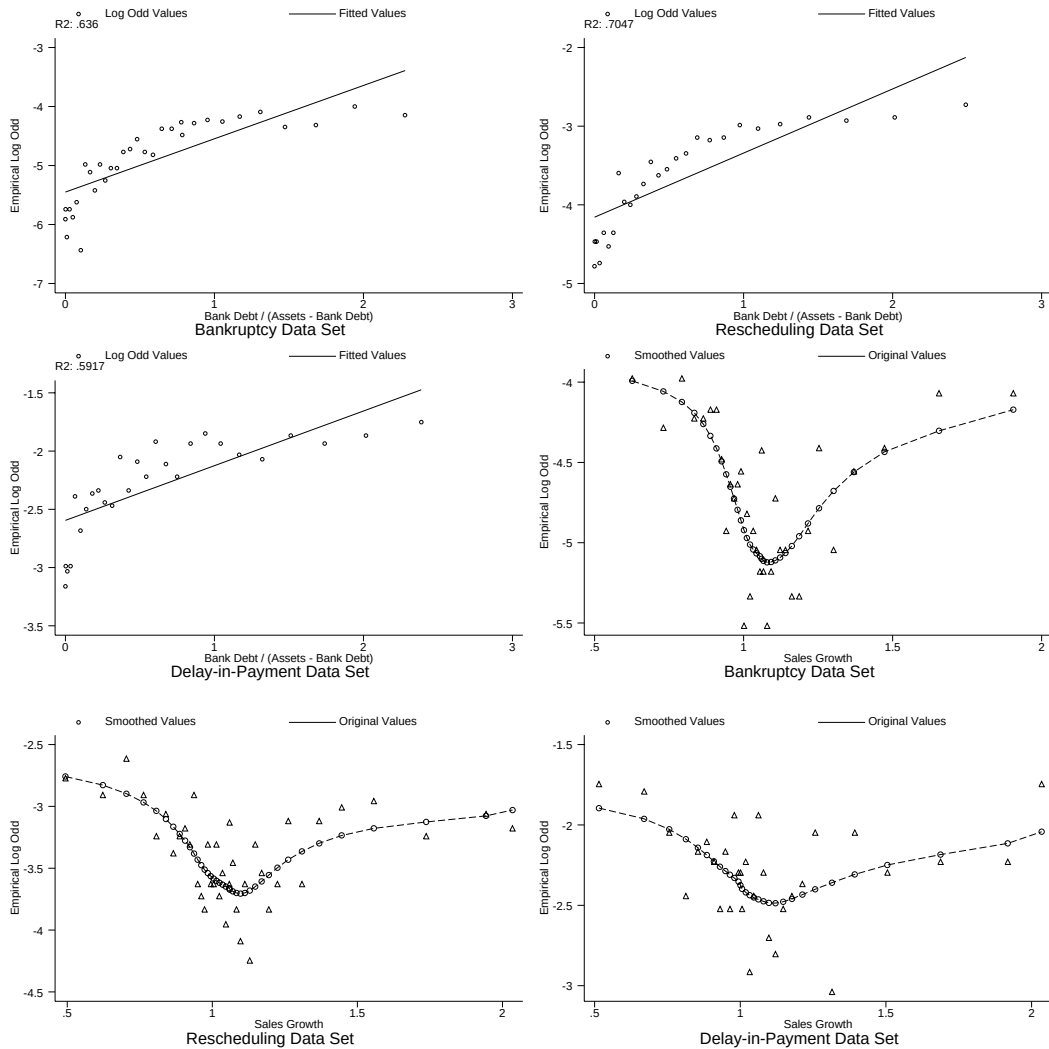
This figure shows that the functional dependence between the log odds and the “EBIT/Assets” ratio is the same for all three default definitions.



One example for the equality of the functional dependence between variables and default probability for all three data sets is depicted in Figure 8. Here the linearity assumption is valid. Further examples for non-linear but monotone and non-monotone behavior are displayed in Figure 9. The functional relationships between accounting ratios and log odds for all variables are recorded in Table 4.

Figure 9. Functional Dependence between “Bank Debt / (Assets - BankDebt)” and “Sales Growth” and the Default Probability for all Three Data Sets

This figure shows that the functional dependence between the log odds and the “Bank Debt / (Assets - Bank Debt)” ratio respectively “Sales Growth” is the same for all three default definitions.



IV.3. Univariate Logit Models

After verifying that the underlying assumptions of a logistic regression are valid, the next step is to estimate univariate logit models to find the most powerful variables per credit risk factor group. Here the data sets are divided into a development sample and a test sample in the way illustrated in Section III. The univariate models are estimated by using exclusively the data of the development samples. However, before one can do so one has to decide which type of logit model should be estimated.

Actually, the data sets at hand are longitudinal or panel data sets as they reveal information about different firms for different points in time. According to Mátyás and Sevestre (1996) panel data sets offer a certain number of advantages over traditional pure cross section or pure time series data sets that should be exploited whenever possible. Amongst other arguments they mention that panel data sets may alleviate the problem of multicollinearity as the explanatory variables are less likely to be highly correlated if they vary in two dimensions. Besides, it is sometimes argued that cross section data would reflect long-run behaviour, while time series data should emphasize short-run effects. By combining these two sorts of information, a distinctive feature of panel data sets, a more general and comprehensive dynamic structure could be formulated and estimated, Mátyás and Sevestre (1996) conclude.

Although these arguments are convincing, the problem with the data sets used in the study at hand is that they are incomplete panel data sets. Not all firms are covered for the whole observation period, on the contrary, as depicted in Section III and Appendix A for a non-negligible number of companies only one accounting statements was gathered at all. What's more, also trend variables shall be included into the analysis. To compute these trend variables balance sheet information of two consecutive years is required, therefore reducing the number of usable observations per firm. Finally, the data is split into an estimation and a validation data set, which again diminishes the amount of time information available. For these reasons the average observation period is reduced to 2.3 years for the bankruptcy and to 1.6 years for the delay-in-payment data set, implying that the panel data almost shranked to a cross section data set.

Besides, some test regressions were run, where the estimation results of univariate logit models assuming cross section data were compared to those of univariate variable effects models exploiting the panel data information. What I found was that the proportion of the total variance contributed by the panel-level variance component was zero (after rounding to 6 decimal places) in all cases. This implies that the panel-level variance component is unimportant and the panel estimator is not different from the pooled estimator where all time-information is neglected and a simple cross-section logit model is estimated. However, the cross-section estimator has the advantage that it is computationally much faster, so that this estimator instead of the panel estimator was used in remaining of this work.

Having decided on that, one can return to look for the accounting ratios with the highest discriminatory power. They can be identified by estimating univariate, cross-sectional logistic

models and then applying the concepts of Cumulative Accuracy Profiles and Accuracy Ratios established by Keenan and Sobehart (1999). These concepts work like the following: to plot Cumulative Accuracy Profiles, companies are first sorted according to their forecasted default probability, from riskiest to safest. Then, for a given fraction of the total number of observations, a Cumulative Accuracy Curve is constructed by calculating the percentage of the defaulters whose default probability is higher or equal to the one of the given fraction.

Figure 10. Cumulative Accuracy Profile of “Liabilities/Assets”

This figure shows an example of a Cumulative Accuracy Profile. The dark curved line shows the performance of the model being evaluated in depicting the percentage of defaults captured by the model at different percentages of the data set, while the thin straight line below represents the naive case of zero information or random assignment of default probabilities. The other thin line represents the case of perfect information, where all defaults are assigned the highest default probabilities. The Accuracy Ratio is the ratio of the performance improvement of the model being evaluated over the naive model to the performance improvement of the perfect model over the naive model. In this example the Accuracy Ratio is 44.174%.

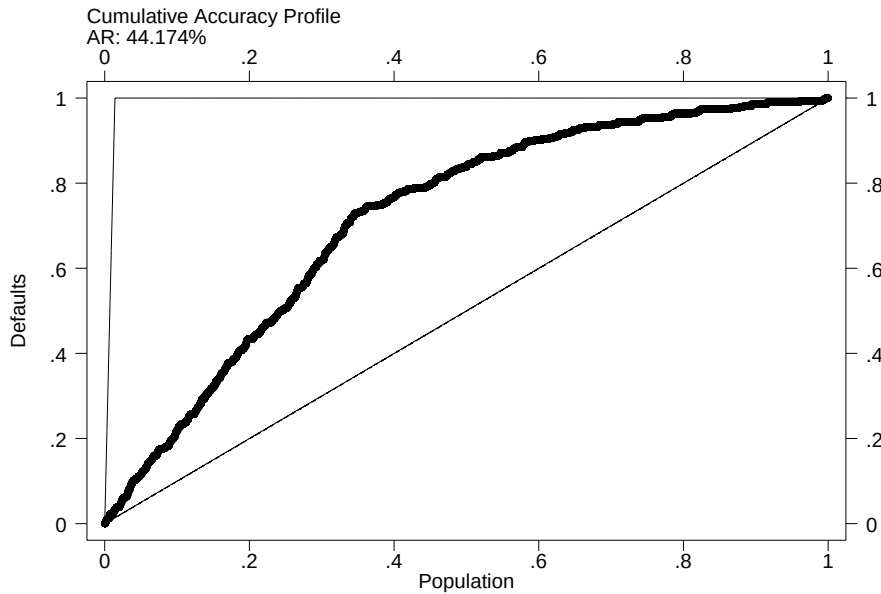


Figure 10 shows the Cumulative Accuracy Profile for the variable “Liabilities / Assets” for the default criterion bankruptcy. The dark curved line shows the performance of the (univariate) model being evaluated in depicting the percentage of defaults captured by the model at different percentages of the data set, while the thin straight line below represents the naive case of zero information or random assignment of default probabilities. The other thin line represents the case of perfect information, where all defaults are assigned the highest default probabilities. The visualized information of the model performance of the Cumulative Accuracy Profile can also be summarized in a single number called Accuracy Ratio. It is the ratio of the performance improvement of the model being evaluated over the naive model to the performance improvement of the perfect over the naive model. For the variable “Liabilities / Assets” the Accuracy Ratio is 44.174%. The Accuracy Ratios for all variables and all data sets are listed in Table 4.

Derivation of the Default Prediction Models

After having calculated the Accuracy Ratios for all candidate input ratios, one possibility to proceed would be to determine the best variable of each of the ten categories leverage, debt coverage, liquidity, activity, productivity, turnover, profitability, firm size, growth rates and leverage development and to combined them to form the basic model for further analysis. However, if one has a look at the correlation between the accounting ratios of one group (as depicted in Appendix B), one can see that for some categories not all variables are highly correlated but that there exist correlation sub-groups. This implies that one probably would run the risk of ignoring important variables if only the accounting ratio with the highest Accuracy Ratio from each category were included in the model building process. Instead, the best variable from each correlation sub-group was selected as long as its Accuracy Ratio was larger than 5%.

Next, backward selection methods were applied to check whether all chosen accounting ratios added statistical significance to the group or whether the logit model could be reduced to a lower number of input variables. Backward elimination is one possible method of statistical stepwise variable selection procedures. It begins by estimating the full model and then eliminates the worst covariates one by one until all remaining input variables are necessary, i.e. their significance level is below the chosen critical level. For this study the analysis was based on the precise likelihood-ratio test, where the significance level was set at 10%.

Of course, there exist some critical voices arguing that statistical stepwise selection procedures would be data mining, as they can yield theoretically implausible models and select irrelevant, or noise, variables. However, as for example Hendry and Doornik (1994) and Hosmer and Lemeshow (1989) respond, stepwise selection procedures are simply necessary if dealing with large sets of possible input factors as in the case at hand. Notice that from a starting list of only 20 variables more than 1 million possible models could be created. Hence, even if the stepwise procedures were indeed data mining, those explicit strategies would be much more preferable than any try-and-error strategies in detecting powerful models. Besides, it is the responsibility of the researcher to take care that the reported final model does not include variables that behave in a counterintuitive way to theory.

Having justified the application of statistical stepwise variable selection procedures, why is the backward elimination procedure not used right from the start? The reason is, that - although there exist some correlation sub-groups - many potential input factors are highly correlated. If all 65 accounting ratios were included into one model to apply backward regression, because of this high correlation the resulting model would probably be of poor quality, with unstable parameter estimates and poor performance of the model when applied outside of the development sample. Furthermore, the weights assigned to the remaining factors can often be counterintuitive in such a case, e.g. it might be possible to have a model in which higher profitability led to higher default rates. Therefore one has to use the procedure described above to select only specific ratios to be included into the backward selection analysis and to exclude those factors that are highly correlated.

However, in addition to the selected accounting ratios also three other factors are included into the model building process, i.e. the size and the legal form of the companies as well as the sector in which they are operating. It is tested whether these variables have any predictive power on their own and whether they interact with the chosen accounting ratios.

Model Validation

Finally, after the completion of the default risk prediction modeling process and the exertion of goodness-of-fit tests, the estimated models are applied to the validation samples to produce out-of-sample and out-of-time forecasts. Then the quality of the forecasts is evaluated with the concepts of Cumulative Accuracy Profiles (CAP) and Accuracy Ratios (AR) described above.

Although the Cumulative Accuracy Profiles is the most popular validation technique currently used in practice, it used to suffer from the major weakness that no confidence intervals could be (analytically) calculated for its summary statistic and hence no rigorous statistical test was available to decide upon the superiority of two competing rating models. However, the Receiver Operating Characteristic Curve (ROC)⁴, a concept similar to the CAP curve and very popular in Medicine, offers these statistical properties. Besides, as proven in Engelmann, Hayden, and Tasche (2003), the Accuracy Ratio is just a linear transformation of the area below the ROC curve. Hence, both concepts contain the same information and all properties of the area under the ROC curve are also applicable to the AR. This implies that confidence intervals for the Accuracy Ratio can be derived by calculating the unbiased estimator $\sigma_{\hat{AR}}^2$ of the variance of $\hat{AR}$ as:

$$\sigma_{\hat{AR}}^2 = \frac{1}{(N_D - 1)(N_{ND} - 1)} [1 + (N_D - 1)\hat{P}_{D,D,ND} + (N_{ND} - 1)\hat{P}_{ND,ND,D} - (N_D + N_{ND} - 1)(\hat{AR})^2], \quad (5)$$

where N_D and N_{ND} are the numbers of the observed defaulters and non-defaulters and $\hat{P}_{D,D,ND}$ and $\hat{P}_{ND,ND,D}$ are estimations of

$$\begin{aligned} P_{D,D,ND} &= P(S_{D,1}, S_{D,2} < S_{ND}) + P(S_{ND} < S_{D,1}, S_{D,2}) \\ &\quad - P(S_{D,1} < S_{ND} < S_{D,2}) - P(S_{D,2} < S_{ND} < S_{D,1}), \\ P_{ND,ND,D} &= P(S_{ND,1}, S_{ND,2} < S_D) + P(S_D < S_{ND,1}, S_{ND,2}) \\ &\quad - P(S_{ND,1} < S_D < S_{ND,2}) - P(S_{ND,2} < S_D < S_{ND,1}). \end{aligned}$$

Here the quantities $S_{D,1}, S_{D,2}$ are independent observations randomly sampled from the distribution of the defaulters, while $S_{ND,1}, S_{ND,2}$ are randomly sampled from the distribution of the non-

⁴ Assume someone has to decide from the rating scores of the debtors which debtors will survive during the next period and which debtors will default. One possibility for the decision-maker would be to introduce a cut-off value (C) and to classify each debtor with a rating score higher than C as a potential defaulter and each debtor with a rating score lower than C as a non-defaulter. Then four decision results are possible. If the rating score is above the cut-off value and the debtor defaults subsequently the decision was correct. Otherwise the decision-maker wrongly classified a non-defaulter as a defaulter. If the rating score is below C and the debtor does not default the classification was correct. Otherwise a defaulter was incorrectly put into the non-defaulters group. The ROC Curve is a plot of the percentage of the defaulters predicted correctly as defaulters (Hit-Rate) vs. the percentage of the non-defaulters wrongly classified as defaulters using all cut-off values that are contained in the range of the rating scores. In contrast to this the CAP is a plot of the Hit-Rate vs. the percentage of all debtors classified as defaulters for all possible cut-off values.

defaulters, and the respective observed rating scores are used to estimate $P_{D,D,ND}$ and $P_{ND,ND,D}$. Besides, it is known that for $N_D, N_{ND} \rightarrow \infty$ $(AR - \hat{AR})/\sigma_{\hat{AR}}^2$ is asymptotically normally distributed with mean zero and standard deviation one, what allows the calculation of confidence intervals at confidence level α for $\hat{AR}$ using the relation

$$P(\hat{AR} - \sigma_{\hat{AR}}^2 \Phi^{-1}(\frac{1+\alpha}{2}) \leq A \leq \hat{AR} + \sigma_{\hat{AR}}^2 \Phi^{-1}(\frac{1+\alpha}{2})) \approx \alpha, \quad (6)$$

where Φ denotes the cumulative distribution function of the standard normal distribution.⁵

The availability of confidence interval now clearly represents an improvement, but nevertheless one still could do better when one wants to decide on the superiority of two competing rating models. Here the simple comparison of confidence intervals could be misleading, as a potential correlation of both models is neglected in this case. So to construct a rigorous test on the difference of the power of two rating models measured by AR_1 and AR_2 it is necessary to calculate the covariance $\hat{\sigma}_{\hat{A}_1 \hat{A}_2}^2$ between $\hat{A}_1$ and $\hat{A}_2$. It can be estimated by

$$\begin{aligned} \hat{\sigma}_{\hat{A}_1 \hat{A}_2}^2 &= \frac{1}{(N_D - 1)(N_{ND} - 1)} [\tilde{P}_{D,D,ND,ND}^{12} + (N_D - 1) \tilde{P}_{D,D,ND}^{12} \\ &+ (N_{ND} - 1) \tilde{P}_{ND,ND,D}^{12} - (N_D + N_{ND} - 1) (\tilde{AR}_1) (\tilde{AR}_2)] \end{aligned} \quad (7)$$

where $\tilde{P}_{D,D,ND,ND}^{12}$, $\tilde{P}_{D,D,ND}^{12}$ and $\tilde{P}_{ND,ND,D}^{12}$ are estimators for $P_{D,D,ND,ND}^{12}$, $P_{D,D,ND}^{12}$ and $P_{ND,ND,D}^{12}$ which are defined as

$$\begin{aligned} P_{D,D,ND,ND}^{12} &= P(S_D^1 > S_{ND}^1, S_D^2 > S_{ND}^2) + P(S_D^1 < S_{ND}^1, S_D^2 < S_{ND}^2) \\ &- P(S_D^1 > S_{ND}^1, S_D^2 < S_{ND}^2) - P(S_D^1 < S_{ND}^1, S_D^2 > S_{ND}^2), \end{aligned} \quad (8)$$

$$\begin{aligned} P_{D,D,ND}^{12} &= P(S_{D,1}^1 > S_{ND}^1, S_{D,2}^2 > S_{ND}^2) + P(S_{D,1}^1 < S_{ND}^1, S_{D,2}^2 < S_{ND}^2) \\ &- P(S_{D,1}^1 > S_{ND}^1, S_{D,2}^2 < S_{ND}^2) - P(S_{D,1}^1 < S_{ND}^1, S_{D,2}^2 > S_{ND}^2), \end{aligned} \quad (9)$$

$$\begin{aligned} P_{ND,ND,D}^{12} &= P(S_D^1 > S_{ND,1}^1, S_D^2 > S_{ND,2}^2) + P(S_D^1 < S_{ND,1}^1, S_D^2 < S_{ND,2}^2) \\ &- P(S_D^1 > S_{ND,1}^1, S_D^2 < S_{ND,2}^2) - P(S_D^1 < S_{ND,1}^1, S_D^2 > S_{ND,2}^2). \end{aligned} \quad (10)$$

Here the quantities S_D^i , $S_{D,1}^i$, and $S_{D,2}^i$ are independent draws from the sample of defaulters. The upper index i indicates whether the score of rating model 1 or the score of model 2 has to be taken. The meaning of S_{ND}^i , $S_{ND,1}^i$, and $S_{ND,2}^i$ is analogous.

To carry out the test on the difference between the two rating methods (where the null hypothesis is equality of both Accuracy Ratios), one has to evaluate the test statistic T which is defined as

$$T = \frac{(\hat{AR}_1 - \hat{AR}_2)^2}{\sigma_{\hat{AR}_1}^2 + \sigma_{\hat{AR}_2}^2 - 2\sigma_{\hat{AR}_1, \hat{AR}_2}^2}. \quad (11)$$

This test statistic T is asymptotically $\chi^2(1)$ -distributed with one degree of freedom. Given a confidence level α , we can calculate the critical values from the $\chi^2(1)$ -distribution for T .

⁵The number of defaults should be at least 50 for the above formula to be sufficiently reliable.

V. Three Rating Models for Austria

In this section the results of the analysis described in the former sections are displayed. In Table 4 one can see that all tested variables behave in an economically meaningful way as the empirical relationship to default corresponds to the theoretically expected dependence between the accounting ratios and the default probability. Besides, when looking at the in-sample Accuracy Ratios for the three data sets bankruptcy, rescheduling and delay-in-payment, it can be noticed that in general variables that are good default predictors for one default definition also are rather powerful when applied to the other two default criteria.

For reasons discussed in Section IV from each credit risk factor category and each correlation sub-group the best accounting ratio, i.e. the one with the highest Accuracy Ratio, was selected for further analysis as long as the Accuracy Ratio was above 5%. This implies that for the bankruptcy data set 19 out of the 65 variables entered into the backward selection process, while for the other two default definitions 21 ratios were chosen. The corresponding variables are again marked in Table 4.

In addition to the selected accounting ratios also three other factors were included into the model building process, i.e. the size and the legal form of the companies as well as the sector in which they were operating. Size and sector were structured as shown in Figure 3 and 4 in Section III, i.e. the variable size grouped the firms according to the sales classes of 5 to 20, 20 to 100, 100 to 500, 500 to 1000 and above 1000 million ATS, while the sector variable divided the companies into the sectors service, trade, manufacturing, construction and agriculture. The variable legal form distinguished between companies with limited and unlimited liabilities. All three factors entered into the analysis as dummy-variables. Besides, it was tested whether they interacted with the chosen accounting ratios.

In Table 5, Table 6 and Table 7 the final rating models for the three default definitions are summarized. Let us start with the bankruptcy model. From Table 5 one can see that the final model contains eight accounting ratios and the dummy variable legal form, i.e. the final model looks like:

$$\begin{aligned} \text{LogOdd} = & -7.41 + 2.66 * \text{Liabilities}/\text{Assts} \\ & + 1.31 * \text{BankDebt}/\text{Assets} \\ & + 1.49 * \text{CurrentLiabilities}/\text{Assets} \\ & - 1.52 * \text{Cash}/\text{CurrentLiabilities} \\ & + 2.84 * \text{AccountsPayable}/\text{NetAssets} \\ & - 0.09 * (\text{NetSales} - \text{MaterialCosts})/\text{PersonnelCosts} \\ & - 2.59 * \text{OrdinaryBusinessIncome}/\text{Assets} \\ & + 0.56 * \text{NetSales}/\text{NetSalesOneYearAgo}(\text{transformed}) \\ & + 1.19 * \text{LimitedLiability}. \end{aligned} \tag{12}$$

This means that in contrast to other studies the size of the company and the sector in which it is operating were not found to be significant predictors of default. However, firms with limited

Table 4
Univariate Regression Results

In this table the results of the analysis described in the former sections are depicted. In column three and four the expected dependence between the accounting ratios and the default probability is compared with the empirical relationship to default. As can be seen all variables behave in an economically meaningful way. In column five the functional form of the relationship between accounting ratio and log odd is listed, where l stands for linear, c for concave and u for an u-shaped empirical dependence. Besides, the univariate Accuracy Ratios for all data sets are shown. Notice that a powerful ratio for the default definition bankruptcy in general also does well in predicting default for the default criteria rescheduling and delay-in-payment. The variables that were selected for the backward selection procedure but that did not enter into the final model are marked with a star, while the “winning” accounting ratios are labeled with an exclamation mark.

Ratio	Factor	Hyp.	Test	Form	AR Bankr.	AR Resched.	AR Delay
1	Leverage	+	+	l	45.56 % !	39.63 %	41.50 %
2	Leverage	-	-	l	44.72 %	44.95 % !	42.62 % !
3	Leverage	-	-	l	44.84 %	44.59 %	40.76 %
4	Leverage	+	+	c	27.08 % *	18.17 % *	15.91 % *
5	Leverage	+	+	l	2.23 %	6.43 % *	17.73 % *
6	Leverage	+	+	l	36.90 % !	39.86 % !	26.41 % !
7	Leverage	+	+	l	23.70 %	28.28 %	13.65 %
8	Leverage	+	+	c	34.91 %	39.64 %	26.22 %
9	Leverage	+	+	c	23.71 %	28.33 %	13.64 %
10	Leverage	+	+	c	28.13 %	33.52 %	19.57 %
11	Debt Coverage	-	-	l	22.79 % *	31.08 % *	40.73 % *
12	Debt Coverage	-	-	l	40.93 % *	36.44 % *	40.74 % *
13	Liquidity	-	-	l	29.86 %	28.91 %	22.42 %
14	Liquidity	-	-	l	22.01 %	28.10 %	25.60 %
15	Liquidity	-	-	l	30.58 % *	29.58 % *	26.31 % *
16	Liquidity	+	+	l	28.38 % !	19.01 % !	11.30 % !
17	Liquidity	-	-	l	-9.50 %	16.89 %	12.40 %
18	Liquidity	-	-	l	22.62 %	22.32 %	18.68 %
19	Liquidity	-/+	-	l	29.69 %	28.64 %	22.59 %
20	Liquidity	-/+	-	l	12.51 %	13.73 %	11.81 %
21	Liquidity	-/+	+	c	18.39 % *	9.41 % *	6.57 % *
22	Liquidity	-/+	+	c	7.44 %	1.12 %	1.35 %
23	Liquidity	+	+	l	1.88 %	17.56 %	-6.19 %
24	Liquidity	-	-	l	25.45 % !	23.89 % *	19.14 % *
25	Liquidity	-	-	l	28.05 %	28.91 %	22.38 %
26	Liquidity	-	-	l	26.62 %	27.91 %	20.56 %
27	Activity	+	+	l	16.31 % *	10.62 % *	8.73 % *
28	Activity	+	+	l	15.35 %	10.20 %	8.37 %
29	Activity	+	+	l	12.18 %	9.66 %	7.53 %
30	Activity	+	+	l	13.82 % *	4.66 %	3.95 %
31	Activity	+	+	l	11.87 %	4.03 %	3.84 %
32	Activity	-	-	l	9.28 %	14.62 %	13.39 %
33	Activity	-	-	l	12.56 %	17.57 % *	16.03 % *

Table 4 continued
Univariate Regression Results

In this table the results of the analysis described in the former sections are depicted. In column three and four the expected dependence between the accounting ratios and the default probability is compared with the empirical relationship to default. As can be seen all variables behave in an economically meaningful way. In column five the functional form of the relationship between accounting ratio and log odd is listed, where l stands for linear, c for concave and u for an u-shaped empirical dependence. Besides, the univariate Accuracy Ratios for all data sets are shown. Notice that a powerful ratio for the default definition bankruptcy in general also does well in predicting default for the default criteria rescheduling and delay-in-payment. The variables that were selected for the backward selection procedure but that did not enter into the final model are marked with a star, while the “winning” accounting ratios are labeled with an exclamation mark.

Ratio	Factor	Hyp.	Test	Form	AR Bankr.		AR Resched.		AR Delay	
34	Activity	+	+	l	4.49 %		0.16 %		3.39 %	
35	Activity	+	+	l	29.71 %		30.85 %		18.58 %	
36	Activity	+	+	l	34.34 %	!	33.89 %	!	26.23 %	!
37	Activity	+	+	l	-5.30 %		7.33 %	*	5.07 %	*
38	Productivity	+	+	l	8.23 %		5.94 %		2.84 %	
39	Productivity	-	-	l	6.33 %		3.98 %		2.14 %	
40	Productivity	-	-	l	17.48 %	!	8.06 %	*	9.38 %	*
41	Productivity	+	+	l	9.86 %		-2.51 %		6.65 %	
42	Turnover	-	-	l	22.50 %		21.53 %		16.49 %	
43	Turnover	-	-	l	27.38 %	*	24.42 %	*	19.34 %	*
44	Turnover	-	-	l	24.49 %		21.06 %		16.20 %	
45	Profitability	-	-	l	33.05 %		29.09 %		36.54 %	
46	Profitability	-	-	l	32.97 %		29.20 %		36.87 %	
47	Profitability	-	-	l	28.52 %		25.13 %		30.47 %	
48	Profitability	-	-	l	29.51 %		27.85 %		29.82 %	
49	Profitability	-	-	l	32.85 %		31.68 %		36.98 %	
50	Profitability	-	-	l	42.50 %	!	39.06 %		41.75 %	!
51	Profitability	-	-	l	39.73 %		34.75 %		39.64 %	
52	Profitability	-	-	l	42.08 %		35.39 %		38.96 %	
53	Profitability	-	-	l	40.10 %		39.33 %		41.49 %	
54	Profitability	-	-	l	41.66 %		41.86 %	!	38.60 %	
55	Profitability	-	-	l	39.88 %		41.40 %		38.34 %	
56	Profitability	-	-	l	38.80 %		40.68 %		37.75 %	
57	Profitability	-	-	l	38.95 %		40.75 %		37.91 %	
58	Profitability	-	-	l	39.92 %		40.31 %		36.63 %	
59	Size	-	-	l	2.21 %		5.43 %		-4.78 %	
60	Size	-	-	l	-5.16 %		3.96 %		-6.78 %	
61	Size	-	-	l	11.18 %	*	15.93 %	*	5.81 %	*
62	Growth Rates	-/+	-/+	u	18.61 %	!	14.54 %		12.87 %	
63	Growth Rates	-/+	-/+	u	18.10 %		14.86 %	*	15.97 %	*
64	Leverage Change	+	+	c	17.16 %	*	13.32 %	*	16.58 %	*
65	Leverage Change	+	+	c	13.41 %	*	16.95 %	*	12.59 %	*

liability seem to have a higher default probability than companies with unlimited liability, which is just what one would expect. Besides, these findings are stable in the sense that they are valid for the other two default definitions, too.

Table 5
The rating model for the bankruptcy data set

This table shows the final model for the bankruptcy default definition. As can be seen from the Hosmer-Lemeshow-Test, the model fits quite well to the in-sample data. Besides, the out-of-sample Accuracy Ratio is almost as high as the in-sample Accuracy Ratio, which implies that the model does a good job when predicting default out-of-sample and out-of-time.

Log likelihood = -1800.075	Prob > chi2	=	0.000
	Pseudo R2	=	0.136

BANKRUPTCY	Coef.	Std. Err.	z	P> z	95% Conf. Int.	
Liabilities / Assets (ar1)	2.658	0.792	3.35	0.001	1.105	4.212
Bank Debt / Assets (ar6)	1.314	0.244	5.38	0.000	0.835	1.792
Current Liab. / Assets (ar16)	1.486	0.288	5.15	0.000	0.920	2.052
Cash / Current Liab. (ar24)	-1.517	0.683	-2.22	0.026	-2.856	-0.177
Acc. Payable / Mat.Costs (ar36)	2.841	0.418	6.79	0.000	2.021	3.661
(Sales-Mat.C.) / Pers.C. (ar40)	-0.092	0.037	-2.45	0.014	-0.166	-0.018
Ord.Bus.Income / Assets (ar50)	-2.589	0.558	-4.64	0.000	-3.684	-1.494
Net Sales / Last N.Sales (ar62)	0.564	0.139	4.06	0.000	0.292	0.837
Legal Form	1.187	0.227	5.22	0.000	0.741	1.633
Constant	-7.410	0.995	-7.45	0.000	-9.361	-5.460

Hosmer-Lemeshow-Test (for 20 groups): Prob > chi2 = **83.22%**

In-Sample Accuracy Ratio: **60.640%**

Out-of-Sample Accuracy Ratio: **58.856%**

From Table 5 one can also learn that the in-sample fit of the model is quite good as shown by the Hosmer-Lemeshow goodness-of-fit test. This test works like the following: First the observations are grouped based on percentiles of the estimated default probabilities. For the study at hand 5% intervals were used i.e. 20 groups were formed. Now for every group the average estimated default probability is calculated and used to derive the expected number of defaults per group. Next this number is compared with the amount of realized defaults in the respective group. The Hosmer-Lemeshow goodness-of-fit statistic is then defined as:

$$\hat{G} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}, \quad (13)$$

where g denotes the total number of groups, n'_k is the number of observations in the k^{th} group,

$$o_k = \sum_{j=1}^{c_k} y_j$$

is the number of defaults among the c_k covariate patterns, i.e. the amount of defaults among the c_k observed, distinct values of the input parameter vector, where y_j denotes the number of positive responses, $y = 1$, per covariate pattern,

$$\bar{\Pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\Pi}_j}{n_k}$$

is the average estimated default probability, and c_k denotes the number of covariate patterns in the k^{th} percentile and m_j denotes the number of observations per covariate pattern. According to Hosmer and Lemenshow (1989) the distribution of the test statistic $\hat{C}$ is well approximated by a chi-square distribution with g-2 degrees of freedom, given that the fitted logistic regression model is the correct model. So in our case of 20 groups the appropriate chi-square distribution has 18 degrees of freedom. The corresponding p-value for the bankruptcy model can then be calculated as 83.22%, which indicates that the model seems to fit quite well.

However, the Hosmer-Lemenshow goodness-of-fit test can also be regarded from another point of view for the application at hand. Until now we only dealt with the development of a model that assigns each corporation a certain default probability or credit score, which leads towards a ranking between the contemplated firms. However, in practice banks usually want to use this ranking to map the companies to an internal rating scheme that typically is divided into about ten to twenty rating grades. The easiest way to do so would be to use the quantiles of the predicted

Table 6
The rating model for the rescheduling data set

This table shows the final model for the restructuring default definition. As can be seen from the Hosmer-Lemenshow-Test, the model fits quite well to the in-sample data. Besides, the out-of-sample Accuracy Ratio is almost as high as the in-sample Accuracy Ratio, which implies that the model does a good job when predicting default out-of-sample and out-of-time.

Log likelihood = -3392.943 Prob > chi2 = 0.000
Pseudo R2 = 0.137

RESCHEDULING	Coef.	Std. Err.	z	P> z 	95% Conf. Int.	
Equity / Assets (ar2)	-1.465	0.135	10.80	0.000	-1.731	-1.199
Bank Debt / Assets (ar6)	1.586	0.144	10.95	0.000	1.302	1.869
Current Liab. / Assets (ar16)	0.787	0.153	5.13	0.000	0.486	1.088
Acc.Payable / Mat.Costs (ar36)	2.732	0.262	10.42	0.000	2.218	3.246
Net Income / Assets (ar54)	-2.787	0.392	-7.10	0.000	-3.557	-2.017
Legal Form	0.703	0.139	5.05	0.000	0.430	0.976
Constant	-5.941	0.201	-29.50	0.000	-6.336	-5.546

Hosmer-Lemenshow-Test (for 20 groups): Prob > chi2 = **59.62%**

In-Sample Accuracy Ratio: **59.947%**

Out-of-Sample Accuracy Ratio: **58.583%**

Table 7
The rating model for the delay-in-payment data set

This table shows the final model for the 90-days-past-due default definition. As can be seen from the Hosmer-Lemenshow-Test, the model fits quite well to the in-sample data. Besides, the out-of-sample Accuracy Ratio is almost as high as the in-sample Accuracy Ratio, which implies that the model does a good job when used to predict default on an out-of-sample and out-of-time basis.

Log likelihood = -3132.075

Prob > chi2 = 0.000
Pseudo R2 = 0.092

DELAY	Coef.	Std. Err.	z	P> z	95% Conf. Int.
Equity / Assets (ar2)	-1.417	0.123	-11.47	0.000	-1.659 -1.175
Bank Debt / Assets (ar6)	0.651	0.137	4.76	0.000	0.383 0.920
Current Liab. / Assets (ar16)	0.710	0.149	4.76	0.000	0.417 1.002
Acc.Payable / Mat.Costs (ar36)	1.665	0.264	6.30	0.000	1.147 2.184
Ord.Bus.Income / Assets (ar50)	-2.083	0.298	-6.97	0.000	-2.668 -1.497
Legal Form	0.305	0.128	2.37	0.018	0.052 0.557
Constant	-3.594	0.190	-18.83	0.000	-3.969 -3.220

Hosmer-Lemenshow-Test (for 20 groups): Prob > chi2 = **77.64%**

In-Sample Accuracy Ratio: **50.315%**

Out-of-Sample Accuracy Ratio: **48.806%**

default probabilities to build groups. If for example 20 rating classes shall be formed, then from all observations the 5% with the smallest default probabilities would be assigned the best rating grade, the next 5% the second and so on till the last 5% with the highest estimated default probabilities would enter into the worst rating class. The Hosmer-Lemenshow test now tells us that, given one would apply the concept described above to form rating classes, overall the average expected default probability per rating grade would fit with the observed default experience per rating class.

What's more, the in-sample Accuracy Ratio is about 60%, which is a reasonable number. Usually the rating models presented in the literature have an Accuracy Ratio between 40% and 70%. As Accuracy Ratios can only be compared reliably for models that are applied to the same data set, because differences in the data set, for example varying relative amounts of defaulters or non-equal data reliability, drive this measure heavily, and in light of the current data set being a mixture of many different data sources, an Accuracy Ratio of about 60% seems quite satisfactory. Furthermore, the out-of-sample Accuracy Ratio, i.e. the number one obtains if one applies the model estimated on the in-sample data to the out-of-sample data, is with about 59% almost as high as the in-sample Accuracy Rate.⁶ This means that the derived model is stable and powerful in the sense that it produces accurate default predictions also on an out-of-sample and out-of-time basis. Again this is also true for the other default definitions.

⁶The only other existing study for the Austrian market is the one by Moody's (see Kocagil, Imming, Glormann, and Escott (2002)), where mainly bankruptcy is used as default definition. Here an out-of-sample Accuracy Ratio of 54.7% is reported.

Another interesting result is that the three models are very similar in that many input variables are the same for all three models. This can be interpreted as another hint that probably a model derived for one default definition can also be used to predict differently defined default events. To test this hypothesis, the model that was estimated using the default criterion bankruptcy, the default event most easily observable, is now applied to the data sets with the default definitions rescheduling and delay-in-payment. Table 8 summarizes the results.

Table 8
Applying the bankruptcy model to other default definitions

This table shows the out-of-sample Accuracy Ratios and their confidence intervals if the bankruptcy model is used to predict default events that have different default definitions. As can be seen one does not lose much prediction power in doing so. This is an interesting result as it suggests that banks could use their “old internal rating models under the new Basel capital accord even if these models were not developed on the detailed default definition provided there.

Default Definition	Model Applied	Accuracy Ratio in %	σ_{AR}	95% conf. interval
Bankruptcy	Bankruptcy	58.856	0.0121	[0.5648, 0.6123]
Restructuring	Restructuring	58.583	0.0196	[0.5474, 0.6243]
Restructuring	Bankruptcy	55.941	0.0212	[0.5179, 0.6010]
Delay-in-payment	Delay-in-payment	48.806	0.0234	[0.4422, 0.5339]
Delay-in-payment	Bankruptcy	48.129	0.0264	[0.4296, 0.5330]

Obviously one does not lose much predictive power if the bankruptcy model is applied to the finer default definitions, as the Accuracy Ratios are only reduced from 58.58% to 55.94% respectively from 48.81% to 48.13%. Besides, one should notice that the confidence intervals of the restructuring respectively the delay-in-payment model heavily overlap with the reported confidence intervals when the model estimated for the bankruptcy data set is applied to those finer default definitions. This again is strong evidence for the power of the bankruptcy model to predict other credit events, too. Finally, Table 9 reports the values of the test statistics and the corresponding p-values for the comparison of the bankruptcy model with the rescheduling respectively the delay-in-payment model. As can be seen, the null hypothesis that the rating model developed with the default criterion bankruptcy is equally powerful in predicting finer default events as the models derived on these criteria can not be rejected in both cases at the 99% confidence level. In view of Basel II these are interesting results as they suggest that banks could use their traditional internal rating models under the new Basel capital accord even if these models were not developed on the detailed default definition provided there.

Table 9
Results of the test for the difference of the power of the three rating models

This table reports the values of the test statistics and the corresponding p-values of the models that are compared. I find that the difference is not significant at a 1% confidence level.

Data Set Used	Models Compared	Test-Statistic	p-value
Restructuring	Bankruptcy vs. Restructuring	5.47	0.0193
Delay-in-payment	Bankruptcy vs. Delay-in-payment	0.14	0.2964

VI. Conclusion

In the second version of the proposed new capital adequacy framework the Basel Committee on Banking Supervision (2001) announced that an internal ratings-based approach could form the basis for setting capital charges for banks with respect to credit risk in the near future. Besides, the Basle Committee on Banking Supervision (2001a) defined default as any credit loss event associated with any obligation of the obligor, including distressed restructuring involving the forgiveness or postponement of principal, interest, or fees and delay in payment of the obligor of more than 90 days. According to the current proposal for the new capital accord banks will have to use this tight definition of default for estimating internal rating-based models. However, historically credit risk models were typically developed using the default criterion bankruptcy, as this information was relatively easily observable. For this reason it was one main purpose of this work to test whether the structure and the performance of credit scoring models is sensitive to the default definitions that are used to derive them.

To answer this question, credit risk rating models based on balance sheet information of Austrian firms using the default definitions of bankruptcy, loan restructuring and 90 days past due were estimated based on logistic regressions and compared. Univariate regressions showed that in general those accounting ratios that are good default predictors for one default definition are also rather powerful when applied to the other two default criteria. Besides, the final models are similar in their structure as many of the selected input variables are the same for all three models. Furthermore, I found that not much prediction power is lost if the model estimated for the bankruptcy default definition is applied to the credit events of restructuring and delay-in-payment instead of the models derived for those finer default definitions, as the corresponding Accuracy Ratios are only reduced from 58.58% to 55.94% respectively from 48.81% to 48.13%. Besides, also the calculation of confidence intervals and the statistical tests based on these confidence intervals asserted that the rating model building process seems to be insensitive to the definition of default. In view of Basel II this is a very interesting result as it means that 'old' rating models developed by banks by using only bankruptcy as default criterion are not automatically outdated but that they have the potential to compete with models derived for the tighter Basel II default definitions in predicting those more complex default events. This implies that banks maybe do not have to re-calibrate their existing rating models but could stick to their traditional ones by just adjusting the default probability upwards to reflect the fact that the Basel II default events occur more frequently than bankruptcies. This possibility would be especially valuable for small banks, as - due to their limited number of clients - they typically face severe problems when trying to collect enough data for being able to statistically reliably update their current rating models within a reasonable time period. Therefore it would be interesting to test whether also scoring models based on other than exclusively accounting information and derived with different methodologies (f.ex. neural networks) are also insensitive with respect to their underlying default definition.

References

- Altman, E., 1968, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* pp. 589–609.
- Altman, E., A. Agarwal, and F. Varetto, 1994, Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks, *Journal of Banking and Finance* 18.
- Altman, E., R. Haderman, and P. Narayanan, 1977, Zeta Analysis, a New Model for Identifying Bankruptcy Risk of Corporation, *Journal of Banking and Finance* pp. 29–54.
- Altman, E., and S. Saunders, 1997, Credit Risk Measurement: Developments over the Last Twenty Years, *Journal of Banking and Finance* 21, 1721–1742.
- Baetge, J., and A. Jerschensky, 1996, Beurteilung der wirtschaftlichen Lage von Unternehmen mit Hilfe von modernen Verfahren der Jahresabschlussanalyse, *Der Betrieb* pp. 1581–1592.
- Barniv, R., A. Agarwal, and R. Leach, 1997, Predicting the Outcome Following Bankruptcy Filing: A Three-State Classification Using Neural Networks, *International Journal of Intelligent Systems in Accounting, Finance and Management* 6.
- Basle Committee on Banking Supervision, 1999a, Credit Risk Modelling: Current Practices and Applications, Bank for International Settlements.
- Basle Committee on Banking Supervision, 1999b, A New Capital Adequacy Framework, Bank for International Settlements.
- Basle Committee on Banking Supervision, 1999c, Principles for the Management of Credit Risk, Bank for International Settlements.
- Basle Committee on Banking Supervision, 2001a, The Internal Ratings-Based Approach, Bank for International Settlements.
- Basle Committee on Banking Supervision, 2001b, The New Basle Capital Accord, Bank for International Settlements.
- Black, F., and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 399–418.
- Blochwitz, S., T. Liebig, and M. Nyberg, 2000, Benchmarking Deutsche Bundesbank's Default Risk Model, the KMV Private Firm Model and Common Financial Ratios for German Corporations, KMV.
- Charitou, A., and C. Charalambous, 1996, The Prediction of Earnings Using Financial Statement Information: Empirical Evidence with Logit Models and Artificial Neural Networks, *Journal of Intelligent Systems in Accounting, Finance and Management* 5.
- Chen, K. H., and T. A. Shimerda, 1981, An Empirical Analysis of Useful Financial Ratios, *Financial Management* pp. 51–60.
- Coenenberg, A., 1993, *Jahresabschluß und Jahresabschlußanalyse*. (Moderne Industrie AG Buchverlag) 14 edn.
- Crosbie, P., 1999, Modelling Default Risk, KMV.

- Crouhy, M., D. Galai, and R. Mark, 2000, A comparative analysis of current credit risk models, *Journal of Banking and Finance* 24, 59–117.
- Crouhy, M., D. Galai, and R. Mark, 2001, Prototype Risk Rating System, *Journal of Banking and Finance* 25, 47–95.
- Deakin, E., 1972, A Discriminant Analysis of Predictors of Business Failure, *Journal of Accounting Research* pp. 167–179.
- Dimitras, A., S. Zanakis, and C. Zopoundis, 1996, A Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications, *European Journal of Operational Research* 90, 487–513.
- Edmister, R., 1972, An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction, *Journal of Financial and Quantitative Analysis* pp. 1477–1493.
- Engelmann, B., E. Hayden, and D. Tasche, 2003, Testing Rating Accuracy, *Risk* pp. 82–86.
- Falkenstein, E., A. Boral, and L. Carty, 2000, RiskCalc Private Model: Moody's Default Model for Private Firms, Moody's Investors Service.
- Gupson, G., C. Finger, and M. Bhatia, 1997, CreditMetrics - Technical Document, JP Morgan.
- Hendry, D., and J. Doornik, 1994, Modelling Linear Dynamic Econometric Systems, *Scottish Journal of Political Economy* pp. 1–33.
- Hodrick, R., and C. Prescott, 1997, Post-War U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking* 29, 1–16.
- Hosmer, D., and S. Lemenshow, 1989, *Applied Logistic Regression*. (John Wiley & Sons).
- Jones, D., and J. Mingo, 1998, Industry Practices in Credit Risk Modeling and Internal Capital Allocations: Implications for a Models-Based Regulatory Capital Standard, Working Paper, Federal Reserve Board.
- Kahya, E., and P. Theodossiou, 1999, Predicting Corporate Financial Distress: A Time-Series CUSUM Methodology, *Review of Quantitative Finance and Accounting* 13, 323–345.
- Kealhofer, S., 1993a, Credit Monitor Overview, KMV.
- Kealhofer, S., 1993b, Portfolio Management of Default Risk, KMV.
- Keenan, S., and J. Sobehart, 1999, Performance Measures for Credit Risk Models, Moody's Risk Management Services.
- Keenan, S., and J. Sobehart, 2000, A Credit Risk Catwalk, *Risk* pp. 84–88.
- Khandani, B., M. Lozano, and L. Carty, 2001, Moody's RiskCalc for Private Companies: The German Model, Moody's Investors Service.
- Kocagil, A., R. Imming, F. Glormann, and P. Escott, 2002, Moody's RiskCalc For Private Companies: Austria, Moody's Investors Service.
- Lettmayr, C., 2001, Bilanzkennzahlen - Handbuch für Praktiker, Österreichisches Institut für Gewerbe- und Handelsforschung.

- Lo, A. W., 1986, Logit Versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies, *Journal of Econometrics* 31, 151–178.
- Mátyás, L., and P. Sevestre, 1996, *The Econometrics of Panel Data*. (Kluwer).
- Merton, R., 1974, On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance* 28, 449–470.
- Nickell, P., W. Perraudin, and S. Varotto, 2000, Stability of Rating Transitions, *Journal of Banking and Finance*.
- Nyberg, M., M. Sellers, and J. Zhang, 2001, Private Firm Model - Introduction to the Modeling Methodology, KMV.
- Sobehart, J., and S. Keenan, 2001, Measuring Default Accurately, *Credit Risk Special Report*, Risk 14, 31–33.
- Sobehart, J., S. Keenan, and R. Stein, 2000a, Benchmarking Quantitative Default Risk Models: A Validation Methodology, Moody's Investors Service.
- Sobehart, J., S. Keenan, and R. Stein, 2000b, Validation Methodologies for Default Risk Models, *Risk* pp. 51–56.
- Tamari, M., 1966, Financial Ratios as a Means of Forecasting Bankruptcy, *Management International Review* pp. 15–21.

A. Appendix: Data Set Descriptions

In Appendix A the composition of the data for the rescheduling and the delay-in-payment data set is depicted. The figures show similar patterns as the figures for the bankruptcy data in Section III.

A.1. The Data Set with Loan-Restructuring as Default Criterion

Table 10
Number of observations and defaults per year for the rescheduling data set

This table shows the total number of the observed balance sheets and defaults per year. The last column displays the yearly default frequency according to the rescheduling data set.

year	observations	in %	defaults	in %	default ratio in %
1992	1,400	2.91	15	1.03	1.07
1993	2,469	5.13	33	2.26	1.34
1994	4,534	9.42	85	5.83	1.87
1995	6,655	13.83	200	13.71	3.01
1996	8,321	17.29	294	20.15	3.53
1997	9,053	18.82	304	20.84	3.36
1998	8,885	18.47	329	22.55	3.70
1999	6,798	14.13	199	13.64	2.93
Total	48,115	100.00	1,459	100.00	3.03

Figure 11. Borrower Counts by Number of Observed Yearly Observations

This figure shows the number of borrowers that have either one or multiple financial statement observations for different lengths of time. Multiple observations are important for the evaluation of the extent to which trends in financial ratios help predict defaults.

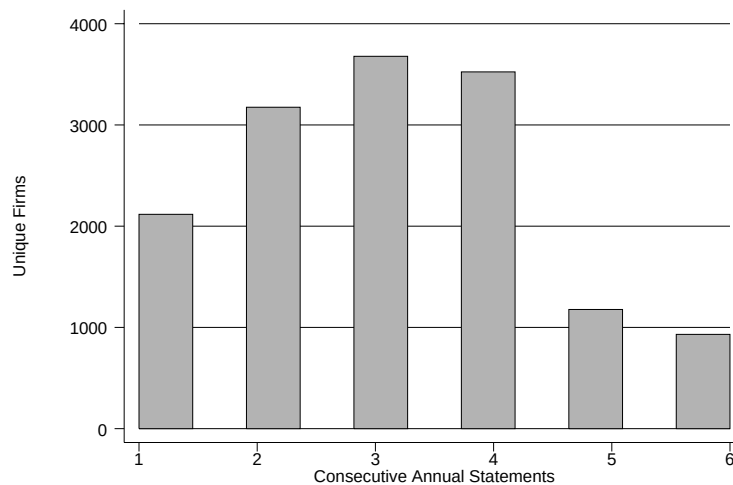


Figure 12. Distribution of Financial Statements by Legal Form

This figure displays the distribution of the legal form. The test sample differs slightly from the estimation sample as its percentage of limited liability companies is a few percentages higher.

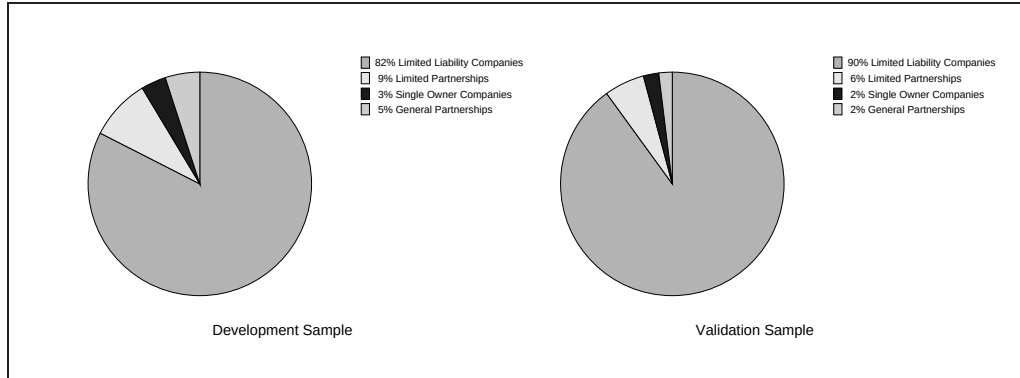


Figure 13. Distribution of Financial Statements by Sales Class

This graph shows the distribution of the accounting statements grouped according to sales classes for the observations in the estimation and the test sample. Differences between those two samples according to this criterion are only marginal.

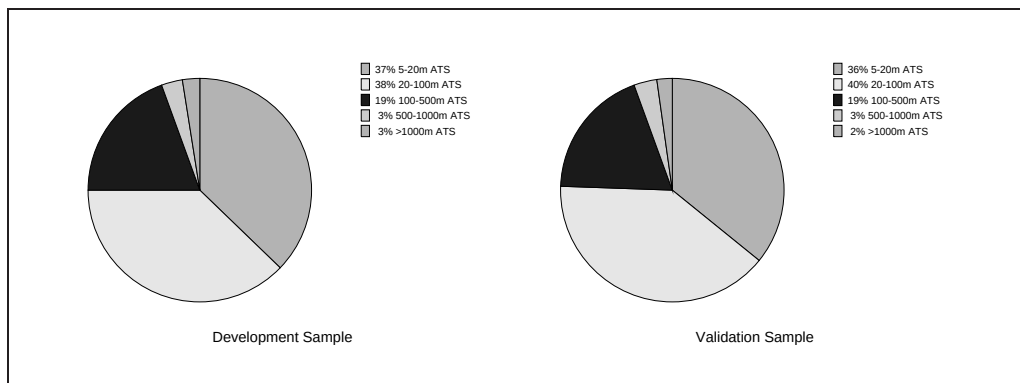
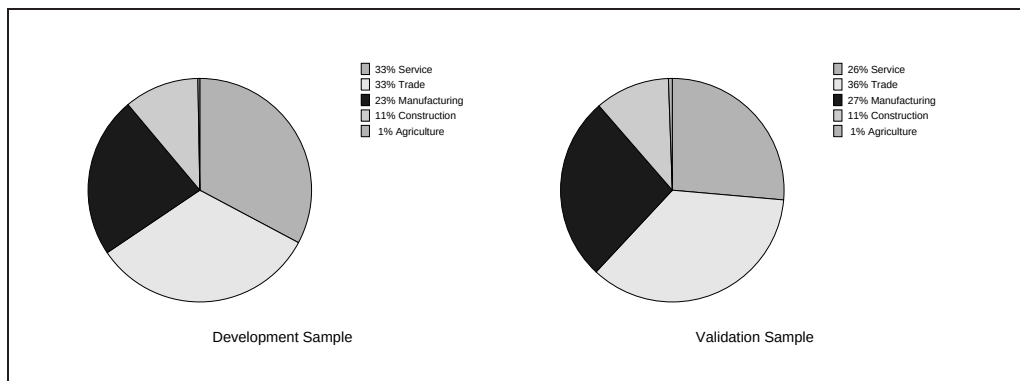


Figure 14. Distribution of Financial Statements by Industry Segments

This figure shows that the distribution of firms by industry differs between the development and the validation sample as there are less service companies in the test sample. This provides a further element of out-of-universe testing.



A.2. The Data Set with 90-Days-Past-Due as Default Criterion

Table 11
Number of observations and defaults per year for the delay-in-payment data set

This table shows the total number of the observed balance sheets and defaults per year. The last column displays the yearly default frequency according to the delay-in-payment data set.

year	observations	in %	defaults	in %	default ratio in %
1992	353	2.10	18	1.12	5.10
1993	776	4.62	46	2.87	5.93
1994	1,469	8.75	100	6.23	6.81
1995	2,605	15.51	268	16.71	10.29
1996	3,486	20.75	380	23.69	10.81
1997	3,459	20.59	359	22.38	10.90
1998	2,918	17.37	316	19.70	10.38
1999	1,731	10.31	117	7.29	6.76
Total	16,797	100.00	1,604	100.00	9.55

Figure 15. Borrower Counts by Number of Observed Yearly Observations

This figure shows the number of borrowers that have either one or multiple financial statement observations for different lengths of time. Multiple observations are important for the evaluation of the extent to which trends in financial ratios help predict defaults.

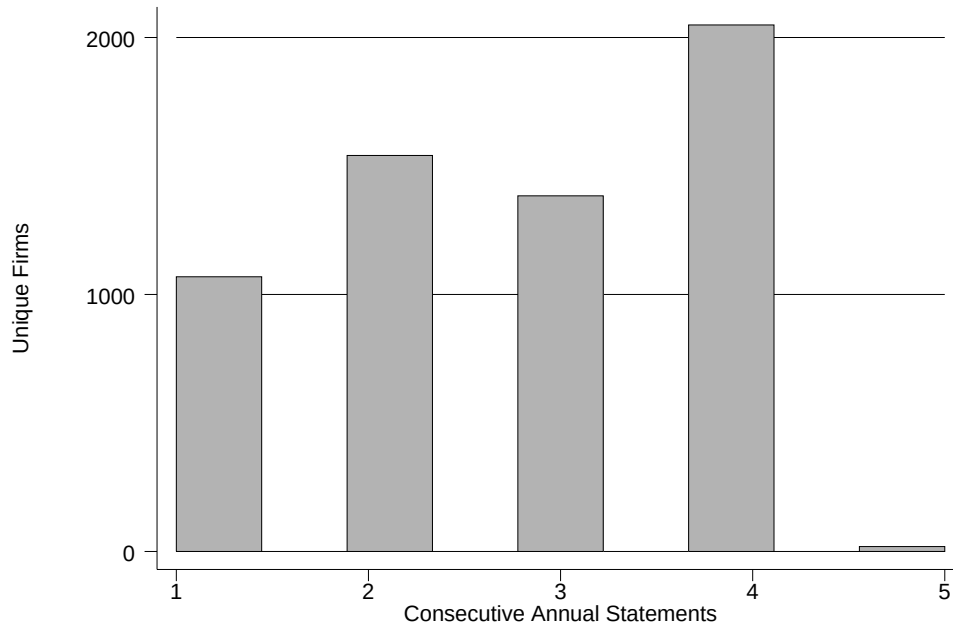


Figure 16. Distribution of Financial Statements by Legal Form

This figure displays the distribution of the legal form. The test sample is almost equal to the estimation sample.

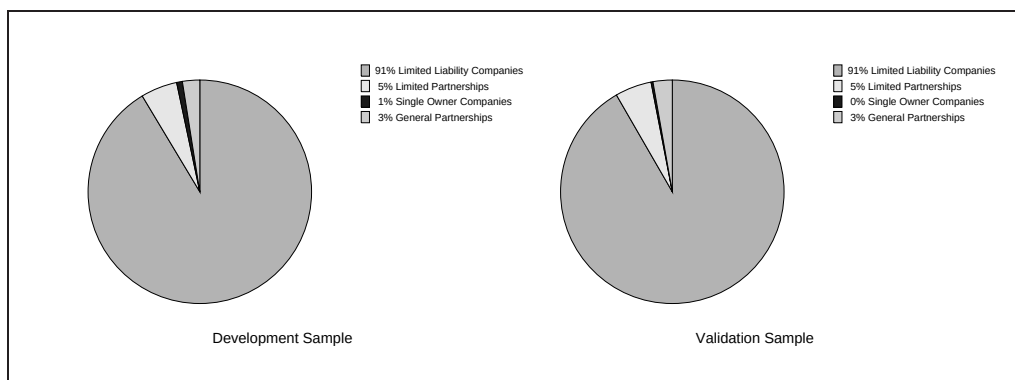


Figure 17. Distribution of Financial Statements by Sales Class

This graph shows the distribution of the accounting statements grouped according to sales classes for the observations in the estimation and the test sample. Differences between those two samples according to this criterion are only marginal.

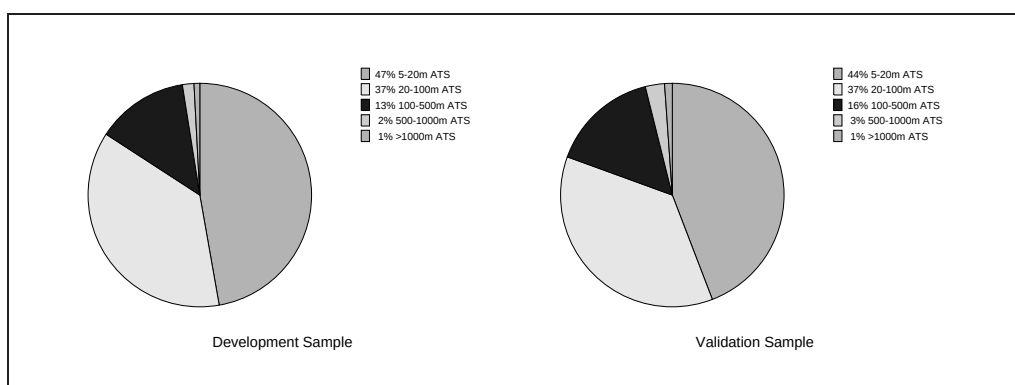
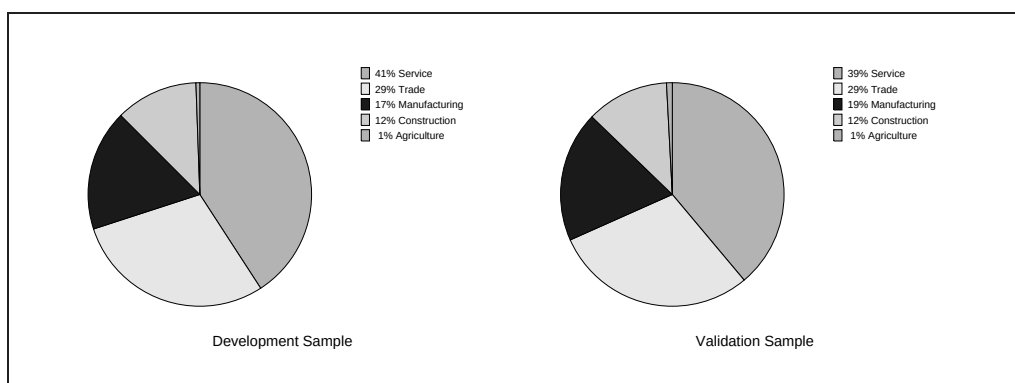


Figure 18. Distribution of Financial Statements by Industry Segments

This figure shows that for the delay-in-payments data set also the distribution of firms by industry is similar for the development and the validation sample.



B. Appendix: Correlations

B.1. Correlations between Accounting Ratios of the Same Credit Risk Factor Group

In Appendix B the pairwise correlations between the accounting ratios of the ten credit risk factor categories leverage, debt coverage, growth rates, leverage development, productivity, turnover, profitability, liquidity, activity and firm size are depicted. One can see that for some categories not all variables are highly correlated, but that there exist correlation sub-groups. Due to space limitations inter-category correlations are not listed, especially as they are in general rather small, just as was expected.

Table 12
Correlations for leverage accounting ratios

Leverage	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ar8	ar9	ar10
ar1	1.000									
ar2	-0.798	1.000								
ar3	-0.787	0.974	1.000							
ar4	0.190	-0.239	-0.239	1.000						
ar5	0.175	-0.159	-0.159	-0.179	1.000					
ar6	0.382	-0.343	-0.341	-0.045	0.351	1.000				
ar7	0.377	-0.345	-0.351	-0.048	0.347	0.985	1.000			
ar8	0.330	-0.320	-0.318	-0.032	0.358	0.913	0.902	1.000		
ar9	0.311	-0.308	-0.313	-0.036	0.357	0.882	0.888	0.984	1.000	
ar10	0.098	-0.114	-0.115	-0.102	0.318	0.943	0.928	0.843	0.815	1.000

Table 13
Correlations for debt coverage accounting ratios

Debt Coverage	ar11	ar12
ar11	1.000	
ar12	0.365	1.000

Table 14
Correlations for growth rate accounting ratios

Growth Rate	ar62	ar63
ar62	1.000	
ar63	0.878	1.000

Table 15
Correlations for leverage change accounting ratios

Leverage Change	ar64	ar65
ar64	1.000	
ar65	0.252	1.000

Table 16
Correlations for productivity accounting ratios

Productivity	ar38	ar39	ar40	ar41
ar38	1.0000			
ar39	-0.7042	1.000		
ar40	-0.5621	0.700	1.000	
ar41	-0.6605	0.505	0.001	1.000

Table 17
Correlations for turnover accounting ratios

Turnover	ar42	ar43	ar44
ar42	1.000		
ar43	0.969	1.000	
ar44	0.992	0.963	1.000

Table 18
Correlations for profitability accounting ratios

Profitability	ar45	ar46	ar47	ar48	ar49	ar50	ar51	ar52
ar45	1.000							
ar46	0.989	1.000						
ar47	0.755	0.738	1.000					
ar48	0.721	0.709	0.898	1.000				
ar49	0.956	0.950	0.722	0.774	1.000			
ar50	0.964	0.960	0.720	0.717	0.950	1.000		
ar51	0.891	0.907	0.652	0.696	0.920	0.905	1.000	
ar52	0.803	0.814	0.689	0.726	0.828	0.808	0.904	1.000
ar53	0.780	0.769	0.846	0.843	0.773	0.822	0.733	0.703
ar54	0.779	0.774	0.629	0.634	0.781	0.818	0.735	0.670
ar55	0.778	0.784	0.623	0.629	0.779	0.816	0.747	0.680
ar56	0.614	0.604	0.698	0.703	0.616	0.653	0.582	0.558
ar57	0.615	0.605	0.694	0.705	0.618	0.655	0.584	0.560
ar58	0.744	0.741	0.596	0.603	0.747	0.780	0.702	0.640

Profitability	ar21	ar22	ar23	ar24	ar25	ar26
ar53	1.000					
ar54	0.715	1.000				
ar55	0.707	0.992	1.000			
ar56	0.838	0.806	0.796	1.000		
ar57	0.841	0.808	0.799	0.995	1.000	
ar58	0.678	0.947	0.942	0.760	0.762	1.000

Table 19
Correlations for liquidity accounting ratios

Liquidity	ar13	ar14	ar15	ar16	ar17	ar18	ar19	ar20
ar13	1.000							
ar14	0.699	1.000						
ar15	0.886	0.742	1.000					
ar16	-0.503	0.019	-0.544	1.000				
ar17	0.538	0.859	0.630	0.308	1.000			
ar18	0.229	0.284	0.222	-0.034	0.220	1.000		
ar19	0.729	0.626	0.778	-0.317	0.588	0.188	1.000	
ar20	0.197	0.128	0.159	-0.145	0.046	0.730	0.049	1.000
ar21	0.281	0.246	0.259	-0.012	0.282	0.047	0.063	0.353
ar22	0.218	0.181	0.191	-0.031	0.188	0.125	0.004	0.420
ar23	-0.141	0.019	-0.180	0.246	0.023	-0.022	-0.135	-0.037
ar24	0.412	0.252	0.338	-0.315	0.091	0.846	0.268	0.750
ar25	0.999	0.707	0.892	-0.501	0.548	0.231	0.736	0.195
ar26	0.785	0.559	0.696	-0.416	0.405	0.334	0.576	0.297

Liquidity	ar21	ar22	ar23	ar24	ar25	ar26
ar21	1.000					
ar22	0.863	1.000				
ar23	-0.004	-0.010	1.000			
ar24	0.050	0.129	-0.084	1.000		
ar25	0.280	0.216	-0.142	0.410	1.000	
ar26	0.229	0.417	-0.117	0.490	0.785	1.000

Table 20
Correlations for activity accounting ratios

Activity	ar27	ar28	ar29	ar30	ar31	ar32	ar33	ar34
ar27	1.000							
ar28	0.996	1.000						
ar29	0.837	0.837	1.000					
ar30	0.030	0.034	0.037	1.000				
ar31	0.025	0.024	0.031	0.988	1.000			
ar32	-0.233	-0.235	-0.255	0.650	0.662	1.000		
ar33	-0.249	-0.250	-0.264	0.584	0.596	0.957	1.000	
ar34	-0.109	-0.109	0.150	0.612	0.619	0.347	0.314	1.0000
ar35	0.058	0.061	0.251	0.188	0.185	-0.112	-0.122	0.4341
ar36	0.251	0.259	0.183	0.267	0.246	-0.040	-0.059	0.0657
ar37	-0.402	-0.399	-0.335	0.318	0.320	0.340	0.335	0.4072

Activity	ar35	ar36	ar37
ar35	1.000		
ar36	0.699	1.000	
ar37	0.124	0.003	1.000

Table 21
Correlations for size accounting ratios

Size	ar59	ar60	ar61
ar59	1.000		
ar60	0.984	1.000	
ar61	0.833	0.833	1.000