



Dresden University of Technology
Faculty of Business Management and Economics

*Dresden Discussion Paper Series
in Economics*



**Verbesserung der Vergleichbarkeit von
Schätzgüteregebnissen von
Insolvenzprognosestudien**

MARTIN BEMMANN

Dresden Discussion Paper in Economics No. 08/05

ISSN 0945-4829

Address of the author:

Martin Bemann
Technische Universität Dresden
Lst. für Wirtschaftspolitik und -forschung
Münchner Platz 3
01069 Dresden
Deutschland

e-mail : martin.bemann@web.de

Editors:

Faculty of Business Management and Economics, Department of Economics

Internet:

An electronic version of the paper may be downloaded from the homepage:
<http://rcswww.urz.tu-dresden.de/wpeconomics/index.htm>

English papers are also available from the SSRN website:
<http://www.ssrn.com>

Working paper coordinator:

Dominik Maltritz
e-mail: wpeconomics@mailbox.tu-dresden.de

Verbesserung der Vergleichbarkeit von Schätzgüteregebnissen von Insolvenzprognosestudien

Martin Bemann
Technische Universität Dresden
Lst. für Wirtschaftspolitik und -forschung
01069 Dresden
martin.bemann@web.de

Abstract:

Ziel des vorliegenden Diskussionspapiers ist es, einen Beitrag zur Verbesserung der Vergleichbarkeit der Schätzgüteregebnisse von Insolvenzprognosestudien zu leisten. Hierzu werden zunächst in der Literatur verwendete kategoriale, ordinale und kardinale Schätzgütemaße vorgestellt. Es wird gezeigt, daß sich die vorgestellten ordinalen Schätzgütemaße für stichprobenübergreifende Schätzgütevergleiche am besten eignen, da sie nicht von den Stichprobenausfallraten beeinflusst werden. Es werden verschiedene Verfahren zur Überführung kategorialer in ordinale Schätzgütemaße entwickelt, die es ermöglichen, die Schätzgüteregebnisse älterer und aktueller Insolvenzprognosestudien auf einer einheitlichen Basis zu vergleichen. Im zweiten Teil des Diskussionspapier werden theorie- und empiriegestützt zahlreiche von der Qualität von Ratingverfahren unabhängige Einflußgrößen identifiziert, welche die betrachteten Schätzgütemaße beeinflussen und damit einen stichprobenübergreifenden Vergleich von Schätzgüteregebnissen erschweren. In diesem Zusammenhang werden Verfahren zur Abschätzung der Informationsverluste, die mit einer Diskretisierung kontinuierlicher Ratingskalen und mit der Vorselektion von Portfolien verbunden sind, entwickelt. Es werden die Schätzgüteregebnisse zahlreicher Insolvenzprognosestudien vorgestellt und mit drei Benchmarks verglichen. Erster Benchmark ist die allein auf Kenntnis von Rechtsform und Branchenzugehörigkeit eines Unternehmens erzielbare Schätzgüte von Insolvenzprognosen. Zweiter Benchmark ist die univariate Prognosegüte einzelner Finanzkennzahlen. Als dritter Benchmark wird das ALTMAN'sche Z-Scoreverfahren, ein multivariates Insolvenzprognosemodell, das in vielen empirischen Studien als Vergleichsverfahren verwendet wird, untersucht. Dabei zeigt sich, daß die mit dem Z-Score-Verfahren erzielbaren Schätzgüteregebnisse so schlecht sind, daß dieses Verfahren nicht als Benchmarkverfahren herangezogen werden sollte. Statt dessen wird die Verwendung der im Diskussionspapier zitierten, vollständig dokumentierten Ratingmodelle empfohlen, die von Modelltestern nachgebaut und auf beliebige Datensätze angewendet werden können. Angewendet auf die jeweiligen Unternehmenszielgruppen entspricht deren Qualität der Qualität kommerzieller Ratingsysteme, wie Auskunfteiscores für Kleinstunternehmen, finanzkennzahlenbasierte Ratingmodelle für KMU und Ratingurteile von Ratingagenturen für große Aktiengesellschaften.

JEL-Classification: G33, C14

Keywords: bankruptcy prediction, forecast validation, accuracy ratio, sample selection, rating granularity

Inhaltsverzeichnis

1	Einleitung	3
2	Schätzgütemaße zur Beurteilung von Insolvenzprognosen.....	6
2.1	Operationalisierung Schätzgüte.....	6
2.2	Schätzgütemaße für kategoriale Insolvenzprognosen	9
2.3	Schätzgütemaße für ordinale Insolvenzprognosen.....	12
2.3.1	Grafische Schätzgütebestimmung.....	17
2.3.2	Quantitative Schätzgütebestimmung	21
2.3.3	Vergleichbarkeit mit Gütemaßen kategorialer Insolvenzprognosen.....	26
2.3.4	Schätzgüteeinbußen durch Verwendung diskreter Ratingskalen.....	29
2.4	Schätzgütemaße für kardinale Insolvenzprognosen.....	32
3	Empirische Befunde zur Schätzgüte von Insolvenzprognosen	38
3.1	Grenzen der Aussagefähigkeit empirischer Vergleiche und Entwicklung von Benchmarks für Schätzgütemaße	38
3.2	Benchmark I: Erzielbare Schätzgüte bei Kenntnis von Rechtsform und Branchenzugehörigkeit eines Unternehmens	51
3.3	Benchmark II: Univariate Trennfähigkeit von Kennzahlen.....	61
3.4	Benchmark III: ALTMAN'S Z-Score	69
3.5	Schätzgütewerte realer Insolvenzprognoseverfahren.....	75
4	Fazit.....	85
	Literaturverzeichnis	86
	Anhang I: Ober- und Untergrenzen Accuracy Ratio.....	96
	Anhang II: Anreizkompatibilität verschiedener Scores.....	111
	Anhang III: Abhängigkeit verschiedener Schätzgütemaße von der durchschnittlichen Ausfallwahrscheinlichkeit.....	114
	Anhang IV: Quantifizierung von Informationsverlusten durch Diskretisierung von Ratingskalen	124
	Anhang V: Der Einfluß der Vorselektion von Portfolios auf die Schätzgüte von Insolvenzprognosen	139

1 Einleitung

Der Fähigkeit zur Erstellung trennscharfer Prognosen von Unternehmensinsolvenzen kommt sowohl aus individueller als auch aus volkswirtschaftlicher Sicht eine große Bedeutung zu. Auf individueller Ebene sind neben den Eigentümern, Mitarbeitern, Kunden, Lieferanten, Wirtschaftsprüfern und sonstigen Geschäftspartnern eines Unternehmens vor allem Banken an präzisen Insolvenzprognosen interessiert, da sie im Falle einer Insolvenz eines Schuldners mit sehr großen eigenen Schäden rechnen müssen.^{2,3} Aus Bankensicht sind trennscharfe Insolvenzprognosen eine wesentliche Voraussetzung dafür, um eine risikogerechte Preissetzung und Ausgestaltung nicht-finanzieller Kreditkonditionen (Limite, geforderte Sicherheiten) umsetzen zu können, Kreditprozesse kostengünstig zu gestalten (durch die Identifikation kritischer Fälle, die einer aufwendigeren Betreuung durch Kreditexperten bedürfen), die eigene Liquidität durch die Schaffung von Verbriefungsmöglichkeiten zu erhöhen und um den ökonomischen bzw. aufsichtsrechtlichen Eigenkapitalbedarf zu ermitteln und zu steuern.⁴

Aus volkswirtschaftlicher Sicht werden trennscharfe Insolvenzprognoseverfahren als eine wichtige Voraussetzung für die Stabilität des Bankensystems gesehen und als eine notwendige Voraussetzung dafür, um mit risikoadäquaten Kreditkonditionen Kreditnehmer zu einem anreizkompatiblen, risikobewußten Verhalten zu motivieren.^{5,6}

Mit der Beurteilung der *Schätzgüte* von Insolvenzprognosen soll ein wichtiges Qualitätsmerkmal von Insolvenzprognoseverfahren bewertet werden, das deshalb auch in der aktuellen Diskussion im Zuge von Basel II einen wichtigen Platz einnimmt.⁷

Weitere *ergebnisbezogene* Qualitätsdimensionen zur Beurteilung von Insolvenzprognoseverfahren, die in diesem Artikel jedoch nicht aufgegriffen werden, sind beispielsweise die zeitliche Stabilität der Prognosen⁸, der zusätzliche Informationsnutzen eines Insolvenzprognoseverfahrens in kombinierter Verwendung mit anderen Verfahren⁹, die Fähigkeit der Verfahren, Insolvenzen theoretisch („kausal“, „strukturell“) erklären zu können¹⁰ oder aber die Fähigkeit

² Siehe auch DIMITRAS, ZANAKIS, ZOPOUNIDIS (1996, S. 488) und BALCAEN, OOGHE (2004, S. 4) für eine Analyse der Interessenten an Insolvenzprognoseverfahren für Unternehmen.

³ FRANKS, SERVIGNY, DAVYDENKO (2004) identifizieren anhand einer empirischen Untersuchung zahlreiche Einflußgrößen auf die Erlösquoten von Bankkrediten bei Eintreten eines Zahlungsausfalls und finden, daß deutsche Banken im Durchschnitt nur rund 60% des ausstehenden Kredits, beispielsweise durch die Verwertung von Sicherheiten, erlösen können (ebenda, S. 4). Siehe hierzu auch BASLER AUSSCHUB (2000b, S. 27f), BASLER AUSSCHUB (2000c, S.7f), S&P (2003a, S.64, S.66) oder MOODY'S (2004a, S. 13). Für weitere empirische Ergebnisse und über 60 Literaturverweise zum Thema Erlösquoten siehe GUPTON und STEIN (2002).

⁴ Zur Relevanz und praktizierten Anwendungsmöglichkeiten von Insolvenzprognosen aus Bankensicht siehe beispielsweise ENGLISH, NELSON (1998, S. 11f.), TREACY, CAREY (2000/1998, S. 897), BASLER AUSSCHUB (2000b, S. 33) und ESCOTT, GLORMANN, KOCAGIL (2001, S.3).

⁵ siehe DEUTSCHE BUNDESBANK (2001, S. 15), BASLER AUSSCHUB (2004, §4), OENB (2004b, S. 33)

⁶ Überhöhte Kreditrisiken sind die häufigste Ursache der über 100 Bankinsolvenzen seit den 60er Jahren in Deutschland, siehe FISCHER (2004, S. 13) und die dort zitierte Literatur. Siehe auch <http://www.the-exit.net/plaza/bankinsolvenzen.de> (22.6.2005) für eine Übersicht aller Bankinsolvenzen in Deutschland seit 1945 (allerdings ohne Angabe der Insolvenzursachen).

⁷ siehe DEUTSCHE BUNDESBANK (2003), ÖSTERREICHISCHE NATIONALBANK (2004), BASLER AUSSCHUB (2005)

⁸ FONS (2002), CANTOR, MANN (2003), LÖFFLER (2003, 2004 a,b), ALTMAN, RIJEN (2004), HAMILTON (2004)

⁹ Siehe SHANNON, (2001/1948, S. 13f.), KEENAN, SOBEHART (1999, S.11f), SOBEHART, KEENAN, STEIN (2000, S. 15) und STEIN ET AL (2003, S. 38). Siehe auch LEHMANN (2003) und GRUNERT, NORDEN, WEBER (2005) für den statistisch signifikant positiven Informationsnutzen bei Einbeziehung von „Softfaktoren“ in das Ratingverfahren deutscher Banken.

¹⁰ Hierin liegt ein wichtiger Vorteil von vollständigen Strukturmodellen und induktiven Verfahren, beispielsweise Scoringmodellen oder Expertensystemen, im Vergleich zu statistischen Modellen. Für eine Übersicht der verschiedenen Verfahren siehe beispielsweise GÜNTHER, GRÜNING (2000, S. 41) oder ÖSTERREICHISCHE NATIONALBANK (2004, S. 32).

der Verfahren, Handlungsempfehlungen zur Beeinflussung der individuellen Insolvenzwahrscheinlichkeit abgeben zu können.^{11,12}

Wünschenswert wäre es, wenn sich die Schätzgüte des Insolvenzprognoseverfahrens direkt in finanziellen Kenngrößen, z.B. der erwarteten Rendite in Basispunkten je Euro Kreditvolumen, ausdrücken ließe.¹³ Dies hätte gegenüber den im folgenden vorgestellten Kenngrößen den Vorteil, daß sich ein derartiges *finanzielles* Schätzgütemaß nicht nur inhaltlich sehr gut interpretieren ließe, sondern sich auch ein direkter Bezug zur Nutzengröße des Entscheiders, beispielsweise der erwarteten Rendite einer Bank, herstellen ließe. Die Annahmen, die hierzu getroffen werden müssen, sind jedoch für den Praxiseinsatz prohibitiv.¹⁴

Die im folgenden vorgestellten Kenngrößen zur Beurteilung der Schätzgüte von Insolvenzprognosen stellen keine derartig hohen Ansprüche an die zur Verfügung stehenden Informationen; und sie lassen sich nur „heuristisch“ rechtfertigen, d.h. nur „*in der Regel*“ – aber leider nicht immer - wird ein Prognoseverfahren, das gemäß einer bestimmten Kennzahl besser ist, auch tatsächlich für einen bestimmten Entscheider einen größeren Nutzen spenden.

Andererseits sind die im folgenden Kapitel vorgestellten Kenngrößen intersubjektiv identisch, da sie nicht auf individuellen (und damit möglicherweise unterschiedlichen) Kosten-/Nutzenkalkulationen aufbauen. Des weiteren sind sie nicht nur zur Beurteilung der Schätzgüte von *Insolvenzprognosen* geeignet, sondern zur Beurteilung von Prognosen beliebiger anderer Anwendungsgebiete, vorausgesetzt es handelt bei den zu prognostizierenden Ereignissen um diskrete Ereignisse mit genau zwei möglichen Ausprägungen.¹⁵ Die vielseitige Anwendbarkeit der Methoden wird auch daraus ersichtlich, daß zentrale Kenngrößen und Begriffe für die Beurteilung der Schätzgüte von Insolvenzprognosen aus anderen Wissenschaftsgebieten übernommen wurden, die sich ebenfalls mit der Prognose unsicherer Ereignisse beschäftigen,

¹¹ Hierin liegt eine wesentliche Schwäche statistischer Verfahren. Sie messen zwar, welche Variablenkombinationen empirisch mit Ausfällen korrelieren, sie geben jedoch keine Kausalitäten wider, so daß sie möglicherweise durch ökonomisch unsinniges Verhalten manipulierbar sind. Dies ist ein Grund dafür, neben dem Schutz geistigen Eigentums (siehe beispielsweise BLÖCHLINGER, LEIPPOLD (2005/2004, S. 20) oder der Verhinderung „selbsterfüllender Prognosen“ (siehe beispielsweise KÜTING, WEBER (2004/1993, S. 350f.), warum Banken ihre Schätzverfahren geheim halten wollen.

Im Gegensatz dazu sind Erklärungen und die Vergabe von Handlungsempfehlungen auf Basis induktiver Verfahren (Scoringmodelle, Expertensysteme) prinzipiell möglich, wobei hier Grenzen darin zu sehen, daß sie nicht aus vollständigen und widerspruchsfreien Modellen abgeleitet werden.

Die Möglichkeit Insolvenzen theoretisch erklären zu können, ist eine der wichtigsten Vorteile von Strukturmodellen. Sie basieren auf vollständigen und widerspruchsfreien – und im Idealfall auch *realistischen* – Modellen. Im Prinzip ist es möglich, mit diesen Modellen quantitative Insolvenzprognosen abzugeben, ohne zuvor empirische Insolvenzdaten als Modellinput verwenden zu müssen. Die Eignung von Strukturmodellen Handlungsanweisungen geben zu können, ist abhängig davon, wie „tief“ die zugrundeliegenden Insolvenzzursachen modelliert werden. Basiert ein Strukturmodell beispielsweise lediglich auf dem Niveau und der Variabilität des Börsenwert eines Unternehmens, bzw. seiner Aktiva (siehe beispielsweise das KMV Public Firm Model in KEALHOFER (2003)), ohne daß die Bestimmungsgründe dieser Variablen dargelegt werden, kann es auch nicht zur Ableitung von Maßnahmen zur Beeinflussung der Insolvenzwahrscheinlichkeit herangezogen werden.

¹² Für weitere, überwiegend technische und organisatorische Anforderungen an Ratingverfahren siehe KRAHNEN, WEBER (2001), CROUHY, GALAI, MARK (2001) oder BASLER AUSSCHUß (2000c, 2004)

¹³ Dieser Ansatz wird in einem Simulationsmodell von JORDÁO und STEIN (2003) gewählt.

¹⁴ Siehe hierzu ebenfalls das Modell von JORDÁO und STEIN (2003). Es erfordert nicht nur die Kenntnis der ratingrelevanten Daten aller potentiellen Kunden. Ebenso müssen die Insolvenzprognosemodelle und Preispolitiken sämtlicher *Mitbewerber* (!) im Detail bekannt sein, um die eigenen monetären Konsequenzen bei Einsatz eines bestimmten Ratingmodells zu ermitteln.

¹⁵ Für zu erklärende Variablen mit mehr als zwei möglichen Ausprägungen stehen effizientere Gütemaße als die im folgenden vorgestellten zur Verfügung. Gleichwohl ist eine Anwendung der vorgestellten Gütemaße technisch möglich, wenn zuvor die zu erklärende Variable so transformiert wird, daß sie nur noch zwei unterschiedliche Werte annehmen kann (beispielsweise 1..für eine überdurchschnittliche Ausprägung und 0..für eine unterdurchschnittliche Ausprägung, siehe beispielsweise GUPTON, STEIN (2005, S. 26ff.))

beispielsweise aus der Signaltheorie¹⁶, der Meteorologie¹⁷ oder der Medizin¹⁸. Im folgenden Abschnitt werden drei Gruppen von Präzisionsmaße zur Beurteilung der Schätzgüte von Insolvenzprognosen vorgestellt. Innerhalb der Gruppen wird jeweils eine Auswahl von Kenngrößen vorgestellt, die zum Teil axiomatisch fundiert sind, sich gut interpretieren lassen oder aber Analogien zu anderen, in ihren ursprünglichen Anwendungsgebieten weithin akzeptierten Kenngrößen aufweisen.

Nach einer Erörterung der Möglichkeiten portfolioübergreifender Vergleiche von Schätzgütemaßen, werden im darauf folgenden Abschnitt empirische Schätzgüteregebnisse verschiedener Ratingverfahren gegenübergestellt. Die Angaben zu den Schätzgütern wurden entweder direkt aus externen Veröffentlichungen Angaben von Ratingagenturen, Auskunftsteien oder wissenschaftlichen Studien entnommen oder werden mittels den im Theorieteil des Diskussionspapier vorgestellten Methoden berechnet. Um neben einer relativen Einordnung der einzelnen Verfahren auch eine Vorstellung über die „absolute“ Qualität der Verfahren zu erhalten, werden drei verschiedene Verfahren herangezogen, mit deren Hilfe eine Untergrenze der gerade noch akzeptablen Qualität von Ratingverfahren gewonnen wird. Erstaunlicherweise können einige der in der Praxis verwendeten Verfahren selbst diese einfachen Hürden nur knapp überwinden, und andere Verfahren scheitern sogar daran.

¹⁶ siehe *Receiver-Operating-Characteristic (ROC)* zur Beschreibung des Fehlerverhaltens bei Datenübertragungen (siehe Abschnitt 2.3.1) oder *(Informations-)Entropie* (siehe Abschnitt 2.4)

¹⁷ siehe Beurteilung der Güte von Regenprognosen mittels des Brier-Scores in BRIER (1950), WINKLER (1994)

¹⁸ siehe SWETS (1973, 997ff.), SWETS (1988, S. 1287ff.) und SWETS, DAWES, MONAHAN (2000, S. 4ff.) für zahlreiche medizinische Anwendungen

2 Schätzgütemaße zur Beurteilung von Insolvenzprognosen

2.1 Operationalisierung Schätzgüte

Unter der *Schätzgüte* eines Insolvenzprognoseverfahrens soll im folgenden der *Grad der Übereinstimmung von Insolvenzprognosen und den tatsächlich eingetretenen Insolvenzen* verstanden werden.

Mathematisch faßbare Präzisierungen des Begriffs *Schätzgüte*, die im folgenden auch vorgestellt werden, müssen berücksichtigen, ob die zu beurteilenden Insolvenzprognosen *kategorialer, ordinaler* oder *kardinaler* Art sind.

- Als **kategoriale Insolvenzprognosen** werden Insolvenzprognosen bezeichnet, die lediglich zwei mögliche Ausprägungen zur Beurteilung der gerateten Unternehmen kennen: „Unternehmen A wird voraussichtlich (innerhalb des nächsten Jahres¹⁹) ausfallen“ vs. „Unternehmen B wird voraussichtlich (innerhalb des nächsten Jahres) *nicht* ausfallen“.
- Im Fall **ordinaler Insolvenzprognosen** werden Urteile über die *relativen Ausfallwahrscheinlichkeiten* der bewerteten Unternehmen abgegeben: „Unternehmen B fällt mit größerer Wahrscheinlichkeit aus als Unternehmen A aber mit geringerer Wahrscheinlichkeit als Unternehmen C“. Zwar könnten ordinale Insolvenzprognosen theoretisch beliebig differenziert sein, in der Praxis haben sich aber ordinale Ratingsysteme durchgesetzt, die ihre Ergebnisse auf eine diskreten, 7- oder 17-stufigen Skala²⁰ in einer von S&P übernommen Notation transformieren.^{21,22}

¹⁹ Im folgenden wird, wenn nicht anders angegeben, die Schätzgüte von Insolvenzprognosen bezogen auf einen Prognosezeitraum von *einem* Jahr untersucht. Wichtiger als die Entscheidung, *welcher* Prognosehorizont zugrundegelegt werden soll, ist die Entscheidung, daß beim Vergleich der Verfahren stets *der gleiche* Prognosehorizont verwendet wird. Da in der Mehrzahl der untersuchten Insolvenzstudien nur 1-Jahres-Ergebnisse veröffentlicht wurden, wurde als Vergleichszeitraum 1 Jahr gewählt. Es lassen sich jedoch, neben der Datenverfügbarkeit, auch inhaltliche Gründe für die Relevanz dieses Prognosezeitraums anführen:

- Ein Verfahren, das viele der auf Einjahressicht stattfindenden Ausfälle korrekt prognostiziert, erfaßt damit automatisch bereits auch viele der auf n-Jahressicht stattfindenden Ausfälle,
- Zumindest ein Teil der Ausfälle ist auf Defizite der Unternehmen zurückzuführen, die sich nicht kurzfristig beheben lassen, beispielsweise eine niedrige Eigenmittelausstattung oder die Abhängigkeit von wenigen Kunden. Unternehmen mit einer über-/unterdurchschnittlichen hohen Ausfallwahrscheinlichkeit innerhalb des nächsten Jahres haben damit tendenziell auch eine über-/unterdurchschnittliche hohe Ausfallwahrscheinlichkeit in den folgenden Jahren,
- Viele Ratingverfahren basieren, zum Teil sogar ausschließlich, auf den, nur einmal pro Jahr aktualisierten, Jahresabschlüssen der Unternehmen. Als Vergleichszeitraum sollte also *mindestens* ein Jahr, bzw. ein vielfaches eines Jahres gewählt werden,
- Gemäß den Regelungen von Basel II sind die Banken verpflichtet, Einjahresausfallwahrscheinlichkeiten zu prognostizieren (BASLER AUSSCHUB (2004, Tz. 285, 331)), sie müssen ihre Ratings einmal jährlich aktualisieren (ebenda Tz. 425) und ihr internes Ratingsystem (mindestens) einmal jährlich überprüfen und validieren (ebenda, insb. Tz. 443, 449).

Siehe auch BASLER AUSSCHUB (2001, S. 12) zu empirischen Befunden hinsichtlich der Zeithorizonte, die Banken ihren Ratingsystemen zugrundelegen: “The ‘time horizon’ over which a rating is expected to be valid (i.e. the forecast horizon of the rating) is mostly described by banks to be one year, [...]. The decision for a one-year horizon is mostly based on annual financial reporting cycles (bank and borrower), frequency of internal review of the rating, and in some cases the uncertainties of projected performance beyond one year.”

²⁰ 17-stufige Skala: 1=“AAA“, 2=“AA+“, 3=“AA“, 4=“AA-“, 5=“A+“, 6=“A“, 7=“A-“, 8=“BBB+“, 9=“BBB“, 10=“BBB-“, 11=“BB+“, 12=“BB“, 13=“BB-“, 14=“B+“, 15=“B“, 16 = „B-“, 17 = „CCC/C“, 7-stufige Skala: “AAA“, “AA“, “A“, „BBB“, „BB“, “B“, “C“.

²¹ In BASLER AUSSCHUB (2000c, S. 23f) werden die Ratingsymbole von 30 Ratingverfahren unterschiedlicher internationalen Ratingagenturen untersucht. 22 Agenturen verwenden Buchstabenkombinationen, um ihr Ratingurteil auszudrücken (ca. 75%), 6 Agenturen kommunizieren ihre Ratings in Form von Noten (20%), nur zwei Agenturen geben ihr Urteil in Form von Ausfallwahrscheinlichkeiten an. Von den 22 „Buchstaben-

- **Kardinale Insolvenzprognosen** ordnen jedem Unternehmen eine Ausfallwahrscheinlichkeit zu.

Die einzelnen Verfahren sind *abwärtskompatibel*: durch beliebige, schwach monotone Transformationen können aus Ausfallwahrscheinlichkeiten ordinal interpretierbare Scorewerte generiert werden, die sich bei Bedarf auch zu einer endlichen Anzahl von Klassen zusammenfassen lassen. Ebenso lassen sich durch Zusammenfassen von Scoreintervallen oder benachbarter Ratingklassen, solange bis nur noch zwei Klassen verbleiben, ordinale Insolvenzprognosen in kategoriale Insolvenzprognosen überführen.

Aber auch der umgekehrte Weg kann - unter Verwendung empirischer Ausfalldaten - beschritten werden. So können beispielsweise die historischen (realisierten) ratingklassenspezifischen *Ausfallquoten* als *Ausfallwahrscheinlichkeiten* kardinaler Insolvenzprognosen verwendet werden.^{23,24,25} Diese Vorgehensweise ist nicht unproblematisch - zur Verbesserung der Prognose individueller und ratingklassenspezifischer Ausfallwahrscheinlichkeiten wurden deshalb zahlreiche Ansätze entwickelt, die *aktuelle* volkswirtschaftliche Indikatoren aber auch individuelle Eigenschaften – zusätzlich zum individuellen Rating – berücksichtigen.²⁶

Die Art in der Insolvenzprognosen vorliegen (kategorial vs. ordinal vs. kardinal), ist verfahrensbedingt.²⁷ So generieren beispielsweise *Diskriminanzanalysen* oder *Neuronale Netze* standardmäßig nur kategoriale Insolvenzprognosen, ihr Output wird in der Praxis jedoch auch ordinal interpretiert.²⁸ Ordinale Insolvenzprognosen ergeben sich beispielsweise aus der Anwendung *subjektiv parametrisierter Scoring- oder Kennzahlenmodelle* während mit *Logit- oder Probit-Regressionensmodellen* kardinale Insolvenzprognosen erstellt werden können.

Die Art, in der Insolvenzprognosen vorliegen *müssen*, ergibt sich aus der beabsichtigten Verwendung der Prognosen: hat der Entscheider lediglich zwei Handlungsoptionen – z.B. Annahme oder Ablehnung eines Kunden, positive oder negative Feststellung der „Notenbankfä-

Ratings“ entsprechen 14 exakt der ausführlichen S&P-Notation (17-stufige Skala), 2 der verkürzten S&P-Notation (7-stufige Skala).

Bei Banken überwiegen hingegen *numerische* Ratingklassenbezeichnungen (ca. 85%) anstelle von Buchstabenkombinationen (ca. 15%), siehe ENGLISH, NELSON (1998, S. 4).

²² S&P verwendet seit 1974 die um Plus- und Minuszeichen modifizierte Ratingklassendarstellung. MOODY'S hingegen erst seit 1982, siehe CANTOR, PACKER (1994, S.2)

²³ Siehe S&P (2004a, S.11): “Many practitioners utilize statistics from this default study and CreditPro® to estimate probability of default and probability of rating transition. It is important to note that Standard & Poor's ratings do not imply a specific probability of default; however, Standard & Poor's historical default rates are frequently used to estimate these characteristics.”

²⁴ Siehe hierzu beispielsweise die regelmäßig aktualisierten Ausfallstatistiken der Ratingagenturen STANDARD AND POOR'S und MOODY'S (vgl. S&P (2004a) und MOODY'S (2004).

²⁵ siehe auch STEIN (2005, S. 1218): „While it is not always the case that powerful models are calibrated accurately to probabilities of default, it is empirically feasible to calibrate a model to real-world default probabilities by performing a default study on the historical behavior of each model score. Thus assuming a bank has historical data and can perform ROC analysis, it can also calibrate a model to a similar level of accuracy using the same machinery.”

²⁶ Vgl. KEENAN (1999): Wachstumsrate des (inflationsbereinigten) Industrieproduktionsindex, absolute (!) Anzahl neuer Speculative Grade Emittenten, Anteil der Speculative Grade Emittenten an allen Emittenten, Rendite 10jähriger staatlicher Schuldverschreibungen, u.a.; BANGIA, DIEBOLD, SCHUERMAN (2002): Zustand der Volkswirtschaft (Expansion/Rezession); S&P (2004b, S.3): Arbeitslosenrate, Anstieg der Zinsstrukturkurve, aggregierte Unternehmensgewinne, Outlook-Verteilung der Speculative Grade Emittenten; HAMILTON (2004): individueller Ratingausblickstatus und individuelle Ratinghistorie.

²⁷ Für eine Übersicht diverser Insolvenzprognoseverfahren siehe beispielsweise GÜNTHER, GRÜNING (2000), ÖSTERREICHISCHE NATIONALBANK (2004).

²⁸ Siehe beispielsweise ALTMAN, SAUNDERS (1998, S. 1737) für eine Zuordnung eines Diskriminanzanalysetrennscores zu Agenturratingklassen.

higkeit“ von Forderungen²⁹ - so genügen kategoriale Insolvenzprognosen. Für eine differenziertere, qualitative Beurteilung von Kunden, beispielsweise als Grundlage zur Entscheidung über die Art und Höhe von jeweils einzufordernden Sicherheiten, werden (mindestens) ordinale Insolvenzprognosen benötigt. Als Grundlage für quantitative Entscheidungen, beispielsweise für die Kredit-, Anleihen- oder Derivatebepreisung oder für die Ermittlung des ökonomischen oder aufsichtsrechtlichen Risikokapitals³⁰ werden hingegen kardinale Insolvenzprognosen benötigt.

²⁹ siehe DEUTSCHE BUNDESBANK (1999)

³⁰ Gemäß den ab Ende 2006 geltenden Neuen Eigenkapitalanforderungen des Basler Ausschuss für Bankenaufsicht müssen die bankinternen Ratingsysteme auf *kardinalen* Insolvenzprognosen (*Ausfallwahrscheinlichkeiten*) basieren, vgl. BASLER AUSSCHUSS (2004, insb. Tz. 461f.).

2.2 Schätzgütemaße für kategoriale Insolvenzprognosen

Kategoriale Insolvenzprognosen teilen die zu beurteilenden Unternehmen in zwei Gruppen, „voraussichtlich insolvent“ vs. „voraussichtlich nicht insolvent“ ein. Keines der heute verwendeten Insolvenzprognoseverfahren ist jedoch auch nur annähernd in der Lage, derart trennscharfe und gleichzeitig stets korrekte Prognosen zu erstellen (siehe hierzu die empirischen Befunde in Abschnitt 3.5).³¹ Abgesehen von „Zufallstreffern“ in kleinen Stichproben, werden kategoriale Insolvenzprognosen deshalb stets *Fehler* aufweisen. Hierbei gibt es zwei mögliche Fehlerarten: Fehler des Typs I –tatsächliche Ausfälle, die als Nicht-Ausfälle prognostiziert wurden (auch *Fehler 1. Art*, α -*Fehler*, *false negative proportion*) - und Fehler des Typs II - tatsächliche Nicht-Ausfälle, die als Ausfälle prognostiziert wurden (auch *Fehler 2. Art*, β -*Fehler*, *false positive proportion*). Üblich ist es, Fehler des Typs I ins Verhältnis zu allen tatsächlichen Ausfällen und Fehler des Typs II ins Verhältnis zu allen tatsächlichen Nicht-Ausfällen zu setzen (siehe Abbildung 1).

	prognostizierte Nichtausfälle	prognostizierte Ausfälle	
tatsächliche Nichtausfälle	✓ korrekte Prognose	✗ Fehler II. Art	$\Sigma=100\%$
tatsächliche Ausfälle	✗ Fehler I. Art	✓ korrekte Prognose	$\Sigma=100\%$

Abbildung 1: Kontingenztabelle

In der Insolvenzprognoseliteratur werden die Terme *100% - Fehler 1. Art* auch als *Trefferquote* (*hit rate*, *true positive proportion*) und *Fehler 2. Art* auch als *Fehlalarmquote* (*false alarm rate*) bezeichnet.^{32,33}

³¹ Da Insolvenzen auch durch *zufällige*, d.h. nach aktuellem Stand der Wissenschaft nicht prognostizierbare – sondern nur durch Wahrscheinlichkeitsfunktionen beschreibbare - Ereignisse ausgelöst werden können (siehe beispielsweise RISKMETRICS (1997, S. 43ff.), HULL (2003)), deren Realisierung erst nach Abgabe einer Prognose bekannt sein werden, beispielsweise durch den Eintritt/ Nichteintritt eines außerordentlichen Schadensereignisses innerhalb des nächsten Jahres oder durch die konkrete Realisierung des Dollarkurses, Ölpreises, Zinssatzes, Inflationsrate innerhalb des nächsten Jahres, folgt daß Insolvenzprognosen bestenfalls stochastischen Charakter haben können.

Daß – zumindest auf Basis von Jahresabschlußangaben – prinzipiell keine deterministischen (und stets wahren) Insolvenzprognosen getroffen werden können, legt auch die „Fehler“-Analyse von OHLSON (1980, S. 129) nahe: „[T]he reports of the misclassified bankrupt firms seems to lack any ‘warning signals’ of impending bankruptcy. All but two of the thirteen companies reported a profit. The two losses were minor [...] and these two companies had strong financial positions [...]. Other ratios analyzed showed the same ‘healthy patterns’. It is not surprising that these firms were misclassified, especially if one considers the profile of the nonbankrupt firms [...]. None of the misclassified bankrupt firms had a ‘going-concern’ qualification or disclaimer of opinion. [...] Some of the firms even paid dividends in the year prior to bankruptcy.”

³² Siehe beispielsweise SWETS (1973, S. 995), ENGELMANN, HAYDEN, TASCHE (2003, S. 13) und OENB (2004c, S. 21).

³³ Eine alternative, möglicherweise intuitivere Definition von *Trefferquote* wäre: Anteil der tatsächlich insolventen Unternehmen an allen Unternehmen, die vom Prognoseverfahren als insolvent prognostiziert wurden. Eine alternative, möglicherweise intuitivere Definition von *Fehlalarmquote* wäre: Anteil aller „Fehlalarme“ (nichtinsolvente Unternehmen, die als insolvent prognostiziert wurden) an allen „Alarmen“ (Prognosen, die einen Ausfall des Unternehmens behaupten). Für eine Übersicht zu Kennzahlen, die aus Kontingenztabelle gebildet werden können siehe beispielsweise SWETS, DAWES, MONAHAN (2000, S. 25f.)

Als Gütemaße können beispielsweise der ungewichtete Mittelwert³⁴ beider Fehlerquoten oder aber ein gewichteter Mittelwert, wobei als Gewichtungsfaktoren beispielsweise die Anteile der Ausfaller und Nichtausfaller an allen Unternehmen des Samples verwendet werden können („Bayesscher Gesamtfehler“³⁵) oder aber die Kosten³⁶, die mit beiden Fehlerarten verbunden sind (Fehler-Typ I: Kreditausfallkosten, Fehler-Typ II: entgangene Kreditmarge, sonstiges „Cross-Selling-Geschäft“³⁷).

Auf eine tiefere Analyse der Vor- und Nachteile der einzelnen Kennzahlen soll an dieser Stelle verzichtet werden, da moderne Insolvenzprognoseverfahren nicht mehr auf *kategorialen* sondern auf ordinalen oder kardinalen Insolvenzprognosen basieren, für die mittlerweile etablierte Gütemaße existieren (siehe insbesondere die folgenden Abschnitte 2.3 und 2.4).³⁸

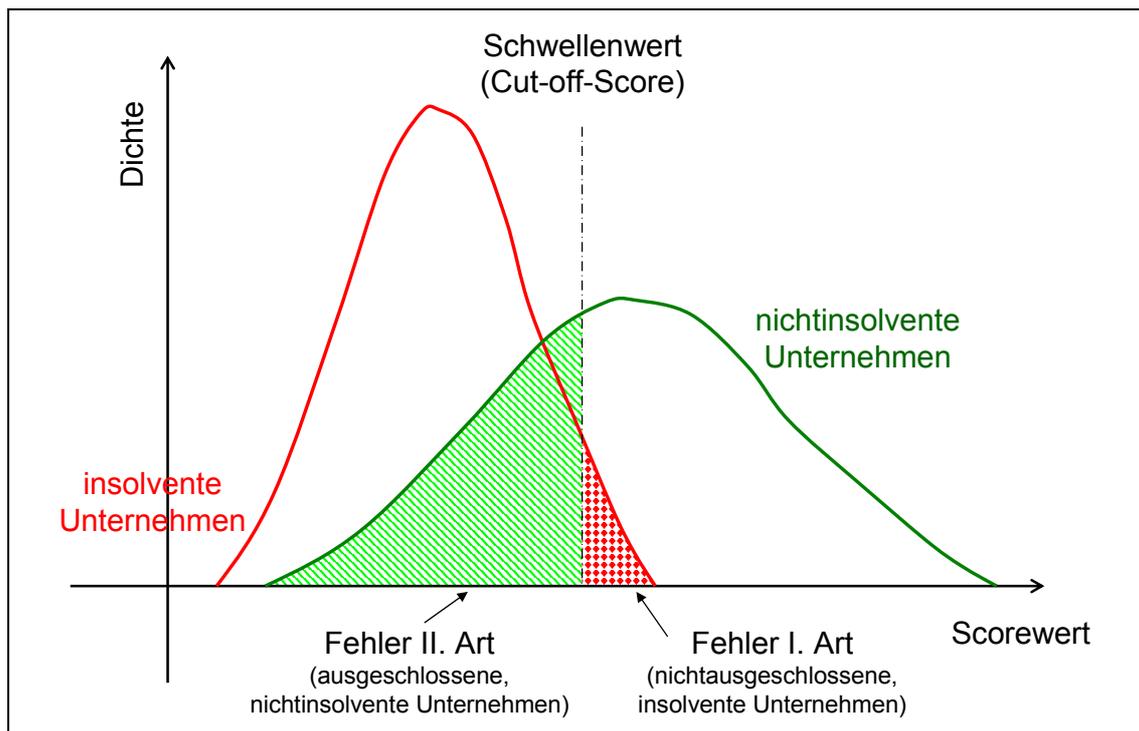


Abbildung 2: Klassifizierungsfehler in Abhängigkeit vom gewählten Schwellenwert und den Score-dichtefunktionen für insolvente und nichtinsolvente Unternehmen³⁹

Des Weiteren besteht bei allen Verfahren, die kategoriale Insolvenzprognosen erstellen, ein Zielkonflikt zwischen Fehlern I und II. Art. Je nach Parametrisierung des Insolvenzprognoseverfahrens kann erreicht werden, daß 0% aller Ausfälle und 100% aller Nichtausfälle - oder umgekehrt - richtig erkannt werden, und auch eine Feinjustierung zwischen diesen Extrem-

³⁴ siehe BALCAEN, OOGHE (2004, S.12) und die dort zitierte Literatur

³⁵ Der „bayessche Gesamtfehler“ gibt an, welcher Anteil der Prognosen falsch ist, ohne dabei zwischen Fehler I. und II. Art zu unterscheiden. Für die Beurteilung der Güte von Insolvenzprognosen ist diese Kenngröße denkbar ungeeignet, da sich bereits mit der naiven Prognose „kein Unternehmen wird je insolvent“ Prognosen abgeben lassen, deren Gesamtfehler nahe bei 0% (in Höhe der durchschnittlichen Insolvenzrate) liegt und damit selbst sehr trennfähige Insolvenzprognosen schlägt, selbst wenn diese nur einen niedrigen Fehler II. Art aufweisen. (OENB (2004a, S. 117ff.))

³⁶ siehe beispielsweise NANDA, PENDHARKAR (2001, S. 155ff.)

³⁷ siehe beispielsweise OENB (2004b, S.33, 80)

³⁸ siehe auch SWETS, PICKETS (1982, S. 24ff., Übersetzung) für eine Übersicht und Diskussion von „häufig verwendeten, aber unzulänglichen, [kategorialen] Schätzgütemaßen“

³⁹ Abbildung in Anlehnung an DEUTSCHE BUNDESBANK (2003, S. 73), ENGELMANN, HAYDEN, TASCHKE (2003, S. 5) und OENB (2004a, S. 107).

punkten ist in der Regel möglich (siehe Abbildung 2).⁴⁰ Die Wahl einer konkreten Fehler-I-II-Kombination – angesichts unendlich vieler alternativer Möglichkeiten - ist demnach willkürlich und somit nur bedingt geeignet, die Qualität eines Verfahrens zu messen.⁴¹

An dieser Stelle setzen Gütemaße zur Beurteilung ordinaler Insolvenzprognosen an: sie bewerten die Klassifikationsleistung eines Insolvenzprognoseverfahren nicht anhand einer einzigen, willkürlich bestimmten, Fehler-I-II-Kombination - sondern sie bewerten sie anhand der Gesamtheit aller möglichen mit dem Prognoseverfahren erzeugbaren Fehler-I-II-Kombinationen.

⁴⁰ siehe beispielsweise OHLSON (1980, S. 124ff.)

⁴¹ Eine denkbare Rechtfertigung für die Beschränkung auf eine einzige Fehler-I-II-Kombination bei der Beurteilung der Schätzgüte eines Verfahrens wäre die Verwendung von „*optimalen*“ („*kostenminimalen*“) Fehlerkombination I. und II. Art. Welche Fehler-I-II-Kombination jedoch „*optimal*“ ist, ist subjektiv verschieden (bei einer Bank sind die Kosten für einen Fehler I. Art im Verhältnis zu einem Fehler II. Art vermutlich höher als bei einem potentiellen Lieferanten des Unternehmens), ist von subjektiv beeinflussbaren Nebenbedingungen abhängig (hier: konkrete Ausgestaltung der Kreditkonditionen (Zinssätze, Sicherheiten, Bürgschaften, ...)) und ist von, zwar subjektiv nicht beeinflussbaren, aber im Zeitverlauf veränderlichen Größen abhängig, beispielsweise von der durchschnittlichen Ausfallquote (zum letzten Punkte siehe auch BALCAEN, OOGHE (2004, S. 15)).

2.3 Schätzgütemaße für ordinale Insolvenzprognosen

Auch wenn ordinale Insolvenzprognosen allgemeiner als kategoriale Insolvenzprognosen sind, genügen für die meisten Anwendungen keine nur vergleichenden Aussagen über das relative Ausfallrisiko von Unternehmen. Benötigt werden vielmehr auch *quantitative Ausfallprognose*, beispielsweise um angeben zu können, ob eine Risikoprämie von 1,5% p.a. für einen endfälligen Kredit mit drei Jahren Laufzeit bei einem Unternehmen angemessen ist, dessen Ratingeinstufung besagt, daß es „*derzeit die Fähigkeit hat, seinen finanziellen Verpflichtungen nachzukommen. Allerdings würden ungünstige Geschäfts-, Finanz- oder gesamtwirtschaftliche Bedingungen vermutlich seine Fähigkeiten und Bereitschaft seinen finanziellen Verpflichtungen nachzukommen beeinträchtigen*“.^{42,43,44}

Trotzdem lohnt sich die Auseinandersetzung mit Schätzgütemaßen für ordinale Insolvenzprognosen:

- Sie entsprechen von der Intention her dem, was die Ratingagenturen zu liefern versprechen.^{45,46}
- Sie stellen mittlerweile eine dominante Methode zur Beurteilung der Qualität von Ratingergebnissen dar.⁴⁷
- Es existieren zahlreiche Möglichkeiten der grafischen Darstellung für die ordinale Schätzgüte von Ratingsystemen, was auch die Kommunizierbarkeit der aus diesen Darstellungen abgeleiteten *Kennzahlen* erleichtert.

⁴² Siehe Definition Ratingstufe B nach S&P(2003b, S.7) für long-term credit ratings: “An obligation rated ‘B’ is more vulnerable to nonpayment than obligations rated ‘BB’, but the obligor currently has the capacity to meet its financial commitment on the obligation. Adverse business, financial, or economic conditions will likely impair the obligor’s capacity or willingness to meet its financial commitment on the obligation.”

⁴³ Siehe auch FRERICHS, WAHRENBURG (2003, S. 13): “Under what circumstances is such a measure [*Area Under the ROC-curve*] useful? The ranking of borrowers is sufficient for credit risk management if banks are not able to charge different credit risk premiums for different customers in the market. In this case, banks maximize their risk-adjusted returns by not granting credit to customers with negative expected returns which is equivalent to defining a minimum credit score. Yet, this line of thought does not lead us to the AUC as a measure of system quality, but to the concept of minimized expected error costs. The AUC measures the quality of the complete ranking and not only of one threshold. Only if the threshold is difficult to define in practice, the AUC may be a sensible measure.”

⁴⁴ Oder kurz und prägnant: “There are no bad loans, only bad prices.”, siehe FALKENSTEIN, BORAL, KOCAGIL (2000, S. 5).

⁴⁵ CANTOR, MANN (2003, S. 6): “MOODY’S primary objective is for its ratings to provide an accurate relative (i.e., ordinal) ranking of credit risk at each point in time, without reference to an explicit time horizon.” und CANTOR, MANN (2003, S. 1, Formatierung hinzugefügt): “Moody’s does *not* target specific default rates for the individual rating categories.”, allerdings auch: “Moody’s also tracks investment-grade default rates and the average rating of defaulting issuers prior to their defaults. These metrics measure Moody’s success at meeting a secondary cardinal or absolute rating system objective, namely that ratings be useful to investors who employ simple rating ‘cutoffs’ in their investment eligibility guidelines.”, siehe ebenda.

⁴⁶ BASLER AUSSCHUB (2000c, S.2) “Most firms report that they rate risk on a relative – rather than absolute – scale, and most indicate that they rate ‘across the business cycle’, suggesting that ratings should in principle not be significantly affected by purely cyclical influences.” Von den 15 Ratingagenturen, die hierzu Angaben machten, gaben 13 an, mit ihrem Rating das “relative Risiko” von Unternehmen zu messen, nur 2 Agenturen messen nach eigenen Angaben das “absolute Risiko” von Unternehmen (KMV Corporation und Upplysningscentralen AB), siehe ebenda (S. 23f).

⁴⁷ Siehe beispielsweise MCQUOWN (1993, S.5ff), KEENAN, SOBEHART (1999, S. 5ff), STEIN (2002, S.5ff.), FAHRMEIR, HENKING, HÜLS (2002, S. 22f), ENGELMANN, HAYDEN, TASCHE (2003, S.3ff), DEUTSCHE BUNDESBANK (2003, S. 71ff.), OENB (2004a, S. 113ff.). Auch die Ratingagenturen STANDARD AND POOR’S (2005, S.19ff.) und MOODY’S (2004c, S.3ff.) messen die Qualität ihrer Ratingsysteme mittels der in diesem Abschnitt vorgestellten Methoden und Kennzahlen.

- Eine gute Trennfähigkeit, wie sie durch Gütemaße für ordinale Insolvenzprognosen gemessen wird, ist auch für die Qualität *kardinaler* Insolvenzprognosen wichtig – und zwar wichtiger als eine korrekte Kalibrierung. Die Sicherstellung *richtiger* (kalibrierter) Ausfallwahrscheinlichkeiten ist leichter zu erreichen, als die Sicherstellung *trennscharfer* Ausfallwahrscheinlichkeiten.^{48,49}
- Beim empirischen Vergleich verschiedener Ratingverfahren auf Basis identischer Stichproben entspricht die Reihenfolge der Güte der einzelnen Verfahren gemäß ordinaler Schätzgütemaße weitgehend der Reihenfolge der Güte bei Zugrundelegung kardinaler Schätzgütemaße.⁵⁰ Somit scheinen Qualitätsunterschiede verschiedener Verfahren weniger in einer unterschiedlich guten Befähigung zu Abgabe *kalibrierter* Ausfallprognosen zu liegen, was nur für kardinale Gütemaße relevant ist, sondern in einer unterschiedlich guten Befähigung zur Abgabe *trennscharfer* Prognosen, was sowohl für ordinale als auch kardinale Schätzgütemaße relevant ist.
- Für diejenigen Aspekte kardinaler Insolvenzprognosen, die nicht bereits mit dem Instrumentarium, das für die Beurteilung ordinaler Insolvenzprognosen entwickelt wurde, gemessen werden können, insbesondere für den Aspekt der *Kalibrierung* von Insolvenzprognosen, stehen derzeit *keine* aussagekräftigen Testverfahren zur Verfügung! Ursächlich für diese Defizite sind *Korrelationen* bezüglich der Ausfallwahrscheinlichkeiten zwischen verschiedenen Unternehmen.^{51,52}

Auch wenn die Ratingagenturen ex-ante mit den von ihnen vergebenen, *ordinalen* Ratingnoten keine *kardinalen* Ausfallprognosen für irgendeinen spezifizierten Zeitraum abgeben wollen⁵³, demonstrieren sie die Güte ihrer Ratingeinschätzungen u.a. mit deren Fähigkeit, Unternehmen mit deutlich unterschiedlichen, monoton steigenden Ausfallhäufigkeiten identifizieren zu können⁵⁴ (siehe Abbildung 3 für die durchschnittlichen ratingklassenspezifischen Ein-

⁴⁸ BLOCHWITZ, LIEBIG, NYBERG (2000, S.3): “It is usually much easier to recalibrate a more powerful model than to add statistical power to a calibrated model. For this reason, tests of power are more important in evaluating credit models than tests of calibration. This does not imply that calibration is not important, only that it is easier to carry out.”, analog STEIN (2002, S.9)

⁴⁹ Die Kalibrierung eines Ratingsystems wird in SOBEHART ET AL (2000, S. 23f.) oder STEIN (2002, S. 8ff.) beschrieben.

⁵⁰ Siehe beispielsweise KRÄMER, GÜTLER (2003) für den Vergleich der Schätzgüte der Ausfallprognosen von S&P und MOODY’S oder SOBEHART, KEENAN, STEIN (2000, S. 14) für den Vergleich der Schätzgüte von sechs unterschiedlichen Verfahren mittels ordinaler und kardinaler Schätzgütekennzahlen.

⁵¹ siehe BASLER AUSSCHUB (2005, S. 31f.): “These [“entropy” based] measures appear to be of limited use only for validation purposes as no generally applicable statistical tests for comparisons are available. [...] The Group [the Validation Group is a subgroup of the Research Task Force (RTF) of the Basel Committee on Banking Supervision] has found that the Accuracy Ratio (AR) and the ROC measure appear to be more meaningful than the other above-mentioned indices because of their statistical properties. For both summary statistics, it is possible to calculate confidence intervals in a simple way. [...] However, due to the lack of statistical test procedures applicable to the Brier score, the usefulness of this metric for validation purposes is limited.” und ebenda, S. 34: “At present no really powerful tests of adequate calibration are currently available. Due to the correlation effects that have to be respected there even seems to be no way to develop such tests. Existing tests are rather conservative [...] or will only detect the most obvious cases of miscalibration [...]”

⁵² Schon bei moderaten Ausfallkorrelationen können die realisierten Ausfallquoten - selbst bei beliebig großen Portfolien - erheblich von den erwarteten Ausfallquoten abweichen, siehe hierzu HUSCHENS, HÖSE (2003, S. 152f.) oder BLOCHWITZ ET AL (2004, S. 10).

⁵³ S&P (2005, S. 28, Formatierung hinzugefügt): „Many practitioners utilize statistics from this default study and CreditPro® to estimate probability of default and probability of rating transition. *It is important to note that Standard & Poor's ratings do not imply a specific probability of default*; however, Standard & Poor's historical default rates are frequently used to estimate these characteristics.“

⁵⁴ siehe beispielsweise MOODY’S (2005, S.7)

jahresausfallraten nach S&P und MOODY'S sowie Abbildung 4 für die entsprechenden Mehrjahresausfallraten).

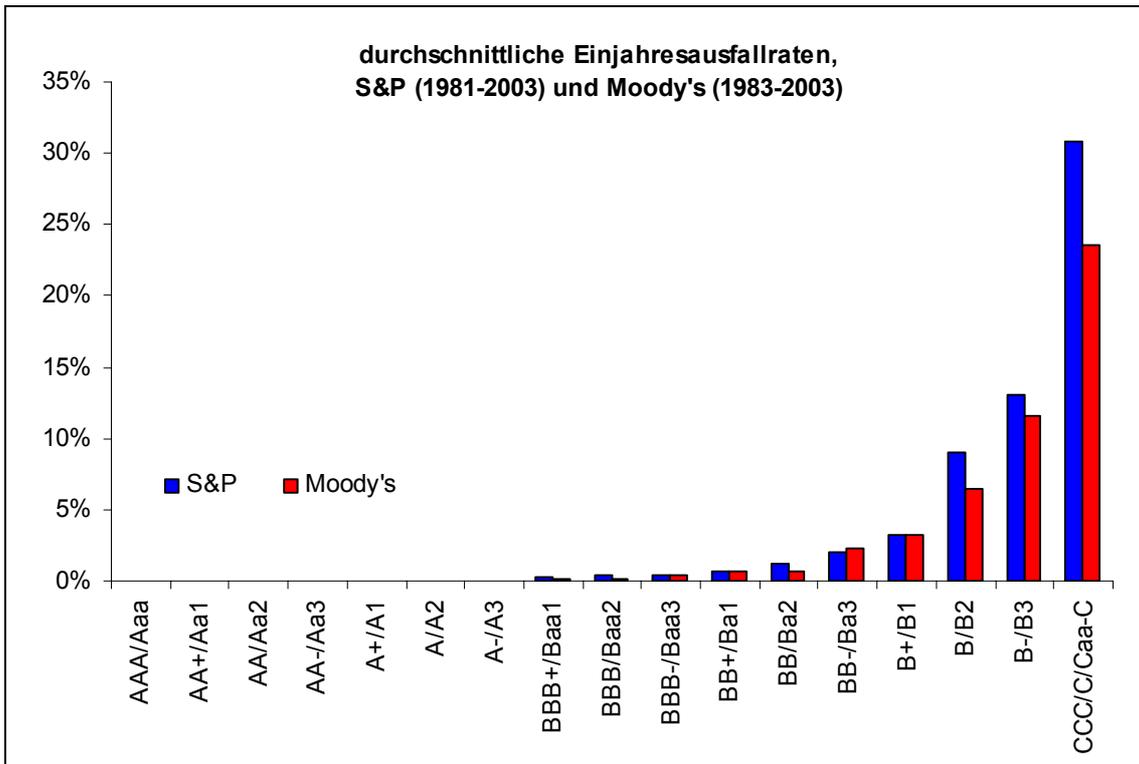


Abbildung 3: durchschnittliche historische Einjahresausfallraten nach modifizierten Ratingklassen gemäß STANDARD & POOR'S (1981-2003) und MOODY'S (1983-2003)⁵⁵

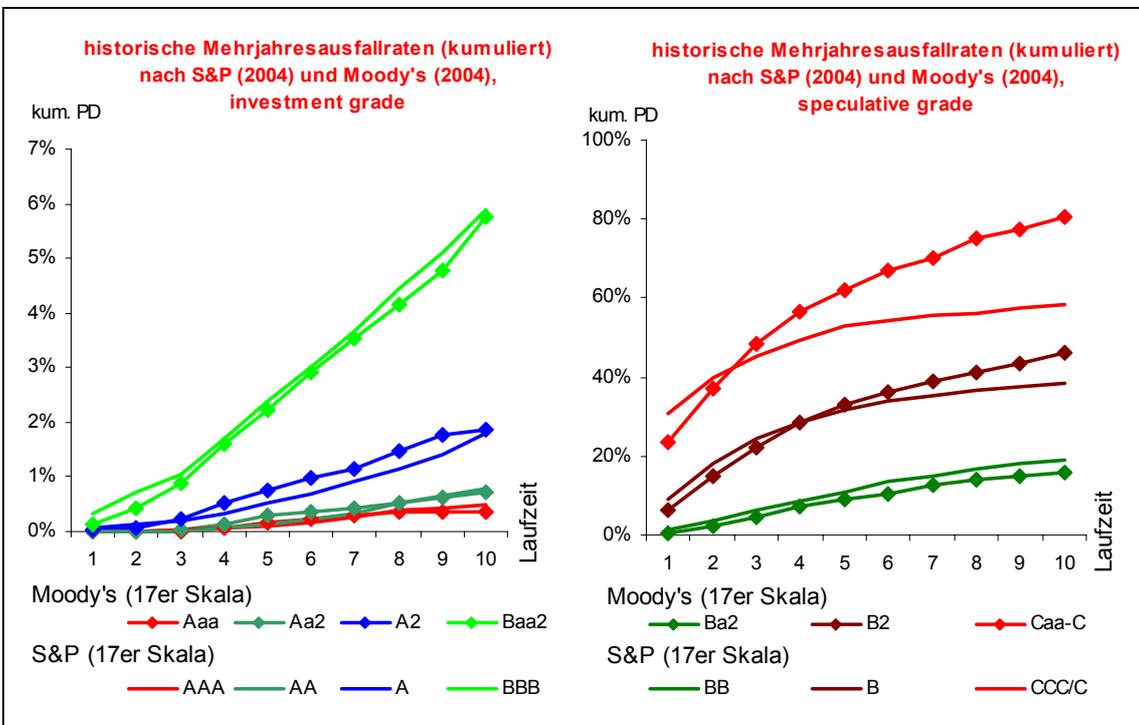


Abbildung 4: durchschnittliche historische Ein- bis Zehnjahresausfallraten nach modifizierten Ratingklassen gemäß STANDARD & POOR'S (1981-2003) und MOODY'S (1983-2003)⁵⁶

⁵⁵ siehe S&P (2004, S.13), MOODY'S (2004, S. 26)

Anmerkung: Die Ratings von S&P und MOODY'S werden von den Finanzmärkten als sehr gute Substitute hinsichtlich ihrer inhaltlichen Bedeutung^{57,58} und Qualität^{59,60} betrachtet. Obwohl beide Agenturen in den Zeiträumen 1981-2003 (S&P) und 1983-2003 (MOODY'S) nahezu identische durchschnittliche Einjahresausfallquoten von 1,76%⁶¹ und 1,86%⁶² verzeichneten, sind die ratingklassenspezifischen Ausfallquoten für sämtliche Ratingstufen (mit Ausnahme bei BB-/Ba3) gemäß MOODY'S niedriger als gemäß S&P (vergleiche Abbildung 3). Hieraus kann zwar geschlossen werden, daß MOODY'S etwas strenger ratet als S&P⁶³, mit Ausnahme bei den besonders schlecht gerateten Unternehmen (CCC/C bzw. Caa/C), die jedoch nur einen sehr geringen Anteil an der Gesamtheit aller Schuldner ausmachen, sind jedoch kaum nennenswerte Unterschiede hinsichtlich der realisierten Ein- und Mehrjahresausfallraten zu verzeichnen.

Abbildung 3 und Abbildung 4 zeigen, daß die Ratingagenturen in der Lage waren, mit ihren ex-ante vergebenen Ratingnoten, ex-post Gruppen von Unternehmen mit sehr unterschiedlichen realisierten Ausfallquoten zu separieren.⁶⁴ Dies ist eine *notwendige* Voraussetzung für trennscharfe Ausfallprognosen, aber keine *hinreichende*. Kompatibel mit den Darstellungen in obigen Abbildungen wären beispielsweise die folgenden Extremfälle⁶⁵:

⁵⁶ Siehe S&P (2004, S.13), MOODY'S (2004, S. 26). Aus Platzgründen wurden für die Ratingstufen AA, A, BBB, BB und B nur die jeweils mittleren Ratingstufen (nach MOODY'S-Notation: Aa2, A2, Baa2, Ba2 und B2) dargestellt, nicht jedoch die mit Modifikatoren (+, - bzw. 1 und 3) versehenen Ratingstufen.

⁵⁷ siehe CANTOR, PACKER (1994, S. 12): „As a practical matter, however, it appears that market participants have historically viewed the Moody's and Standard and Poor's scales as roughly equivalent [...]“

⁵⁸ Unternehmen die sowohl von S&P als auch von MOODY'S geratet werden, erhalten auf einer 17-gliedrigen Skala in 40%-45% exakt das gleiche Rating; in weiteren 40%-45% unterscheiden sich die Ratingurteile um genau einen Punkt und in 10%-15% um genau zwei Punkte. Ratingdifferenzen von mehr als 2 Stufen treten nur in 2,5%-5% aller Fälle auf, siehe GÜTLER (2004, S.13 und die dort zitierten Studien). Auf einer 7-stufigen Skala stimmen die Bewertungen von S&P und MOODY'S in 71%, aller Fälle exakt überein, in 28% aller Fälle unterscheiden sie sich um eine Note, in 1,1% um zwei Noten und (nur) in 0,1% aller Fälle um drei oder mehr Noten (eigene Auswertungen auf Basis von GÜTLER (2004, Appendix B)).

⁵⁹ siehe ELTON ET AL (2004, S. 2755f.) und STEINER, HEINKE (2000, S. 560f.)

⁶⁰ Auf Basis von Untersuchungen mit *identischen* Stichproben von Unternehmen, die sowohl von S&P als auch von MOODY'S geratet wurden, finden KRÄMER, GÜTLER (2003) und GÜTLER (2004) eine geringfügige Überlegenheit des Ratingsystems von MOODY'S hinsichtlich unterschiedlicher Prognosezeiträume und Validierungskennzahlen.

⁶¹ S&P(2004, S. 16)

⁶² Hinweis: Der in MOODY'S (2004, S. 26) angegebene Wert von 1,24% ist falsch, wie auf Nachfrage bei MOODY'S bestätigt wurde. In MOODY'S (2005, S.17) wird die durchschnittliche jährliche Ausfallrate aller Unternehmen von 1983-2004 mit 1,79% angegeben.

⁶³ Siehe hierzu auch GÜTLER (2004, S.13) und die dort zitierten Studien: bei Unternehmen, die sowohl von S&P und MOODY'S geratet wurden und bei denen sich die Ratingurteile von S&P und MOODY'S unterscheiden, vergibt MOODY'S in rund 60% aller Fälle das schlechtere Rating. Uneinheitliche Ergebnisse erbringen die Studien dahingehend, ob diese Negativverzerrung eher Bewertungsunterschieden im Investment- oder Speculative-Grade-Bereich zuzurechnen ist.

⁶⁴ Den Ausfallstudien von S&P und MOODY'S liegen *Emittentenratings* (auch *issuer ratings, corporate credit ratings, implied senior-most rating, default ratings, natural ratings, estimated senior ratings*) zugrunde, welche ein Maß für die erwartete Ausfallwahrscheinlichkeit der Unternehmen darstellen sollen, ohne jedoch einen expliziten Prognosehorizont anzugeben, siehe S&P (2003b, S. 3ff., 61ff.), CANTOR, MANN (2003, S. 6f.), MOODY'S (2004b, S. 8) und MOODY'S (2005, S. 39). Die Ratings konkreter Verbindlichkeiten von Unternehmen, sogenannte *Emissionsratings* (*issue rating*), berücksichtigen neben der Ausfallwahrscheinlichkeit auch die erwartete Ausfallschwere und können deshalb als ein Maß für die erwarteten Ausfallkosten interpretiert werden. Je nach Rang der bewerteten Verbindlichkeit innerhalb der Kapitalstruktur des Unternehmens, gewährten Sicherheiten und weiteren Einflußfaktoren ergibt sich das Emissionsrating durch Zu- oder Abschläge (von meist 1 bis 2 Punkten) aus dem Emissionsrating. Das *Emittentenrating* eines Unternehmens entspricht in der Regel dem *Emissionsrating* erstrangiger, unbesicherter Verbindlichkeiten, siehe ebenda.

⁶⁵ siehe hierzu auch CANTOR, MANN (2003, S. 14)

- Extremfall I: Das Ratingsystem ordnet fast alle Unternehmen in eine mittlere Ratingstufe (siehe Abbildung 3) ein, beispielsweise BB, bzw. Ba2, und nur sehr wenige Unternehmen in andere Ratingstufen.
- Extremfall II: Das Rating ordnet fast alle Unternehmen in die extremen Ratingstufen, d.h. entweder AAA oder CCC/C, ein und nur wenige in die mittleren Ratingstufen.

Im Fall I wäre das Rating nahezu wertlos, da es praktisch keine Differenzierung zwischen den verschiedenen Schuldnern erlaubte. Im Fall II hingegen wäre der Informationsnutzen immens: das Rating würde stets „extreme Prognosen“ stellen – also entweder eine extrem niedrige oder eine extrem hohe Ausfallwahrscheinlichkeit vorhersagen – und die Prognosen würden meist zutreffen - einem AAA-Rating würde innerhalb eines Jahres fast nie und einem CCC/C-Rating würde zumindest in 25%-30% aller Fälle ein Ausfall folgen.

Um die ordinale Qualität eines Ratingsystems zu bestimmen, müssen somit nicht nur die ratingklassenspezifischen Ausfallquoten bekannt sein, sondern auch die Verteilung der Unternehmen auf die einzelnen Ratingklassen.

2.3.1 Grafische Schätzgütebestimmung

Die Fähigkeit eines Ratingsystems „gute“ und „schlechte“ Schuldner mit großer Zuverlässigkeit zu trennen, kann beispielsweise durch ROC-⁶⁶ oder CAP-Kurven⁶⁷ visualisiert werden und durch verschiedene hierauf basierende und ineinander überführbare Kennzahlen quantifiziert werden.⁶⁸ ROC-Kurven werden bereits seit Anfang der 50er Jahre im Bereich der experimentellen Psychologie verwendet.⁶⁹ Weitere Bezeichnungen für ROC- oder CAP-Kurven sind *Aufklärungsprofil*, *Powerkurve*, *Lorenzkurve*, *Ginikurve*, *lift-curve*, *dubbed-curve* oder *ordinal dominance graph*.⁷⁰

Die ROC-Kurve eines Verfahrens ergibt sich aus der Menge sämtlicher Kombinationen von Trefferquoten (100% - Fehler I. Art) und Fehlalarmquoten (Fehler II. Art), die ein Insolvenzprognoseverfahren bei Zugrundelegung verschiedener Trennwerte (cut-off scores), d.h. bei Überführung von ordinalen in kategoriale Insolvenzprognosen, liefern kann (siehe Abbildung 5, links). Bei einem „zu scharf gestellten“ Trennwert würden sämtliche Unternehmen als voraussichtlich insolvent prognostiziert (100% Trefferquote (= 0 % Fehler I. Art); 100% Fehlalarmquote (=100% Fehler II. Art)), bei einem „zu lax eingestellten“ Trennwert, würde keines der Unternehmen als voraussichtlich insolvent prognostiziert (0% Trefferquote; 0% Fehlalarmquote). Auch zwischen diesen Extremfällen gibt es einen Ausgleich zwischen Treffer- und Fehlalarmquoten. Die Qualität ordinaler Insolvenzprognosen zeigt sich gerade in der Art dieses Ausgleichs. Ein perfektes Prognoseverfahren müßte keinen einzigen Schuldner zu Unrecht ausschließen, um 100% aller Ausfälle zu erfassen (vertikaler Verlauf der ROC-Kurve von (0%; 0%) nach (0%; 100%)), eine anschließende Verschärfung des Trennwertes würde nur zu einer Erhöhung der Fehlalarmquote führen (horizontaler Verlauf der ROC-Kurve von (0%; 100%) nach (100%; 100%)). Die ROC-Kurve eines Ratingverfahrens hingegen, dessen Bewertungen rein zufällig erfolgten, würde entlang der Diagonalen verlaufen – jeder Prozentpunkt bei der Verbesserung der Trefferquote müßte mit einem Prozentpunkt bei der Verschlechterung der Fehlalarmquote erkauft werden.

ROC-Kurven realer Insolvenzprognoseverfahren weisen eine konkave Form auf (siehe Abbildung 5, siehe auch Anhang I für zahlreiche empirische ROC-Kurven). Die Konkavität einer ROC-Kurve impliziert, daß die realisierten Ausfallraten mit besserer Bonität geringer werden, d.h. daß das zugrundeliegende Ratingsystem „semi-kalibriert“ ist.⁷¹

CAP-Kurven (siehe Abbildung 5, rechts) resultieren aus der Anwendung eines nur geringfügig modifizierten Konstruktionsprinzips wie das der ROC-Kurven. Auf der X-Achse werden

⁶⁶ “A curve [...] is called a ‘ROC’ – sometimes short for *Receiver* Operating Characteristic, especially in the field of signal detection, and sometimes short for *Relative* Operating Characteristic, in generalized applications.”, siehe SWETS (1988, S. 1287, Formatierung hinzugefügt).

⁶⁷ CAP ... cumulative accuracy profile

⁶⁸ Für alternative grafische Darstellungsmöglichkeiten für ROC-Kurven und hierauf basierende Kennzahlen siehe SWETS, PICKET (1982, S. 31ff.), die zur Visualisierung und Kennzahlenbildung „binormale“ ROC-Kurven vorschlagen, d.h. ROC-Kurven, bei denen die Abszissen- und Ordinatenwerte für Quantilswerte zwischen 1% und 99% gemäß der Umkehrverteilung der Standardnormalverteilung transformiert werden. Gemäß SWETS, PICKET (1982, S. 31f.) lassen sich empirische binormale ROC-Kurven gut durch lineare Funktionen approximieren - was sich jedoch bei eigenen Untersuchungen anhand der in Anhang I vorgestellten empirischen ROC-Kurven nicht bestätigen ließ. *Ließen* sich binormale ROC-Kurven tatsächlich gut durch lineare Funktionen annähern, so wäre eine gute Inter- und Extrapolation binormaler ROC-Kurven auf Basis von *zwei* Fehler-I-II-Kombinationen möglich und könnten ROC-Kurven eindeutig durch zwei Parameter beschrieben werden.

⁶⁹ siehe SWETS (1988, S. 1287)

⁷⁰ siehe BLOCHWITZ, LIEBIG, NYBERG (2000, S. 33): power curve, lift-curve, dubbed-curve, receiver-operator curve; FALKENSTEIN, BORAL, KOCAGIL (2000, S.25): Gini curve, Lorenz curve, ordinal dominance graph; SCHWAIGER (2002, S. 27): Aufklärungsprofil

⁷¹ siehe KRÄMER (2003, S. 403)

hier nicht die Fehlalarmquoten (Fehler II. Art) abgetragen, sondern der Anteil von Unternehmen, die vom Prognoseverfahren ausgeschlossen werden müssen, unabhängig davon, ob es sich um tatsächliche Ausfälle oder Nicht-Ausfälle handelt, um eine bestimmte Trefferquote (100%-Fehler I. Art) zu erzielen. Die CAP-Kurve eines perfekten Ratings verläuft ausgehend vom Punkt (0%; 0%) steil nach rechts oben – jedoch nicht vertikal, da mindestens PD% aller Unternehmen ausgeschlossen werden müssen, um alle Ausfälle zu erfassen (siehe die Strichpunktlinie in Abbildung 5, rechts).

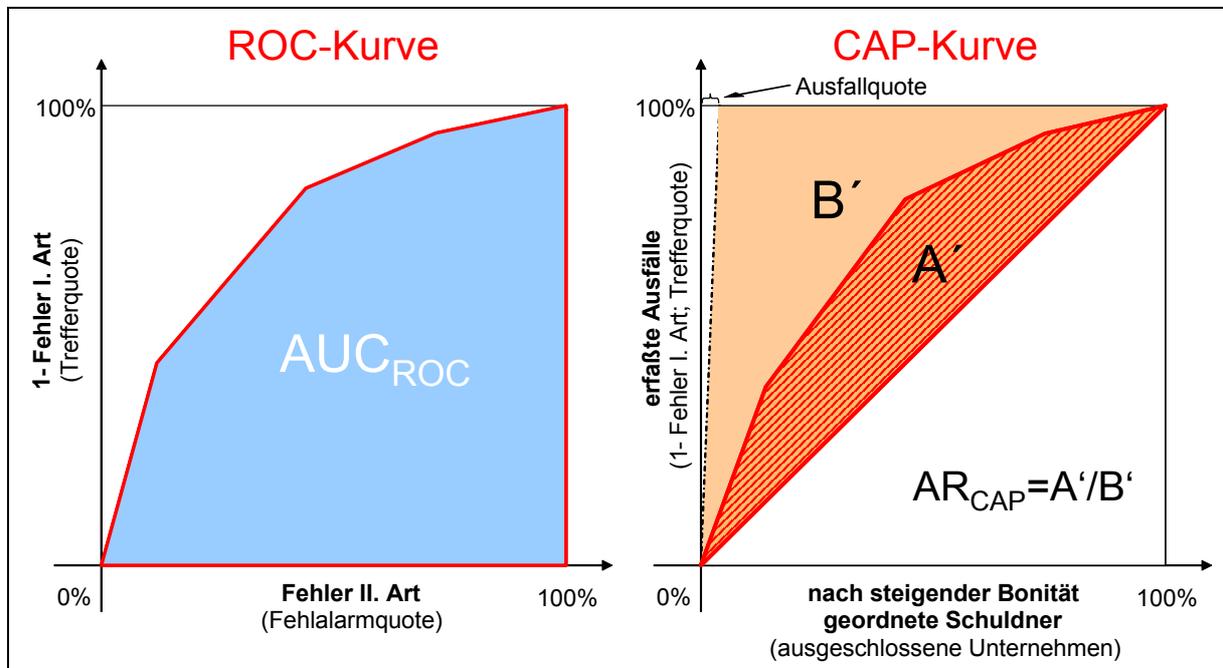


Abbildung 5: ROC- und CAP-Kurve

Liegt die ROC-/CAP-Kurve eines Verfahrens A_1 bei Beurteilung der gleichen Unternehmen links oberhalb der ROC-/CAP-Kurve eines anderen Verfahrens A_2 , so liefert Verfahren A_1 für jeden denkbaren Trennwert bessere Prognosen als Verfahren A_2 (vgl. Abbildung 6, links). Falls die relative Vorteilhaftigkeit eines Verfahrens nur von dessen relativer Präzision abhängt⁷², wäre Verfahren A_1 damit Verfahren A_2 stets vorzuziehen - unabhängig von individuellen Spezifika des Entscheidungsträgers, beispielsweise hinsichtlich seiner Kosten für Fehler I. und II. Art. Schneiden sich hingegen die ROC-Kurven zweier Verfahren B_1 und B_2 (siehe Abbildung 6, rechts), so ist keines der beiden Verfahren dem anderen objektiv überlegen.⁷³ Im gewählten Beispiel ist Verfahren B_1 besser als B_2 in der Lage, innerhalb der „guten Unternehmen“ zu differenzieren, während Verfahren B_2 im Bereich der „schlechten Unternehmen“ eine höhere Trennschärfe aufweist.⁷⁴

⁷² Ein weiteres Kriterium für die relative Vorteilhaftigkeit der Verfahren könnte in den *Kosten* der Verfahren selbst begründet liegen. Während Banken für die Einstufung von kleineren Krediten häufig auf automatisierte (und tendenziell wenig trennscharfe) Verfahren zurückgreifen, lohnt sich bei größeren Krediten auch arbeitsintensivere Bewertungsverfahren, siehe TREACY, CAREY (2000/1998, S. 905) und BASLER AUSSCHUSS (2000b, S.18f.)

⁷³ siehe beispielsweise BLOCHWITZ, LIEBIG, NYBERG (2000, S.7f)

⁷⁴ Nach KRÄMER (2003, S. 402) besteht beim Vergleich von Ratingsystemen der „Normalfall“ in sich schneidenden ROC- bzw. CAP-Kurven: „Insofern hilft das Konzept der Ausfalldominanz in vielen Anwendungen nicht weiter. [...] Das Konzept der Ausfalldominanz [empfiehlt] sich vor allem zum Aussortieren von Substandard-Systemen.“ Trennschwächere (ausfalldominierte) Prognosen lassen sich aus trennstärkeren Prognosen ableiten, siehe KRÄMER (2003, S. 397f.)

Ist der Entscheider nicht gezwungen, entweder nur das eine oder nur das andere Verfahren zu verwenden, kann er möglicherweise die Verfahren B_1 und B_2 so zu einem dritten Verfahren B_3 kombinieren, daß dieses sowohl B_1 als auch B_2 überlegen ist. Gleiches gilt auch für die Verfahren A_1 und A_2 : obwohl A_2 strikt von A_1 dominiert wird, läßt sich aus der Kombination von A_1 und A_2 *möglicherweise* ein Verfahren A_3 generieren, das nicht nur A_2 sondern auch A_1 überlegen ist. So ließ sich in empirischen Studien beispielsweise die Prognosequalität von ausschließlich kennzahlenbasierten Bankenratings durch die Einbeziehung von „weichen Faktoren“ verbessern, obwohl die Prognosegüte der kennzahlenbasierten Ratings besser war, als die Prognosegüte der Softfaktorratings.⁷⁵

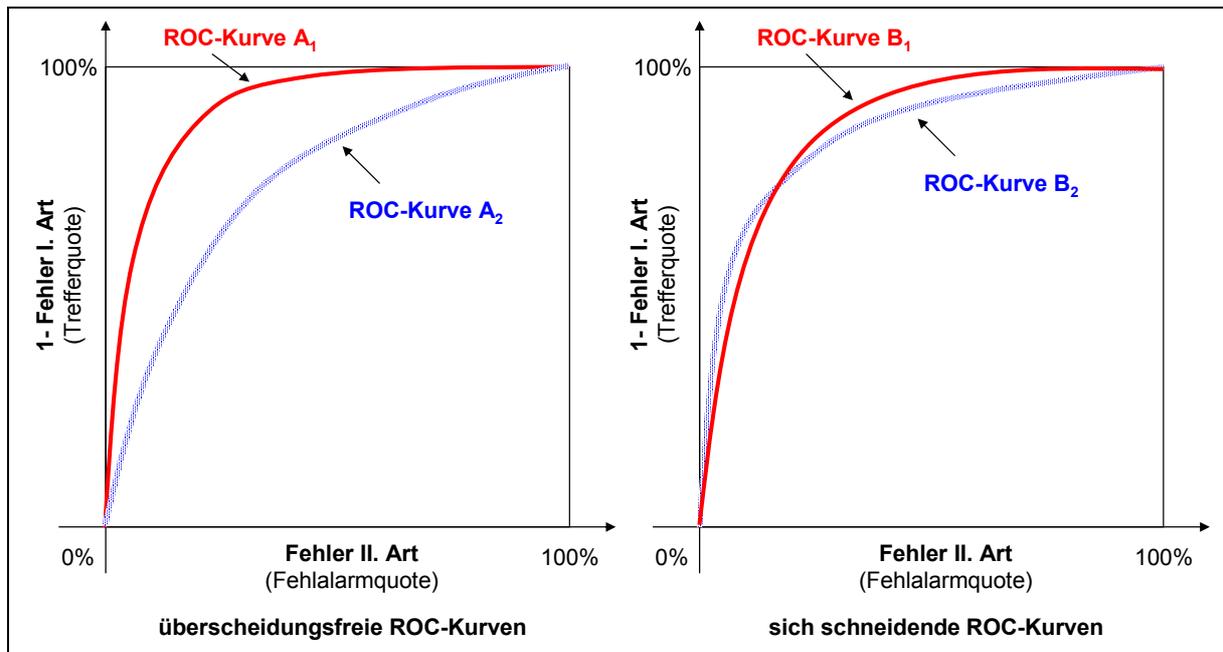


Abbildung 6: überschneidungsfreie und sich schneidende ROC-Kurven

Neben grafischen Darstellungen werden für den praktischen Einsatz bei der Messung der Schätzgüte von Insolvenzprognosen jedoch auch *Kennzahlen* benötigt, um die in CAP- oder ROC-Kurven enthaltenen Informationen möglichst kompakt darzustellen.

Als Mindestanforderungen an eine derartige Kennzahl ist zu stellen, daß diese Kennzahl für alle ROC-Kurven definiert sein sollte und daß, wenn die ROC-Kurve eines Verfahrens A_1 links oberhalb der ROC-Kurve A_2 liegt, das A_1 zugehörige Gütemaß einen größeren Wert aufweisen sollte, als das A_2 zugehörige Gütemaß.⁷⁶

Im folgenden Abschnitt sollten zwei ineinander überführbare Gütemaße vorgestellt werden, welche dieser Forderung genügen.

⁷⁵ Siehe hierzu LEHMANN (2003, S.21). Siehe auch GRUNERT, NORDEN, WEBER (2005), die allerdings nur aggregierte Kennzahlen für die unterschiedlichen Modelle (Kennzahlenrating, Softfaktorrating, kombiniertes Rating) ausweisen und keine CAP- oder ROC-Kurven. In letzterer Studie war die Prognosefähigkeit des Softfaktorratings sogar größer als die des Kennzahlenratings (siehe ebenda, S. 519). Für die LEHMANN (2003)-Studie siehe auch Abschnitt 3.5. Die GRUNERT, NORDEN, WEBER (2005)-Studie wurde aufgrund der geringen Stichprobengröße (340 Nichtausfaller, 69 Ausfaller) *nicht* in die Übersicht in Abschnitt 3.5 aufgenommen.

⁷⁶ Die Umkehrung der Mindestforderung („Wenn $A_1 > A_2$, dann liegt die ROC-Kurve von A_1 links oberhalb von A_2 “) gilt nicht. Ausgeschlossen wird lediglich, daß die A_1 zugehörige CAP-/ROC-Kurve rechts unterhalb der A_2 zugehörigen CAP-/ROC-Kurve liegt. Denkbar ist aber auch, daß sich die CAP-/ROC-Kurven von A_1 und A_2 schneiden, wenn $A_1 > A_2$. CANTOR, MANN (2003, S. 12): „Although the accuracy ratio is a good summary measure, not every increase in the accuracy ratio implies an unambiguous improvement in accuracy.“

Aus den Anstiegen der ROC- oder CAP-Kurven kann auf die relativen Ausfallquoten der erfaßten Unternehmen geschlossen werden: je steiler die Kurven verlaufen, desto höher sind die Ausfallquoten der in den Kurventeilstücken jeweils erfaßten Unternehmen.⁷⁷ Bei Kenntnis der durchschnittlichen Ausfallquote können aus den ROC- oder CAP-Kurven sogar Aussagen über die absoluten Ausfallhäufigkeiten abgeleitet werden.

Mit def ... absolute Anzahl von Ausfallern und $ndef$... absolute Anzahl von Nicht-Ausfallern müssen an einem bestimmten Punkt x der CAP-Kurve $\Delta x \cdot (def+ndef)$ zusätzliche Unternehmen ausgeschlossen werden, die *absolute Anzahl* an korrekt klassifizierten Ausfallern um $\Delta CAP(x) \cdot def$ zu erhöhen. Somit ergibt sich:

$$F 1) PD(x) = \frac{\Delta CAP(x) \cdot def}{\Delta x \cdot (def + ndef)}$$

$$F 2) PD(x) = \frac{\partial CAP(x)}{\partial x} \cdot PD$$

mit $\partial CAP(x)/\partial x =$ Tangens des Anstiegs der CAP-Kurve am Punkt x ,

PD ...durchschnittliche Stichprobenausfallrate

$PD(x)$...Ausfallrate am Punkt x [hier: der CAP-Kurve]

Ein analoges Vorgehen bei der ROC-Kurve ergibt:

$$F 3) PD(x) = \frac{\Delta ROC(x) \cdot def}{\Delta x \cdot ndef + \Delta ROC(x) \cdot def}$$

$$F 4) PD(x) = \frac{\Delta ROC(x) \cdot PD}{\Delta x \cdot (1 - PD) + \Delta ROC(x) \cdot PD}$$

$$F 5) PD(x) = \frac{PD}{\frac{\Delta x \cdot (1 - PD)}{\Delta ROC(x)} + PD}$$

$$F 6) PD(x) = \frac{PD}{(1 - PD) \cdot \left(\frac{1}{\left(\frac{\partial ROC(x)}{\partial x} \right)} \right) + PD}$$

mit $\partial ROC(x)/\partial x =$ Tangens des Anstiegs der ROC-Kurve am Punkt x ,

Besteht eine ROC- oder CAP-Kurve aus linearen Teilstücken, wie in Abbildung 5, so bedeutet dies, daß *innerhalb* der einzelnen Intervalle keinerlei Risikodifferenzierung zwischen den darin enthaltenen Unternehmen vorgenommen wird, d.h. für alle Unternehmen des Intervalls sieht sich der Anwender einem identischen Ausgleich zwischen Fehlern I. und II. Art – und damit einer identischen Ausfallwahrscheinlichkeit – gegenüber. Ursache hierfür ist möglicherweise, daß die entsprechende ROC- oder CAP-Kurven nicht auf kontinuierlichen Scores basiert sondern auf diskreten Ratingklassen.⁷⁸

⁷⁷ siehe auch FALKENSTEIN, BORAL, CARTY (2003/2000, S. 30f)

⁷⁸ Siehe hierzu beispielsweise S&P(2005, S.19ff.). In den dort abgebildeten CAP-Kurven werden zu den linearen Teilstücken auch die zugehörigen Ratingklassen angegeben.

2.3.2 Quantitative Schätzgütebestimmung

Das am häufigsten im Zusammenhang mit ROC-Kurven verwendete Gütemaß ist der Inhalt der Fläche unter der ROC-Kurve:

$$\text{F 7) } AUC_{\text{ROC}} = \int_0^1 CAP(x) dx \quad (\text{AUC, area under the [ROC] curve})^{79,80}$$

Die Fläche unter der ROC-Kurve entspricht der Wahrscheinlichkeit, mit der für zwei zufällig ausgewählte Individuen, wobei jeweils eines aus den Ausfall- und eines aus den Nicht-Ausfallunternehmen gezogen wurde, das Nicht-Ausfall-Unternehmen richtigerweise einen besseren Scorewert als das Ausfallunternehmen aufweist.⁸¹ Bei einem perfekten Ratingsystem beträgt AUC stets 100%, bei einer rein zufälligen Scorevergabe (naives Ratingsystem) liegt der Erwartungswert für AUC bei 50%.^{82,83}

Um den Wertebereich des Gütemaßes auf das Intervall [-100%; +100%] anstelle [0%; 100%] zu skalieren und um „naive Prognosen“ mit einem Erwartungswert von 0% anstelle von 50% auszuweisen, wird auf Basis von AUC_{ROC} die Kenngröße Accuracy Ratio⁸⁴ (AR_{ROC}) wie folgt ermittelt:

$$\text{F 8) } AR_{\text{ROC}} = 2 \cdot (AUC_{\text{ROC}} - 0,5).$$

Die Kenngröße AR läßt sich als einen Spezialfall anderer, gebräuchlicher ordinaler Kenngrößen darstellen.⁸⁵

Im Gegensatz zum Vorgehen bei ROC-Kurven ist es *nicht* üblich, die auf Basis von CAP-Kurven berechneten AUC-Maße auszuweisen. Üblich ist hier nur die gegenüber der ROC-Kurve leicht modifizierte Berechnung der CAP-Accuracy Ratio mit:

$$\text{F 9) } AR_{\text{CAP}} = A'/B'$$

wobei A' für die Fläche zwischen CAP-Kurve und Diagonale steht und B' für die Fläche zwischen der Diagonale und der CAP-Kurve, die ein perfektes Ratingverfahren maximal erzielen könnte (siehe auch Abbildung 5, rechts). Es läßt sich zeigen, daß AR_{ROC} und AR_{CAP} identisch sind.⁸⁶

⁷⁹ siehe beispielsweise DEUTSCHE BUNDESBANK (2003, S.71ff.)

⁸⁰ Für weitere, gut interpretierbare Kennzahlen siehe LEE (1999).

⁸¹ Siehe LEE (1999, S. 455)

⁸² Der niedrigste mögliche Wert, 0%, würde von einem Ratingsystem erzielt, dessen Prognosen *immer* falsch sind. Durch Negation der Prognosen ließe sich dieses Rating jedoch leicht in ein perfektes Rating überführen.

⁸³ Eine alternative Bezeichnung für AUC_{ROC} ist CoC ... Coefficient of Concordance, vergleiche LEHMANN (2003, S. 12).

⁸⁴ Weitere Bezeichnung für Accuracy Ratio sind: *Gini-Index* (siehe BLOCHWITZ, LIEBIG, NYBERG (2000)), *LORENZ-MÜNZER-Konzentrationsmaß* (siehe DVFA (2004, S. 599), *Power-Statistic* (siehe FAHRMEIR, HENKING, HÜLS (2000, S. 27))

⁸⁵ Siehe hierzu HAMERLE, RAUHMEIER, RÖSCH (2003, S. 21f.), die zeigen, daß die Accuracy Ratio ein Spezialfall der allgemeineren Kennzahl SOMER'S D ist, die auch für zu erklärende Variablen mit mehr als zwei möglichen Ausprägungen (Ausfall vs. Nichtausfall) definiert ist. Siehe auch SOMER (1962, S. 804f.) für eine Darstellung der Relation von SOMER'S D zu anderen ordinalen Gütemaßen wie KENDALL'S tau oder GOODMAN and KRUSKAL'S gamma.

⁸⁶ siehe ENGELMANN, HAYDEN, TASCHE (2003, S. 23)

Es gilt:

$$\text{F 10) } AR_{\text{CAP}} = \frac{\int_0^1 \text{CAP}(x) dx - \frac{1}{2}}{\int_0^1 \text{CAP}_{\text{perfekt}}(x) dx - \frac{1}{2}} \quad \text{mit}$$

$$\text{F 11) } \int_0^1 \text{CAP}_{\text{perfekt}}(x) dx = 1 - \frac{\text{PD}}{2} \quad \text{folgt}$$

$$\text{F 12) } AR_{\text{CAP}} = \frac{\int_0^1 \text{CAP}(x) dx - \frac{1}{2}}{1 - \frac{\text{PD}}{2} - \frac{1}{2}}$$

$$\text{F 13) } AR_{\text{CAP}} = \frac{2 \cdot \int_0^1 \text{CAP}(x) dx - 1}{1 - \text{PD}}$$

Für ein Ratingsystem mit g diskreten Klassen läßt sich das Integral $\int \text{CAP}(x) dx$ in je g drei- und rechteckige Teilflächen mit bekannten Seitenlängen zerlegen (siehe Abbildung 7). In der Abbildung wird der Flächeninhalt für jeweils eine rechteckige und eine dreieckige Fläche angegeben.

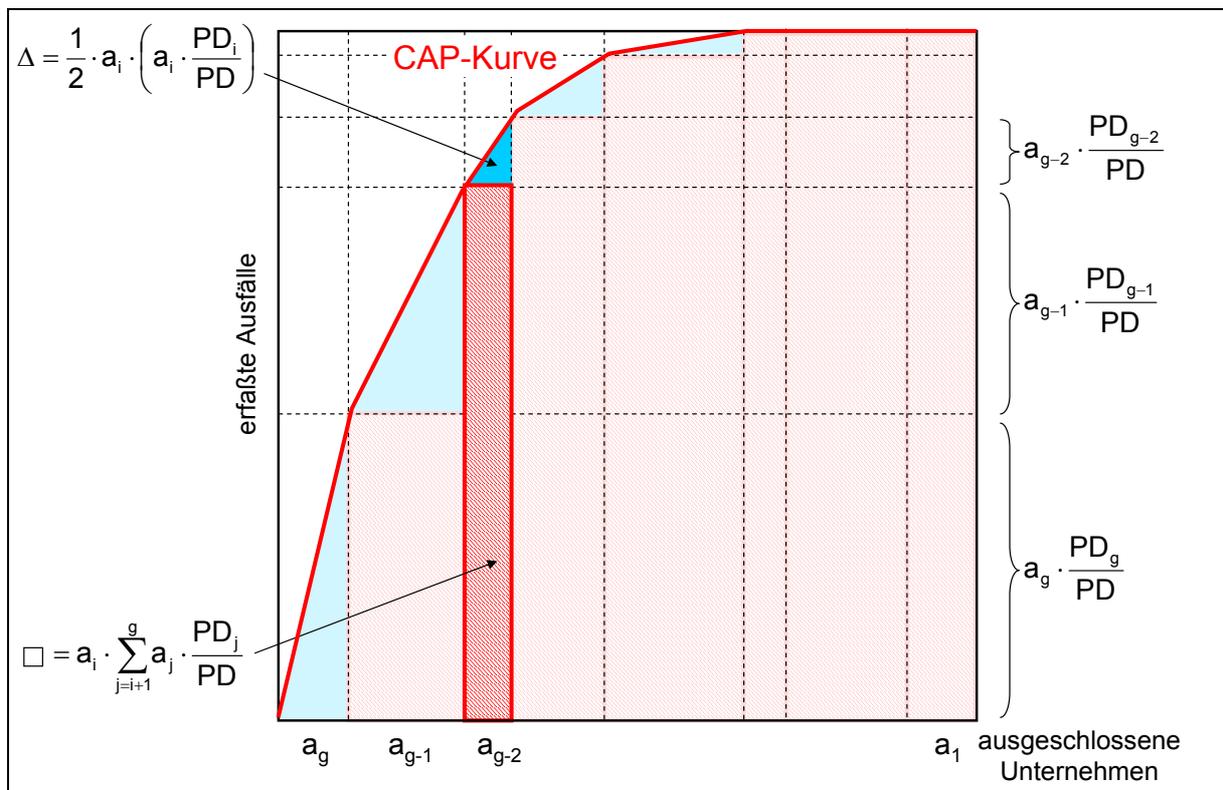


Abbildung 7: Vertikale Dekomposition der Fläche unter der CAP-Kurve

Insgesamt ergibt sich damit:

$$\text{F 14) } \int_0^1 \text{CAP}(x) dx = \sum_{i=1}^g \left(\frac{1}{2} \cdot a_i \cdot \left(a_i \cdot \frac{\text{PD}_i}{\text{PD}} \right) + a_i \cdot \left(\sum_{j=i+1}^g a_j \cdot \frac{\text{PD}_j}{\text{PD}} \right) \right) \quad \text{mit}$$

- g...Anzahl der diskreten Ratingklassen,
- a_j ... Anteil der Unternehmen in Ratingklasse j,
- PD_j ... realisierte Ausfallhäufigkeit in Ratingklasse j,
- PD .. durchschnittliche Ausfallwahrscheinlichkeit

Unter der defintitorischen Gleichung

$$F 15) \text{ kumPD}_i \equiv \frac{\sum_{j=1}^i a_j \cdot PD_j}{PD} \quad \text{wobei } \text{kumPD}_i \text{ für den Anteil der Ausfaller in den Ratingklassen } 1..i \text{ an der Gesamtheit aller Ausfaller steht und}$$

$$F 16) \frac{a_i \cdot PD_i}{PD} = \text{kumPD}_i - \text{kumPD}_{i-1} \text{ sowie}$$

$$F 17) \text{kumPD}_g = 1 \quad \text{ergibt sich eingesetzt in Formel F 14 ergibt sich somit}$$

$$F 18) \int_0^1 CAP(x)dx = \sum_{i=1}^g a_i \cdot \left(\frac{\text{kumPD}_i - \text{kumPD}_{i-1}}{2} + (1 - \text{kumPD}_i) \right) \text{ und damit}$$

$$F 19) \int_0^1 CAP(x)dx = \sum_{i=1}^g a_i \cdot \left(1 - \frac{\text{kumPD}_i + \text{kumPD}_{i-1}}{2} \right) \quad \text{mit } \sum_{i=1}^g a_i = 1$$

$$F 20) \int_0^1 CAP(x)dx = 1 - \frac{1}{2} \sum_{i=1}^g a_i \cdot (\text{kumPD}_i + \text{kumPD}_{i-1})$$

Eingesetzt in Formel F 13 ergibt sich:

$$F 21) AR_{CAP} = \frac{2 \cdot \left(1 - \frac{1}{2} \sum_{i=1}^g a_i \cdot (\text{kumPD}_i + \text{kumPD}_{i-1}) \right) - 1}{1 - PD} \quad \text{und damit}$$

$$F 22) \boxed{AR_{CAP} = \frac{1 - \sum_{i=1}^g a_i \cdot (\text{kumPD}_i + \text{kumPD}_{i-1})}{1 - PD}}$$

Die vorgestellten Gütemaße sind gleichermaßen für kontinuierliche Scores wie für Scores mit einer endlichen Anzahl möglicher Ausprägungen (*Ratingklassen*) definiert.⁸⁷ Die Zusammenfassung kontinuierlicher Scores zu diskreten Ratingklassen ist mit nur mit geringfügigen Informationsverlusten verbunden (siehe Abschnitt 2.3.4), so daß die Qualität ordinaler Insolvenzprognosen allein anhand der relativen Häufigkeiten der einzelnen Ratingklassen sowie der Angaben zu den realisierten Ausfallquoten je Ratingklasse bestimmt werden kann. Diese Daten werden beispielsweise von den Ratingagenturen S&P und MOODY'S veröffentlicht, sind aber auch für andere Ratingverfahren, beispielsweise den CREDITREFORM-Bonitätsindex und für zahlreiche andere Verfahren verfügbar (siehe auch die Abschnitte 3.4 und 3.5).

Hinweis: Den Veröffentlichungen der Ratingagentur MOODY'S liegt eine von obigen Darstellungen abweichende Accuracy-Ratio-Defintion zugrunde (welche zu *niedrigeren* Werten führt als gemäß der üblichen Definition, insbesondere bei Mehrjahresprognosen):⁸⁸

⁸⁷ Im Fall kontinuierlicher Scores entspräche g der Anzahl der gerateten Untenehmen mit $a_i=1/g$ für alle i, d.h. jedes Unternehmen würde eine separate Ratingklasse bilden.

⁸⁸ siehe beispielsweise MOODY'S (2005, S.11)

$$F\ 23) \ AR_{CAP,MOODY'S} = A' / 0,5$$

Die MOODY'S-Accuracy-Ratio ändert sich somit nur dann, wenn sich auch die zugrundeliegende CAP-Kurve ändert; d.h. sie ist invariant gegenüber simultanen, proportionalen Änderungen der Ausfallwahrscheinlichkeiten sämtlicher Ratingklassen. Hiervon versprechen sich die Autoren eine im Zeitverlauf stärkere Unabhängigkeit der Accuracy Ratio von der durchschnittlichen Ausfallrate.⁸⁹

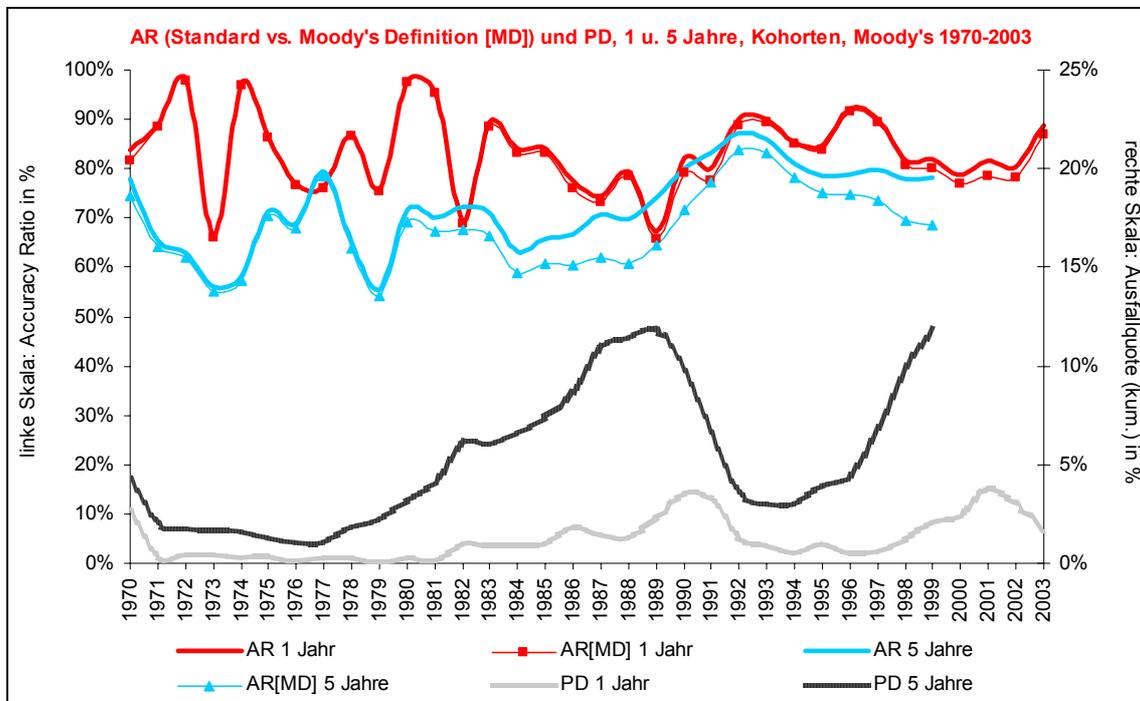


Abbildung 8: Zeitreihen Accuracy Ratio (Standard vs. MOODY'S Definition [MD]) und Ausfallraten, MOODY'S Kohorten, 1970-2003, Prognosehorizont: 1 Jahr und 5 Jahre

Die Forderung nach Invarianz der Accuracy Ratio gegenüber proportionalen Änderungen der Ausfallwahrscheinlichkeiten aller Ratingklassen kann jedoch *nicht* sinnvoll sein: ein imperfektes Ratingsystem, das 90% der Schuldner in eine Klasse mit einer Ausfallwahrscheinlichkeit von 0% und die restlichen 10% der Unternehmen mit einer Ausfallwahrscheinlichkeit von 20% einteilt, würde gemäß der MOODY'S-Kennzahl als äquivalent zu einem perfekten Rating eingestuft – das aus einer Verfünffachung aller ratingklassenspezifischen resultiert, das also ebenfalls 90% der Unternehmen in eine Klasse mit 0% Ausfallwahrscheinlichkeit und die restlichen 10% der Unternehmen in eine Klasse mit 100% Ausfallwahrscheinlichkeit einstuft (durchschnittliche Ausfallwahrscheinlichkeit: 10 %).⁹⁰ Ferner läßt sich empirisch und

⁸⁹ „The benefit of not including the perfect foresight comparison is that the accuracy ratio will be invariant to changes in the aggregate default rate and will only change due to changes in the rating distribution and the distribution of default rates. In practical terms, though, there is little empirical difference between the accuracy ratio of SOBEHART ET AL ([2000]) and that used in this Special Comment.“ CANTOR, MANN (2003, S.11) *“The accuracy ratio measures only relative accuracy, not absolute accuracy, and is invariant to proportional changes in marginal default rates. The marginal default rate is the percent of issuers in any given rating category that subsequently default. If the marginal default rates for all rating categories change proportionally, neither the CAP plot nor the accuracy ratio changes at all.”*, CANTOR, MANN (2003, S.12).

⁹⁰ Beide Ratingverfahren erzeugen die gleiche CAP-Kurve (allerdings unterschiedliche ROC-Kurven): es müssen jeweils 10% der Unternehmen ausgeschlossen werden, um 100% der Ausfälle zu erfassen, somit ergibt sich in beiden Fällen eine identische MOODY'S-Accuracy-Ratio von 90,0% $(=(10\% \cdot 100\% / 2 + 90\% \cdot 100\% - 0,5) / 0,5)$. Bei der üblichen AR-Berechnung (siehe Formel F 13) erfolgte im ersten Fall eine Normierung mit dem Faktor 1/98% und im zweiten Fall mit dem Faktor 1/90%, so daß sich die AR-Werte 91,8% und 100% ergäben. Das perfekte Rating erhielte somit einen besseren AR-Wert als das imperfekte Rating und es erhielte den höchstmöglichen AR-Wert von 100%.

theoretisch zeigen, daß - entgegen den genannten Zielsetzungen - die gemäß MOODY'S ermittelten Schätzgütekennzahlen sogar stärker (negativ) mit den durchschnittlichen Ausfallraten korrelieren als die „normalen“ Accuracy Ratios.⁹¹

Als weitere Nachteile sind zu nennen, daß die MOODY'S-modifizierte Accuracy Ratio inkompatibel zur ROC-Kurven-Accuracy Ratio ist und daß perfekte Ratingsysteme gemäß der MOODY'S AR keine Werte von 100% erreichen können. Bezüglich der 1-Jahreswerte weisen die „üblichen“ und die gemäß MOODY'S modifizierten AR-Werte zwar nur geringe Unterschiede auf (die Werte unterscheiden sich um den Faktor $1/(1-PD)$). Größere und im Zeitverlauf erheblich variierende Unterschiede sind jedoch bezüglich der Mehrjahres-Accuracy-Ratios verzeichnen (siehe auch Abbildung 8).⁹²

⁹¹ Quelle: eigene Untersuchungen. Basierend auf den MOODY'S-Daten ergibt sich im Zeitraum 1983-2003 ein Korrelationskoeffizient von -0,591 / -0,682 für den Zusammenhang zwischen der MOODY'S AR (Prognosezeitraum 1 Jahr / 5 Jahre) und den durchschnittlichen Ausfallraten der einzelnen Jahre. Die Korrelationskoeffizienten der „normalen“ AR mit den durchschnittlichen Ein- bzw. Fünfjahresausfallrate beträgt -0,494 / -0,437.

Ferner wurde ein Simulationsexperiment mit 10.000 Simulationsläufen basierend auf einem Unternehmensportfolio mit 5.000 Unternehmen durchgeführt (was ungefähr der derzeitigen Anzahl von Unternehmen entspricht, die ein über ein S&P Rating verfügen, siehe S&P (2005, S. 26)), deren Verteilung über die Ratingklassen und deren Ausfallwahrscheinlichkeiten (5 Jahre, kumuliert) den Ausfallstatistiken von S&P entnommen wurden. Im Gegensatz zu den empirischen Studien kann im Simulationsexperiment sichergestellt werden, daß die ratingklassenspezifischen Ausfallwahrscheinlichkeiten (im Zeitverlauf bzw. zwischen den einzelnen Simulationsläufen) konstant sind. In jedem Simulationslauf wurden die Zufallsrealisierungen der individuellen Ausfälle und die darauf basierenden Realisierungen der durchschnittlichen Ausfallraten und der Accuracy Ratios ermittelt. Während die „normalen“ AR und die durchschnittlichen Ausfallraten der 10.000 Simulationsläufe praktisch unkorreliert waren, betrug die Korrelation der MOODY'S-Accuracy Ratio mit den durchschnittlichen Ausfallraten -18,3%.

⁹² Aufgrund der zu Beginn der 70er Jahre sehr geringen Anzahl gerateter Unternehmen weisen die Accuracy-Ratio-Werte zunächst sehr große, im Zeitverlauf abnehmende jährliche Schwankungen auf. 1970 betrug die Zahl der von MOODY'S gerateten Unternehmen nur 1.031, 2004 hingegen fast fünfmal so viel (4.810), siehe MOODY'S (2005, S.22f.).

2.3.3 Vergleichbarkeit mit Gütemaßen kategorialer Insolvenzprognosen

Wie in den Abschnitten 2.3.1 und 2.3.2 dargestellt, basieren letztlich nicht nur Gütemaße für kategoriale Insolvenzprognosen sondern auch ROC- /CAP-Kurven und die hieraus abgeleiteten Kennzahlen für ordinale Insolvenzprognosen, wie Area-under-the-curve (AUC) und Accuracy Ratio (AR), auf einer Analyse der Fehler I. und II. Art.⁹³

Auch wenn sich für jede Kombination von Fehlern I./II. Art unendlich viele ROC-Kurven mit unterschiedlichen AUC-Werten finden lassen welche diese Kombination beinhalten⁹⁴, lassen sich zumindest exakte Ober- und Untergrenzen der mit diesen ROC-Kurven verbundenen Schätzgütemaße (AUC, AR) angeben (für eine Herleitung der Ergebnisse dieses Abschnitts siehe Anhang I). Da die durch die Ober- und Untergrenzen aufgespannten Intervalle in der Regel sehr groß sind, wurde zusätzlich ein heuristischer Schätzer für die Accuracy-Ratio entwickelt (und empirisch getestet), der eine bestimmte funktionale Form der ROC-Kurven unterstellt. Dies hat den Vorteil, daß sich auch die Ergebnisse älterer Insolvenzprognosestudien (siehe hierzu die Übersichten in den Abschnitten 3.3 und 3.4), die häufig nur einzelne Kombinationen von Fehlern I. und II. Art vermelden oder die nur einzelne CAP-Koordinaten angeben auf einer *einwertigen* Basis untereinander⁹⁵ und auf einer *einheitlichen* Basis mit den Ergebnissen neuerer Studien vergleichen lassen.

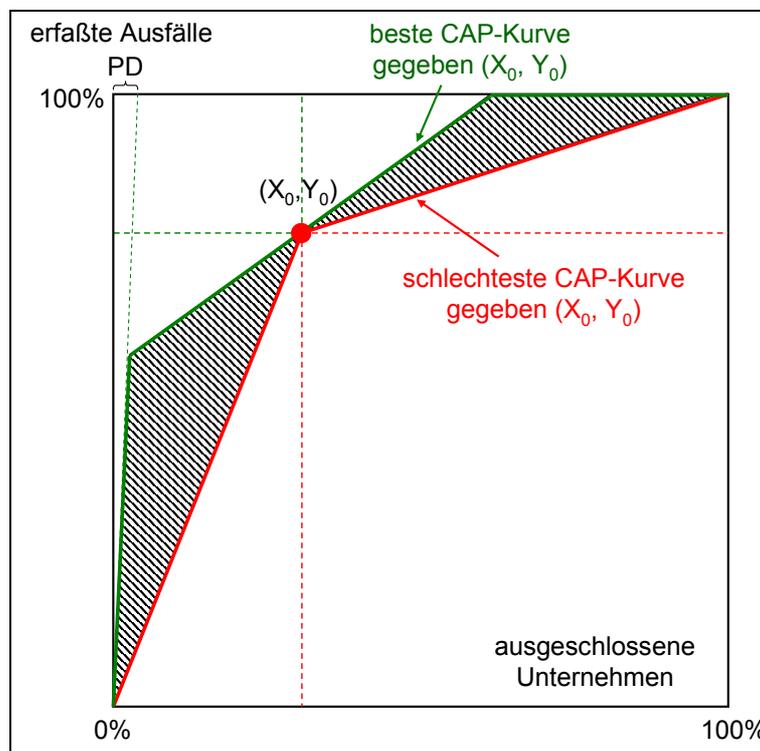


Abbildung 9: beste und schlechteste CAP-Kurve gegeben (X_0, Y_0)

Unter plausiblen Annahmen läßt sich zeigen, daß die „schlechtestmögliche“, durch einen gegebenen Punkt (X_0, Y_0) im CAP-Diagramm verlaufende CAP-Kurve der Linienzug $[0\%; 0\%]$

⁹³ Im Gegensatz zur Beurteilung kategorialer Insolvenzprognosen erfolgt die Beurteilung ordinaler Prognosen jedoch nicht anhand einer einzigen Kombination der Fehler I./II. Art sondern anhand der Gesamtheit aller Kombinationen von Fehlern I./II. Art, die mit der Anwendung eines Insolvenzprognoseverfahrens bei Zugrundelegung verschiedener Trennwerte erzeugt werden können.

⁹⁴ Einzige Ausnahme ist die Fehler-I/II-Kombination $(0\%; 0\%)$, die nur von der „perfekten“ (entlang der Außenseiten des ROC-Diagramms verlaufenden) ROC-Kurve beinhaltet wird.

⁹⁵ Ein weiterer Vorteil ist, daß selbst ein Vergleich *zwischen* älteren Insolvenzprognosestudien und aktuellen Übersichtsstudien erleichtert wird, da die zweiwertigen Fehler-I-II-Kombinationen in einwertige Gütemaße umgewandelt werden.

– $[X_0; Y_0] - [100\%; 100\%]$ ist. Diese CAP-Kurve korrespondiert mit einem ordinalen Prognoseverfahren, das nur zwischen den beiden durch X_0 und Y_0 getrennten Gruppen differenzieren kann, innerhalb der Gruppen jedoch keinerlei Trennfähigkeit besitzt.

Die „bestmögliche“ durch $(X_0; Y_0)$ verlaufende CAP-Kurve muß folgende Bedingungen erfüllen: ausgehend von $[0\%; 0\%]$ verläuft sie zunächst entlang der gestrichelten PD-Gerade, d.h. die zuerst ausgeschlossenen Unternehmen weisen eine Ausfallwahrscheinlichkeit von 100% auf, anschließend verläuft sie linear durch $(X_0; Y_0)$ bis sie die 100%-Trefferquoten-Gerade schneidet (siehe Abbildung 9). Anstieg und Absolutglied dieser Gerade werden durch einen Optimierungsansatz ermittelt. Anschließend verläuft die CAP-Kurve waagrecht bis sie $(100\%; 100\%)$ erreicht, d.h. die zuletzt ausgeschlossenen Unternehmen haben eine Ausfallwahrscheinlichkeit von 0%.

Die Formeln für die Ermittlung der best- und schlechtestmöglichen AR-Werte, gegeben eine Fehlerkombination I. und II. Art oder gegeben ein CAP-Koordinatenpaar $(X_0; Y_0)$, werden in Tabelle A dargestellt. Neben den Ober- und Untergrenzen für die Accuracy-Ratio, die basierend auf *einer* Fehler-I-II-Kombination ermittelt werden, werden hier auch vier heuristische Schätzer für die Accuracy-Ratio angegeben: AR_{MW} (Mittelwert aus Ober- und Untergrenze der erreichbaren Accuracy Ratio), AR_α , AR_β und $AR_{\alpha\&\beta}$, dem Mittelwert aus AR_α und AR_β . Zur Herleitung zu empirischen Untersuchungen zur Eignung dieser Kenngrößen siehe Anhang I und IV. Bei Fehler-I-II-Kombinationen mit identischen Fehlern I. und II. Art liefern alle vier Verfahren sehr präzise Schätzwerte für die tatsächliche Accuracy Ratio, bei asymmetrischen Fehlern I. und II. Art werden die besten Ergebnisse mit $AR_{\alpha\&\beta}$ erzielt.

	CAP-Koordinaten	Fehler I. und II. Art
AR_{min}	$\frac{Y_0 - X_0}{1 - PD}$	$1 - (F_1 + F_2)$
AR_{max}	$\frac{1 - PD - 4X_0 + 4 \cdot X_0 \cdot Y_0 + 4 \cdot PD \cdot Y_0 - 4PD \cdot Y_0^2}{1 - PD}$	$1 - 4 \cdot F_1 \cdot F_2$
AR_{MW}	$\frac{AR_{max} + AR_{min}}{2}$	$\frac{AR_{max} + AR_{min}}{2}$
AR_α	$\frac{\ln(X_0 - Y_0 \cdot PD) - \ln Y_0 - \ln(1 - PD)}{\ln(X_0 - Y_0 \cdot PD) + \ln Y_0 - \ln(1 - PD)}$	$\frac{\ln F_2 - \ln(1 - F_1)}{\ln F_2 + \ln(1 - F_1)}$
AR_β	$\frac{\ln(1 - Y_0) - \ln(1 - PD - X_0 + Y_0 \cdot PD) + \ln(1 - PD)}{\ln(1 - Y_0) + \ln(1 - PD - X_0 + Y_0 \cdot PD) - \ln(1 - PD)}$	$\frac{\ln F_1 - \ln(1 - F_2)}{\ln F_1 + \ln(1 - F_2)}$
$AR_{\alpha\&\beta}$	$\frac{AR_\alpha + AR_\beta}{2}$	$\frac{AR_\alpha + AR_\beta}{2}$

Tabelle A: Formeln zur Ermittlung exakter Ober- und Untergrenzen und heuristischer Schätzer für die Accuracy Ratio basierend auf CAP-Koordinaten/ Fehler I. und II. Art

Beispiel: In einer Studie von ALTMAN (1968, S. 599; siehe auch Abschnitt 3.4), wird ein Insolvenzprognoseverfahren vorgestellt, das bei einer Stichprobe von je 33 insolventen und nicht-insolventen Unternehmen mit einem Prognosehorizont von einem Jahr 31 der insolventen und 32 der nicht-insolventen Unternehmen richtig klassifizieren konnte. Der Fehler I. Art beträgt somit 6,1%, der Fehler II. Art 3,0%. Die Accuracy Ratio eines Verfahren, das so niedrige Fehler I. und II. Art liefert, liegt damit zwischen $1 - (6,1\% + 3,0\%) = 90,9\%$ und $1 - 4 \cdot 6,1\% \cdot 3,0\% = 99,3\%$ (Mittelwert Ober- und Untergrenze: 95%), $AR_\alpha = 96,5\%$ und $AR_\beta = 97,8\%$, $AR_{\alpha\&\beta} = 97,2\%$. Im Gegensatz zu vielen anderen Studien wird bei ALTMAN

(1968, S. 603) zusätzlich eine vollständige Übersicht der erzielten Scorewerte für alle prognostizierten Unternehmen angeben, so daß *hier* sogar die vollständige ROC-Kurve des Verfahrens reproduziert werden und der zugehörige Accuracy-Ratio-Wert direkt ermittelt werden können. Als exakter AR-Wert ergibt sich hier 98,7%.

Zum Vergleich: die Einjahres-Schätzgütewerte (Accuracy Ratio) der Ratings von S&P und MOODY'S betragen durchschnittlich „nur“ ca. 85% (siehe auch Abschnitt 3.5).⁹⁶ Allerdings basieren die Schätzgütewerte für ALTMAN'S Verfahren auf der Analyse einer sehr kleinen und sehr alten Stichprobe (die Unternehmensdaten stammen aus den Jahren 1946-1965 (ebenda, S.593)). Des weiteren wurden die Parameter des von ALTMAN verwendeten Modells auf Basis der gleichen Stichprobe kalibriert auf der anschließend die Prognosegüte ermittelt wurde („In-Sample“), so daß es sich bei obigen Schätzgüteergebnissen nicht um *Prognosen* im eigentlichen Sinne handelt, sondern um *nachträgliche* Klassifikationen („ex-post Prognosen“).

In späteren Vergleichsstudien anderer Autoren erzielte das ALTMAN-Verfahren deutlich schlechtere Schätzergebnisse (siehe Abschnitt 3.4).

⁹⁶ Die erreichten Accuracy-Ratio-Werte von ca. 85% beziehen sich auf die Grundgesamtheiten aller von S&P oder MOODY'S gerateten Unternehmen. Die Schätzgüte der Agenturratings bei US-amerikanischen Unternehmen des Verarbeitenden Gewerbes, siehe die Stichprobendefinition von Altman (1968) in Abschnitt 3.4, ist deutlich schlechter, siehe hierzu auch Abschnitt 3.5.

2.3.4 Schätzgüteeinbußen durch Verwendung diskreter Ratingskalen

Von Ratingagenturen und Banken werden kontinuierliche (oder „quasikontinuierliche“⁹⁷) Scores in diskrete Ratingklassen zusammengefaßt, beispielsweise um historische Ausfall- oder Migrationshäufigkeiten zu ermitteln.^{98,99,100} Aus Sicht externer Anwender, denen nur die Ratingklasseninformationen und nicht die kontinuierlichen Scores selbst zur Verfügung stehen, ist dies jedoch mit einem Qualitätsverlust der Insolvenzprognosen verbunden, der sich in niedrigeren Schätzgütwerten niederschlägt, da sämtliche Informationen über die relativen Ausfallwahrscheinlichkeiten von Unternehmen *der gleichen* Ratingklasse verloren gehen.
101,102

Untersuchungen zur Quantifizierung der Größenordnungen dieser Informationsverluste sind nicht bekannt. In zwei Untersuchungen wird mittels empirischer Daten und Simulationsstudien lediglich gezeigt, daß die Verwendung kontinuierlicher Scores statt diskreter Ratingklassen

⁹⁷ Beispielsweise können mit einem von MOODY'S-KMV entwickelten Ansatz bis zu 1.000 unterschiedliche „Ratingklassen“ (Ausfallwahrscheinlichkeitenintervalle) differenziert werden, siehe KEENAN, SOBEHART (1999, S.12)). Mit dem CREDITREFORM-Bonitätsindex können bis zu 500 verschiedene „Ratingklassen“ unterschieden werden, siehe SCHWAIGER (2002, S.16). Die KMU-Ratingmodellfamilie RiskCalc, für die zahlreiche lokale Versionen existieren, liefert sogar *kontinuierliche* Modellergebnisse (Ausfallwahrscheinlichkeiten), siehe beispielsweise KOCAGIL ET AL. (2003, S. 30).

⁹⁸ Von den 46 größten US Finanzinstituten hatten Ende der 90'er Jahre nur vier Institute Ratingsysteme mit mehr als acht Ratingstufen für performende Kredite und vier Banken hatten sogar nur ein bis drei Ratingstufen. Hinzu kommt, daß ein Drittel der untersuchten Banken mehr als die Hälfte, teilweise sogar mehr als 80% ihrer Kunden einer einzigen Ratingkategorie zuordnete. Nur bei 15% aller Banken waren weniger als ein Drittel der Kunden in der am stärksten belegten Kategorie vertreten, siehe TREACY, CAREY (2000/1998, S. 902).

Basierend auf einer Untersuchung von mehr als einhundert US-Geschäftsbanken unterschiedlicher Größen finden ENGLISH, NELSON (1998, S. 5), daß bei den Banken ca. 2/3 aller Neukunden in die jeweils stärkste Ratingklasse eingeordnet werden, wobei allerdings eine von der Zentralbank vorgegebene Fünfklassenskala vorgegeben wurde.

Gemäß aktuelleren Untersuchungen haben deutsche Geschäftsbanken derzeit bankinterne Ratingverfahren mit 8 Ratingstufen (Dresdner Bank, 2002), 12 RS (Commerzbank, 2003), 25 RS (Volks- und Raiffeisenbanken, 2002), siehe FISCHER (2004, S. 165 und die dort zitierte Literatur).

⁹⁹ “Internal rating systems with larger numbers of grades are more costly to operate because of the extra work required to distinguish finer degrees of risk. Banks making heavy use of ratings in analytical activities are most likely to choose to bear these costs because fine distinctions are especially valuable in such activities (however, at least a moderate number of Pass grades is useful even for internal reporting purposes).”
TREACY, CAREY (2000/1998, S. 902)

¹⁰⁰ KRAHNEN, WEBER (2001, S.13): “[...] the central question for the definition of a rating system now remains, how fine a rating system should be, i.e., how many categories it should have. It could be as fine as the POD itself, being basically identical to POD, or it could map PODs into a finite number of categories. Of course, a rating system which models POD would be the most exact one. However, for quite a number of situations a less fine rating system would be sufficient and more appropriate in an organizational context. The fineness of a rating system cannot be considered independently from Backtesting [...]. There is no use in defining a large number of rating categories, if a bank is not able to back-test consistently, due to lack of data.”, siehe auch Basler Ausschuß (2001, S. 15): “There are attractions to using a pooled PD concept [instead of using issuer specific probabilities of default] in terms of its interaction with other elements of the IRB framework. [...] An example is the granularity adjustment.”

¹⁰¹ MCQUOWN (1993, S. 8): “There are 19 different gradations at S&P, including the pluses and minuses. S&P's precision could be, therefore, no greater than 1 in 19. [...] We suspect that the resolution of EDFs may be nearer 1 in 100. Banks, typically, use fewer than ten gradations, of which three may be non-performing.”

¹⁰² MILLER (1998): “While the intensive quantification of market risk has led to measurements accurate to the basis point (and beyond), difficulties in quantifying credit risk have resulted in the practice of measuring this risk with far less precision. Indeed, financial institutions that develop their own internal measures of credit risk usually employ a ‘1’ to ‘9’ scale of creditworthiness for their exposures, [...] Even with the further refinement of ‘notches’ designated with a ‘+’ or ‘-’ the vast universe of credit risk is reduced to at most thirty buckets.”

– unter bestimmten Umständen - keine *statistisch signifikanten* Verbesserungen der Schätzgüte der Verfahren zur Folge hat.^{103,104} Diese Ergebnisse sind insofern unbefriedigend, als daß sich *bei Simulationsstudien* durch die Wahl der Stichprobengröße beliebig große oder kleine Unterschiede hinsichtlich der interessierenden Größen (hier: Abhängigkeit der Schätzgüte eines Verfahrens von der Feinheit der verwendeten Ratingskala) als *statistisch signifikant* oder *statistisch insignifikant* identifizieren lassen können.¹⁰⁵

In Anhang IV werden zwei Verfahren vorgestellt, mit denen die erwartete Größenordnung der Beeinträchtigung der Schätzgüte eines Verfahrens bei Diskretisierung kontinuierlicher Scorewerte abgeschätzt wird.

Zunächst vermittelt jedoch Abbildung 10 anhand eines empirischen Beispiels eine erste Vorstellung über die Relevanz des Themas. In dieser Abbildung sind die Schätzgütwerte der Ratings von MOODY'S (Accuracy Ratio, 1 Jahr Prognosehorizont) für den Zeitraum 1970-2003 in drei unterschiedlichen Aggregationsniveaus dargestellt:¹⁰⁶

- 2-er Skala: Einteilung der Unternehmen in zwei Gruppen („Investment Grade“ (Aaa-Baa3) vs. Speculative Grade (Ba1-Caa/C)),
- 7er-Skala: „einfache Ratingbuchstaben“ (Aaa ... Caa/C),
- 17er-Skala: modifizierte Ratingbuchstaben (Aaa, Aa1, Aa2, A3, ..., CCC/C).

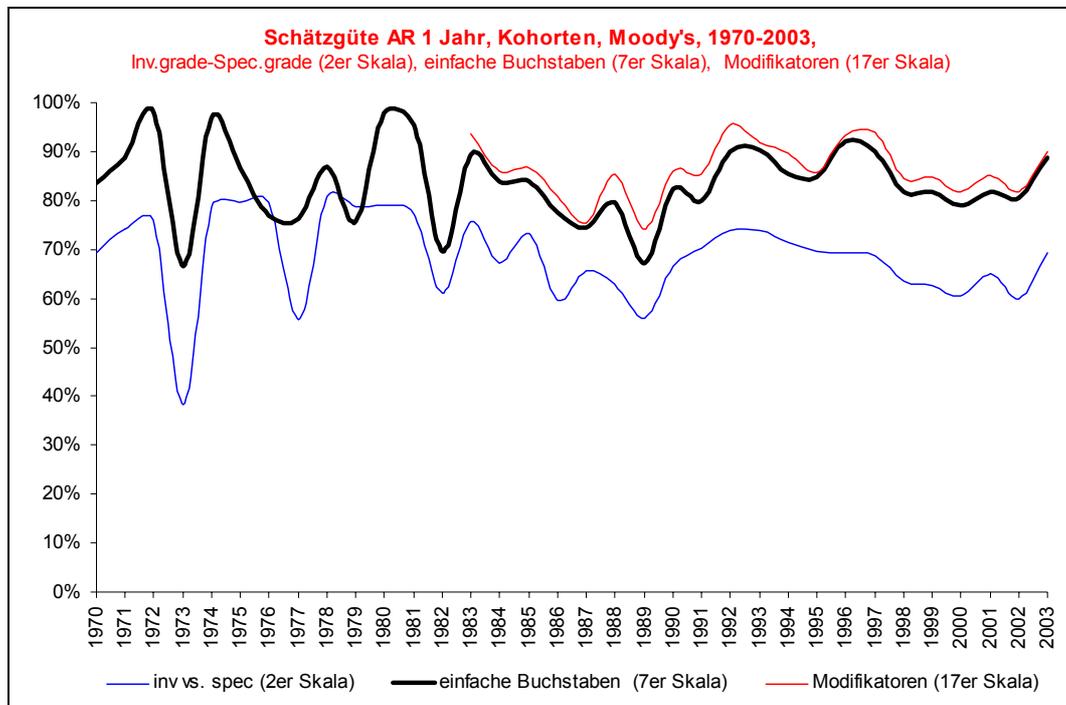


Abbildung 10: Schätzgüte MOODY'S Ratings (1 Jahr Prognosezeitraum, Accuracy Ratio), 1970-2003, bei unterschiedlich feinen Ratingskalen (Quelle: eigene Berechnungen)

¹⁰³ siehe KEENAN/ SOBEHART (1999, S. 12f.), FRERICHS, WAHRENBURG (2003, insb. S. 3, 35)

¹⁰⁴ Eine weitere Studie untersucht den Zusammenhang zwischen Feinheit der Ratingskala und regulatorischen Eigenkapitalanforderungen, siehe JANKOWITSCH, PICHLER, SCHWAIGER (2003).

¹⁰⁵ Siehe zur Unzweckmäßigkeit des statistischen Signifikanzbegriffs im Rahmen von Simulationsmodellen SCHMEISER (2001, S. 43).

¹⁰⁶ Quelle: eigene Berechnungen auf Basis von Kohortenangaben von MOODY'S (2004, S. 27-38)

In allen Jahren des Zeitraums 1983-2003 waren die Schätzgütwerte des 7er Skalenratings besser als die des 2er Skalenratings¹⁰⁷ und in allen Jahren waren die Schätzgütwerte des 17er Skalenratings besser als die 7er Skalenratings.

Im Zeitraum 1983-2003 betrug die durchschnittliche ermittelte Meßgüte auf Basis des 2er-Skalen-Ratings immerhin 66,9% ($\sigma = 5,4\%$)¹⁰⁸, des 7-er-Skalen-Ratings 83,1% ($\sigma = 6,0\%$) und des 17er-Skalen-Ratings 86,2% ($\sigma = 5,7\%$).

Der Übergang vom 2er zum 7er Skalenrating, also die Hinzufügung von 5 zusätzlichen Ratingklassen, führte zu einer Erhöhung der Schätzgüte um durchschnittlich 16,2 Prozentpunkte ($\sigma = 3,8$ PP). Der Übergang vom 7er- zum 17er-Skalenrating, also die Hinzufügung zehn weiterer Ratingklassen, führte nur noch zu einer Verbesserung der Schätzgüte um 3,2 Prozentpunkte ($\sigma = 1,7$ PP). Mit den in Anhang IV dargestellten Verfahren läßt sich ferner abschätzen, daß beim Übergang von einem 17-stufigen Rating zu einem Rating mit kontinuierlicher Skala der Informationsgehalt des Ratings um weitere 0,5 Prozentpunkte zunehmen würde, bzw. umgekehrt ausgedrückt, die Verwendung einer 17-stufigen anstelle einer kontinuierlichen Skala führt im gewählten Beispiel zu erwarteten Informationsverlusten (gemessen in AR) von rund 0,5 Prozentpunkten. Die Verwendung einer 30stufigen Ratingskala hingegen würde zu Informationsverlusten von nur ca. 0,1 Prozentpunkten führen.

Die auftretenden Informationsverluste resultieren überwiegend aus Informationsverlusten, die in der schlechtesten Ratingklasse (CCC/C bzw. Caa/C) auftreten. Die Diskretisierung der Scores im Bereich der anderen Ratingklassen, insbesondere im Investmentgradebereich, hat praktisch keine Auswirkungen auf den Informationsgehalt des Ratings.

¹⁰⁷ Hierbei handelt es sich *nicht* um rein konstruktionsbedingte Eigenschaften des verwendeten Schätzgütemaßes. So waren beispielsweise in den Jahren 1976 und 1979 die 2er-Skalen Accuracy Ratio höher als die 7er-Skalen Accuracy Ratio. Beide Jahre waren durch eine außergewöhnlich niedrige Anzahl an Ausfällen geprägt (1976: 2 Ausfälle, 1979: 1 Ausfall), die zufälligerweise die beste Speculative-Grade-Klasse Ba betrafen. Die Zusammenfassung der „ganzen“ Ratingklassen AAA, Aa, A und Baa zu einer einzigen Ratingklasse, Investment Grade, führte in beiden Jahren zu keiner Verschlechterung des Informationsgehalt des Ratings, da in keiner der drei Klassen Ausfälle zu verzeichnen waren, während die Zusammenfassung der Ratingstufen Caa/C, Ba und B sogar zu einer *Verbesserung* des Prognosewerts der Ratings führten, da die *Angabe keiner Sortierung* besser ist als die *Angabe einer falschen Sortierung* [der Ausfallraten innerhalb der speculative grade-Klassen].

¹⁰⁸ Siehe auch CANTOR, MANN (2003, S. 19): “Much of the information content of the rating system comes from the ability to determine whether credits are investment grade or speculative grade.” Gemäß eigenen Untersuchungen auf Basis von *ganzen* Ratingstufen liegt die Accuracy-Ratio-maximierende dichotome Ratingklassentrennung für Moody’s Kohortendaten nicht für den ganzen Zeitraum 1971-2003 zwischen Aaa-Baa vs. Ba-Caa/C (investment grade vs. speculative grade) sondern häufig, und seit 1992 sogar ununterbrochen, zwischen Aaa-Ba vs. B-Caa/C. Auf Basis der Pooldaten liegt die trennschärfste dichotome Grenze von 1971-1999 zwischen Aaa-Baa vs. Ba-Caa/C (investment grade vs. speculative grade) und ab 2000 zwischen Aaa-Ba vs. B-Caa/C.

Auf Basis der *modifizierten Ratingklassen* liegt die AR-maximale Ratingklassentrennung für die Kohorten in den einzelnen Jahren von 1983-2003 zwischen Aaa-Baa2 vs. Baa3-Caa/C und Aaa-B1 vs. B2-Caa/C, letztere Trennung ist seit 2001 ununterbrochen AR-maximal. Für die Pooldaten liegt die AR-maximale Trennung der modifizierten Ratingklassen seit 1986 ununterbrochen bei Aaa-Ba2 vs. Ba3-Caa/C.

Soll auf Basis von Ratinginformationen eine möglichst trennscharfe (hier im Sinne von *Accuracy-Ratio-maximale*) Zweiklasseneinteilung von Unternehmen vorgenommen werden, so liefert die übliche Investment- vs. Speculative Grade-Einteilung in keinem Fall (Pools vs. Kohorten, ganze vs. modifizierte Ratingklassen) mehr optimale Ergebnisse. Statt dessen sollten mindestens die beiden „besten“ modifizierten Speculative-Grade-Klassen, Ba1 und Ba2, dem besseren der beiden Segmente zugerechnet werden. Aus Praktikabilitätsgründen (Identität der Trenngrenzen zwischen ganzen und modifizierten Ratingklassen) sollte zusätzlich mindestens auch die Klasse Ba3 dem besseren der beiden Segmente zugerechnet werden.

2.4 Schätzgütemaße für kardinale Insolvenzprognosen

Während ordinale Insolvenzprognosen lediglich eine Reihung von Unternehmen entsprechend den erwarteten Ausfallwahrscheinlichkeiten vornehmen, ordnen kardinale Insolvenzprognoseverfahren jedem Unternehmen eine Ausfallwahrscheinlichkeit zu.

Da sich Ausfallwahrscheinlichkeiten auch als Reihungskriterium interpretieren lassen, können kardinale Insolvenzprognosen hinsichtlich aller Gütekriterien bewertet werden, die für ordinale Insolvenzprognosen entwickelt wurden:^{109,110}

- *Auflösung (resolution)* mißt, wie stark differenziert die *realisierten* Ausfallquoten bezogen auf die unterschiedlichen abgegebenen Prognosen sind. Minimale Auflösung ist dann gegeben, wenn für alle Prognosen die gleichen realisierten Ausfallquoten zu verzeichnen sind. Maximale Auflösung ist dann gegeben, wenn nur 0%- oder 100%-Ausfallquoten zu verzeichnen sind (für die grafische Darstellung der *Auflösung* eines Prognoseverfahrens siehe Abbildung 3),
- *Trennfähigkeit (discrimination)* mißt, wie stark sich die Prognosen bei Ausfällen/ Nichtausfällen unterscheiden.

Zusätzlich können aber auch Kriterien geprüft werden, bei denen die ex-ante Angabe von Ausfallwahrscheinlichkeiten zwingend erforderlich ist:

- *Kalibrierung* mißt, inwiefern die einzelnen prognostizierten Ausfallwahrscheinlichkeiten mit den realisierten Ausfallquoten übereinstimmen,
- *systematische Verzerrung (unconditional bias)*: gibt an, wie stark sich die durchschnittliche prognostizierte Ausfallwahrscheinlichkeit von der tatsächlichen durchschnittlichen Ausfallquote unterscheidet,
- *Feinheit (refinement)* mißt, wie stark differenziert die Ausfallprognosen sind. Minimale Feinheit ist dann gegeben, wenn stets eine identischen Ausfallwahrscheinlichkeit prognostiziert wird; maximale Feinheit ist dann gegeben, wenn nur 0%- oder 100%-Prognosen abgegeben werden.

Kennzahlen die simultan von allen oder einigen dieser Eigenschaften kardinaler Insolvenzprognosen determiniert werden, werden im folgenden als Maße für die *Präzision (accuracy)* eines Verfahrens bezeichnet. Kennzahlen, welche die Präzision eines Prognoseverfahrens ins Verhältnis zur Präzision eines bestimmten Referenzverfahrens setzen, werden als Maße der *Relativen Präzision* (auch *skill scores* oder *relative accuracy*) bezeichnet.¹¹¹

Kennzahlen die nur einzelne der oben aufgeführten Aspekte kardinaler Insolvenzprognosen messen (insbesondere *Kalibrierung*), werden im folgenden nicht weiter diskutiert. Vertreter hierfür sind beispielsweise:¹¹²

¹⁰⁹ Siehe MURPHY, WINKLER (1992, S. 440) für die formale Definitionen der im folgenden vorgestellten Gütekriterien *resolution*, *discrimination*, *calibration*, *refinement*, *unconditional bias*, *accuracy* und *skill*.

¹¹⁰ Auch die in Abschnitt 2.3.2 vorgestellten Kenngrößen Accuracy Ratio und area-under-the-ROC-curve werden von der *Auflösung* und *Trennfähigkeit* der Prognosen beeinflusst.

¹¹¹ siehe MURPHY, WINKLER (1992, S. 440)

¹¹² Für weitere Testverfahren, u.a. Binomialtest, χ^2 -Test, Normalverteilungstest, die ausschließlich die Korrektheit der Kalibrierung einzelner oder aller Ratingklassen überprüfen, siehe BASLER AUSSCHUB (2005, S. 47ff.).

- *Gruppiertes Brierscore*^{113, 114} :

$$\text{F 24) } \text{Brier}_{\text{gruppiert}} = \frac{1}{g} \sum_{i=1}^g (\text{PD}_{i,\text{prog}} - \text{PD}_{i,\text{tat}})^2$$

mit $\text{PD}_{i,\text{prog}/\text{tat}}$... für Ratingklasse i prognostizierte/ realisierte Ausfallrate,
 g ... Anzahl der Ratingklassen^{115,116},

- „*Rommelfanger-Index*“¹¹⁷

$$\text{F 25) } F = \sum_{i=1}^g \Delta_i \cdot \frac{a_i + \hat{a}_i}{2} \cdot f_i$$

mit $\Delta_i = \max(0; \text{PD}_{i,\text{tat}} - \text{PD}_{i,\text{prog}})$ für $i=1..g-1$, bzw. $\Delta_i = \max(0; \text{PD}_{i,\text{prog}} - \text{PD}_{i,\text{tat}})$ für $i=g$, a_i ,
 \hat{a}_i ... relatives Volumen aller Kredite in der Validierungsstichprobe/ Lernstichprobe,
 f_i ... „geeignetes Gewicht“^{118,119}

Die beiden im folgenden näher vorgestellten Präzisionsmaße kardinaler Insolvenzprognosen basieren auf einem einheitlichen Grundprinzip: sie vergleichen die individuellen prognostizierten Ausfallwahrscheinlichkeiten $\text{PD}_{i,\text{prog}}$ mit den realisierten Ausfallergebnissen Θ_i (mit $\Theta_i=1/ \Theta_i=0$ falls Schuldner i ausgefallen/ nicht ausgefallen ist) und belegen die dabei auftretenden Differenzen mit unterschiedlichen „Strafen“.

Im Gegensatz zu kategorialen Insolvenzprognosenverfahren, die nur die Extremprognosen „Ausfall“ vs. „Nichtausfall“ erlauben, ist bei stochastischen Ausfallprognosen (kardinalen Ausfallprognosen) möglicherweise zunächst nicht ersichtlich, warum Abweichungen der individuellen Prognosen (Ausfallwahrscheinlichkeiten) und Ausfallrealisierung als Fehler „bestraft“ werden sollten. Schließlich können die Prognosen beliebige Werte zwischen 0% und 100% annehmen, während die Ausfallrealisierungen nur die Extremwerte 1 („Ausfall“) oder 0 („Nichtausfall“) annehmen können: Selbst wenn die prognostizierten Ausfallwahr-

¹¹³ Siehe beispielsweise FRERICHS, WAHRENBURG (2003, S. 16, eigene Notation). In einer Simulationsstudie finden die Autoren, daß der gruppierte Brierscore nicht als Validierungskenngröße für Ratingsysteme geeignet ist, da er nicht in der Lage ist, von der Konstruktion her „unterlegene“ Ratingsysteme zuverlässig zu identifizieren.

¹¹⁴ Trotz einer gewissen formalen Ähnlichkeit unterscheiden sich der *gruppierte Brierscore* und der *Brierscore*, der im folgenden vorgestellt wird, fundamental. Der gruppierte Brierscore wird nur von der Güte der *Kalibrierung* eines Ratingverfahrens beeinflusst, während die *Trennschärfe* des Ratingsystems nicht berücksichtigt wird.

¹¹⁵ Sinnvoller als die Gleichgewichtung der ratingklassenspezifischen quadrierten Differenzen der prognostizierten und realisierten Ausfallquoten bei der Ermittlung des Scores wäre sicher eine Berücksichtigung der relativen Belegungsstärken der einzelnen Ratingklassen:

$$\text{Brier}_{\text{gruppiert}}^* = \sum_{i=1}^g a_i (\text{PD}_{i,\text{prog}} - \text{PD}_{i,\text{tat}})^2$$

mit a_i ... Anteil der Unternehmen in Ratingklasse i an allen Unternehmen,

¹¹⁶ Als Ergebnis einer Simulationsstudie finden die Autoren, daß der gruppierte Brierscore – im Gegensatz zu anderen untersuchten Kenngrößen - als (alleiniges) Gütekriterium nicht in der Lage ist, schlechte von guten Ratingsystemen zu unterscheiden, vgl. FRERICHS, WAHRENBURG (2003, S.17).

¹¹⁷ DVFA (2004, S. 600, eigene Notation)

¹¹⁸ siehe DVFA (2004, S. 599). Es wird keine Aussage getroffen, wie die „geeigneten Gewichte“ beschaffen sein müssen.

¹¹⁹ Weitere Kritikpunkte an diesem Index, neben seiner ausschließlichen Fokussierung auf den Aspekt *Kalibrierung*, sind die Abhängigkeit von irrelevanten Größen (Struktur der Lernstichprobe) und die Setzung von Anreizen zur systematischen Fehlprognose: da in den Klassen $1..g-1$ nur zu hohe und in der Klasse g nur zu niedrige Ausfallwahrscheinlichkeiten „bestraft“ werden, besteht ein Anreiz, alle Prognosen systematisch zu hoch (Ratingklasse $1..g-1$) bzw. zu niedrig (Ratingklasse g) anzusetzen.

scheinlichkeiten „richtig“ sind, d.h. *korrekt kalibriert* sind, wenn also beispielsweise 5% aller Unternehmen ausfallen, bei denen das Verfahren eine Ausfallwahrscheinlichkeit von 5% vorhergesagt hat und 10% aller Unternehmen ausfallen, bei denen das Verfahren eine Ausfallwahrscheinlichkeit von 10% vorhergesagt hat, usw. werden die Verfahren „bestraft“, d.h. erhalten nicht die bestmögliche Ausprägung. „Bestraft“ wird in diesen Fällen jedoch die nicht perfekte *Trennfähigkeit* der Verfahren: ein Verfahren, das bei allen deutschen Unternehmen im Jahr 2003 eine Insolvenzwahrscheinlichkeit von 1,35% vorausgesagt hätte, wäre zwar perfekt kalibriert, würde aber eine hohe „Strafe“ für seine nicht-trennfähigen Prognosen erhalten; ein Verfahren hingegen, daß bei 1,35% dieser Unternehmen eine Insolvenzwahrscheinlichkeit von 100% und bei den restlichen 98,65% eine Ausfallwahrscheinlichkeit von 0% vorhergesagt *und* mit diesen Prognosen auch immer recht gehabt hätte, hätte die bestmögliche Bewertung erhalten.¹²⁰

Folgende beide Präzisionsmaße für kardinale Insolvenzprognosen, die sich hinsichtlich ihrer „Straffunktionen“ unterscheiden, werden betrachtet:

- **logarithmische „Strafffunktion“:** CIE (conditional information entropy, bedingte Informationsentropie)¹²¹:

- **F 26)**
$$CIE = -\frac{1}{n} \sum_{i=1}^n \ln(|PD_{i,prog} + \theta_i - 1|)$$
 mit $n \dots$ Anzahl der Schuldner,

Anmerkung: nicht definiert ist CIE nur für die Fälle, in denen ein Ausfall eintritt, obwohl er mit Sicherheit ausgeschlossen wurde ($\theta_i=1$ und $PD_{i,prog}=0$) oder in denen kein Ausfall eintritt, obwohl er mit Sicherheit prognostiziert wurde ($\theta_i=0$ und $PD_{i,prog}=1$).

- **F 27)**
$$E(CIE) = -\frac{1}{n} \sum_{i=1}^n PD_{i,tat} \cdot \ln PD_{i,prog} + (1 - PD_{i,tat}) \cdot \ln(1 - PD_{i,prog})$$

- im Fall von g diskreten Ratingklassen ergibt sich:

- **F 28)**
$$E(CIE) = -\sum_{i=1}^g a_i (PD_{i,tat} \cdot \ln PD_{i,prog} + (1 - PD_{i,tat}) \cdot \ln(1 - PD_{i,prog}))$$

mit $a_i \dots$ Anteil der Unternehmen in Ratingklasse i an allen Unternehmen

- **F 29)**
$$Skill_{CIE} = CIER = \frac{CIE_{PD} - CIE}{CIE_{PD}}$$

- **F 30)**
$$CIE_{PD} = -(PD \cdot \ln PD + (1 - PD) \cdot \ln(1 - PD))$$

mit $CIER \dots$ conditional information entropy *ratio*¹²²

mit $CIE_{PD} \dots$ CIE-Wert einer „naiven“ Referenzprognose, die stets die Wahrscheinlichkeit PD prognostiziert^{123,124}

¹²⁰ siehe beispielsweise KRÄMER (2003, S. 396f.)

¹²¹ Die *Entropie*, stellt ein aus der Thermodynamik entlehntes Konzept dar, welches das Ausmaß an Unordnung eines Systems messen soll. Im Kontext von Insolvenzprognosen soll *Entropie* das Ausmaß an Unsicherheit quantifizieren, das mit einer bestimmten von einem Ratingmodell für eine Stichprobe von Unternehmen behaupteten Ausfallwahrscheinlichkeitsverteilung verbunden ist, vgl. SOBEHART, KEENAN, STEIN (2000, S. 14). Siehe SHANNON (2001/1948, S. 11f.) für eine axiomatische Rechtfertigung für die Verwendung der Logarithmusfunktion als „Straffunktion“ – allerdings ist das letzte der der Axiome im Fall von Unternehmensinsolvenzen mit nur zwei möglichen Ausprägungen „Ausfall“ vs. „Nicht-Ausfall“ nicht sinnvoll anwendbar. Siehe ferner MATHESON, WINKLER (1976), KEENAN, SOBEHART (1999, S.9), und BASLER AUSSCHUB (2005, S.44) für Formel F 27 (eigene Notation). Siehe KRÄMER, GÜTTLER (2003, S. 12) für Formel F 26.

¹²² siehe KEENAN, SOBEHART (1999, S. 10)

Anmerkung: Der Term $CIE_{PD} - CIE$ wird auch als KULLBACK-LEIBLER-Distanz¹²⁵ oder Wealth-Growth-Rate-Pickup¹²⁶ bezeichnet.

- **quadratische „Strafffunktion“:** Brierscore:¹²⁷

- **F 31)** $BS = \frac{1}{n} \sum_{i=1}^n (PD_{i,prog} - \theta_i)^2$ ¹²⁸

- **F 32)** $E(BS) = \frac{1}{n} \sum_{i=1}^n PD_{i,tat} \cdot (1 - PD_{i,prog})^2 + (1 - PD_{i,tat}) \cdot PD_{i,prog}^2$

- im Fall von g diskreten Ratingklassen vereinfacht zu:

- **F 33)** $E(BS) = \sum_{i=1}^g a_i \cdot (PD_{i,tat} \cdot (1 - PD_{i,prog})^2 + (1 - PD_{i,tat}) \cdot PD_{i,prog}^2)$

- **F 34)** $Skill_{BS} = \frac{BS_{naiv} - BS}{BS_{naiv}}$ ¹²⁹ mit

- **F 35)** $BS_{naiv} = PD \cdot (1 - PD)^2 + (1 - PD) \cdot PD^2 = PD \cdot (1 - PD)$

Die „Strafffunktionen“ sind in beiden Fällen *willkürlich* in dem Sinne gewählt, als daß sich kein direkter Bezug zur letztendlich interessierenden Nutzengröße des/der Anwender/s der Prognosen herstellen läßt (siehe auch Abschnitt 1). Die Kenngrößen zeigen jedoch ein „plausibles“ Verhalten, so daß eine Korrelation mit den Nutzengrößen der potentiellen Anwender der Prognosen zumindest vermutet werden kann: Beide Scores „belohnen“ richtig kalibrierte^{130,131} und trennfähige¹³² Prognosen - und durch Umformungen der resultierenden Scorewer-

¹²³ CIER ... conditional information entropy ratio, SOBEHART, KEENAN, STEIN (2000, S. 14): “The CIER compares the amount of ‘uncertainty’ regarding default in the case where we have no model (a state of more uncertainty about the possible outcomes) to the amount of ‘uncertainty’ left over after we have introduced a model (presumably, a state of less ignorance),

¹²⁴ Angesichts der im Zeitverlauf sehr volatilen Ausfallraten (zumindest bei Ausfallstudien von Unternehmensanleihen), ist selbst für die Prognose der durchschnittlichen Ausfallraten mehr als nur „Naivität“ erforderlich. Zur Schätzung künftiger Ausfallraten siehe beispielsweise KEENAN (1999), S&P (2004b, S.3).

¹²⁵ siehe BASLER AUSSCHUB (2005, S. 30)

¹²⁶ siehe CANGEMI, SERVIGNY, FRIEDMAN (2003, S. 40)

¹²⁷ Für die Definition des Brierscores siehe BRIER (1950, S. 1), MURPHY, WINKLER (1992, S. 439, Formel 7), KRÄMER, GÜTLER (2003, S. 11), FRERICHS, WAHRENBURG (2003, S.14), OeNB (2004a, S. 123ff.), GRUNERT, NORDEN, WEBER (2005, S.517)

¹²⁸ In der im Kontext von Regressionsanalysen verwendeten Notation entspricht der Brierscore damit der mit n dividierten Summe der quadrierten Residuen (RSS), mit $RSS = \sum Y_i^* - Y_i$, mit Y_i^* ... prognostizierter Wert der erklärten Variable Y und Y_i ... tatsächliche Ausprägung der erklärten Variable Y, siehe beispielsweise GUJARATI (1999/1992, S. 170ff.).

¹²⁹ In der im Kontext von Regressionsanalysen verwendeten Notation entspricht BS_{naiv} der mit n dividierten Summe der absoluten Variation der zu erklärenden Variable (TSS). Somit gilt $Skill_{BS} = (TSS - RSS) / TSS$. Somit gilt $Skill_{BS} = r^2$, mit r^2 ... sample coefficient of determination und $r^2 = ESS/TSS$ und $ESS = TSS - RSS$, siehe beispielsweise GUJARATI (1999/1992, S. 170ff.).

¹³⁰ Diese Aussage ist nicht trivial. Wird als Strafffunktion beispielsweise $|PD_{i,prog} - \theta_i|$ verwendet, so führt $PD_{i,prog} = 0\%$ für $E(PD_{i,tat}) < 50\%$ und $PD_{i,prog} = 100\%$ für $E(PD_{i,tat}) > 50\%$ zu geringeren erwarteten Strafen als $PD_{i,prog} = E(PD_{i,tat})$, siehe Anhang II. Siehe ebenfalls Anhang II für den Beweis der Anreizkompatibilität des Brierscores und des Entropiescores.

¹³¹ Bereits BRIER (1950, S.2) führte die *Anreizkompatibilität* als einen Vorteil seines Scores an: „[the forecaster] is encouraged to state unbiased estimates of the probability of each event when he cannot forecast perfectly.“

¹³² Beide Schätzgütemaße erreichen ihre günstigsten Ausprägungen, wenn ein Verfahren stets Ausfallwahrscheinlichkeiten von 0% oder 100% prognostiziert und die Prognosen auch eintreffen.

te lassen sich auch Bezüge zu den anderen Gütekriterien für kardinale Insolvenzprognosen, wie Auflösung, Feinheit, systematische Verzerrung, herstellen.¹³³

$$BS = \underbrace{PD \cdot (1 - PD)}_{\text{Varianz/naiver BS}} + \underbrace{\sum_{i=1}^g a_i \cdot (PD_{i,prog} - PD_{i,tat})^2}_{\text{Kalibrierung}} - \underbrace{\sum_{i=1}^g a_i \cdot (PD - PD_{i,tat})^2}_{\text{Auflösung (resolution)}}$$

Abbildung 11: Dekomposition Brierscore in Komponenten Varianz – Kalibrierung – Auflösung¹³⁴

Aus der in Abbildung 11 dargestellten Dekomposition des Brierscores wird jedoch eine problematische Eigenschaft der vorgestellten kardinalen Gütemaße ersichtlich: ihre Abhängigkeit von der durchschnittlichen Ausfallrate der Grundgesamtheit. Je größer die Varianz der Umgebung ($PD \cdot (1 - PD)$), desto *größer* (=schlechter) ist der Brierscore, den ein Verfahren in der jeweiligen Umgebung erhält (für eine weiterführende Analyse siehe Anhang III). Um diese unerwünschte Umweltabhängigkeit kardinaler Gütemaße zu beseitigen, wird die Verwendung von Skillmaßen vorgeschlagen, die den ermittelten Gütewert in Relation zur Güte von naiven Prognosen in der gleichen Umgebung betrachten.^{135,136} Unerwünscht ist diese Abhängigkeit, weil sie den Performancevergleich unterschiedlicher Verfahren stark beeinträchtigt, wenn die Leistungsfähigkeit der Verfahren auf Grundgesamtheiten mit unterschiedlichen durchschnittlichen Ausfallhäufigkeiten gemessen werden.¹³⁷ Empirisch und modelltheoretisch läßt sich jedoch zeigen, daß auch Skillscores umweltabhängig sind – während der Brierscore (für $PD_i < 50\%$) mit zunehmenden Ausfallwahrscheinlichkeiten immer „schlechter“ wird, werden die zugehörigen Skillscores aber paradoxerweise immer „besser“. ^{138,139} Die im Abschnitt 2.3 vorgestellten Schätzgütemaße für ordinale Insolvenzprognosen

¹³³ vgl. MURPHY, WINKLER (1992)

¹³⁴ Quelle: MURPHY, WINKLER (1992, S. 439, Formel 10, eigene Notation).

¹³⁵ WINKLER (1994, S. 1397): “The development of so called ‘skill-scores’ has been motivated by the desire to produce average scores that reflect the relative ability of forecaster rather than some combination of the forecaster’s ability to and the situation’s difficulty. These skill scores attempt to neutralize the contribution of the situation by comparing a forecaster’s average score to the average score that an unsophisticated forecasting scheme would have obtained for the same set of forecasting situations.”

¹³⁶ siehe auch KRÄMER (2003, S. 406)

¹³⁷ Unterschiede der Ausfallraten zwischen verschiedenen Stichproben können beispielsweise aufgrund unterschiedliche langer Beobachtungszeiträume resultieren (die kumulierte Ausfallwahrscheinlichkeit steigt näherungsweise linear mit der Zeit an), durch die Abdeckung unterschiedlicher Phasen des Konjunkturzyklus oder durch die Betrachtung von Unternehmensgruppen, die sich hinsichtlich ihrer Ausfallrisiken unterscheiden oder aber auch durch die gezielte Auswahl von Stichproben mit einem deutlich überhöhten Anteil ausgefallener Unternehmen, was vor allem bei Insolvenzprognosestudien mit einer vergleichsweise geringen Anzahl an einbezogenen Unternehmen üblich ist

Von den 31 bei FALKENSTEIN, BORAL, CARTY (2003/2000, S. 14) aufgeführten Insolvenzprognosestudien basieren 18 auf der Auswertung von weniger als 100 Unternehmen, 9 auf der Auswertung von 100-1.000 und nur 4 auf der Auswertung von mehr als 1.000 Unternehmen. In 14 der 18 (=78%) der Insolvenzprognosestudien mit weniger als 100 Unternehmen bestanden exakt 50% der enthaltenen Unternehmen aus Ausfallern, gleiches gilt für 5 der 9 (=56%) der Studien mit 100-1.000 Unternehmen und 0 von 4 der Studien mit über 1.000 Unternehmen. In letzteren Fällen lagen die Ausfallanteile bei 5,0%, 6,6%, 9,7% und 14,1%.

¹³⁸ Bei einer Untersuchung der Prognoseergebnisse von 20 Wetterstationen von Standorten mit deutlich unterschiedlichen regionalen Regenwahrscheinlichkeiten, fand WINKLER (1994, S. 1401f) eine sehr hohe Korrelation zwischen den jeweiligen Brierscores und durchschnittlichen Niederschlagshäufigkeiten ($r = +0,87$) – was zu erwarten war („In der Wüste ist es leicht, einen regenfreien Tage vorherzusagen.“). Fast ebenso stark korreliert waren jedoch auch die Skill-Brierscores und Niederschlagshäufigkeiten (r ca. $+0,80$) („In der Wüste ist es schwer, einen hohen Prozentsatz der wenigen Regentage vorherzusagen.“) – was der Intention der

vorgestellten Schätzgütemaße für ordinale Insolvenzprognosen weisen diesen Nachteil nicht auf.¹⁴⁰

Vereinzelt werden obige Gütemaße auch unter der Fiktion einer richtigen Kalibrierung verwendet, d.h. ex-post wird $PD_{i,prog} = PD_{i,tat}$ für alle i gesetzt.¹⁴¹ Die Formeln F 28 und F 33 vereinfachen sich dann zu:

$$F\ 36) \ CIE_{kal} = -\sum_{i=1}^g a_i (PD_i \cdot \ln PD_i + (1 - PD_i) \cdot \ln(1 - PD_i))$$

$$F\ 37) \ BS_{kal} = \sum_{i=1}^g a_i \cdot (PD_i \cdot (1 - PD_i)^2 + (1 - PD_i) \cdot PD_i^2)$$

$$F\ 38) \ BS_{kal} = \sum_{i=1}^n a_i \cdot PD_i \cdot (1 - PD_i).$$

Die so erhaltenen Gütemaße sind unempfindlich gegenüber möglichen Fehlkalibrierungen (oder gar *fehlenden* Kalibrierungen, wie bei ordinalen Insolvenzprognosen) - der mittlere Term in Abbildung 11 entfällt - und messen somit lediglich die *Varianz der Umwelt* und die *Auflösung* der Prognosen. Diese Kennzahlen sind auch als Obergrenzen der Schätzgüten der jeweiligen Verfahren interpretierbar, die sich bei richtiger Kalibrierung und gegebener Auflösung erreichen lassen.

Für den portfolioübergreifenden Vergleich von Insolvenzprognoseverfahren sind sie aufgrund der Abhängigkeit von der durchschnittlichen Ausfallrate ungeeignet. Beim Vergleich verschiedener Prognoseverfahren auf Basis identischer Portfolien sind sie zwar nicht informativer als die bereits vorgestellten Kenngrößen für ordinale Insolvenzprognosen (AUC_{ROC} und AR), können aber ergänzend, insbesondere im Fall sich schneidender CAP-Kurven, herangezogen werden, da sich letztlich keines der diversen Qualitätsmaße für die Schätzgüte von Insolvenzprognosen direkt in Beziehung zum Nutzen der Prognosen aus Sicht unterschiedlicher Anspruchsgruppen bringen läßt. Deuten beim direkten Vergleich zweier Prognoseverfahren alle Indikatoren auf die Überlegenheit des gleichen Verfahrens, so wird der Entscheidungsträger – wenn er sich für eines der beiden Verfahren entscheiden muß, und nicht beispielsweise den Informationsgehalt beider Verfahren kombinieren kann – in seiner Sicherheit bestärkt, das richtige Verfahren zu wählen. Geben die verschiedenen Indikatoren widersprüchliche Signale, ist anzunehmen, daß der Entscheidungsträger keinen „großen Fehler“ macht, wenn er sich zufällig für eines der beiden Verfahren entscheidet oder sekundäre Entscheidungskriterien (z.B. Entwicklungskosten, Transparenz/ Nachvollziehbarkeit, ...) heranzieht.

Skillscores (siehe oben) offensichtlich zuwiderläuft. [Anmerkung: größere Brierscores korrespondieren mit *schlechteren* Prognosen; größere Skillbrierscores mit *besseren* Prognosen.]

¹³⁹ Siehe Anhang III für eine modelltheoretische Herleitung der Ergebnisse. Siehe auch CANTOR, MANN (2003, S. 12) und DVFA (2004, S. 599) bezüglich der (weitgehenden) Invarianz der Accuracy Ratio bezüglich der durchschnittlichen Ausfallrate.

¹⁴⁰ siehe Anhang III

¹⁴¹ vgl. KRÄMER, GÜTLER (2003, S. 12)

3 Empirische Befunde zur Schätzgüte von Insolvenzprognosen

3.1 Grenzen der Aussagefähigkeit empirischer Vergleiche und Entwicklung von Benchmarks für Schätzgütemaße

Rein theoretisch läßt sich zeigen, daß bei Anwendung eines Ratingverfahrens auf unterschiedliche Stichproben von Unternehmen erhebliche und systematisch variierende Schätzgüten ermittelt werden können.¹⁴² Dies bedeutet, daß die Qualität eines Ratingverfahrens nicht eindeutig anhand eines konkreten Zahlenwertes beschreibbar ist – und daß die relative Qualität von Ratingverfahren nicht eindeutig aus einem stichprobenübergreifenden Vergleich von Schätzgütemaßen abgeleitet werden kann, wenn sich die Stichprobensamensetzungen hinsichtlich „schätzgüterrelevanter“ Aspekte unterscheiden. In der Literatur wurden auf Basis theoretischer und empirischer Untersuchungen zahlreiche solcher von der Qualität eines Ratingverfahrens *unabhängiger* Einflußgrößen identifiziert. Als Determinanten für derartige schätzgüterrelevante „strukturelle Unterschiede“ in der *Zusammensetzung von Stichproben* wurden unter anderem genannt:¹⁴³

- *Pool vs. Kohortendaten*: Je nachdem, ob bei der Überprüfungen eines Verfahrens nur Datensätze von Unternehmen einer identischen Periode (Kohorten) verwendet werden oder aber Datensätze von Unternehmen verschiedener Perioden (Pools), sind Unterschiede bei den gemessenen Schätzgüten zu erwarten. Theoretisch sind bessere Werte auf Basis von Kohortendaten zu erwarten, da ein Ratingverfahren hier lediglich *zeitpunktbezogene* trennscharfe Reihungen zwischen den Unternehmen vornehmen muß, um eine gute Bewertung zu erreichen, während ein Ratingverfahren bei Pooldaten zusätzlich auch im *Zeitverlauf konsistente Reihungen* vornehmen muß.¹⁴⁴ Verschärft wird der Unterschied, wenn das Ratingverfahren im Zeitverlauf möglichst individuell *stabile* Ratings generieren soll (Point-in-time- (Current-Condition-) (PIT) vs. Through-the-cycle-Ansatz (TTC)^{145,146}). Empirisch scheint das Problem jedoch vergleichsweise unbedeutend zu sein.¹⁴⁷ Außerdem

¹⁴² Siehe beispielsweise HAMERLE, RAUHMEIER, RÖSCH (2003).

¹⁴³ OENB (2004, S. 137f.) stellt folgende Anforderungen an die Qualität von Validierungsbenchmarkdaten: vergleichbare Datenqualität, inhaltliche Übereinstimmung der Daten (insb. bei qualitativen Inputs), Konsistenz der Zielgrößen (Ausfalldefinition), strukturelle Konsistenz (vergleichbare Verteilung hinsichtlich Größenklassen, regionaler Verteilung, Branchenstruktur und Rechtsformen).

¹⁴⁴ Beispiel: Damit ein Insolvenzprognoseverfahren auf Basis der 2004er Unternehmenskohorte eine möglichst hohe Schätzgüte erreichen kann, müssen die Unternehmen, die zum 1.1.2004 über ein BBB-Rating verfügten, eine möglichst deutlich niedrigere Ausfallquote im Jahr 2004 aufweisen als Unternehmen, die zum 1.1.2004 mit BB geratet waren. Damit ein Verfahren auf Basis eines gepoolten Datensatzes mit Unternehmensdaten der Jahre 1970-2004 eine möglichst hohe Schätzgüte erreichen kann, muß zusätzlich noch beispielsweise gelten, daß die Einjahresausfallquoten von 1990 mit BBB gerateten Unternehmen möglichst deutlich niedriger sind als die Einjahresausfallquoten der 2004 mit BB gerateten Unternehmen ist. Letztlich bedeutet dies, daß poolbasierte Ratings im Zeitverlauf stabile ratingklassenspezifische Ausfallwahrscheinlichkeiten aufweisen müssen, um eine gute Bewertung zu erhalten.

¹⁴⁵ Auf Basis eines Simulationsmodells findet LÖFFLER (2004a, S. 709) dramatische Performanceunterschiede zwischen TTC- und PIT-Ratings.

¹⁴⁶ Für eine ausführliche Beschreibung der beiden Ansätze siehe BASLER AUSSCHUB (2005, S.10 ff.).

¹⁴⁷ So erreichen beispielsweise die MOODY's-Ratings im Zeitraum 1983-2002 Einjahres-/ Fünfjahres-Accuracy Ratios auf Basis der gepoolten Daten von 82,6% / 71,0%, während die durchschnittlichen emittentengewichteten Kohorten-Accuracy-Ratios im gleichen Zeitraum mit 83,5% / 72,9% nur geringfügig höher lagen, CANTOR, MANN (2003, S.19).

Bezieht man auch die MOODY's-Daten der Jahre 1970-1982 in die Auswertungen ein, kehrt sich die Reihenfolge sogar um: die emittentengewichtete 1970-2003er Kohorten-Accuracy-Ratio ist mit 83,4% sogar geringfügig kleiner als die 1970-2003er Pool-Accuracy-Ratio mit 83,5%. Quelle: eigene Untersuchungen [Hinweis: Im Gegensatz zu CANTOR, MANN (2003) erfolgte die Berechnung der AR-Werte auf Basis eines 7-

basieren praktisch alle Insolvenzprognosestudien auf gepoolten Daten (um Engpässe, insbesondere hinsichtlich der Anzahl an Ausfallunternehmen zu vermeiden).¹⁴⁸

- *Positionierung der Beobachtungsperiode im Ausfallzyklus:* Festgestellt wurde, daß die Qualität von Insolvenzprognosen leider immer dann besonders schlecht ist, wenn sie besonders wichtig ist: nämlich in Zeiten hoher gesamtwirtschaftlicher Ausfallraten (siehe auch Abbildung 8). Eine bildhafte Begründung hierfür lautet, daß in einem Portfolio aus „weißen, grauen und schwarzen Schafen“ in wirtschaftlich guten Zeiten tendenziell nur „schwarze Schafe“ insolvent würden, in wirtschaftlich schwierigeren Zeiten hingegen sowohl „schwarze“ als auch „graue Schafe“. ^{149,150,151} Da die meisten Ausfallstudien auf gepoolten Daten langer Zeiträume basieren (siehe oben) und damit verschiedene Ausfallzyklusphasen abdecken, dürfte die empirische Relevanz des Problems beim Vergleich der Schätzwertwerte verschiedener Verfahren gering sein.
- *Vorselektierte Portfolien (insbesondere Bankportfolien):* Da in Bankportfolien sowohl *besonders ausfallgefährdete* Unternehmen („schwarze Schafe“) als auch *besonders stabile* Unternehmen („weiße Schafe“) unterrepräsentiert sind, wird die Trennschärfe von Insolvenzprognoseverfahren negativ beeinflusst, da diese Verfahren somit tendenziell nur noch zwischen „hell- und dunkelgrauen Schafen“ differenzieren müssen.¹⁵² Besonders ausfallgefährdete Unternehmen sind deshalb unterrepräsentiert, da Banken keine *Neukunden* akquirieren, bei denen sie von einer hohen Ausfallgefährdung ausgehen.¹⁵³ Ebenso werden *Bestandskunden*, deren Bonität sich seit Aufnahme der Kundenbeziehung so stark verschlechtert hat, daß die Bank nunmehr mit einer hohen Ausfallgefährdung rechnet aus dem „performenden Portfolio“ entfernt und in ein spezielles „Problemkreditportfolio“ überführt. Erfolgt die Beurteilung der Prognosegüte eines Verfahrens anschlie-

7-stufigen Ratings. Ferner wurde die „normale“ AR-Definition und nicht die MOODY’S-AR-Definition zugrundegelegt (siehe Abschnitt 2.3.2)]. Die Differenz zwischen den Kohorten- und Pool-AR ist sehr sensitiv bezüglich der Werte einzelner Jahre. Beispielsweise wäre ohne die Daten der Jahre 1973, 1976, 1977 oder 1979 die emittentengewichtete Kohorten-AR größer als die Pool-AR. Alle vier Jahre waren durch stark unterdurchschnittliche Ausfallraten und Accuracy-Ratios geprägt, wovon die Kohorten-Performance stärker negativ betroffen war als die Pool-Performance.

¹⁴⁸ In einer Metastudie von AZIZ, DAR (2004, S. 35ff.) basiert nur eine von 82 Insolvenzprognosestudien auf den Daten einer Kohorte. 69 Studien (84%) basieren auf gepoolten Daten von mindestens fünf Jahren und 35 (42%) sogar auf gepoolten Daten von mindestens 10 Jahren.

¹⁴⁹ „Recessions both increase and broaden the base of defaulters, lowering the measured power of default models in these periods. [...] In [good] times, the really bad firms (C) default, while moderate firms (B) and excellent firms (A) don’t default. In bad times, however, section C still shows more defaults, but now section B is a gray area; a new class of firms that were previously almost never defaulting are now defaulting at low, but significant rates. This adds gray to a situation that was previously black and white, which the lower power of the model reflects.” FALKENSTEIN, BORAL, KOCAGIL (2000, S. 22f)

¹⁵⁰ Siehe auch S&P (2004b, S. 14): „Trends in the one-year Gini ratio emerge during periods of both extremes in default pressure, [...] In periods of high defaults, there tends to be greater variation with respect to how the defaults are distributed across the ratings spectrum, which reduces the Gini.“, siehe auch BALCAEN, OOGHE (2004, S. 31 und die dort zitierte Literatur).

¹⁵¹ Basierend auf den MOODY’S-Daten ergibt sich für den Zeitraum 1983-2003 ein Korrelationskoeffizient von -0,494 / -0,437 für den Zusammenhang zwischen den Accuracy-Ratio-Werten der einzelnen Kohorten (Prognosezeitraum 1 Jahr / 5 Jahre) und den durchschnittlichen 1- und kumulierten 5-Jahres-Ausfallraten der einzelnen Jahre, Quelle: eigene Untersuchungen.

¹⁵² Neben der *Validierung* sind auch die anderen Entwicklungsphasen von Ratingmodellen durch die Vorselektion von Stichproben betroffen. Wird ein Ratingmodell nur auf Basis von Daten der nicht-abgelehnten Kunden eines Portfolios entwickelt, ist mit falsch kalibrierten Ausfallprognosen bei Anwendung des Verfahrens auf abgelehnte Kunden zu rechnen, siehe FEELDERS (2000, S. 1ff.).

¹⁵³ Zum Aspekt der „zensierten“ Neukundenakquisition und den quantitativen Auswirkungen auf die Qualität von Insolvenzprognosen siehe KRAFT, KROISANDT, MÜLLER (2004, S. 7f.) und die dort zitierte Literatur.

ßend unter Ausschluß dieser „Problemkunden“ ist mit einer weiteren Verschlechterung der gemessenen Trennschärfe zu rechnen.^{154,155,156}

Eine Unterrepräsentation von Unternehmen mit niedrigen Ausfallwahrscheinlichkeiten in Bankportfolien ist deshalb anzunehmen, da es sich dabei tendenziell um solche Unternehmen handelt, die keine (oder nur unterdurchschnittlich viele) Bankkredite aufnehmen müssen, weil sie beispielsweise nur über einen geringen Kapitalbedarf verfügen, sich vollständig mit Eigenmitteln finanzieren können oder weil ihnen alternative Fremdfinanzierungsmöglichkeiten offenstehen, beispielsweise die Anleihenemission. Bei Ausschluß der 10% der schlechtesten Unternehmen ist mit Schätzgüteeinbußen auf Basis der Restportfolien von 10% bis über 50% (!) zu rechnen, bei Ausschluß der 10% der besten Unternehmen mit Einbußen von nur 2,5% oder weniger.¹⁵⁷

- *Durchschnittliche Ausfallraten:* Ein Nachteil der in Abschnitt 2.4 vorgestellten Schätzgütemaße für *kardinale* Insolvenzprognosen ist deren Abhängigkeit von den durchschnittlichen Stichprobenausfallraten. Die vorgestellten Schätzgütemaße für *ordinale* Insolvenzprognosen sind hiervon *nicht* betroffen, d.h. mit ihnen ist ein stichprobenübergreifender Vergleich möglich, selbst wenn sich die Ausfallraten der verschiedenen Stichproben unterscheiden (siehe auch die Abschnitte 2.3 und 2.4 und Anhang III).
- *Branchenzugehörigkeit:* Besonders trennscharfe Prognosen erzielen die Ratingagenturen bei der Bewertung von Finanzunternehmen. Während die MOODY'S-Ratings bei Finanzunternehmen eine Einjahres-Accuracy-Ratio von 92,3% erreichen, erzielen sie bei Nicht-Finanzunternehmen mit 80,5% eine deutlich schlechtere Performance.¹⁵⁸ Interessanterweise werden gerade Finanzunternehmen bei den meisten Insolvenzprognosestudien und -verfahren ausgeschlossen (siehe auch die Abschnitte 3.3 bis 3.5). Eine weitere Branchendifferenzierung innerhalb der *Nicht-Finanzunternehmen* ist vermutlich nicht mit nennenswerten Prognosegüteunterschieden verbunden.^{159,160,161}
- *Unternehmensgröße:* Bei „großen“ Unternehmen erzielen MOODY'S-Ratings mit einer durchschnittlichen Accuracy Ratio von 84,4% eine deutlich bessere Trennschärfe (Accu-

¹⁵⁴ Zum Aspekt des Ausschlusses von Problemkunden aus der Schätzstichprobe siehe LEHMANN (2003, S.8)

¹⁵⁵ FALKENSTEIN, BORAL, CARTY (2003/2000, S. 23): “Many institutions transfer credits to special asset groups once a credit is placed in any of the regulator criticized asset categories. Once there, many institutions do not continue to spread the financial statements associated with these high risk borrowers, or the borrowers no longer submit them.”

¹⁵⁶ FALKENSTEIN, BORAL, KOCAGIL (2000, S. 7): “Much of the dearth in default data is due to the vagaries of data storage within financial institutions. Defaulting companies are often purged from the system after their troubles begin, which creates a sample bias in that the default probability implicit in current bank databases is invariably low, even for a non-recessionary period.”

¹⁵⁷ Siehe Anhang V für die zugehörigen formalen und empirischen Untersuchungen.

¹⁵⁸ siehe MOODY'S (2004c, S.2). Die AR-Werte beziehen sich auf den „historischen Durchschnitt“ (vermutlich 1983-2003).

¹⁵⁹ In einem Sample mit 30.000 Unternehmen fanden BLOCHWITZ, LIEBIG, NYBERG (2000, S. 28ff.) (siehe auch Abschnitt 3.3) nur geringfügige Trennschärfeunterschiede zwischen den Branchenaggregaten *Handel*, *Verarbeitendes Gewerbe* und *sonstige Unternehmen*. Die AR-Werte der branchenspezifischen Diskriminanzanalysemodelle betragen 55,8%, 60,0% und 54,4%.

¹⁶⁰ In einem Sample mit 50.000 Unternehmen konnten DWYER, KOCAGIL, STEIN (2004, S. 17f.) die Trennschärfe ihres Ratingverfahrens nur minimal (aber statistisch signifikant) von 54,4% auf 55,1% durch die Einbeziehung branchenspezifischer Kennzahlen für neun unterschiedliche Branchenaggregate erhöhen. (Die Trennschärfe *innerhalb* der Branchenaggregate wurde nicht angegeben).

¹⁶¹ Mit einem Ratingmodell, das auf einer Stichprobe von 19.500 Unternehmen basiert und keine branchenspezifischen Kennzahlen enthält, finden KOCAGIL ET AL. (2003, S. 20) folgende, nur geringfügig variierende branchenspezifische Prognosegüten: AR Bauunternehmen = 58,6%, AR Produzenten=59,1%, AR Dienstleister = 54,0%, AR Handel = 57,1%.

racy Ratio) als bei „kleinen“ Unternehmen mit 74,0%.¹⁶² Noch größere Trennschärfeunterschiede sind in zahlreichen empirischen Studien festzustellen (siehe Abschnitt 3.3).¹⁶³ Als ursächlich für die schlechtere Trennfähigkeit der Verfahren bei kleinen Unternehmen, die seltener börsennotiert sind oder über ein Agenturrating verfügen, wird deren schlechtere formale Qualität der *Jahresabschlußdaten*^{164,165,166} (siehe auch den nächsten Punkt), die Nichtverfügbarkeit von Kapitalmarktdaten,¹⁶⁷ aber auch die schlechtere Qualität von *Ausfallinformationen*^{168,169} genannt.^{170,171}

- *Datenqualität*: Eine „schlechte Datenqualität“, d.h. das Vorliegen unvollständiger oder gar falscher Informationen über die finanziellen Verhältnisse von Unternehmen oder gar

¹⁶² siehe MOODY'S (2004c, S.2). Die AR-Werte beziehen sich auf den „historischen Durchschnitt“ (vermutlich 1983-2003). Es wurde nicht angegeben, welche Unternehmen als „groß“ und welche als „klein“ klassifiziert wurden.

¹⁶³ Siehe auch KOCAGIL ET AL. (2003, S. 20). Mittels des gleichen Modells wurden AR-Werte für Unternehmen mit Umsätzen zwischen 0,5 und 5 Mio. EUR von 51,1% erzielt, 59,3% für Unternehmen mit 5-25 Mio. EUR und 64,6% für Unternehmen mit mehr als 25 Mio. EUR Umsatz.

¹⁶⁴ STEIN ET AL (2003, S. 5): „An important result of these structural differences, are differences in the availability of good quality data on which to develop and test default models. Because most middle-market firms are not issuers of public securities, they are not required to report details of their financial statements on a regular basis as public firms do. These firms typically report such information to their lenders but it is not generally available to the marketplace in most countries. Furthermore, the quality of these reported financial statements both with respect to data accuracy and accounting rigor is typically inferior to that in the public markets.“

¹⁶⁵ FALKENSTEIN, BORAL, CARTY (2003/2000, S. 77): “Accounting statements are less noisy for rated companies, and this is reflected in the far greater number of audited statements for public companies as opposed to private companies. Adding noise to the input variables clearly weakens the power of any model to predict from them.”

¹⁶⁶ BOHN, AVORA, KORABLEV (2005, S. 14) begründen die *verschlechterte (!)* Prognoseleistung eines kennzahlenbasierten Ratingmodells bei Anwendung auf größere Unternehmen wie folgt: “larger firms have more sophisticated financial statements since they usually operate in multiple segments. This makes their financial ratios more difficult to interpret.”

¹⁶⁷ STEIN ET AL (2003, S.5). “[...] Since, by definition, private firms do not have publicly traded equity and debt, price series of these financial assets are not available for individual firms. This implies that even if the firm-specific details of these companies were publicly known, price discovery reflecting the incorporation of these risks does not take place. Thus, various asset pricing-based approaches to default risk that have enjoyed wide success and acceptance for public firms cannot be directly applied to private ones.”

¹⁶⁸ FALKENSTEIN, BORAL, CARTY (2003/2000, S. 77): “Yet, some of this loss of power can also be explained by the fact that defaults are measured better for rated firms than unrated public firms, and for public firms vs. private firms. More of the 'goods' in the unrated universes are mislabelled, but unfortunately we do not know which ones.”

¹⁶⁹ Durch den Zusammenschluß der Datenbanken von MOODY'S und KMV ließ sich beispielsweise abschätzen, daß in der MOODY'S-Datenbank ca. 2/3 (!) der Ausfälle von Unternehmen mit einem Umsatz von weniger als 1 Mio. US-\$ *nicht* erfaßt waren, siehe DWYER, STEIN (2003, S. 7, 9). Die MOODY'S-Datenbank wird für die Entwicklung, Kalibrierung und Validierung von Insolvenzprognosemodellen für kleine und mittelständische Unternehmen genutzt.

¹⁷⁰ FALKENSTEIN, BORAL, KOCAGIL (2000, S. 12): “Size is a notorious correlate with various inputs, most significantly the quality of financial statements and our measurement of default. Larger companies tend to have audited statements that are of better quality. More importantly, perhaps, is that our measure of default is more accurate for the larger firms. While we do our best to make sure that companies in our database are truly defaulted or nondefaulted companies, and in fact exclude more data than we use because of this effort, inevitably we do make some misidentifications. Therefore, size and data quality correlate positively.“

¹⁷¹ Andererseits argumentieren BOHN, ARORA, KORABLEV (2005, S. 14), daß die Finanzkennziffern von großen Unternehmen schwieriger zu interpretieren seien, da diese häufig in mehreren Segmenten operierten, was die Leistungsfähigkeit kennzahlenbasierter Ratingmodelle einschränken würde.

über deren Ausfallstatus beeinträchtigt die Qualität von Insolvenzprognosemodellen, die mit diesen Daten entwickelt werden und verfälscht die Validierungsergebnisse.^{172,173}

- *Regionale Herkunft:* Die Prognosekraft der Agenturratings variiert erheblich mit der regionalen Herkunft der beurteilten Unternehmen. So sind die Agenturratings für europäische Unternehmen deutlich trennschärfer als für US-amerikanische Unternehmen¹⁷⁴. Ob diese Unterschiede tatsächlich auf *regionenspezifische* Eigenarten zurückzuführen sind, beispielsweise auf Unterschiede in den nationalen Rechnungslegungsstandards oder Insolvenzgesetzgebungen¹⁷⁵, ob sich die Qualität der Ratingverfahren zwischen den einzelnen regionalen Niederlassungen der Ratingagenturen unterscheidet¹⁷⁶ - oder ob sich die gefundenen Unterschiede *vollständig* auf die verschiedenen bereits genannten anderen „strukturelle Stichprobenunterschiede“ (Abdeckung unterschiedlicher Phasen des Ausfallzyklus, Unternehmensgrößen oder Branche¹⁷⁷, ...) zurückführen läßt, ist unbekannt.

Aber auch methodische Aspekte des *Prognoseverfahrens* (in Abgrenzung zu „strukturellen“ Aspekten der verwendeten *Stichproben*) können die gemessene Prognosequalität, gemessen in Accuracy Ratio, beeinflussen. Hierzu zählen beispielsweise:

- *Die zugrundegelegten Ausfallfalldefinitionen:*¹⁷⁸ Die ermittelte Schätzgüte von Insolvenzprognoseverfahren kann je nach zugrundegelegter *Ausfalldefinition* variieren, wobei die empirische Relevanz dieses Problems vermutlich nicht sehr groß ist^{179,180,181} - zumindest

¹⁷² siehe STEIN ET AL (2003, S. 30f.)

¹⁷³ Für die möglichen quantitativen Auswirkungen von Datenbereinigungen auf die ermittelte Schätzgüte siehe DWYER, KOCAGIL, STEIN (2004, S. 8, 19ff.): hier wurden Verbesserungen der Einjahres-(Fünfjahres-)Accuracy Ratio eines Verfahrens von 48,2% (40,1%) auf 51,7% (45,5%) erreicht. Siehe auch ESCOTT, GLORMANN, KOCAGIL (2001b, S.19), die sogar Verbesserungen der Einjahres-Accuracy-Ratio von 59,7% auf 70,9% erzielten. Hier ist es jedoch sehr fraglich, inwiefern durch die Bereinigungen tatsächlich *Qualitätsmängel* der Daten behoben wurden und nicht etwa nur die Datenbasis in einer für das Modell vorteilhaften Weise verzerrt wurde, indem bestimmte Gruppen von Unternehmen, bei denen das Modell zuvor häufig falsche Prognose abgab, ausgeschlossen wurden.

¹⁷⁴ Die Ein-/Fünfjahres-Accuracy Ratios für US-amerikanische Unternehmen mit S&P-Rating betragen 82%/74%, die für europäische Unternehmen hingegen 94%/84%, siehe S&P (2004b, S. 11). Für MOODY'S gelten folgende Werte: US-amerikanische Unternehmen 81,0%/66,6% [Accuracy Ratio nach MOODY'S Definition] und für europäische Unternehmen 95,3%/92,4%, siehe MOODY'S (2004c, S. 2).

¹⁷⁵ CANTOR (2004, S.3): „Regional distinctions, however, are critically important when applying this global methodology. Local expertise is likely to be quite valuable in rendering judgments about the meaning of financial statements, the macroeconomic and financial environment, and potential sources of support. Moreover, regional bankruptcy regimes may influence the incentives of issuers to service their debts and the incentives of other market participants to provide financial support in times of distress.”

¹⁷⁶ CANTOR (2004, S.3): “Why have European ratings provided more powerful rank orderings of credit risk compared to American ratings? Is it because Moody's analytical practices are better in Europe or because relative risk is simply easier to judge in Europe? The following exhibit presents data which suggests the correct answer may be ‘some of each.’”

¹⁷⁷ Siehe die Angaben in CANTOR (2004, S. 9). Abgesehen von Aaa-gerateten Unternehmen, sind europäische MOODY'S-geratete produzierende Unternehmen in der Regel deutlich größer (zwei bis viermal) als US-amerikanische MOODY'S-geratete produzierende Unternehmen. Gemäß den Angaben in BASLER AUSSCHUB (2000c, S. 33f.) ist der Anteil der *Finanzunternehmen* an allen von MOODY'S gerateten US-amerikanischen Unternehmen mit 30% viel kleiner, als der entsprechende Anteil bei europäischen Unternehmen mit 66%! Die entsprechenden Quoten für S&P betragen 52% und 69%.

Beide Faktoren (Unternehmensgröße, Branchenherkunft) können in empirischen (univariaten) Untersuchungen Unterschiede bezüglich der Trennfähigkeit von Ratings erklären.

¹⁷⁸ Siehe hierzu insbesondere BALCAEN, OOGHE (2004, S. 21ff.).

¹⁷⁹ In einer Studie auf Basis von 35.000 mittelständischen österreichischen Unternehmen zeigt HAYDEN (2003, S. 33), daß statistische Verfahren, bei deren Kalibrierung lediglich *Insolvenzen* als Ausfallereignisse betrachtet wurden, eine praktisch identische Trennschärfe bei der Prognose von *Kreditrestrukturierungen* oder *Zah-*

bei Verwendung der verschiedenen üblichen *objektiven* Ausfallkriterien (Zahlungsverzug [typischerweise mindestens 90 Tage], Kreditstrukturierung, Insolvenz).^{182,183,184} Aus theoretischen Gründen wäre es sinnvoller, den Tatbestand des *Zahlungsverzugs* nicht als *Ausfallereignis* zu definieren, da ein *Zahlungsverzug* an sich keine Schäden für die Gläubiger verursacht.^{185,186} Bestenfalls könnten *Zahlungsverzüge* gute Indikatoren für spätere tatsächliche Forderungsausfälle der Gläubiger darstellen, was aber eher eine Verwendung als *erklärende* denn als *zu erklärende* Variable nahelegt. Unter praktischen Gesichtspunkten stehen der Verwendung von *Zahlungsverzügen* (aber auch anderen Ereignissen, beispielsweise *Kreditrestrukturierungen*) als Ausfallereignis vor allem Fragen der Datenverfügbarkeit im Wege – selbst bei Zugriffsmöglichkeit auf bankinterne Unterlagen.¹⁸⁷ Noch problematischer ist aber die Ausfalldefinition gemäß Basel II, bei der ein Ausfall unter anderem bereits dann als gegeben gilt, wenn die Bank davon ausgeht, daß ein Schuldner „seinen Kreditverpflichtungen [...] *mit hoher Wahrscheinlichkeit* nicht nach-

lungverzögerungen erreichen wie statistische Modelle, die speziell auf diese Ausfalldefinitionen kalibriert wurden.

¹⁸⁰ siehe auch GRICE, DUGAN (2001, S. 154ff.)

¹⁸¹ Wichtiger als die *Art* der Ausfalldefinition scheint zu sein, ob nur die individuellen *ersten* Ausfälle prognostiziert werden sollen: LEHMANN (2003, S.8): „The analysis was also carried out with (not first-time) LLP [loan-loss-provision] [...] as definition of default. A number of ‘easily classifiable’ observations entered, the performance of the rating system rose by a considerable amount. Yet, this is rather trivial. Usually, defaulted loans enter a separate monitoring process. The bank is most of all interested in the ‘surprises’ in its non-default loan portfolio. The true capabilities of a credit rating system show in the prediction of first-time LLP, not the extrapolation of past LLP. The definition of the default criterion has great impact on the results. Therefore, studies with a different default criterion cannot be compared easily.“

¹⁸² Zu den verschiedenen üblicherweise verwendeten Zahlungsausfalldefinitionen siehe S&P (2004, S. 7f), MOODY’S (2004, S. 3) und die Tabellen in den Abschnitten 3.3 bis 3.5.

¹⁸³ Die *ökonomischen Folgen* eines Ausfalls scheinen jedoch sehr stark von der Ausfallart abzuhängen, siehe VARMA, CANTOR (2005, S. 32f., 43) für Anleiherlösquoten nach sieben unterschiedlichen (*ursprünglichen*) Ausfallereignissen. Im Fall von *Restrukturierungen* oder *Tilgungsausfällen (missed principal)* erzielen Anleihergläubiger im Durchschnitt um 35 Prozentpunkte höhere Erlösquoten als im Fall von *chapter 7 – Insolvenzen*. Fast 85% der untersuchten 1.084 Ausfälle betreffen jedoch *Zinsausfälle* oder *chapter-11-Insolvenzen*, bei denen sich die durchschnittlichen Erlösquoten - bei einer univariaten Analyse - mit rund 10% nicht erheblich unterscheiden (noch geringer sind die Unterschiede auf Basis multivariater Analysen).

¹⁸⁴ KOCAGIL, AKHAVEIN (2001, S. 5, Formatierung hinzugefügt): „The discussion about the definitions of default included within the proposals appears to have centered around when a firm would be considered to have defaulted, and hence the impact on aggregate default rate numbers and PDs. *There has been less discussion on how different default definitions might impact the variables used within internal rating tools. Our understanding is that this is because, as our own experience shows, the factors that can predict default are generally the same, whether the definition of default is 90 days past due or bankruptcy* (in fact many of the definitions contained within BIS II are steps on the road to bankruptcy/insolvency).“

¹⁸⁵ In empirischen Untersuchungen finden beispielsweise GUPTON, STEIN (2005, S. 22), daß bei Bankkrediten zwischen 20% und 50% aller „Zahlungsverzugs-Ausfälle“ keinerlei ökonomische Schäden für die betroffenen Banken verursachen, ohne daß diese beispielsweise eventuell vorhandene Sicherheiten verwerten oder die Kredite restrukturieren müßten. Aus diesen Gründen werden derartige Vorfälle von vielen Banken weder als „Ausfälle“ angesehen noch als solche erfaßt und dokumentiert. Dieses bankübergreifend uneinheitliche Datenerfassungsverhalten führt jedoch zu widersprüchlichen Messungen hinsichtlich der realisierten Ausfallquoten und zu Inkompatibilitäten hinsichtlich weiterer Basel-II-Risikoparameter, wie beispielsweise den LGD-Quoten (Verlustquoten bei Eintritt eines Ausfalls), siehe hierzu auch BASLER AUSSCHUB (2005, S. 63) und NORDEN, WEBER (2005, S. 48f.).

¹⁸⁶ siehe hierzu auch KOCAGIL, AKHAVEIN (2001, S. 5), KOCAGIL ET AL. (2003, S. 6)

¹⁸⁷ siehe FALKENSTEIN, BORAL, KOCAGIL (2000, S. 7): “If a company has its loans restructured in such a way that there is an adverse effect upon the lender, such as moving payments back in time without any compensation, in general we do not capture this as a default. This is not our intention, as adverse restructurings are part of Moody’s corporate definition of default. It is a limitation of the data: rarely are such restructurings properly recorded in internal systems.”

kommt“.¹⁸⁸ Als Hinweis auf eine drohende Zahlungsunfähigkeit gilt dabei unter anderem, wenn die Bank „eine *Wertberichtigung* oder Abschreibung aufgrund einer *deutlichen Verschlechterung der Kreditqualität* seit der Hereinnahme des Kredits [bucht]“. Bei der Ermittlung der Schätzgüte derartiger Ausfallprognosen wird somit letztlich nicht mehr geprüft, wie gut die Banken Insolvenzen oder andere Zahlungsausfälle prognostizieren können, sondern wie gut sie ihre eigenen *Ausfallprognosen prognostizieren* können (siehe *hohe Wahrscheinlichkeit, verschlechterte Kreditqualität*), was eher einem Test der *Stabilität* als der *Prognosefähigkeit* von Ratings entspricht, oder aber ihre eigenen künftigen Handlungen (Bildung von *Wertberichtigungen*) prognostizieren können. Wird die Bildung von Rückstellungen durch künftige Ausprägungen des gleichen Ratingsystems ausgelöst, dessen Prognosefähigkeit überprüft werden soll, so kann die Prognose von Rückstellungsbildungen ebenfalls eher als ein Test der Stabilität als der Prognosefähigkeit der Ratings interpretiert werden!

- *Die angestrebte Stabilität der Ratingergebnisse:*^{189,190} Durch die Verfolgung anderer, konkurrierender Zielstellungen, insbesondere durch eine „künstliche“ Stabilisierung von Ratingurteilen (siehe auch die obigen Anmerkungen zum OTC-Verfahren), kann die Präzision von Insolvenzprognoseverfahren erheblich negativ beeinflusst werden.¹⁹¹
Anmerkung: Nicht zu verwechseln ist die angestrebte *Stabilität individueller Ratingnoten* mit der *Stabilität des Ratingsystems*. Mit letzterem Gütekriterium wird gemessen, ob die Prognosegüte des Ratingsystems systematisch im Zeitverlauf (aber bei identischem Prognosehorizont) nachläßt¹⁹², siehe auch die Anmerkungen zu *Pool- vs. Kohortendaten*.
- *Die Feinheit der Ratingskala:* Je unpräziser der Rater sein Ratingurteil kommuniziert, desto geringer ist die Trennfähigkeit seiner Insolvenzprognosen. Bei sieben Ratingklassen (AAA, AA, A, ..., CCC/C) ist mit einem Informationsverlust, gemessen in Accuracy Ratio, gegenüber einem Rating mit kontinuierlicher Skala in Höhe von ca. 2% - 3% (1 bis 2,5 Prozentpunkte) zu rechnen, bei einem 17-stufigen Ratingsystem hingegen nur mit einem Informationsverlust von 0,5% (< 0,5 Prozentpunkte) (siehe die Abschnitte 2.3.4 und Anhang IV).

¹⁸⁸ BASLER AUSSCHUB (2004, Tz. 452f., Formatierungen hinzugefügt)

¹⁸⁹ “Many financial market participants - investors, regulators and issuers - desire stable ratings. However, while reflecting an aversion to volatility per se, their desire for rating stability also reflects the view that *more stable* ratings are *more accurate* ratings with respect to the relative fundamental credit risk of a borrower”, siehe CANTOR, MANN (2003, S. 15), “Some investors [...] highly value rating stability to avoid unexpected portfolio revisions.”, siehe FONS (2002, S. 4), “Moody’s believes that our ratings system-management practices, as set forth above, are desired by both issuers and investors. Issuers want stability in ratings and the opportunity to make changes in their financial condition, if possible, to avoid changes in ratings.”, siehe ebenda (S. 12).

¹⁹⁰ CANTOR, MANN (2003, S. 1): “Moody’s corporate bond ratings are intended to be ‘accurate’ and ‘stable’ measures of relative credit risk, as determined by each issuer’s relative fundamental creditworthiness and without reference to explicit time horizons. Moody’s performance should therefore be measured by both rating accuracy (the correlation between ratings and defaults) and rating stability (the frequency and magnitude of ratings changes).

¹⁹¹ Als positive Folge erhöhter Ratingstabilität wird eine Verringerung der Transaktionskosten von Investoren gesehen, die - beispielsweise aus aufsichtsrechtlichen Gründen - ratingabhängige Investitionsstrategien verfolgen (müssen), siehe auch LÖFFLER (2004a).

¹⁹² siehe NORDEN, WEBER (2005, S. 41)

SWETS (1988, 1289ff.) nennt vier Voraussetzungen für die Validität und Zuverlässigkeit von ermittelten Schätzwerten, welche durch einige der aufgeführten Beispiele verletzt werden:

- (1) „*Angemessenheit der Wahrheit*“: Ex-post sollten die wahren Ausprägungen der zu prognostizierenden Variablen mit Sicherheit bekannt sein. Herrscht ex-post Unsicherheit über den tatsächlichen Ausfallstatus der zu beurteilenden Elemente der Teststichprobe, ist damit zu rechnen, daß die Schätzwerte zu niedrig ausgewiesen werden. Problematisch ist dies vor allem hinsichtlich der Aspekte *Datenqualität* und *Ausfalldefinition*.
- (2) *Unabhängigkeit von Wahrheitsbestimmung und Systembetrieb*: Der Prozeß der Wahrheitsfindung sollte unabhängig von der Einstufung der Testfälle vom System ablaufen. Andernfalls ist damit zu rechnen, daß die Schätzwerte systematisch zu gut ausgewiesen werden. Problematisch ist dieser Punkt vor allem bei *subjektiven Ausfalldefinitionen*.
- (3) *Unabhängigkeit von Stichprobe und Wahrheitsermittlung*: Das Verfahren der Wahrheitsermittlung sollte nicht zu einer Verzerrung der Stichprobe führen. In diesem Zusammenhang sei auf die Ausführungen zu den *vorselektierten Portfolien* verwiesen.
- (4) *Repräsentativität der Teststichprobe*: Hier sind die Aspekte *Positionierung im Ausfallzyklus* und *Zusammensetzung der Teststichproben* hinsichtlich der Aspekte *Rechtsform*, *Branchenzugehörigkeit*, *Unternehmensgröße* und *regionale Herkunft* zu nennen.

Generell sollte ein Vergleich der Schätzwerte unterschiedlicher Ratingverfahren nach Möglichkeit auf Basis des gleichen Unternehmensportfolios und unter identischen methodischen Prämissen (verwendete Ausfalldefinition, angestrebte Stabilisierung, ...) erfolgen. Beim empirischen Vergleich unterschiedlicher Verfahren steht diese Option jedoch nur selten zur Verfügung, insbesondere dann nicht, wenn die Ratingergebnisse nicht von unabhängiger Stelle reproduziert werden können, beispielsweise weil bewußt Details zu den Ratingverfahren geheim gehalten werden¹⁹³, das Rating durch subjektive Entscheidungen beeinflusst wird¹⁹⁴ oder das Ratingverfahren auf der Verarbeitung nicht öffentlich zugänglicher Informationen beruht.¹⁹⁵ In diesen Fällen sollen mit einem portfolio- und verfahrenübergreifenden Vergleich zumindest Tendenzaussagen, unter Berücksichtigung aller bekannten Störgrößen, getroffen werden können.

In vielen Untersuchungen werden jedoch nicht nur die Ergebnisse *eines* Verfahrens präsentiert, sondern auch die - auf den gleichen Datensätzen beruhenden - Ergebnisse anderer Verfahren, so daß zumindest hier ein direkter Vergleich der Prognosegüte der verschiedenen Verfahren möglich ist. Als ein Benchmark hat sich hier das Z-Score-Verfahren von ALTMAN durchgesetzt (siehe Abschnitt 3.4). Als weitere Benchmarks werden häufig auch die univariaten Prognoseleistungen einzelner Kennzahlen vermeldet (siehe Abschnitt 3.3).

¹⁹³ siehe beispielsweise ALTMAN, HALDEMAN, NARAYANAN (1977, S.12): „The actual coefficients [...] for the seven variables cannot be reported due to the proprietary nature of the ZETA model [...]“ oder ALTMAN, MARCO, VARETTO (1994, S. 512): „The coefficients of all the functions are protected by secrecy for the purpose of safeguarding the investments [...] made in research, testing and database creating.“

¹⁹⁴ S&P (2003b, S. 17): “There are no formulae for combining scores to arrive at a rating conclusion. Bear in mind that ratings represent an art as much as a science. A rating is, in the end, an opinion.”

¹⁹⁵ Ein Insolvenzprognoseverfahren von PLATTNER (2002, S.50f.) verwendet beispielsweise als eine von 27 (!) erklärenden Variablen eine Dummyvariable, die dann den Wert 1 annimmt, wenn die Hausbank des Unternehmens glaubt, daß das Unternehmen „temporäre Liquiditätsprobleme“ hat. Für einen Forscher außerhalb der Bank besteht damit praktisch keine Möglichkeit, die individuellen Ausprägungen dieser Variable für die von ihm betrachteten Unternehmen zu spezifizieren. Dies ist um so problematischer, als daß es sich bei dieser Dummyvariable sogar um die einflußreichste aller im Modell betrachteten Variablen handelt, siehe ebenda (S. 46).

Ferner ist zu beachten, daß keines der in der Praxis verwendeten Ratingverfahren für sich in Anspruch nehmen kann, sämtliche Insolvenzen und Nicht-Insolvenzen perfekt im Vorhinein erkennen zu können. Bestenfalls können Aussagen stochastischer Natur über das Ausfallverhalten der gerateten Unternehmen getroffen werden. Damit müssen jedoch die empirisch ermittelten Schätzwerte als Realisationen von Zufallszahlen interpretiert werden, für die sich formal approximative oder durch numerische Verfahren ermittelte Konfidenzintervalle angeben lassen.

Eine sehr konservative¹⁹⁶, im folgenden als *Näherung 1* bezeichnete, Schätzung für die Standardabweichung der Accuracy Ratio eines Portfolios (auf deren Basis in einem nächsten Schritt die Ermittlung von Konfidenzintervallen erfolgt) mit einer erwarteten AUC_{ROC} A , N_D Ausfallern und N_{ND} Nicht-Ausfallern ist wie folgt gegeben: ¹⁹⁷ Durch Ersetzen von AUC_{ROC} durch AR gemäß Formel F 8

$$F 39) \hat{\sigma}_{AR}^2 \leq 4 \cdot \frac{AUC_{ROC} \cdot (1 - AUC_{ROC})}{\min(N_D, N_{ND})}$$

$$F 40) \hat{\sigma}_{AR}^2 \leq \frac{1 - AR^2}{\min(N_D, N_{ND})} \quad (\text{Näherung 1})$$

Eine effizientere, formal nur geringfügig aufwendigere Näherung für die Standardabweichung der Accuracy Ratio, im folgenden als *Näherung 2* bezeichnet, ist wie folgt gegeben:¹⁹⁸

$$F 41) \hat{\sigma}_{AR}^2 \leq \frac{4}{3 \cdot N_D \cdot N_{ND}} \cdot \left[(2N_{ND} + 1) \cdot AUC_{ROC} \cdot (1 - AUC_{ROC}) - (N_{ND} - N_D)(1 - AUC_{ROC})^2 \right] \text{ und}$$

$$F 42) \hat{\sigma}_{AR}^2 \leq \frac{1}{3 \cdot N_D \cdot N_{ND}} \cdot \left[(2N_{ND} + 1) \cdot (1 - AR^2) - (N_{ND} - N_D)(1 - AR)^2 \right] \quad (\text{Näherung 2})$$

Eine unverzerrte Schätzung für die Standardabweichung der Accuracy Ratio kann mit folgendem, als „E-H-T (2003)“¹⁹⁹ bezeichneten, relativ umständlich zu implementierenden Verfahren ermittelt werden:

$$F 43) \hat{\sigma}_{AR}^2 = \frac{1}{(N_D - 1)(N_{ND} - 1)} \left[1 + (N_D - 1) \cdot \hat{P}_{D,D,ND} + (N_{ND} - 1) \cdot \hat{P}_{ND,ND,D} - 4(N_D + N_{ND} - 1) \left(A - \frac{1}{2} \right)^2 \right]$$

$$F 44) \hat{\sigma}_{AR}^2 = \frac{1}{(N_D - 1)(N_{ND} - 1)} \left[1 + (N_D - 1) \cdot \hat{P}_{D,D,ND} + (N_{ND} - 1) \cdot \hat{P}_{ND,ND,D} - (N_D + N_{ND} - 1) \cdot AR^2 \right] \quad \text{E-H-T (2003)}$$

Die Terme $P_{D,D,ND}$ und $P_{ND,ND,D}$ und ihre Determinanten berechnen sich wie folgt:

$$F 45) P_{D,D,ND} = P(S_{D,1}, S_{D,2} < S_{ND}) + P(S_{ND} < S_{D,1}, S_{D,2}) - P(S_{D,1} < S_{ND} < S_{D,2}) - P(S_{D,2} < S_{ND} < S_{D,1}) \text{ und}$$

$$F 46) P_{ND,ND,D} = P(S_{ND,1}, S_{ND,2} < S_D) + P(S_D < S_{ND,1}, S_{ND,2}) - P(S_{ND,1} < S_D < S_{ND,2}) - P(S_{ND,2} < S_D < S_{ND,1})$$

wobei $S_{D,1}$ und $S_{D,2}$ für Ratingscorewerte von zwei zufällig aus der Grundgesamtheit aller Ausfallern gezogenen Unternehmen stehen und $S_{ND,1}$ und $S_{ND,2}$ für die Ratingscorewerte von zwei zufällig aus der Grundgesamtheit aller Nichtausfallern gezogenen Unternehmen.

¹⁹⁶ Konservativ bedeutet hier, daß der gemäß obiger Formel ermittelte Wert größer ist als die tatsächliche Standardabweichung.

¹⁹⁷ siehe STEIN (2002, S. 19), BASLER AUSSCHUB (2005, S.41) mit $\hat{\sigma}_{AR}^2 = 4 \cdot \hat{\sigma}_{AUC}^2$

¹⁹⁸ siehe STEIN (2002, S. 19) mit $\hat{\sigma}_{AR}^2 = 4 \cdot \hat{\sigma}_{AUC}^2$

¹⁹⁹ siehe ENGELMANN, HAYDEN, TASCHKE (2003, S. 10, Formel 10) und BASLER AUSSCHUB (2005, S.40) mit $\hat{\sigma}_{AR}^2 = 4 \cdot \hat{\sigma}_{AUC}^2$

- mit $P(S_{D,1}, S_{D,2} < S_{ND}) = \sum_{i=1}^g$ (Anteil der Nichtausfaller mit einem Rating von i) *
(Anteil der Ausfaller mit einem Rating schlechter als i)², also
F 47) $P(S_{D,1}, S_{D,2} < S_{ND}) = \sum_{i=1}^g \left(a_i \cdot \frac{1 - PD_i}{1 - PD} \cdot \left(\sum_{j=i+1}^g a_j \cdot \frac{PD_j}{PD} \right)^2 \right)$ (eigene Berechnungen)
- mit $P(S_{ND} < S_{D,1}, S_{D,2}) = \sum_{i=1}^g$ (Anteil der Nichtausfaller mit einem Rating von i) *
(Anteil der Ausfaller mit einem Rating besser als i)², also
F 48) $P(S_{ND} < S_{D,1}, S_{D,2}) = \sum_{i=1}^g \left(a_i \cdot \frac{1 - PD_i}{1 - PD} \cdot \left(\sum_{j=1}^{i-1} a_j \cdot \frac{PD_j}{PD} \right)^2 \right)$ (eigene Berechnungen)
- mit $P(S_{D,1} < S_{ND} < S_{D,2}) = P(S_{D,2} < S_{ND} < S_{D,1})$ und $P(S_{D,1} < S_{ND} < S_{D,2}) = \sum_{i=1}^g$ (Anteil der Nichtausfaller mit einem Rating von i) * (Anteil der Ausfaller mit einem Rating besser als i) *
(Anteil der Ausfaller mit einem Rating schlechter als i), also
F 49) $P(S_{D,1} < S_{ND} < S_{D,2}) = P(S_{D,2} < S_{ND} < S_{D,1}) = \sum_{i=1}^g \left(a_i \cdot \frac{1 - PD_i}{1 - PD} \cdot \left(\sum_{j=1}^{i-1} a_j \cdot \frac{PD_j}{PD} \right) \cdot \left(\sum_{j=i+1}^g a_j \cdot \frac{PD_j}{PD} \right) \right)$ (eigene Berechnungen)

Analog gilt:

- **F 50)** $P(S_{ND,1}, S_{ND,2} < S_D) = \sum_{i=1}^g \left(a_i \cdot \frac{PD_i}{PD} \cdot \left(\sum_{j=i+1}^g a_j \cdot \frac{1 - PD_j}{1 - PD} \right)^2 \right)$ (eigene Berechnungen)
- **F 51)** $P(S_D < S_{ND,1}, S_{ND,2}) = \sum_{i=1}^g \left(a_i \cdot \frac{PD_i}{PD} \cdot \left(\sum_{j=1}^{i-1} a_j \cdot \frac{1 - PD_j}{1 - PD} \right)^2 \right)$ (eigene Berechnungen)
- **F 52)** $P(S_{ND,1} < S_D < S_{ND,2}) = P(S_{ND,2} < S_D < S_{ND,1}) = \sum_{i=1}^g \left(a_i \cdot \frac{PD_i}{PD} \cdot \left(\sum_{j=1}^{i-1} a_j \cdot \frac{1 - PD_j}{1 - PD} \right) \cdot \left(\sum_{j=i+1}^g a_j \cdot \frac{1 - PD_j}{1 - PD} \right) \right)$ (eigene Berechnungen)

In einem Simulationsexperiment, basierend auf einem Portfolio mit 10.000 Unternehmen und 200 Simulationsläufen, wurden für 15 unterschiedliche Ausfallwahrscheinlichkeitenvektoren²⁰⁰ $PD_1..PD_g$ und einer gegebenen Verteilung der Unternehmen $a_1..a_g$ über die Ratingklassen 1..g die Standardabweichungen der Accuracy-Ratio-Verteilungen, auf Grundlage von je 200 Meßwerten, ermittelt und mit den gemäß *Näherung 1*, *Näherung 2* und *E-H-T (2003)* berechneten Werten verglichen (siehe Abbildung 12).

Die mit *Näherung 1* ermittelten Standardabweichungen sind, bei dem zugrundeliegenden Portfolio, um den Faktor 2,3 bis 1,8 zu groß, die mit *Näherung 2* ermittelten Standardabweichungen 1,8 bis 1,4 mal zu groß. Die mit *E-H-T (2003)* bestimmten Standardabweichungen entsprechen den per Simulation ermittelten Standardabweichungen.

²⁰⁰ Für die Simulationen wurden die Unternehmen in 17 Ratingklassen eingeteilt. Die Verteilung der Unternehmen über die einzelnen Ratingklassen wurde entsprechend der Verteilung der von S&P gerateten Unternehmen (Stand: Jahresbeginn 2003) gewählt, als ratingklassenspezifische Ausfallwahrscheinlichkeiten wurden die historischen, kumulierten 1-15-Jahresausfallraten gemäß S&P (2004, S. 13) verwendet.

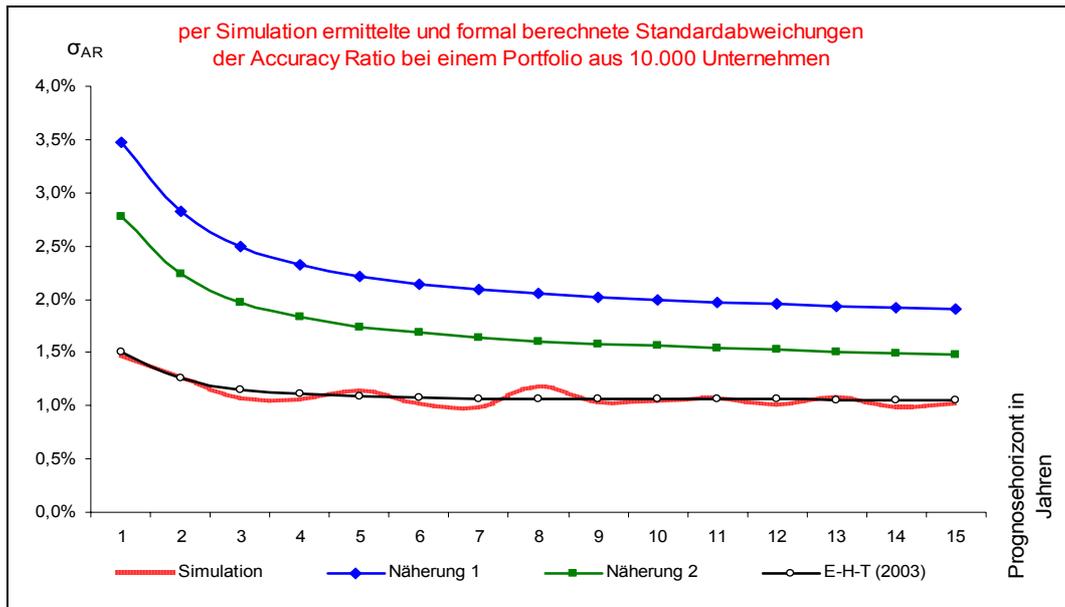


Abbildung 12: per Simulation und formal (exakt und approximativ) berechnete Standardabweichungen der Accuracy Ratio eines Ratingverfahrens basierend auf einem Portfolio von 10.000 Unternehmen, 15 Prognosehorizonten und je 200 Simulationsläufen

Anmerkung: Wie aus den entsprechenden Formeln ersichtlich ist, werden für die beiden *Näherungsverfahren* lediglich Angaben über die erwartete Präzision des Ratingmodells (gemessen in AUC_{ROC} oder AR) und die Anzahl der Ausfall- und Nicht-Ausfaller benötigt, während das exakte (unverzerrte) Verfahren E-H-T(2003) ratingklassenspezifische Angaben hinsichtlich Ausfallwahrscheinlichkeiten und Häufigkeiten benötigt, die jedoch in den meisten Insolvenzprognosestudien nicht veröffentlicht werden.

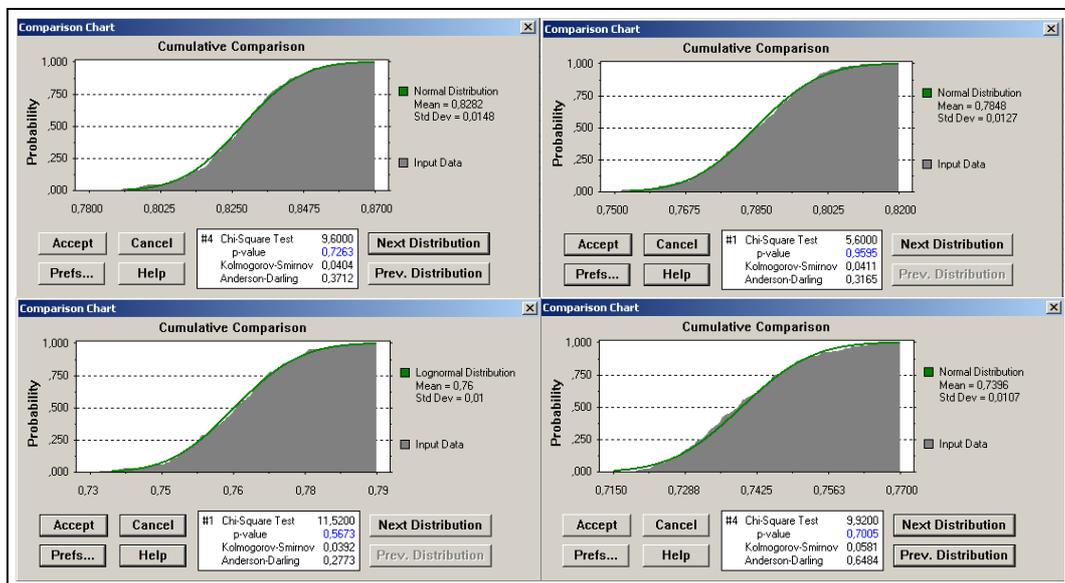


Abbildung 13: Per Simulation ermittelte Accuracy-Ratio-Verteilungen (graue Flächen) und per Maximum-Likelihood-Schätzung angepaßte Normalverteilungen (Linienzüge)

Sind die Accuracy-Ratio-Werte näherungsweise normalverteilt²⁰¹, so lassen sich Konfidenzintervalle zum Vertrauensniveau α wie folgt konstruieren:²⁰²

$$F 53) KI_{\alpha} = \left[AR \pm \sigma_{AR} \cdot \Phi^{-1}\left(\frac{1+\alpha}{2}\right) \right]$$

mit Φ^{-1} ... Umkehrfunktion der Standardnormalverteilung

Für die gewählten, mit 10.000 Unternehmen sehr großen Portfolien ist die Normalverteilungsapproximation für die Accuracy Ratio angemessen (siehe Abbildung 13)²⁰³, so daß hier Konfidenzintervalle für die Accuracy Ratio auf formalem Wege (mittels der Formeln F 44 bis F 53) bestimmt werden können.

Für kleinere Portfolien, bzw. Portfolien mit sehr kleinen Ausfallwahrscheinlichkeiten ist die Annahme normalverteilter (oder *poissonverteilter*²⁰⁴) Ausfallquoten und Accuracy Ratios möglicherweise nicht angemessen²⁰⁵, so daß Konfidenzintervalle nur noch numerisch, d.h. per Simulation, bestimmt werden können.

Beim empirischen Vergleich von Ratingsystemen ist das Ausmaß an Unsicherheit bezüglich der Qualität von Ratingsystemen zu berücksichtigen, das sich allein durch zufällige Realisierungen individueller Ausfälle ergeben kann. Bei dem gewählten Beispielportfolio mit 10.000 Unternehmen ergibt sich für den Einjahreshorizont eine *erwartete* Accuracy Ratio von 82,9%. Mit einer Wahrscheinlichkeit von 5% sind aber Accuracy-Ratio-Realisierungen zu verzeichnen, die um mehr als 3 Prozentpunkte über oder unter diesem Wert liegen.²⁰⁶

²⁰¹ Eine exakte Normalverteilung ist nicht denkbar, da die Accuracy Ratios in jedem Simulationslauf konstruktionsbedingt zwischen +1 und -1 liegen müssen.

²⁰² siehe beispielsweise ENGELMANN, HAYDEN, TASCHE (2003, S. 10)

²⁰³ Dargestellt sind in den vier Grafiken in Abbildung 13 die empirischen Verteilungen der in den Simulationen ermittelten Accuracy-Ratio-Werte für die ersten 4 der 15 PD-Vektoren (ratingklassenspezifische Ausfallwahrscheinlichkeiten für die Prognosehorizonte 1 bis 4 Jahre). In der Abbildung sind auch die Testgütemaße für die Qualität der Anpassungen gemäß drei verschiedenen Testverfahren, χ^2 -, KOLMOGOROV-SMIRNOV- und ANDERSON-DARLING-Test, dargestellt. Gemäß dem χ^2 -Test gelten Anpassungen mit p-Werten größer als 0,5 als gut, gemäß dem KOLMOGOROV-SMIRNOV-Test p-Werte kleiner als 0,03 und gemäß dem ANDERSON-DARLING-Test p-Werte kleiner als 1,5 (siehe DECISIONEERING (2000, S. 141)). Gemäß dem χ^2 -Test und dem ANDERSON-DARLING-Test werden alle vier in obiger Abbildung dargestellten Anpassungen als *gut* bewertet, gemäß dem KOLMOGOROV-SMIRNOV-Test werden alle Anpassungen (knapp) als *nicht gut* bewertet.

²⁰⁴ Die Poissonverteilung kann zur Modellierung *seltener Ereignisse* verwendet werden. Im Fall von Portfolien mit sehr kleinen ratingklassenspezifischen Ausfallwahrscheinlichkeiten und „hinreichend großen“ (HARTUNG (1991, S. 122)) Ratingklassen können unter Ausnutzung des Poissonschen Grenzwertsatzes die ratingklassenspezifischen Ausfälle auch durch Poissonverteilungen, mit $\lambda_i = n_i \cdot PD_i$, modelliert werden. Die Summe von poissonverteilten Zufallszahlen ist ebenfalls poissonverteilt mit $\lambda_{\Sigma} = \sum \lambda_i$ und für $\lambda \geq 9$ läßt sich die Poisson- gut durch eine Normalverteilung approximieren, siehe HARTUNG (1991, S. 213).

²⁰⁵ Bestünde ein Portfolio aus n Unternehmen mit *identischen* Ausfallwahrscheinlichkeiten PD, so ließe sich die tatsächlich binomialverteilte Anzahl der Ausfälle nach dem Theorem von DE MOIVRE/ LAPLACE gut durch eine Normalverteilung mit $\mu = n \cdot PD$ und $\sigma = \sqrt{n \cdot PD \cdot (1 - PD)}$ approximieren, falls $n \cdot PD \cdot (1 - PD) \geq 9$, siehe HARTUNG (1991, S. 201).

Für die entsprechenden *Ausfallquote* gälte somit $\mu_{\text{Quote}} = PD$ und $\sigma_{\text{quote}} = \sqrt{\frac{PD \cdot (1 - PD)}{n}}$

Für kleine PD, d.h. $\mu = n \cdot PD \approx n \cdot PD \cdot (1 - PD)$, ist eine gute Normalverteilungsapproximation der Ausfälle also dann möglich, wenn der Erwartungswert der Ausfälle mindestens 9 beträgt. Für PD=1% ergäbe sich beispielsweise eine Mindestportfoliogröße von $\frac{9}{1\% \cdot 99\%} \approx \frac{9}{1\%} = 900$ Unternehmen (mit identischen Ausfallwahrscheinlichkeiten).

²⁰⁶ Mit $\alpha=95\%$, $-\Phi^{-1}(2,5\%) = \Phi^{-1}(97,5\%) \approx 1,96$ und $\sigma_{AR} = 1,5\%$ ergibt sich ein Konfidenzintervall $KI_{95\%}$ mit $KI_{95\%} = [82,9\% \pm 2,9\%] = [80,0\%;85,8\%]$

Bei einem Portfolio identischer Struktur mit nur 5.000 Unternehmen, was ungefähr der derzeitigen Anzahl der von MOODY'S oder S&P gerateten Unternehmen entspricht²⁰⁷, ergäben sich näherungsweise um den Faktor $\sqrt{2}$ größere 95%-Konfidenzintervalle ($\pm 4,2\%$); bei einem Portfolio von „nur“ 1.600 Unternehmen, was ungefähr der Anzahl der von S&P und MOODY'S gerateten Unternehmen Mitte der Achtziger Jahre entspricht, sogar um den Faktor 2,5 größere Konfidenzintervalle ($\pm 7,4\%$).

Trotz der genannten Einschränkungen hinsichtlich der Vergleichbarkeit empirisch ermittelter Schätzgütemaße verschiedener Ratingverfahren sollen in den folgenden Abschnitten die entsprechenden Werte verschiedener Insolvenzprognosestudien gegenübergestellt werden, wobei – soweit Angaben hierzu vorhanden waren – auch die Ausprägungen der oben aufgeführten *strukturellen Einflußgrößen* auf die erreichbaren Schätzgütern (Umfang der Datenbasis, zeitliche und branchenmäßige Herkunft der untersuchten Unternehmen, Unternehmensgrößen, Datenquelle: Bankportfolio ja/nein) aufgeführt werden.

Die Leistungsfähigkeit der verschiedenen Verfahren soll aber nicht nur untereinander verglichen werden (siehe Abschnitt 3.5), sondern auch in Relation zu drei *Benchmarkverfahren* beurteilt werden, die bereits mit einfachsten Mitteln umgesetzt werden können:

- In Abschnitt 3.2 wird hierfür gezeigt, welche Schätzgüte sich bereits mit Insolvenzprognosen erreichen läßt, die lediglich auf der Kenntnis von Rechtsform und Branchenzugehörigkeit eines Unternehmens basieren.
- In Abschnitt 3.3 werden empirische Schätzgütern von Insolvenzprognosen vorgestellt, die auf der Kenntnis *einzelner* aus den Jahresabschlüssen der Unternehmen berechneter Kennzahlen beruhen.
- In Abschnitt 3.4 wird die Leistungsfähigkeit eines *multivariaten* Insolvenzprognosemodells untersucht, das in vielen empirischen Studien als Benchmarkverfahren verwendet wird.

²⁰⁷ Zum 1. Januar 1985 (2004) verfügten 1.620 (4.810) Unternehmen über ein Rating von MOODY'S und 1.650 (5.189) Unternehmen über ein Rating von S&P, siehe MOODY'S (2005, S.35) und S&P (2005, S. 26)

3.2 Benchmark I: Erzielbare Schätzgüte bei Kenntnis von Rechtsform und Branchenzugehörigkeit eines Unternehmens

In diesem Abschnitt wird dargelegt, welche Schätzgüte sich mit ordinalen Insolvenzprognosen erreichen läßt, die lediglich auf den öffentlich zugänglichen Insolvenzstatistiken beruhen. In den Insolvenzstatistiken des STATISTISCHEN BUNDESAMTES werden univariate Aufschlüsselungen von Insolvenzhäufigkeiten nach Branchen, Rechtsformen, Bundesländern sowie Unternehmensalter und -größe (Anzahl der Beschäftigten) veröffentlicht. Zusätzlich werden für die Unternehmen verschiedener Branchen, Rechtsformen und Bundesländer univariate *Insolvenzquoten* veröffentlicht.²⁰⁸ Sind Insolvenzhäufigkeiten *und* -quoten bekannt, so lassen sich hieraus die Unternehmensbestände der verschiedenen Gruppen berechnen.²⁰⁹ Sortiert man die einzelnen Gruppen nach realisierten Insolvenzquoten, läßt sich mit diesen Parametern (Insolvenzquoten, Häufigkeiten) der univariate *ex-post*-Informationsnutzen (Accuracy Ratio) obiger Kriterien ermitteln.

Für *Prognosezwecke* muß jedoch eine *ex-ante*, d.h. zu Beginn derjenigen Periode, für welche Ausfälle prognostiziert werden sollen, zur Verfügung stehendes Sortierkriterium vorliegen. Im folgenden werden hierfür die *realisierten* Insolvenzraten des Vorjahres verwendet. Ausgefeiltere Verfahren sind durchaus denkbar, jedoch kann die Qualität *ordinaler* Insolvenzprognosen hierdurch bestenfalls marginal verbessert werden (siehe im Vorgriff auf die Ergebnisse dieses Abschnitts Abbildung 18).

Um auf Basis von Informationen hinsichtlich Branchen-, Rechtsform- oder Bundeslandzugehörigkeit eines Unternehmens trennfähige Insolvenzprognosen ableiten zu können, müssen: (I) nennenswerte Unterschiede der Insolvenzquoten von Unternehmen der verschiedener Gruppen und (II) eine hohe zeitliche Stabilität dieser Unterschiede gegeben sein. Diese Aspekte werden im folgenden untersucht. Abbildung 14 stellt hierfür die Entwicklung der Insolvenzquoten, untergliedert nach 14 Wirtschaftsbereichen, für den Zeitraum 1994-2003 dar, Abbildung 15 die Entwicklung der Insolvenzquoten untergliedert nach Rechtsformen von 1980-2003 und Abbildung 16 untergliedert nach Bundesländern für 1980-2003.²¹⁰ Insgesamt werden in den diversen Auswertungen 3,3 Mio. Unternehmen (Stand: 2003) erfaßt.²¹¹

²⁰⁸ Siehe beispielsweise STATISTISCHES BUNDESAMT (2004a, 2004c). Zusätzlich finden sich hier auch die für Ausfallprognosezwecke eher uninteressanten Untergliederungen von Insolvenzen nach Antragstellern (Gläubiger vs. Schuldner) und Eröffnungsgründen (Überschuldung, Zahlungsunfähigkeit, ...).

²⁰⁹ $\text{Anzahl Unternehmen}_i = \text{Insolvenzen}_i / \text{Insolvenzquote}_i$, falls $\text{Insolvenzquote}_i > 0\%$.

²¹⁰ Datenquellen für die univariaten Auswertungen: 1980-1998 GÜNTERBERG, WOLTER (2003) und für 1999-2003 STATISTISCHES BUNDESAMT (2004b).

²¹¹ Mit zwei Ausnahmen (siehe unten), entsprechen die Unternehmensbestandsangaben im wesentlichen denen der Umsatzsteuerstatistik, in der alle steuerpflichtigen Unternehmen mit einem Jahresumsatz von mindestens 16.617 Euro erfaßt werden. Zusätzlich berücksichtigt werden 350.000 GmbHs und 7.000 AGs (Stand: 2003), die insbesondere als Beteiligungsgesellschaften nicht der Umsatzsteuerpflicht unterliegen und somit nicht in der Umsatzsteuerstatistik erfaßt werden (STATISTISCHES BUNDESAMT (2004a, S. 20f)). Bei der Berechnung der rechtsformabhängigen Insolvenzquoten werden die nicht in der Umsatzsteuerstatistik erfaßten GmbHs und AGs vom STATISTISCHEN BUNDESAMT korrekt berücksichtigt (ebenda) – nicht jedoch bei der Berechnung von branchenspezifischen oder der gesamtwirtschaftlichen Insolvenzquote, wie sich durch Rückrechnung auf Basis der Insolvenzhäufigkeiten und -quoten (S. 19, 20, 41) nachvollziehen läßt, die damit um ca. 12% zu *hoch* ausfallen, was bei einer telefonischen Anfrage beim Statistischen Bundesamt auch bestätigt wurde. Die gesamtwirtschaftliche Insolvenzquote für Deutschland beträgt somit für 2003 nicht 1,35% (ebenda, S. 20) oder 1,34% (STATISTISCHES BUNDESAMT (2004c) sondern 1,19% (eigene Berechnungen). Zur Vermeidung dieses Fehlers wurde bei den eigenen Auswertungen auf Basis der Relationen im Jahr 2003 ein Hochrechnungsfaktor für den GmbH-Bestand aus der Umsatzsteuerstatistik von 1,83 und für AGs von 2,07 ermittelt und für alle anderen Kalenderjahre zur Korrektur der branchenspezifischen Unternehmensbestandsangaben angewendet. Größere Abweichungen gegenüber den (fehlerhaften) branchenspezifischen Insolvenzquoten des STATISTISCHEN BUNDESAMTES gibt es in Branchen mit einem überdurchschnittlich hohen

Zu Abbildung 14: Es sind deutliche Unterschiede zwischen den Insolvenzquoten der verschiedenen Branchen²¹² zu verzeichnen und die Relationen sind relativ stabil im Zeitverlauf: von den fünf Branchen mit den höchsten Insolvenzquoten im Jahr 1994 sind selbst 2003 noch vier unter den fünf Branchen mit den höchsten Insolvenzquoten. Von den fünf „besten“ Branchen (Branchen mit den niedrigsten Insolvenzquoten) aus dem Jahr 1994 sind auch 2003 noch drei unter den fünf „besten“ Branchen.²¹³

Anhaltende, deutlich überdurchschnittlich hohe Insolvenzquoten sind nur im *Baugewerbe* zu verzeichnen; nur im Jahr 2003 war in einer anderen, *zahlenmäßig* jedoch relativ unbedeutenden²¹⁴ Branche (*Kredit- und Versicherungsgewerbe*²¹⁵), eine noch höhere Insolvenzquote festzustellen. Ständig unterdurchschnittlich niedrige Insolvenzquoten sind beispielsweise in den Branchen *Energie- und Wasserversorgung*, *Erziehung und Unterricht* sowie *Erbringung sonstiger öffentlicher und persönlicher Dienstleistungen* zu verzeichnen.

In Tabelle B werden die in den einzelnen Jahren erzielbaren Accuracy-Ratio-Werte von ordinalen Insolvenzprognosen (Prognosedauer: 1 Jahr) dargestellt, deren einzige erklärende Variable die Branchenzugehörigkeit eines Unternehmens ist.²¹⁶ Dargestellt werden sowohl die „Prognosewerte“, die sich bei Kenntnis der tatsächlichen Reihenfolge der branchenspezifischen Ausfallquoten in den jeweiligen Jahren ergeben hätte („ex-post“) und die Prognosewerte bei Unterstellung der gleichen Reihenfolge wie im Vorjahr („ex-ante“).

GmbH-Anteil: Bergbau und Gewinnung von Steinen und Erden: 1,07% (STATISTISCHES BUNDESAMT (2004c)) vs. 0,75% (eigene Berechnungen) und Verarbeitendes Gewerbe: 1,60% (STATISTISCHES BUNDESAMT (2004c)) vs. 1,30% (eigene Berechnungen). Geringfügige Abweichungen sind in Branchen mit unterdurchschnittlichem GmbH-Anteilen zu verzeichnen: Gastgewerbe 1,25% vs. 1,20% oder Gesundheitswesen: 1,43% vs. 1,35%.

²¹² Amtliche Wirtschaftszweigbezeichnung und ggf. eigene verwendete Abkürzung: A: Land- u. Forstwirtschaft („Landwirtschaft“), B: Fischerei u. Fischzucht („Fischerei“), C: Bergbau und Gewinnung von Steinen und Erden („Bergbau“), D: Verarbeitendes Gewerbe, E: Energie- und Wasserversorgung („Energie“), F: Baugewerbe, G: Handel; Instandhaltung und Reparatur von Kfz und Gebrauchsgütern („Handel“), H: Gastgewerbe, I Verkehr und Nachrichtenübermittlung („Verkehr“), J: Kredit- und Versicherungsgewerbe („Kredit“), K: Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von wirtschaftlichen Dienstleistungen, („Vermietung“), M: „Erziehung und Unterricht („Erziehung“), N: Gesundheits-, Veterinär- und Sozialwesen („Gesundheit“), O: Erbringung sonstiger öffentlicher und persönlicher Dienstleistungen („sonst_DL“).

²¹³ Da im folgenden nur *Einjahres*ausfallprognosen erstellt werden sollen, ist die Stabilität der Ausfallquoten auf Einjahressicht entscheidend: die durchschnittliche Rangänderung beträgt 0,94 Ränge je Branche und Jahr. In 45% aller Fälle blieb der Rang der Branche im Vorjahresvergleich konstant, in 32% trat eine Änderung um einen Rang, in 15% um zwei Ränge und nur in 7% aller Fälle eine Änderung um mehr als 2 Ränge im Vorjahresvergleich auf. Bei zufälligen Rangordnungen betrüge die durchschnittliche Rangänderung 4,64 Ränge je Branche und Jahr und in 67% aller Fälle wäre eine Rangänderung von mehr als zwei Rängen zu verzeichnen gewesen.

²¹⁴ Nur 0,5% aller rund 3,3 Mio. Unternehmen gehören der Branche „Kredit- und Versicherungsgewerbe“ an, hingegen gehören 11,4% aller Unternehmen dem Baugewerbe an.

²¹⁵ Hierunter fallen beispielsweise Kreditinstitute (WZ 65.12), Kapitalanlagegesellschaften (WZ 65.23.1), das Versicherungsgewerbe (WZ 66) oder Versicherungsvertreter (WZ 67.20.1).

²¹⁶ Zur Berechnung der Accuracy Ratio werden neben der Sortierreihenfolge und den in Abbildung 14 dargestellten branchenspezifischen Ausfallquoten zusätzlich nur noch die Anteile der Branchen an der Gesamtzahl aller Unternehmen benötigt.

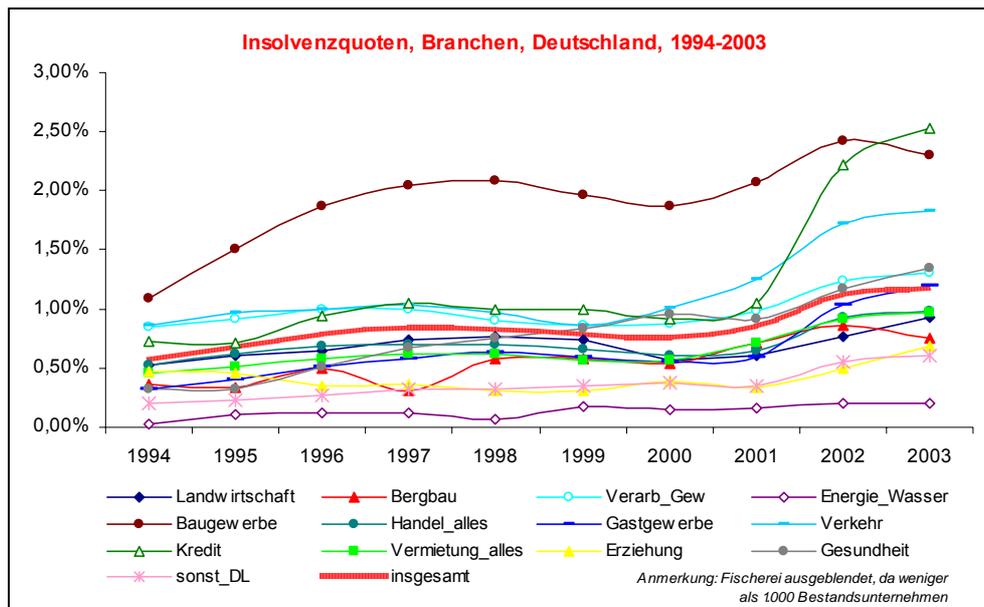


Abbildung 14: Insolvenzzquoten nach Wirtschaftsbereichen, Deutschland, 1994-2003²¹⁷

Acc. Ratio	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	MW
ex-ante		25,5%	26,7%	25,8%	25,4%	25,4%	24,7%	25,0%	20,9%	19,8%	24,3%
ex-post	23,3%	25,6%	26,7%	25,9%	25,6%	25,4%	25,1%	26,0%	21,9%	19,8%	24,5%

Tabelle B: Univariate Schätzgüte (Accuracy Ratio) Branchenzugehörigkeit, ex-post-Sortierung (realisierte Ausfallraten des aktuellen Jahres) und ex-ante-Sortierung (realisierte Ausfallraten des Vorjahrs)

In der folgenden Abbildung 15 sind die Insolvenzzquoten für fünf verschiedene Rechtsformen (GmbH, AG (inkl. KGaA), Personengesellschaften, Einzelunternehmen und sonstige Rechtsformen²¹⁸) für den Zeitraum 1980-2003 dargestellt. Auffallend sind die anhaltend und weit überdurchschnittlichen hohen Insolvenzzquoten von GmbHs. Obwohl ihr Anteil an allen Unternehmen nur 25% beträgt (2003), entfallen auf sie 51% aller Insolvenzen (2003). Ihre Insolvenzzquote ist damit ungefähr doppelt so hoch, wie die durchschnittliche Insolvenzzquote aller Unternehmen und dreimal so hoch wie die durchschnittliche Insolvenzzquote aller *anderen* Unternehmen (in anderen Jahren ist sie sogar bis zu 5,5 mal so hoch). Noch höhere Insolvenzzquoten sind, zumindest ab 2001, nur bei den Aktiengesellschaften (inkl. KGaA) zu verzeichnen, deren Anteil an allen Unternehmen aber vergleichsweise unbedeutend ist (0,4%).

²¹⁷ Quelle: eigene Berechnungen basierend auf GÜNTERBERG, WOLTER (2003, S. 142f) 1994-1998 (Herausrechnung Kleingewerbetreibende) und STATISTISCHES BUNDESAMT (2004b) 1999-2003, branchenspezifische Korrektur Untererfassung GmbH und AG

²¹⁸ Hierunter zählen beispielsweise Vereine oder Genossenschaften. Nur 1,6% aller erfassten Unternehmen haben diese Rechtsform gewählt.

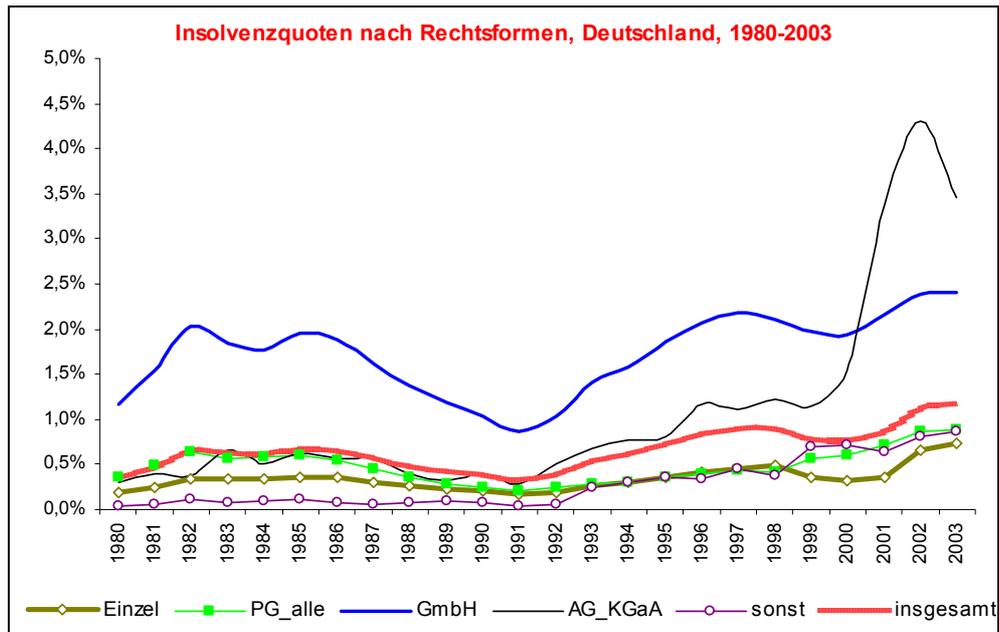


Abbildung 15: Insolvenzquoten nach Rechtsformen, Deutschland, 1994-2003²¹⁹

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	...
ex-post		41,4%	40,6%	37,9%	37,9%	38,3%	38,4%	38,0%	38,4%	38,9%	38,6%	38,6%	...
ex-ante	39,8%	41,4%	40,6%	37,9%	37,9%	38,3%	38,4%	38,0%	38,4%	38,9%	38,6%	38,6%	...

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	MW
	40,0%	38,8%	39,1%	39,3%	37,9%	37,5%	36,1%	35,2%	41,5%	42,2%	30,6%	28,1%	38,0%
	40,0%	38,8%	39,1%	39,3%	38,4%	37,5%	36,1%	39,4%	41,5%	42,5%	30,6%	28,1%	38,3%

Tabelle C: Univariate Schätzigüte (Accuracy Ratio) Rechtsform, ex-post-Sortierung und ex-ante-Sortierung

Als letztes univariates Erklärungskriterium für Unternehmensinsolvenzen wird in Abbildung 16 die Bundeslandzugehörigkeit betrachtet. Auffallend ist hier die divergente Entwicklung des Insolvenzgeschehens zwischen Neuen und Alten Bundesländern im Zeitraum von 1990 bis 1997. Ab 1997 stabilisiert sich das Ost-West-Gefälle bei den Insolvenzquoten, ab ca. 1999 steigen die Insolvenzquoten in den Alten und ab 2003 fallen in den Neuen Bundesländern.

Die Rangfolge der Insolvenzquoten innerhalb der Neuen und innerhalb der Alten Bundesländer ist relativ stabil. Innerhalb der Neuen Bundesländer erzielen Thüringen und Brandenburg vergleichsweise niedrige und Sachsen-Anhalt und Berlin (Ost) vergleichsweise hohe Insolvenzquoten. Innerhalb der Alten Bundesländer zeichnet sich ein Nord-Süd-Gefälle bei den Insolvenzquoten ab: besonders gut schneiden Baden-Württemberg, Bayern, Hessen und Rheinland-Pfalz ab - und ab 1999 auch das Saarland, besonders schlecht (neben Berlin-West) Nordrhein-Westfalen, Schleswig-Holstein, Bremen, Niedersachsen - und ab 2001 auch Hamburg.

²¹⁹ Quelle: eigene Berechnungen basierend auf GÜNTERBERG, WOLTER (2003, S. 144f) für 1994-1998 (Herausrechnung Kleingewerbetreibende) und STATISTISCHES BUNDESAMT (2004b) für 1999-2003.

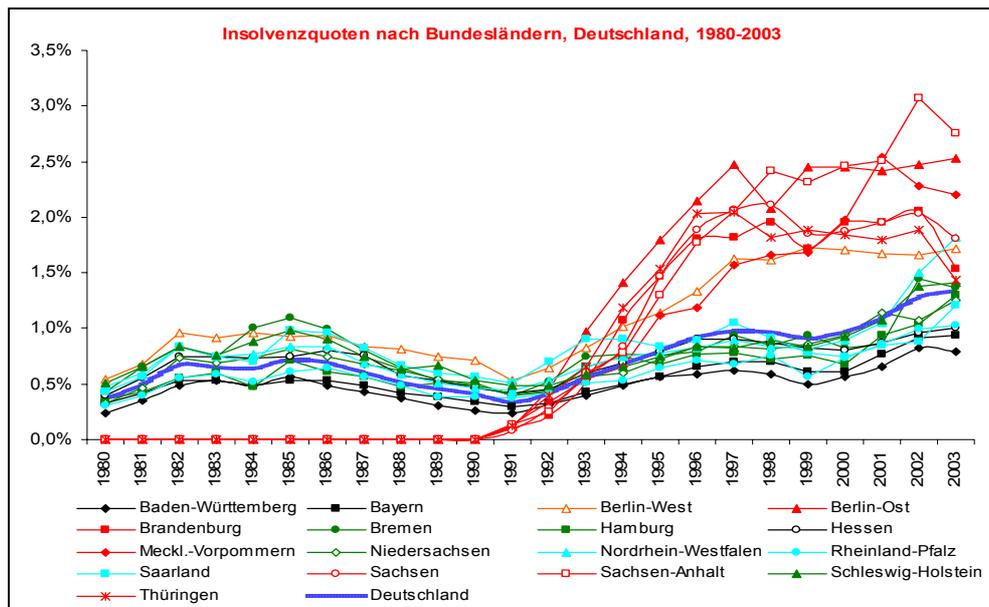


Abbildung 16: Insolvenzquoten nach Bundesländern, Deutschland, 1980-2003²²⁰

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	...
ex-post		11,7%	12,0%	8,7%	12,3%	10,6%	12,0%	11,1%	10,8%	12,7%	12,7%	19,1%	...
ex-ante	13,2%	11,8%	12,1%	8,7%	12,8%	11,7%	12,6%	11,7%	11,4%	13,0%	12,9%	19,8%	...

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	MW
	13,1%	9,7%	11,5%	16,8%	20,4%	21,2%	22,0%	24,4%	23,8%	21,5%	18,2%	18,5%	15,4%
	14,0%	12,0%	14,4%	17,8%	20,5%	21,6%	22,4%	25,0%	24,5%	21,6%	19,9%	19,4%	16,0%

Tabelle D: Univariate Schätzgüte (Accuracy Ratio) Bundesland, ex-post-Sortierung und ex-ante-Sortierung

Neben der *univariaten* ist insbesondere die *multivariate* Prognosegüte der einzelnen Variablen von Interesse. Hierfür müßten Bestands- und Insolvenzdaten von Gruppen von Unternehmen mit jeweils identischen Ausprägungen hinsichtlich der Kriterien Branche, Rechtsform und Bundesland vorliegen - bzw. individuelle Datensätze mit obigen drei Kriterien. Derartige Daten können nur von den Statistischen Landesämtern, nicht jedoch vom Statistischen Bundesamt geliefert werden. Vom Statistischen Bundesamt zur Verfügung gestellt werden konnte jedoch eine Kreuztabellierung von Unternehmensbestands- und Insolvenzdaten nach Branchen und Rechtsformen für die Jahre 1999-2003.²²¹

Das weitere Vorgehen ist analog dem oben beschriebenen Vorgehen bei den univariaten Auswertungen, mit dem Unterschied, daß nun nicht mehr nur 14 Gruppen (Branchen) oder 6 Gruppen (Rechtsformen)²²² sondern 84 (=14*6) Gruppen (Kombinationen aus Branchen und

²²⁰ Quelle: eigene Berechnungen basierend auf GÜNTERBERG, WOLTER (2003, S. 151ff) 1994-2001 (Herausrechnung Kleingewerbetreibende) und STATISTISCHES BUNDESAMT (2004b) 2001-2003, Extrapolation Aufteilung Insolvenzen Berlin-gesamt in Berlin-West und Berlin-Ost ab 2001, Extrapolation Unternehmensbestand Berlin-West mit westdeutscher Wachstumsrate Unternehmensbestand ab 1991

²²¹ STATISTISCHES BUNDESAMT (2004b)

²²² Im Gegensatz zu den Daten, die für die univariaten Auswertungen zur Verfügung standen (1980-2003), konnte bei den Daten für die multivariaten Auswertungen (1999-2003) eine Untergliederung von Unternehmensbestands- und -insolvenzdaten für unterschiedliche Rechtsformen innerhalb der Personengesellschaften vorgenommen werden. Aufgrund der für die restlichen Personengesellschaften untypisch hohen Insolvenzquoten bei den GmbH&Co KG (vergleiche die entsprechenden Werte in Abbildung 17) wurde für diese Unternehmen eine separate Gruppe gebildet.

Rechtsformen) zur Verfügung stehen. Bei rund 3,3 Mio. Unternehmen und rund 40.000 Insolvenzen pro Jahr (2003) entfallen somit im *Durchschnitt* 39.000 Unternehmen und 470 Insolvenzen pro Jahr auf eine Unternehmensgruppe. Zur Vermeidung von Ausreißerwerten und der Division mit Null wurden sämtliche Gruppen, die im Jahr 1999 über weniger als 1.000 Bestandsunternehmen verfügten innerhalb einzelner Rechtsformen zusammengefaßt²²³, so daß sich die Gruppenzahl auf 61 reduzierte.²²⁴

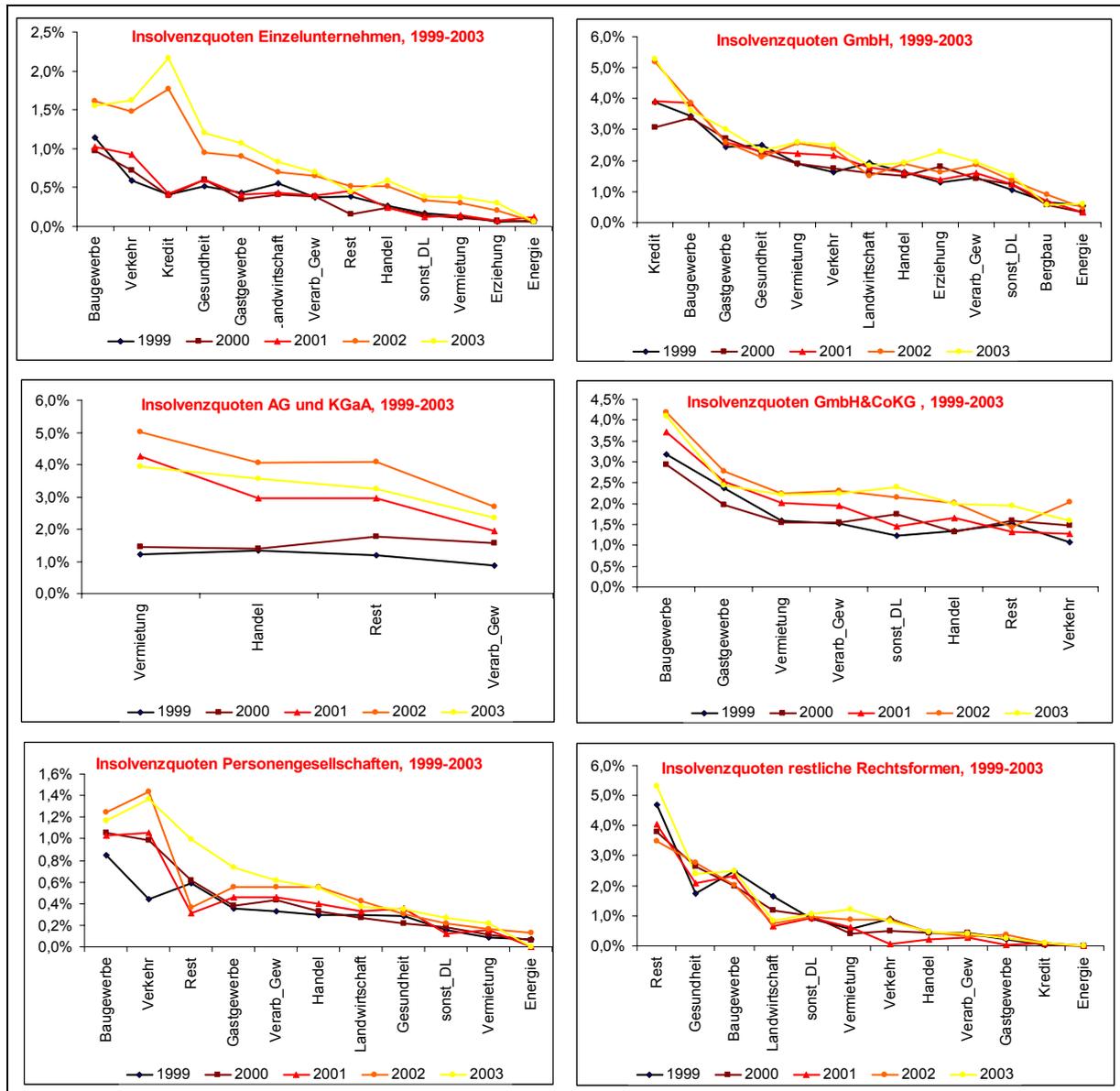


Abbildung 17: Insolvenzquoten nach Rechtsformen und Branchen, Deutschland, 1999-2003

Mit 517.130 Unternehmen (15,6% von 3,3 Mio.) war *Einzelunternehmen - Handelsgewerbe* die zahlenmäßig am stärksten vertretene Gruppe, am schwächsten war mit 479 Unternehmen (0,015%) die „Restegruppe“ *sonstige Rechtsformen – sonstige Branchen* vertreten. Nachfol-

²²³ Die Zusammenfassung erfolgte entlang der Variable *Rechtsform*, da diese einen größeren univariaten Erklärungsgehalt hatte als die Variable *Branchenzugehörigkeit*. Beispielsweise wurden die Gruppen *Einzelunternehmen - Fischerei* (716 Unternehmen) und *Einzelunternehmen - Bergbau* (591) zur Gruppe *Einzelunternehmen – sonstige Branchen* zusammengefaßt.

²²⁴ Von den Zusammenfassungen waren insgesamt nur 0,3% aller Unternehmen betroffen. Schwach besetzt sind vor allem die Branchen *Fischerei* und *Bergbau* sowie die Rechtsformen *GmbH&CoKG* und *AG/KGaA*.

gende Abbildung 17 stellt für den Zeitraum 1999-2003 die Insolvenzquoten für alle Unternehmen untergliedert nach Rechtsformen und Branchen dar.²²⁵

Die Darstellung zeigt, daß:

- deutliche Unterschiede der Ausfallraten von Unternehmen unterschiedlicher Branchen bei gleicher Rechtsform auftreten (siehe den steil fallenden Verlauf der einzelnen Graphen für alle Rechtsformen (außer AG/KGaA²²⁶)),
- deutliche Unterschiede der Ausfallraten von Unternehmen gleicher Branchen bei unterschiedlichen Rechtsformen auftreten (siehe insbesondere auch die Skalierungsunterschiede der Y-Achsen der verschiedenen Rechtsformdiagramme),
- für alle Rechtsformen die branchenspezifischen *relativen* Ausfallraten im Zeitverlauf sehr stabil sind (siehe den weitgehend *schneidungsfreier* Verlauf der Jahresgraphen innerhalb der einzelnen Diagramme) und
- für GmbHs, Personengesellschaften und *restliche Rechtsformen* auch die absoluten Ausfallraten im Zeitverlauf relativ konstant sind. Für Einzelunternehmen scheint zwischen 2001 und 2002 ein Strukturbruch bei den *absoluten* Insolvenzraten vorzuliegen, da hier ein erheblicher und weitgehend proportionaler Anstieg der Insolvenzraten über alle Branchen hinweg, möglicherweise verursacht durch Änderungen des Insolvenzrechts 1999^{227,228}, zu verzeichnen ist. Bei GmbH&Co.KG und AG/ KGaA sind hingegen im Zeitverlauf relativ proportional über alle Branchen hinweg zunehmende Insolvenzquoten zu verzeichnen.

Die Reihenfolge der Insolvenzquoten der 61 Branchen-Rechtsform-Gruppen ist im untersuchten Zeitraum trotz vereinzelter Ausreißer sehr stabil, die Korrelationskoeffizienten der Ränge benachbarter Jahre liegen zwischen 0,925 und 0,975 und erreichen damit eine fast perfekte Stabilität und sind auch statistisch hoch hochsignifikant.²²⁹

Von den zahlenmäßig bedeutenderen Gruppen (Anteil an allen Unternehmen > 1%) werden anhaltend schlechte Plazierungen von den *GmbH – Baugewerbe* (Rang_{MW 1999-2003}=58,0 von 61, PD_{MW 1999-2003}=3,6%, Anteil_{Unternehmen}=3,9%), *GmbH – Vermietung [und Dienstleistungen]*²³⁰ (Rang=51,4; PD=2,2%, Anteil=6,6%) und *GmbH – Verkehr und Nachrichtenübermittlung* (Rang=49,8; PD=2,1%, Ant.=1,1%) erzielt. Anhaltend gute Plazierungen werden hingegen von *Personengesellschaften (ohne GmbH&CO.KG) – Vermietung [und Dienstleistungen]* (Rang = 6,8; PD=0,15%, Ant.=3,9%), *Einzelunternehmen – Vermietung [und*

²²⁵ Das Sortierkriterium bei der Anordnung der Branchen innerhalb der beiden Darstellungen waren die durchschnittlichen Insolvenzquoten im Zeitraum 1999-2003.

²²⁶ Bei den Aktiengesellschaften waren gerade diejenigen Branchen zu schwach besetzt, um separat ausgewiesen zu werden, die bei den anderen Rechtsformen typischerweise mit überdurchschnittlich hohen (Baugewerbe, Gastgewerbe, Kredit- und Versicherungsgewerbe) oder niedrigen (Energie- und Wasserwirtschaft, sonstige öffentliche und private Dienstleistungen) Insolvenzquoten verbunden sind.

²²⁷ siehe beispielsweise STATISTISCHES BUNDESAMT (2004, S.5ff)

²²⁸ siehe auch PLATTNER (2002, S. 37): „Erschwert wird die Interpretation der Entwicklung seit der Insolvenzrechtsreform, weil ein Teil der neu eingeführten Verbraucherinsolvenzen die Folge dahinter stehender wirtschaftlicher Aktivitäten ist. Privatkonkurse können somit im ökonomischen Sinn Unternehmensinsolvenzen sein. (Mit der Reform der Reform, die am 1. Dezember 2001 in Kraft getreten ist, ist das Verbraucherinsolvenzverfahren stärker auf Privatpersonen beschränkt.) Darüber hinaus werden jetzt mehr Insolvenzen erfasst: So sind nunmehr auch GbRs insolvenzfähig.“

²²⁹ Auf Basis von 100.000 Simulationsläufen wurde als 99%-Quantil der Rangkorrelationen bei *zufälligen* Rängen ein Wert von 29,8% ermittelt. Das 95%-Quantil liegt bei 21,3%.

²³⁰ Zur Branchengruppe „Vermietung [und Dienstleistungen]“ gehören die Wirtschaftszweige Grundstücks- und Wohnungswesen (WZ 70), Vermietung beweglicher Sachen (WZ 71), Datenverarbeitung und Datenbanken (WZ 72), Forschung und Entwicklung (WZ 73), Erbringung von wirtschaftlichen Dienstleistungen (WZ 74).

Dienstleistungen] (Rang=8,4; PD=0,21%, Ant.=15,5%) und Einzelunternehmen – sonstige Dienstleistungen (Rang=14,6; PD=0,23%, Ant. =6,5%) erzielt.

AR	1999	2000	2001	2002	2003	MW
ex-post		53,8%	54,7%	44,5%	44,5%	49,4%
ex-ante	53,0%	54,1%	55,1%	45,2%	45,2%	50,5%

Tabelle E: Multivariate Schätzgüte (Accuracy Ratio) Branche und Rechtsform, ex-post- und ex-ante-Sortierung

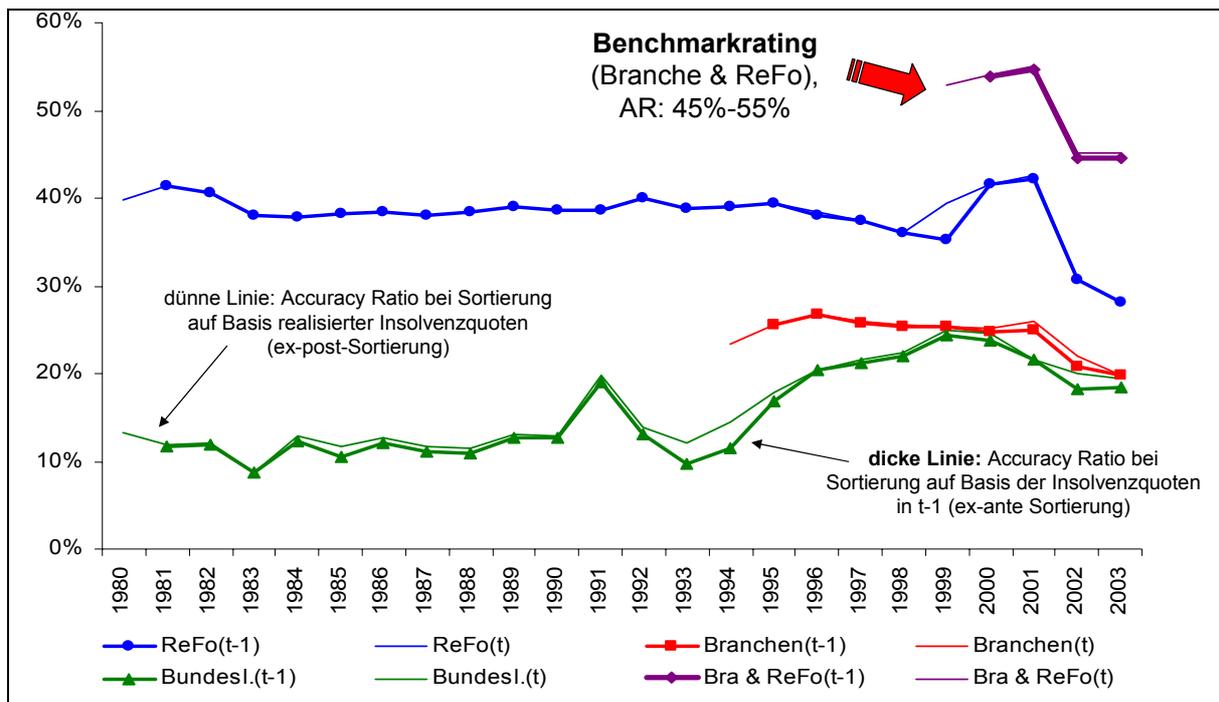


Abbildung 18: Zeitreihen Accuracy Ratio für Prognosegüte Branchen, Rechtsformen und Bundesländer (univariat) und Kombination Branche und Rechtsform bei ex-ante- und ex-post-Sortierung (t-1 vs. t)

In Tabelle E werden die in den einzelnen Jahren bei Kenntnis von Branche *und* Rechtsform eines Unternehmens erzielbaren Accuracy-Ratio-Werte ordinaler Insolvenzprognosen – bei einem Prognosehorizont von einem Jahr - dargestellt. Dargestellt werden auch hier sowohl die „Prognosewerte“, die sich bei Kenntnis der tatsächlichen Reihenfolge der gruppenspezifischen Ausfallquoten in den jeweiligen Jahren („ex-post“) ergeben würden und die Prognosewerte bei Unterstellung der gleichen Reihenfolge wie im Vorjahr („ex-ante“). In Abbildung 18 werden Accuracy-Ratio-Zeitserien für die verschiedenen vorgestellten univariaten Prognosen und das multivariate Branchen-und-Rechtsform-Rating dargestellt und in Abbildung 19 (ex-post) CAP-Kurven²³¹ für das Branchen-und-Rechtsform-Rating.

²³¹ Den CAP-Kurven wurden die „ex-post-Prognosen“ zugrunde gelegt, da nur geringfügige Qualitätsunterschiede (gemessen in Accuracy Ratio) zwischen ex-post- und ex-ante-Prognosen zu verzeichnen sind und für 1999 nur ex-post-Prognosen erstellt werden konnten, da die für die ex-ante-Prognosen benötigten realisierten gruppenspezifischen Ausfallraten für 1998 nicht zur Verfügung standen.

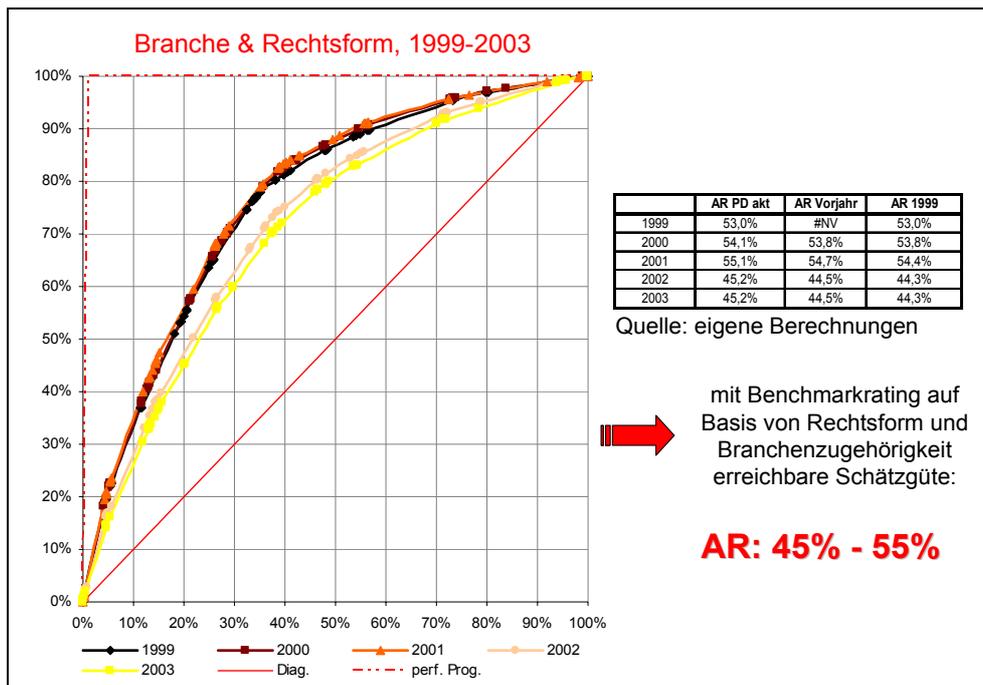


Abbildung 19: Einjahresprognosen-CAP-Kurven (ex-post) 1999-2003 und Accuracy-Ratio-Werte (ex-post und ex-ante) für Benchmarkrating Branche & Rechtsform

Insgesamt zeigt sich, daß sich allein aus der Kenntnis der Branche und Rechtsform eines Unternehmens individuelle Ausfallprognosen mit einer Accuracy Ratio von 45%-55% (Mittelwert 1999-2003: 50%) erstellen lassen. Für diejenigen Ratingverfahren, die auf der Bewertung der gleichen Grundgesamtheit basieren (deutsche Unternehmen sämtlicher Rechtsformen und Branchen) – also insbesondere für Bonitätsurteile von Auskunftsteilen, beispielsweise den CREDITREFORM-Bonitätsindex (siehe Abschnitt 3.5) – stellen sie eine sinnvollere Untergrenze der gerade noch akzeptablen Prognosequalität dar als ein fiktives „naives Zufallsrating“ mit einer erwarteten Accuracy Ratio von 0%.

Anmerkungen: Verbesserungen des Branchen-und-Rechtsform-Ratings können *möglicherweise* durch eine detailliertere Branchenuntergliederung erreicht werden. Die wichtigsten Hebel hierfür dürften in einer stärkeren Untergliederung der größten – und vermutlich hinsichtlich des Insolvenzgeschehens *heterogenen*²³² – Branchengruppen Verarbeitendes Gewerbe (10,6% aller Unternehmen), Baugewerbe (11,4%), Handel (24,3%), Vermietung, Dienstleistungen (27,0%) und sonstige öffentliche und private Dienstleistungen (9,1%) liegen.

Ferner ist zu vermuten, daß der Informationsgehalt des Ratings durch die Einbeziehung der Bundeslandinformationen der Unternehmen verbessert werden kann.²³³

Die Berücksichtigung weiterer, zu Beginn des Abschnitts genannter Variablen, wie Unternehmensgröße (siehe Abbildung 20^{234,235}) und Alter der Unternehmen, ist im Rahmen des hier

²³² Selbst wenn diese Gruppen völlig homogen wären, würde die Qualität der Analyse hierunter nicht leiden. Einzig die Anzahl der Gruppen würden steigen.

²³³ So sind beispielsweise in Sachsen bei gleichen Rechtsformen und Branchenzugehörigkeiten der Unternehmen durchgehend andere (höhere) Insolvenzquoten zu verzeichnen als im gesamten Bundesgebiet (eigene Auswertungen auf Basis von Daten des STATISTISCHEN LANDESAMT DES FREISTAATES SACHSEN (2004)).

²³⁴ Die Darstellungen basieren auf den Daten von BLOCHWITZ, LIEBIG, NYBERG (2000, S. 12 und Anhang 2, S.3). Die Untersuchung basiert auf 140.000 Jahresabschlüssen deutscher Unternehmen, deren Bilanzen im Zeitraum 1994-1999 bei der Deutschen Bundesbank zur Feststellung der „Diskontfähigkeit“ eingereicht wurden.

Obwohl sich die Insolvenzquoten zwischen den einzelnen Unternehmensgrößenklassen bis um den Faktor 10 unterscheiden, ist der Informationsnutzen der Unternehmensgröße relativ gering [auf einer unternehmensgewichteten Basis] (AR = 15,2%) und entspricht ungefähr nur dem Informationsnutzen der Bundes-

vorgestellten Verfahrens *in dieser Form* nicht sinnvoll, da die Gruppengrößen dann zu klein würden.²³⁶ Entweder müssten einzelne Gruppen zusammengefaßt²³⁷ werden, oder die Berücksichtigung zusätzlicher Kriterien müßte mit alternativen statistischen Methoden erfolgen, beispielsweise mittels linearen Regressions- oder Logit-Analysen²³⁸ (mit deutlich weniger Interaktionstermen als sich durch die Kombination der Ausprägungen der verschiedenen Variablen ergäben).

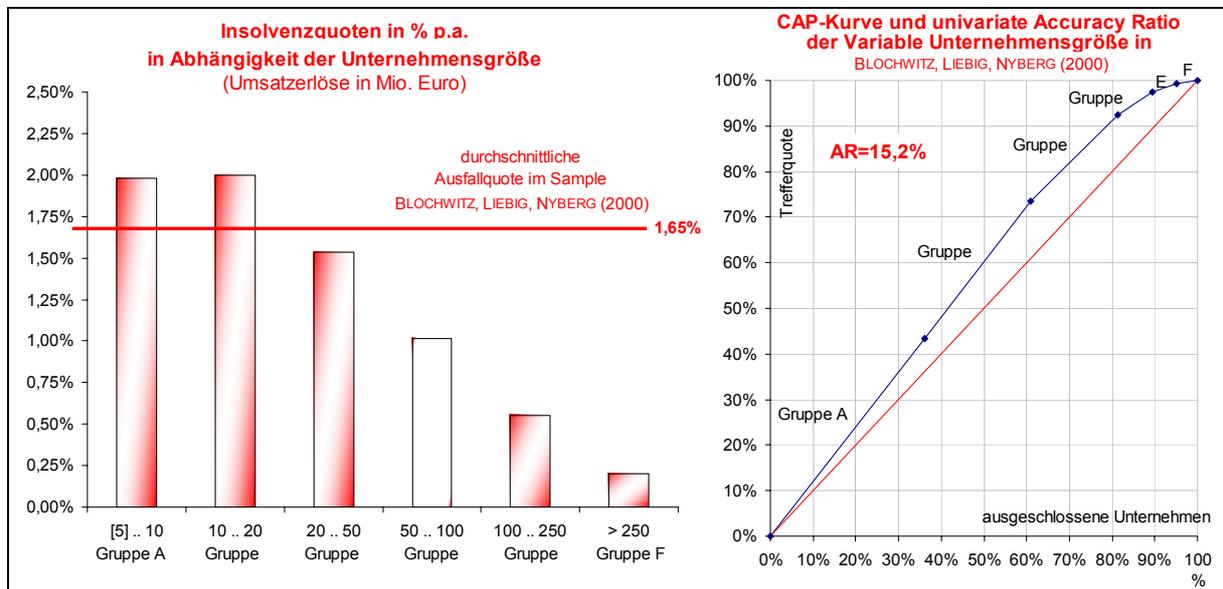


Abbildung 20: Insolvenzquoten in Abhängigkeit der Unternehmensgröße (Umsatzerlöse in Mio. Euro) (linke Seite); CAP-Kurve und univariate Trennfähigkeit des Kriteriums Unternehmensgröße in einer empirischen Studie (rechte Seite), Quelle: eigene Darstellungen

landzugehörigkeit der Unternehmen (vergleiche Tabelle D). Der Grund hierfür liegt darin, daß die meisten Unternehmen des Samples nur leicht überdurchschnittliche (Gruppen A und B) oder leicht unterdurchschnittliche (Gruppe C) Insolvenzquoten aufweisen. Die Gruppen mit extremen Insolvenzquoten sind nur vergleichsweise schwach vertreten. Siehe auch Abbildung 20, rechts: über 80% der Unternehmen entfallen auf die Gruppen A, B und C.

Vermutlich ist der Informationsnutzen der *Unternehmensgröße* auf einer Euro-gewichteten Basis erheblich größer als auch einer unternehmensgewichteten Basis.

²³⁵ Ähnliche Werte für die univariate Trennfähigkeit der Unternehmensgröße findet HAYDEN (2003) in einem Sample von 35.000 österreichischen Unternehmen mit Umsätzen zwischen 0,4 Mio. bis 75 Mio. Euro (siehe ebenda, Variable 61 (Umsatz/CPI), AR = 11% (bei Insolvenzprognose)).

²³⁶ Bei 14 Branchenaggregaten, 6 Rechtsformtypen, 16 Bundesländern, (beispielsweise) 5 Unternehmensgrößenklassen und (beispielsweise) 3 Altersstufen ergeben sich 20.000 Gruppen. Bei 3,3 Mio. Unternehmen und 40.000 Insolvenzen entspräche dies nur durchschnittlich 165 Unternehmen und 2,0 Insolvenzen pro Gruppe und Jahr.

²³⁷ Aufgrund vergleichbarer branchenspezifischer Insolvenzquoten (siehe Abbildung 17) könnten beispielsweise die Rechtsformengruppen GmbH&Co.KG und GmbH zusammengefaßt werden. Ebenso wäre möglicherweise eine Einteilung der Bundesländer in nur 3 statt 16 Gruppen ausreichend (Neue Länder, Süddeutschland, Norddeutschland). Ebenso könnte eine Betrachtung von nur je 2 Unternehmensgrößen- und Altersgruppen ausreichend sein.

²³⁸ Dies wäre vermutlich insbesondere bei den kardinal skalierten Variablen *Unternehmensgröße* und *Alter* sinnvoll, da hier Gruppenbildungen stets mit Informationsverlusten verbunden sind.

3.3 Benchmark II: Univariate Trennfähigkeit von Kennzahlen

Neben unterschiedlich stark beschränkten Einzelfallentscheidungen durch Kreditexperten, formalisierten Auswertungen „weicher Faktoren“ wie *Managementqualität* oder *Brancheinschätzung*²³⁹, basieren die Ratingverfahren von Banken typischerweise zu erheblichen Teilen auf der Auswertung „harter Fakten“, insbesondere auf einer Analyse von Kennzahlen, die aus den Daten der Bilanzen und Gewinn- und Verlustrechnungen der Kreditnehmer berechnet werden.^{240,241}

Auch die meisten in wissenschaftlichen Studien vorgestellten Insolvenzprognosemodelle sowie die Prognosemodelle kommerzieller Anbietern, die insbesondere zur kostengünstigen Bonitätsbeurteilung von mittelständischen Unternehmen entwickelt wurden (siehe auch Abschnitt 3.5) beschränken sich von vornherein ausschließlich auf die Analyse derartiger *Finanzkennzahlen*. In einigen Ratingmodellen werden zusätzlich auch die im vorherigen Kapitel genannten Variablen, wie *Unternehmensgröße*, *Branche*, *Rechtsform* oder aber *regionale Herkunft*, berücksichtigt.²⁴²

Auch wenn die Ratingagenturen offiziell *nicht* einräumen, daß die von ihnen vergebenen Urteile im wesentlichen auf formalen Kennzahlenanalysen beruhen²⁴³, wurde in empirischen Studien zumindest gezeigt, daß sich die Ratingurteile (oder auch *Änderungen* von Ratingurteilen) renommierter Agenturen allein anhand von (wenigen) Kennzahlen mit einer hohen Übereinstimmungsrate *prognostizieren*²⁴⁴ lassen, oder daß sich allein anhand von einfachen Kennzahlenmodellen Insolvenzprognosen mit einer Qualität erstellen lassen, die der Prognosequalität der Ratingurteile renommierter Agenturen entsprechen oder diese sogar übertreffen.^{245,246}

²³⁹ siehe beispielsweise GRUNERT, NORDEN, WEBER (2005)

²⁴⁰ siehe beispielsweise BASLER AUSSCHUB (2000b, S. 17ff.)

²⁴¹ ROMEIKE, WEHRSPORN (2004b, S. 9): „Grundsätzlich gilt: Je größer das Finanzierungsvolumen einer Bank, desto größer wird die Bedeutung qualitativer weicher Faktoren auf des Rating. Aber auch im hohen Millionenbereich behalten weiche Faktoren den Charakter eines Modifikators. Die Basis für das Rating wird nach wie vor durch die (mathematisch-)methodische Analyse der quantitativen Daten, insbesondere der Finanzdaten des Unternehmens, gelegt.“

²⁴² siehe auch ESCOTT, GLORMANN, KOCAGIL (2001b, S. 20)

²⁴³ S&P (2003, S. 53). “The ratio medians are purely statistical, and are not intended as a guide to achieving a given rating level. [...] Ratios are helpful in broadly defining a company’s position relative to rating categories. They are not intended to be hurdles or prerequisites that should be achieved to attain a specific debt rating. Caution should be exercised when using the ratio medians for comparisons with specific company or industry data because of major differences in method of ratio computation, importance of industry or business risk, and impact of mergers and acquisitions.“ usw. oder S&P (2003, S. 17) “There are no formulae for combining scores to arrive at a rating conclusion. Bear in mind that ratings represent an art as much as a science.”

²⁴⁴ Obwohl das von AMATO, FURFINE (2004, S. 2666, Panel A) verwendete Verfahren schlecht kalibriert ist, es werden zu selten extreme Prognosen wie AAA, AA oder B und CCC/C getroffen, gelingt es ihnen auf Basis von nur sieben Kennzahlen die Ratingurteile von STANDARD AND POOR’S auf einer siebenstufigen Ratingskala mit hoher Präzision zu bestimmen (n=10.002): in 53,0% aller Fälle wird das S&P-Rating korrekt bestimmt, in 43,5% aller Fälle tritt eine Abweichung um genau eine Stufe, in 3,3% um genau zwei Stufen und nur in 0,1% aller Fälle eine Abweichung um mehr als 2 Stufen auf. Zum Vergleich: Auf einer 7-stufigen Skala stimmen die Bewertungen von S&P und MOODY’S in 71%, aller Fälle exakt überein, in 28% aller Fälle unterscheiden sie sich um eine Note, in 1,1% um zwei Noten und (nur) in 0,1% aller Fälle um drei oder mehr Noten (eigene Auswertungen basierend auf GÜTTLER (2004, Anhang B)).

Auf einer vierstufigen Skala, es werden nur die Investment-Grade-Klassen AAA, AA, A und BBB betrachtet, können BLUME, LIM, MACKINLAY (1998, S. 1404, n = 7.324) auf Basis von sieben Kennzahlen (zuzüglich Jahresdummies) in 57,2% - 40,2% - 2,3% - 0,3% das Ratingurteil von S&P mit 0 – 1 – 2 – 3 Stufen Abweichung bestimmen. Zum Vergleich: Auf der gleichen Skala stimmen die Ratingurteile von S&P und MOODYS’S in 78,9% - 20,5% - 0,5% - 0,0% überein (eigene Auswertungen basierend auf GÜTTLER (2004, Anhang B)).

²⁴⁵ siehe CAREY, HRYCAY (2001), ALTMAN, RIJEN (2004), FONS, VISWANATHAN (2004)

Als zweiter Benchmark – neben dem im vorherigen Abschnitt vorgestellten Branche-und-Rechtsform-Rating, sollen in diesem Abschnitt deshalb Kennzahlen vorgestellt werden, die sich in verschiedenen Studien als besonders trennfähig erwiesen haben. Ebenso soll deren Trenngüte *quantifiziert* werden. Mit Ausnahme der Studie von BEAVER (1966), der als Begründer der (univariaten) kennzahlenbasierten Insolvenzprognose gilt²⁴⁷, werden in der folgenden Übersicht nur Studien berücksichtigt, die auf der Auswertung der Daten mehrerer tausend Unternehmen beruhen, so daß die statistische Qualität der Ergebnisse dieser Studien als vergleichsweise zuverlässig angenommen werden kann.

Folgende Angaben können den einzelnen Spalten von Tabelle F entnommen werden:

- *Studie*: Autoren und Veröffentlichungsjahr der Studie,
- *Anzahl Unternehmen*: Die Anzahl der in die Auswertungen einbezogenen Unternehmen wird separat für ausgefallene und nicht-ausgefallene Unternehmen ausgewiesen. Ferner werden die Stichprobengrößen der Validierungsstichprobe²⁴⁸, falls vorhanden, ausgewiesen, ebenfalls getrennt nach Ausfall- und Nicht-Ausfallbeobachtungen, da die Zuverlässigkeit (Breite der Konfidenzintervalle) des ermittelten Schätzgütemaßes, Accuracy Ratio, hiervon abhängt (und nicht nur von der Gesamtzahl der untersuchten Unternehmen).
- *Datenbasis*: Es werden Angaben zu den verwendeten Datenquellen (Bankportfolien, externe Datenbanken, Veröffentlichungen von Ratingagenturen, usw.) gemacht und die geographischen Herkunft, Branchen- und Rechtsformzugehörigkeit der untersuchten Unternehmen angegeben.
- *sonstiges*: In dieser Spalte werden unter anderem Angaben zur verwendeten *Ausfalldefinition* gemacht.
- *Prognosegüte*: In dieser Spalte werden die Schätzgüten der in den Studien untersuchten Kennzahlen bzw. Ratingmodelle, gemessen in Accuracy-Ratio, angegeben. Lagen hierzu keine Angaben vor, so wurden sie, i.d.R. mittels des $AR_{\alpha\&\beta}$ -Verfahrens (siehe

²⁴⁶ Fraglich ist, ob Ratingagenturen wirklich, entsprechend einer häufig geäußerten Ansicht, bei der Erstellung ihrer Ratings auch nicht-öffentlich zugängliche Informationen verarbeiten (siehe beispielsweise WHITE (2001, S. 13ff.) oder FONS (2002, S. 5)), und damit prinzipiell in die Lage versetzt werden, präzisere Insolvenzprognosen zu erstellen als externe Analysten. Siehe hierzu MOODY'S (2000, S. 9): "A common misconception is that rating agencies, through privileged access to senior management of an issuer, always gain insights that are not available through any other means. While Moody's meets with management of most issuers of securities it rates, it does not view this practice as indispensable to the formation of an accurate rating opinion. Moody's finds that meetings with senior management tend to accelerate the analytic process, because much of the information needed can be gathered quickly from one source and the answers to certain questions easily obtained, but not to change the substance of such process. If such a meeting is not available, the analyst usually has access to public filings, industry publications, and information from the issuer's competitors, suppliers, and customers through its normal course of business."

Siehe hierzu auch MAHONEY (2002b, S. 4): "Rating agencies routinely request nonpublic data in the course of their surveillance activities. However, unlike accounting firms, rating agencies have no authority to demand such data, and indeed many firms do not provide requested data. [...] They can only work with the information which has been disclosed or which management has elected to provide."

²⁴⁷ RÖSLER (1988, S. 102), FALKENSTEIN, BORAL, CARTY (2003/2001, S. 9), BALCAEN, OOGHE (2004, S. 8)

²⁴⁸ Typischerweise erfolgt die Validierung von Ratingmodellen auf Grundlage von Daten, die nicht bereits für die Parametrisierung der Modelle verwendet wurden. Eine Validierung auf Basis der Lerndaten eines Ratingmodells testet strenggenommen nicht die *Prognosefähigkeit* des Modells, sondern seine Fähigkeit, die Lerndaten zu reproduzieren. Je nach Flexibilität der Modellstruktur (Anzahl der einbezogenen Variablen und Parameter, zusätzliche Restriktionen [beispielsweise hinsichtlich der Vorzeichen, die bestimmte Parameter annehmen/ nicht annehmen dürfen] kann eine gute Reproduktion der Lernstichprobe mit einer wesentlich schlechteren tatsächlichen („out-of-sample“) Prognoseleistung der Modelle einhergehen („over-fitting“), siehe u.a. STEIN (2002, S. 14f.), OENB (2004a, S. 48), DWYER, KOCAGIL, STEIN (2004, S. 26).

Abschnitt 2.3.3 und Anhang I), geschätzt. Geschätzte AR-Werte werden mit dem Symbol (*) gekennzeichnet. Ebenso werden für jede Studie die auf *Näherung 2* (siehe Abschnitt 3.1) basierenden 95%-Konfidenzintervalle der Accuracy-Ratio-Werte angegeben.²⁴⁹

²⁴⁹ Auf *Näherung 2* wird zurückgegriffen, da mit den zur Verfügung stehenden Daten der exakte Wert für die Standardabweichung der Accuracy-Ratio, siehe E-H-T-(2003) in Abschnitt 3.1, *nicht* ermittelt werden kann. Basierend auf den ebenfalls in Abschnitt 3.1 vorgestellten Untersuchungen, werden die per *Näherung 2* ermittelten Werte mit dem Faktor 1,6 dividiert. Für AR in Formel F 42 wird die empirisch ermittelte Accuracy Ratio, maximal jedoch ein Wert von 80%, verwendet.

Tabelle F: Übersicht Studien zur univariaten Trennfähigkeit von Variablen; *Fortsetzung auf den folgenden Seiten*

Studie	Anzahl Unternehmen	Datenbasis	sonstiges	Prognosegüte
BEAVER (1966) ²⁵⁰	79 Ausfaller, 79 Nicht-Ausfaller (gepaarte Stichprobe (Branche, Größe)), 706 Unt.-jahre, keine separate Validierungsstichprobe	Ausfaller: „MOODY’S Industrial Manual“ (überwiegend große Kapitalgesellschaften), Nichtausfaller: aus „12,000 Leading U.S. corporations“ 1954-1964, USA, Verarbeitendes Gewerbe (industrial firms)	Gesamtfehler für 30 Knz., 7 Knz.-grp, (S. 106), FI-FII-Komb. für 7 Knz. (S.90, 109), Progn.: 1..5 Jahre, Def. Ausfall: Insolvenz (59/79), Anleiheausfall (3/79), Kontoüberziehung (1/79), Nichtzahlung Dividenden für Vorzugsaktien (16/79)	cash flow / total debt: FI: 21,5%, FII: 5,1% (1 Jahr), → AR: 73,4%-95,6%, $AR_{\alpha\&\beta}=89\%$, $AR_{\text{tatsächlich}} = 93,1\% (*)$ ²⁵¹ , net income / total assets: FI: 16,5%; FII: 11,4% (1 Jahr) → AR: 72,1%-92,5%, $AR_{\alpha\&\beta}=86\% (*)$ current ratio: FI: 30,4%, FII: 8,9% (1 Jahr) → AR: 60,7%-89,2%, $AR_{\alpha\&\beta}=80\% (*)$ $KI_{95\%,AR} = \pm 7 PP (*)$
KEENAN, SOBEHART (1999),	530 Ausfaller, 9.000 Unt. 54.000 Unt.-jahr; (walk-forward-testing ²⁵²) siehe auch SOBEHART, KEENAN, STEIN (2000)	1989-1999, (große) Aktiengesellschaften („MOODY’S proprietary databases“)	AR-Angaben für eine Kennzahl und drei/ fünf Ratingmodelle, Prognose: 1 Jahr Def. Ausfall: ?	AR distance to default (“Merton model variant”): 67% AR return on assets: 53% $KI_{95\%,AR} = \pm 3 PP (*)$
BLOCHWITZ, LIEBIG, NYBERG (2000)	2.300 Ausfaller, mind. 30.000 Nicht-Ausfaller, 139.685 Unt.-jahre (Lernstichprobe: Daten 1994-1996, keine separate Validierungsstichprobe ?)	Bilanzdatenbank DEUTSCHE BUNDESBANK, “representative for German middle market borrowing” (S.12), Umsatz mind. 5 Mio. Euro, 1993 – 1999; Verarb. Gew.: 44%, Handel: 42,6%, Bau: 5,5%, Landwirt. 0,5%, andere: 7,3%, verschiedene Rechtsformen	6 Kennzahlen, branchenabhängige Untersuchungen (S.28-30), 3 Ratingmodelle, Prognose: 1 Jahr Def. Ausfall: Eröffnung Insolvenzverfahren	AR capital recovery ratio: 53,7% (manufacturing), 44,0% (other enterprises), 47,9% (trade), AR equity ratio: 44,8% (other, trade), AR return on equity: 40,4% (manuf.), 37,5% (trade), 24,9% (other) AR ROCE: 45,2% (manuf.) AR net interest rate: 44,4% (manuf.) $KI_{95\%,AR} = \pm 1,5 PP (*)$

²⁵⁰ Anmerkung: Im Fall univariater Untersuchungen sind bei Abwesenheit einer Validierungsstichprobe keine dramatischen Probleme wegen Überanpassungen an die Lernstichprobe („overfitting“) – wie im Fall multivariater Analysen - zu erwarten, da keine Modellparameter geschätzt werden müssen. Wenn jedoch sehr viele Variablen untersucht werden, dann ist diejenige Variable, die bei einer gegebenen Stichprobe die beste Prognosegüte erzielt, mit hoher Wahrscheinlichkeit eine Variable, deren realisierte Schätzgüte über der erwarteten Schätzgüte liegt.

²⁵¹ Siehe Abschnitt 2.3.3 und Anhang I für die Verfahren zur Ermittlung der Ober- und Untergrenzen für die Accuracy-Ratio sowie des $AR_{\alpha\&\beta}$ -Wertes. Nur für die Kennzahl Cash Flow to Total debt sind bei BEAVER (1966, S. 92ff.) zusätzlich für die Prognosehorizonte 1..5 Jahre Häufigkeitsprofile für insolvente und nicht-insolvente Unternehmen untergliedert in 22 Gruppen gegeben. Auf Basis dieser Werte ergibt sich sogar eine Accuracy Ratio von 93,1%.

²⁵² „Walk-forward-testing“ bezeichnet ein datensparsames Verfahren, bei dem es keine strikte Trennung zwischen Lern- und Validierungsstichproben gibt. Zunächst wird ein Modell basierend auf historischen Daten bis zu einem Zeitpunkt t_0 parametrisiert und dessen Leistungsfähigkeit anhand der Ausfallinformationen für das Jahr t_1 validiert. Anschließend wird das Modell – unter Einbeziehung der t_1 -Daten - erneut geschätzt und auf t_2 -Daten validiert, usw. usf., siehe SOBEHART, KEENAN, STEIN (2000, S. 2f.) und STEIN (2002, S. 15f.). Getestet wird mit diesem Ansatz letztlich die Prognosefähigkeit einer bestimmten Schätzmethodik (z.B. *lineare Diskriminanz-* oder *Logitanalyse*), nicht jedoch die Leistungsfähigkeit eines konkreten, parametrisierten Ratingmodells.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>sonstiges</i>	<i>Prognosegüte</i>
FALKENSTEIN, BORAL, CARTY (2003/2000)	Public Firms (große AGs): 15.805 Unt., 1.529 Ausfälle, 130.019 Unt.-jahre; Private Firms (mittelständische Unt.): 24.718 Unt. 1.621 Ausfaller 115.351 Unt.-jahre (Lernstichprobe: bis einschließlich 1995, enthält ca. 25% d. Ausfaller u. 50% d. Nicht-Ausfaller (S.23))	public firms: MOODY'S Default Database: US, Kanada, Compustat, 1980-1999; (CRD) private company data: "predominantly private, middle market corporations provided by financial institutions" (S.22), USA, Kanada (?), Branchen: Verarb. Gewerbe 22%, Dienstleistungen 27%, Handel 31%, Bau 6%, andere 14%	AR-Angaben für 2 Knz., PD-Profile für 10 Knz., ca. 10 verschiedene Ratingmodelle/ Modellvariationen Prognose: 1 und 5 Jahre, Def. Ausfall: 90 Tage Zahlungsverzug, Vornahme von Abschreibungen durch Bank, Insolvenz	AR liabilities / assets: 61,9% (public fi.) AR liabilities / assets: 44,8% (private fi.) AR net income / assets: 61,5% (public) AR net income / assets: 39,9% (private) [alle Angaben bezogen auf 1 Jahr Prognosezeitraum] (S.86) $KI_{95\%,AR} = \pm 2 PP (*)$ [sowohl bei public firms als auch bei private firms, Basis: komplette Stichprobe ?]
SOBEHART, STEIN, MIKI-TYANSKA, LI (2000)	1.406 Ausfaller, 14.447 Unternehmen („walk-forward-testing“)	MOODY'S proprietary default database, MOODY'S proprietary ratings database, COMPUSTAT, IDC, non-financial US firms [untergliedert in 30 Branchenaggregate], (große) Aktiengesellschaften, 1980-1999	quantilsabhängige Ausfallquoten für 9 verschiedene Variablen (S.11), AR-Angaben für eine Kennzahl und 5 Ratingmodelle, Prognose: 1 Jahr; Def. Ausfall: Insolvenz, Kreditrestrukturierung, Nichtzahlung Dividenden, Zahlungsausfall Zinsen/ Tilgung	AR distance to default ("Merton model variant"): 67% AR return on assets: 53% $KI_{95\%,AR} = \pm 2 PP (*)$
ESCOTT, GLORMANN, KOCAGIL (2001a,b)	Entwicklung: 485 Ausfaller, 4.866 Unt. 11.427 Unt.-jahre, (1987-1992) „außer-gewöhnlich gute Erfassungsqualität“ (S.5) Validierung: 1.000 Ausfaller, 20.000 Unt., 100.000 Unt.-jahre (1992-1999)	Datenbasis: Banken [Entwicklung + Validierung] (S. 4), nicht-börsennotierte Unternehmen div. Branchen (keine Finanzdienstleister), Deutschland, Validierungsdaten: „repräsentativ und realitätsnah“ (S.5); konzernunabhängig, nicht staatsabhängig, keine Besitzges. oder Bauträger; Verarb. Gew.: 52%, Handel. 27%, Dienstleistungen: 13%, Bau: 5%, sonst 2% [Validierung];	grafische Darstellungen relative Häufigkeiten Ausfaller/ Nichtausfaller je acht Gruppen; für 9 Kennzahlen (S.11-15) Prognose: 1 Jahr, 5 Jahre Def. Ausfall: Konkurs, Vergleich, Moratorium, Scheck-/Wechselprotest	AR Finanzkraft: 56% (*) , AR Nettoverschuldungsquote: 54% (*) , AR Kapitalbindungsdauer: 53% (*) , AR EBITD-ROI: 48% (*) , AR Fremdkapitalstruktur: 44% (*) AR Umsatzrentabilität: 43% (*) AR Eigenkapitalquote: 39% (*) AR Personalaufwandsquote: 3% (*) AR Umsatzwachstum: 1% (*) ²⁵³ (alle Angaben beziehen sich auf den Entwicklungsdatenbestand), $KI_{95\%,AR} = \pm 3,5 PP (*)$

²⁵³ Da insolvente Unternehmen sowohl bei stark unter- als auch stark überdurchschnittlichen hohen Wachstumsraten relativ häufiger vertreten sind als nicht-ausfallende Unternehmen, ließe sich die (monoton) Trennschärfe der Variable Umsatzwachstum durch eine geeignete Transformation (Quadrierung, absoluter Betrag, ..) deutlich erhöhen.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>sonstiges</i>	<i>Prognosegüte</i>
SHUMWAY (2001)	300 Ausfaller, 3.182 Unternehmen, 39.745 Unt.-jahre (Lernstichprobe: 1962-1983 [1.822 Unt., 118 Ausfaller])	Schnittmenge aus Compustat Industrial File und CRSP Daily Stock Return File for NYSE and AMEX stocks, Ausfaller: Wall Street Journal Index, Capital Changes Reporter, Compustat Research File, Directory of Obsolete Securities, Nexis, 1962 -1992, keine Finanzunternehmen;	Dezilangaben Trefferquoten, 1 Kennzahl, 9 Ratingmodelle (S. 118, 120, 122) Prognose: 1 Jahr Def. Ausfall: Insolvenzantrag	AR net income / total assets: 66,4% (*) KI _{95%,AR} = ± 5,5 PP (*) [Basis: Validierungsstichprobe ?]
KOCAGIL, IMMING, GLORMANN, ESCOTT (2003)	2.156 Ausfaller, 19.524 Unternehmen, 83.613 Beobachtungen, Aufteilung in Lern- und Validierungsstichprobe?	“kombinierte Datenbestände der zwei größten österreichischen Banken”, überwiegend KMU (0,5-1 Mio. EUR: 29%, 1-5 Mio. EUR: 41%, 5-25 Mio. EUR: 20%, ...), gleiche Ausschlüsse wie in ESCOTT, GLORMANN, KOCAGIL (2001a,b), Verarb. Gewerbe: 27%, Handel: 33%, Dienstleistungen: 17%, Bau: 12%, sonstige: 11%;	grafische Darstellungen relative Häufigkeiten Ausfaller/ Nichtausfaller je acht Gruppen; für 8 Kennzahlen (S.13-17) Prognose: 1 Jahr, 5 Jahre Def. Ausfall: unklar (siehe S.6ff): Insolvenz, Wertberichtigung ?	AR Eigenkapitalkennzahl 43% (*) AR Verbindlichkeitsstruktur 25% (*) AR Verkaufserlös 37% (*) AR Schuldendienst 33% (*) AR Cashflow/ Verbindlichkeiten 39% (*) AR Verbindlichkeitskennzahl 37% (*) AR Umsatzwachstum 3% (*) [untransformiert] AR Struktur Umlaufvermögen 22% (*) (bezüglich Entwicklungsdatenbestand), ²⁵⁴ KI _{95%,AR} = ± 3 PP (*)
ENGELMANN, HAYDEN, TASCHE (2003)	3.000 Ausfaller, 300.000 Unt.-jahre (Validierung: 825 Ausfaller, 200.000 Unt.-jahre)	Bilanzdatenbank Deutsche Bundesbank (siehe oben), Deutschland, 1994-1999	AUC-Angaben für 3 Knz. (S. 15, 16) und 3 Ratingmodelle, Konfidenzintervalle AUC Prognose: 1 Jahr Def. Ausfall: Insolvenz	AR ordinary business income / total assets: 57,7% (*) AR change in (netsales / total assets): 22,5% (*) AR current assets/ total assets: 7,1% KI _{95%,AR} = ± 1,5 PP (*) [Basis: vollständige Stichprobe?]

²⁵⁴ Es wurde angenommen, daß - wie bei ESCOTT, GLORMANN, KOCAGIL (2001a,b) – 1/3 der Ausfälle auf den Entwicklungs- und 2/3 der Ausfälle auf den Validierungsdatenbestand entfallen.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>sonstiges</i>	<i>Prognosegüte</i>
HAYDEN (2003)	Ausfaller – Unt.jahre (insg. vs. Validierg. [1998-1999]): Insolvenzsample: 1.024 -124.479 (393 – 27.388) Restrukturierung: 1.459 – 48.115 (528 – 15.683) Zahlungsverzug: 1.604 – 16.797, (433 – 4.649)	3 österr. Banken, ÖNB, KMU Forschung Austria; Ausschluß zahlreicher inkonsistenter oder unvollständiger Datensätze, keine Großunternehmen, keine Unt. mit Umsatz < 0.36 Mio. Euro, 1987-1999, 81% GmbH, 14% „limited partnerships“, 4% single-ownership-companies, 1% general partnerships / 29% Verarb. Gewerbe, 25% Dienstleistungen, 33% Handel, 12% Bau, 1% Landwirtschaft	AR-Angaben für 65 Kennzahlen (!) aus 10 Kennzahlengruppen und 3 verschiedene Ausfalldefinitionen Prognose: 1 Jahr Def. Ausfall: (1) Insolvenz vs. (2) Restrukturierung vs. (3) Zahlungsverzug > 90d	AR equity / assets: 44,1% (*) , AR ordinary business income / assets: 41,1% (*) AR ordinary business income / operating income: 40,3% (*) AR cash flow / (liabilities-advances): 39,4% (*) AR retained earnings/assets: 39,0% (*) [Mittelwert der AR-Werte der drei Prognosen (Insolvenz, Kreditrestrukturierung, Verzug)] KI _{95%,AR} = ± 2,5 PP, ± 2 PP, ± 2 PP (Insolvenz, Restrukturierung, Verzug) (*) [Basis: vollständige Stichprobe?]

(*) Quelle Accuracy Ratio- Angaben: eigene Berechnungen basierend auf in den Studien veröffentlichten Angaben zu Gesamtfehlern, Fehlern I. und II. Art o.ä.

Unter Vernachlässigung der Studie von BEAVER (1966), die auf einem sehr kleinen Datensatz und einer heute eher ungebräuchlichen Ausfalldefinition beruhte, lassen sich folgende Aussagen zur Trennfähigkeit einzelner Kennzahlen treffen:

- Die trennschärfsten Kennzahlen für mittelständische Unternehmen erreichen in den einzelnen Studien Accuracy Ratios von mindestens rund 45%, teilweise bis 55%. Es handelt sich überwiegend um Kennzahlen aus den Bereichen *Kapitalstruktur*, *Rentabilität* und *Finanzkraft* (debt coverage).
- Für große Aktiengesellschaften sind um etwa 10 Prozentpunkte bessere Accuracy-Ratio-Werte zu verzeichnen (rund 55% bis 65%). Die größte Trennschärfe wurde dabei von einer Kennzahl erreicht („distance to default“²⁵⁵), die nicht auf Buchwert- sondern auf Marktwertangaben beruht und die – als einzige Kennzahl – auch *Risiken* (hier: Volatilität des Marktwertes) des Unternehmens mißt.

Vergleich mit den Ergebnissen des Branchen-und-Rechtsform-Ratings:

Die Prognosegüte des Branchen-Rechtsform-Ratings bei Beurteilung der Gesamtheit aller (deutschen) Unternehmen entspricht der in den verschiedenen Studien ermittelten univariaten Prognosequalitäten der trennschärfsten Kennzahlen – zumindest bei mittelständischen Unternehmen (nicht jedoch bei Aktiengesellschaften). Würde das Branchen-Rechtsform-Rating allerdings auf die in den einzelnen Studien verwendeten Stichproben angewendet – und nicht auf die Grundgesamtheit *aller* (deutschen) Unternehmen – wäre seine Prognosequalität geringer als in Abschnitt 3.2 ermittelt, da die den zitierten Studien zugrundeliegenden Stichproben deutlich homogener, im Fall der Aktiengesellschaften-Samples sogar *perfekt* homogen hinsichtlich des Kriteriums *Rechtsform* sind, das die wichtigste erklärende Variable des Branchen-Rechtsform-Ratings darstellt. Unterrepräsentiert sind vor allem Personengesellschaften und Einzelunternehmen.

Auch hinsichtlich der *Branchenzugehörigkeit* wäre eine weitere Verschlechterung der Prognosegüte des Branchen-Rechtsform-Ratings zu erwarten: zwar umfaßt nur eine Studie ausschließlich Unternehmen des Verarbeitenden Gewerbes - während in den anderen Studien Unternehmen fast aller Branchen einbezogen werden; explizit ausgeschlossen werden meist nur Unternehmen der Finanzbranche. Allerdings sind in den vorgestellten Studien Branchen mit „extremen“ Ausfallwahrscheinlichkeiten - sowohl hohen (Baugewerbe) als auch niedrigen (Dienstleistungen) - unterrepräsentiert, während Branchen mit durchschnittlichen Ausfallraten, insbesondere das Verarbeitende Gewerbe, zum Teil deutlich überrepräsentiert sind.²⁵⁶

Insgesamt kann deshalb davon ausgegangen werden, daß auf Basis der trennschärfsten einzelnen Kennzahlen bessere Insolvenzprognosen getroffen werden können als bei ausschließlicher Kenntnis von Rechtsform und Branche eines Unternehmens. Um die entsprechenden Kennzahlen berechnen zu können, müssen jedoch die Jahresabschlüsse des zu beurteilenden Unternehmens vorliegen – was eine viele stärkere Restriktion darstellt, als die Kenntnis von Branche und Rechtsform, auf die häufig bereits aus dem Unternehmensnamen geschlossen werden kann.

²⁵⁵ siehe SOBEHART ET AL (2000, S. 9): distance to default = $(MVA - DP) / Vol_{MV}$, mit MVA .. market valued assets, DP .. default point, Vol_{MV} .. Volatilität des Marktwertes mit $DP = \text{kurzfristige} + 0,5 * \text{langfristige Verbindlichkeiten}$.

²⁵⁶ Gemäß den in Abschnitt 3.2 verwendeten Daten sind jeweils rund 11% aller 3,3 Mio. deutschen Unternehmen in den Branchen Verarbeitendes Gewerbe und Bau tätig, 24% im Handelsgewerbe, 46% in diversen Dienstleistungsbranchen (Branchen H, K-O) und 7% in anderen Branchen.

3.4 Benchmark III: ALTMAN's Z-Score

Nach der Veröffentlichung der in Abschnitt 3.3 vorstellten univariaten Insolvenzprognosestudie BEAVERS (1966) lag eine Weiterentwicklung der kennzahlenbasierten Insolvenzprognoseverfahren hin zu *multivariaten* Modellen nahe²⁵⁷. Bereits kurze Zeit später wurde diese Entwicklung durch das von ALTMAN (1968) entwickelte Z-Score-Modell, das als erstes *multivariates* Insolvenzprognosemodell für Unternehmen gilt,^{258,259} vollzogen. Im Gegensatz zu univariaten Modellen sind multivariate Insolvenzprognosemodelle nicht nur auf die Trennschärfe *einzelner* Kennzahlen beschränkt - sondern können sich auch die gemeinsame Prognoseleistung aller Kennzahlen zunutze machen; auch von Kennzahlen, die in einem univariaten Kontext nur über eine geringe oder gar statistisch insignifikante Prognoseleistung verfügen.²⁶⁰ Und, obwohl sie auf keiner explizit formulierten, konsistenten und vollständigen Theorie basieren, liefern sie zumindest standardisierte Methoden zur Auswahl und Gewichtung der einzubeziehenden Kennzahlen.²⁶¹

Da sich das ALTMANSche Z-Score-Verfahren in der finanzwissenschaftlichen Literatur immer noch großer Beliebtheit erfreut²⁶², selbst heute noch Grundlage einiger kommerzieller Ratingverfahren ist²⁶³ und da zahlreiche empirische Studien existieren, die seine Leistungsfähigkeit in einer Vielzahl von Modellumgebungen getestet haben, soll im folgenden seine Eignung als Benchmark zur Beurteilung der Prognosequalität von Ratingsystemen untersucht werden. Tatsächlich handelt es sich bei dem Z-Score-Verfahren um eine *Familie* von Verfahren, da mittlerweile bereits drei Z-Scoreversionen existieren (siehe unten).

Das ALTMANSche Z-Score-Verfahren wurde mittels der Multivariaten Linearen Diskriminanzanalyse²⁶⁴ parametrisiert und ordnet strenggenommen die zu beurteilenden Unternehmen lediglich in zwei Gruppen ein („voraussichtlich solvent“ vs. „voraussichtlich insolvent“).²⁶⁵ Die

²⁵⁷ Bereits BEAVER (1966) sah diese Möglichkeit, konnte jedoch nach eigenen Angaben keine befriedigenden Lösungen finden: „Suggestions for future research: The analysis conducted here has been a univariate analysis – that is, it has examined the predictive ability of ratios, one at a time. It is possible, that a multiratio analysis, using several different ratios [...] would predict even better than the single ratios. Some preliminary efforts have undertaken to develop multiratio models, but the results have not been very encouraging in the sense, that the best single ratio appears to predict about as well as the multiratio models.“ (ebenda, S. 100)

²⁵⁸ Siehe beispielsweise SOBEHART ET AL (2000, S.6), FALKENSTEIN, BORAL, CARTY (2003/2000, S.9), FRIEDRICH, WAHRENBURG (2003, S.4), BALCAEN, OOGHE (2004, S.11).

²⁵⁹ Multivariate Analyseverfahren wurden jedoch zu diesem Zeitpunkt bereits auf dem Gebiet der *Privatkundenbonitätsanalyse* verwendet, siehe ALTMAN (1968, S. 591).

²⁶⁰ ALTMAN (1968, S. 594): “The variable finally established did not contain the most significant variables, amongst the twenty-two original ones, measured independently.” und ebenda (S. 595f.) bezüglich der Variable X_5 – die in späteren Modellversionen nicht mehr berücksichtigt wurde: „This final ratio is quite important because [...] it is the least significant ratio on a univariate basis. In fact, based on the statistical significance measure, it would not have appeared at all. However, because of its unique relationship to other variables in the model, the Sales/ Total assets ratio ranks second in its contribution to the overall discriminatory ability of the model.”

²⁶¹ vgl. ALTMAN (2002, S.8)

²⁶² FALKENSTEIN, BORAL, CARTY (2003/2000, S. 74): “The most well-known quantitative model for private firms in the U.S. is Altman's Z-score. Virtually every accounting or financial analysis book uses Z-score to demonstrate how financial statement data can be translated into an equation that helps predict default. [...]”

²⁶³ Das Z-Scoreverfahren liegt den Ratings der Agenturen CONFIRM GmbH und IKU zugrunde, die ihre Dienstleistungen für 1.200,- Euro bzw. 312,- Euro pro Rating anbieten, siehe ROMEIKE, WEHRSPHON (2004, S. 18, 27, 29).

²⁶⁴ siehe beispielsweise KÜTING, WEBER (2004/1993, S. 363), DIMITRAS, ZANAKIS, ZOPOUNIDIS (1996), DEUTSCHE BUNDESBANK (1999, S. 58), ÖSTERREICHISCHE NATIONALBANK (2004, S. 41f.).

²⁶⁵ ALTMAN (1968) leitet aus den Ergebnissen sogar eine Dreiteilung der Modellprognosen ab. Neben den Modellaussagen „solvent“ vs. „insolvent“ definiert er ein mittleres Z-Scoreintervall als „Grauzone“ („zone of ignorance“ oder „gray area“, siehe ebenda, S. 606). Im von der DEUTSCHEN BUNDESBANK (1999, S.55) ver-

Trennfähigkeit des Verfahrens wird jedoch anhand einer ordinalen Interpretation des Diskriminanzwertes („Z-Score“) ermittelt (siehe auch Abschnitt 2.3).²⁶⁶

Auf Basis einer sehr kleinen Stichprobe von 33 insolventen und 33 solventen Unternehmen (für weitere Details siehe Tabelle G) gelang es ALTMAN (1968) bei einem Prognosehorizont von einem Jahr 31 der insolventen und 32 der nicht-insolventen Unternehmen richtig zu klassifizieren. Die Zuordnung basierte auf der Diskriminanzfunktion:

$$\text{F 54: } Z\text{-Score} = 1,2X_1 + 1,4X_2 + 3,3X_3 + 0,6X_4 + 1,0X_5 \quad ^{267} \text{ mit}$$

X_1 ... working capital/total assets,
 X_2 ... retained earnings/total assets,
 X_3 ... earnings before interest and taxes/total assets,
 X_4 ... market value equity/book value of total liabilities,
 X_5 ... sales/total assets.²⁶⁸

Unternehmen mit einem Z-Score-Wert von weniger als 1,81 gelten gemäß dieses Verfahrens auf Einjahressicht als hochgradig insolvenzgefährdet, Unternehmen mit einem Z-Scorewert größer als 2,67 bzw. 2,675²⁶⁹ als ungefährdet.

Um das Z-Scoreverfahren auch auf nicht-börsennotierte Unternehmen anwenden zu können, wurde in einer späteren Modellüberarbeitung bei der Kennzahldefinition für X_4 der *Marktwert* des Eigenkapitals durch den *Buchwert* des Eigenkapitals ersetzt und alle Koeffizienten neu geschätzt:

$$\text{F 55: } Z'\text{-Score} = 0,717X_1 + 0,847X_2 + 3,107X_3 + 0,420X'_4 + 0,998X_5 \quad ^{270} \text{ mit}$$

X_1 ... X_3 und X_5 .. siehe oben,
 X'_4 ... *book value equity*/book value of total liabilities

Da die Kennzahl X_5 als zu branchenabhängig angesehen wurde, wurde sie im Rahmen einer dritten Modellversion entfernt, deren Ziel die Sicherstellung der Anwendbarkeit des Modells auf Unternehmen außerhalb des Verarbeitenden Gewerbes war. Die Koeffizienten der übrigen Variablen wurden erneut geschätzt:

$$\text{F 56: } Z''\text{-Score} = 6,56X_1 + 3,26X_2 + 6,72X_3 + 1,05X'_4 \quad ^{271} \text{ mit}$$

X_1 ... X_4 ... siehe Z' -Score.

Gemäß diesem Score gelten Unternehmen mit einem Z'' -Wert kleiner als 1,10 als insolvenzgefährdet.²⁷² Die folgende Tabelle G vermittelt einen Überblick über empirische Studien, in denen die Prognoseleistung des Z-Score-Verfahrens, in einer seiner drei Versionen, gemessen wurde. Wo vorhanden, wurden auch die univariaten Trennfähigkeiten der jeweils besten Kennzahlen und die Prognoseleistungen der jeweils besten Verfahren angegeben.

wendeten Diskriminanzanalyseverfahren wird dieser Bereich als „B-Bereich“ oder „Indifferenter Bereich“ bezeichnet.

²⁶⁶ Siehe beispielsweise auch ALTMAN, SAUNDERS (1998, S.1737) für ein Zuordnungsverfahren von Z-Scores zu S&P-Ratingnoten auf Basis von 750 gerateten US-Unternehmen.

²⁶⁷ In ALTMAN (1968, S. 594) wird folgende Formel angegeben: $Z\text{-Score} = 0,012X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,999X_5$. In dieser Formel müssen jedoch die Variablen X_1 bis X_4 als absolute Prozentwerte (beispielsweise 33 statt 33%) eingegeben werden (siehe hierfür und für Formel F 54 ALTMAN (2000/1968, S. 12f)).

²⁶⁸ Die Variablen X_2 , X_3 , and X_4 stehen stellvertretend für die „historische, gegenwärtige und zukünftige Profitabilität“ der Unternehmen, siehe ALTMAN, RIJKEN (2004, S. 2686, Übersetzung).

²⁶⁹ vgl. Altman (1968, S. 602) vs. Altman (2002, S.18)

²⁷⁰ ALTMAN (2000/1968, S.25)

²⁷¹ ALTMAN, SAUNDERS (1998, S. 1737f.): „The Z'' -score model is a four variable version of the Z-score approach. It was designed to reduce distortions in credit scores for firms from different industries.“

²⁷² siehe ALTMAN (2002, S. 22).

Tabelle G: Übersicht Studien zur Trennfähigkeit der verschiedenen Z-Scoreversionen, *Fortsetzung auf folgender Seite*

Studie	Anzahl Unternehmen	Datenbasis	sonstiges	Z/Z'/Z''	Prognosegüte
ALTMAN (1968), ALTMAN (2000/ 1968)	33 Ausfaller, 33 Nicht-Ausfaller (gepaarte Stichprobe [Branche, Größe]; Lernstichprobe ²⁷³)	Verarbeitendes Gewerbe, USA, 1946-1965, Bilanzsummen: 0,7 Mio. -25 Mio. US\$ (MW 6,4 Mio. US\$), AGs	Def. Ausfall: Insolvenzantrag (1968), Zahlungsausfall „public held debt“ (2000/1968)	Z Z' Z''	FI.: 6,1%, FII. 3,0%, → AR Z-Score: 90,9% -99,3% , $AR_{\alpha\&\beta}=97\%$, AR_{tats.ächlich}=98,7% (*) FI.: 9,1% FII.: 3,0%, → AR Z'-Score: 87,9% - 98,9% , $AR_{\alpha\&\beta}=96\%$ (*), AR Z''-Score: 70% - 80% (?) ²⁷⁴ $KI_{95\%,AR} = \pm 10,5$ PP (*)
KEENAN, SOBEHART (1999), SOBEHART, KEENAN, STEIN (2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	Z' Z''	AR Z'-Score: 43% AR Z''-Score: 53% AR beste einzelne Kennzahl: 53% (return on assets [=NI/assets]) AR bestes Modell: 73% (“Nonlinear Model”, “MOODY’s Model”) $KI_{95\%,AR} = \pm 3$ PP (*)
SOBEHART, STEIN, MIKITY-ANSKA, LI (2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	Z(?) Z''(?) (S. 14)	AR Z-Score (?): 56% AR Z''-Score (?): 53% AR beste einzelne Kennzahl: 53% (return on assets) AR bestes Modell: 73% (MOODY’S Public Firm) $KI_{95\%,AR} = \pm 2$ PP (*)

²⁷³ Die angegebene Klassifizierungsgüte bezieht sich auf die *Lernstichprobe* und ist deshalb, auch wegen des *geringen Umfangs* des Datensatzes – insbesondere angesichts der großen Anzahl an untersuchten Variablen (22) (ibid., S. 594) - positiv verzerrt. In ALTMAN (1968, S. 601f.) wurden zwar diverse Validierungstichproben betrachtet – mit denen allerdings keine simultanen Untersuchungen hinsichtlich der Fehler I. und II. Art durchgeführt wurden. Die Validierungstichprobe zur Untersuchung der Fehler 2. Art ist verzerrt (nicht-insolvente Unternehmen *mit Ertragsproblemen*).

²⁷⁴ In ALTMAN (2002, S. 18) werden drei weitere eigene Studien zitiert bei denen für verschiedene Stichproben insolventer Unternehmen die Fehler I. Art des Z-Score-Verfahrens bei Zugrundelegung von cut-off-Scores von 2,675 und 1,81 angegeben werden (1969-1975: 86 Unternehmen, Fehler I. Art: 18%/25%; 1976-1995: 110 Unternehmen, Fehler I. Art: 15%/22%; 1997-1999: 120 Unternehmen, Fehler I. Art: 6%, 16%). Angaben zur Herkunft der insolventen Unternehmen werden nicht gemacht (Rechtsform, Branche, Größe). Ebenso fehlen Angaben zur Anzahl und Herkunft der untersuchten nicht-insolventen Unternehmen (vermutlich wurden keine nicht-insolventen Unternehmen in die Studien einbezogen). Ferner wird erwähnt: „In 1999, the proportion of U.S. industrial firms that had Z-Scores below 1.81 was over 20%.“ (ebenda, S.18). Unter den Annahmen, daß auch in den anderen Untersuchungszeiträumen 20% der solventen Unternehmen des verarbeitenden Gewerbes in den USA Z-Scores unter 1,81 aufwiesen und die in den Studien zugrundegelegten Ausfallunternehmen zufällig aus den Ausfallunternehmen des Verarbeitenden Gewerbes ausgewählt wurden, ergeben sich Accuracy Ratios für die drei Studien von: 55%-80% ($AR_{\alpha\&\beta}=71\%$), 58%-82% ($AR_{\alpha\&\beta}=74\%$) und 64%-87% ($AR_{\alpha\&\beta}=79,5\%$) [eigene Berechnungen auf Basis der Angaben zu den Fehlern I. und II. Art].

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Sonstiges</i>	<i>Z/Z'/Z''</i>	<i>Prognosegüte</i>
FALKENSTEIN, BORAL, CARTY (2003/2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	Z' Z'' (S.75)	public firms, 1 Jahr (S.86f.): AR Z'-score: 51,3% AR beste einzelne Kennzahl: 61,5% (net income / assets) AR Z''-score: 62,5% AR bestes Modell: 78,5% „percentiles and their squares“ private firms, 1 Jahr (S.86f.): AR Z'-score: 33,1% AR beste einzelne Kennzahl: 44,8% (net income / assets) AR Z''-score: 45,5% AR bestes Modell: 54,1% („RiskCalc“) KI _{95%,AR} = ± 2 PP / ± 2 PP [public firm, private firm] (*)
FALKENSTEIN, BORAL, KOCA-GIL (2000)	Ausfaller/ Unt./ Unt.-jahre: US private: 1.393/ 33.964/ 139.060; Australian private: 1.447 / 27.712 / 79.877; Canada private: 271 / 4.472 / 18.538; (resampling ²⁷⁵)	mittelständische Unternehmen, Branchen: Verarb. Gewerbe: 25%, Handel: 35%, Dienstleistungen: 17%, andere 11%); häufig fehlende Brancheninformationen, 1989-1999	Def. Ausfall: Zahlungsverzug > 90d, Insolvenz, Einordnung als Problemkredit, Abschreibung	Z'' (S.20)	AR Z''-Score 42,0% (US, 1-2 Jahre) AR bestes Modell: 53,7% („RiskCalc US“) (US, 1-2 Jahre) AR Z''-Score: 27,7% (Australien, 1-2 Jahre) AR bestes Modell: 39,7% („RiskCalc Australia“) (Austr., 1-2 J.) AR Z''-Score: 49,7% (Kanada, 1-2 Jahre) AR bestes Modell: 58,3% („RiskCalc Canada“) (Kanada, 1-2 J.) KI _{95%,AR} = ± 2 PP / ± 2 PP / ± 4,5 PP [US, Australia, Can.] (*) ²⁷⁶
KOCAGIL, AKHAVEIN (2001)	1,143 Ausfaller, 41.557 Unt., 170.503 Unt.-jahre, keine separate Validierungsstichpr. ?	Japanische nicht-börsennotierte Unternehmen, Ausschluß von Finanz- und Immobilienunt., staatliche. Unt., 1994-2000	Def. Ausfall: Zahlungsverzug > 90d, Insolvenz, Einordnung als Problemkredit, Abschr.	Z' Z''	AR Z'-Score: 41.8% AR Z''-Score: 30.9% AR bestes Model: 69.4% (RiskCalc Japan) KI _{95%,AR} = ± 2 PP / ± 2 PP

²⁷⁵ Das Resampling, hier auch als „cross-validation“ bezeichnet, ist ein datensparsames Verfahren, bei dem es keine strikte Trennung zwischen Lern- und Validierungsstichproben gibt. Per Zufallsauswahl wird eine Teilmenge der Daten der gesamten Stichprobe ausgewählt (hier 80%) und zur Schätzung des Modells verwendet (Lernstichprobe). Die Validierung erfolgt anschließend auf Basis der übrigen (hier: 20%) Unternehmen. Durch sukzessives Wiederholen der beschriebenen Prozeduren können Konfidenzintervalle der interessierenden Größe, insbesondere der Accuracy Ratio, ermittelt werden und die Stabilität des Modells (enthaltene Variablen, quantitative Ausprägungen der Koeffizienten) überprüft werden. „It is not a test of [the model] directly, but instead the general methodology, and gives a good indication of its robustness.“, see *ibid*, S. 20.

Zum Thema resampling/ bootstrapping siehe auch STEIN (2002, S. 17ff.)

²⁷⁶ FALKENSTEIN, BORAL, KOCA-GIL (2000, S. 18) geben folgende Werte als „Standardfehler“ der Accuracy Ratio an (meinen aber wahrscheinlich die *Standardabweichung* der AUC, siehe Formel bei *ibid*, S. 26): 1,8 PP, 1,5 PP und 3,6 PP [US, Australia, Canada].

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Sonstiges</i>	<i>Z/Z'/Z''</i>	<i>Prognosegüte</i>
SHUMWAY (2001)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	Z (S. 27)	AR Z-Score: 52,4% (*) AR beste einzelne Kennzahl: 66,4% (*) (net income / total assets) AR bestes DA-Modell mit ALTMAN Variablen: 70,4% (*) AR bestes Modell mit ALTMAN –Variablen: 76,6% (*) AR bestes Modell: 83,7% (*) („Accounting and Market“) KI _{95%,AR} = ± 4,5 PP (*)
STEIN, KOCAGIL, BOHN, AK-HAVEIN (2003)	3.123 Ausfaller, 43.950 Unt., 326.216 Unt.-jahre (USA und Kanada) (bis 1999 keine Trennung in Lern- und Validierungsstichprobe)	MOODY'S KMV's CRD (Quelle: 20 Banken aus USA (76%) und Kanada (24%)), mittelständische Unternehmen: Umsätze 0,5 Mio. – 8,8 Mio. - 287,5 Mio. US\$ (5%-50%-95% Quantil), 1986-2001, Branchen: VG: 26%, Handel: 34%, DL: 15%, Bau: 9%, Landwirtschaft 6%	Def. Ausfall: Zahlungsverzug >90d, Insolvenz, Einordnung als Problemkredit, Abschreibung	(Z) Z'' (S.8)	AR Z'-Score: 39% (*) (aus graf. Darstellung, S.12) AR bestes Modell: 52,5% (*) (RiskCalc US/Kanada) KI _{95%,AR} = ± 1,5 PP (*) ebenda, S. 8: "Note that there are actually two Z-score models, a four variable model and a five variable model. In general, we have found the four variable model to outperform the five variable model and thus we present only the four variable model's results."
KOCAGIL, IMMING, GLORMANN, ESCOTT (2003)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Z'' (?)	AR Z'-Score (?): 34,5% AR beste einzelne Kennzahl: 43% (*) (Eigenkapitalkennzahl) AR bestes Modell: 54,7% (RiskCalc Österreich)
DWYER, KOCAGIL, STEIN (2004)	3.764 Ausfaller, 51.000 Unt., 225.000 Unt.-jahre (USA u. Kanada) (walk-forward-test)	MOODY'S KMV'S CRD (siehe oben), 1989-2002, ohne Finanzunternehmen, Versicherungen, Immobilienunt., ohne nichtgewinnorientierte und staatliche Unternehmen	Def. Ausfall: wahrscheinlich identisch mit obiger Definition	Z''(?)	AR Z'-Score (?): 42,3% , AR bestes Modell: 57,0% (EDF RiskCalc V3.1) KI _{95%,AR} = ± 1,25 PP (*)

(*) Quelle Accuracy Ratio- Angaben: eigene Berechnungen basierend auf in den Studien veröffentlichten Angaben zu Gesamtfehlern, Fehlern I. und II. Art o.ä.

Mit Ausnahme der Originalstudie von ALTMAN (1968) und den zitierten späteren Untersuchungen ALTMANS (2002) zeigen die empirischen Ergebnisse durchgehend sehr schlechte Prognoseleistungen der diversen Z-Scoreverfahren – nicht nur im Vergleich zu anderen Ratingverfahren, sondern selbst im Vergleich zu den Prognoseleistungen einfacher Kennzahlen. Auch wenn das Z-Score-Modell ursprünglich auf US-amerikanische Aktiengesellschaften mit Daten von (aus heutiger Sicht) 40 bis 60 Jahre alten Jahresabschlüssen angepaßt wurde, sind die univariaten Prognoseleistungen der verwendeten Kennzahlen, mit Ausnahme der Kennzahl X_5 , die in der späteren Z“-Score Modellversion nicht mehr verwendet wurde, selbst bei aktuellen Jahresabschlüssen internationaler²⁷⁷ und deutscher/österreichischer mittelständischen Unternehmen keineswegs schlecht.^{278,279} Schlecht ist aber die Kalibrierung der Koeffizienten.²⁸⁰ Sie ist offensichtlich so schlecht, daß mit diesem *multivariaten* Modell keine besseren Prognosen getroffen werden als mit der *univariaten* Analyse einfacher Kennzahlen, wie beispielsweise net income / total assets (Jahresüberschuß/ Bilanzsumme).

Insgesamt ist das Z-Score-Verfahren damit *nicht* als Benchmarkverfahren zur Beurteilung der Leistungsfähigkeit von Insolvenzprognoseverfahren geeignet und wird deshalb in den Übersichten in Abschnitt 3.5 nicht mehr als Referenzverfahren herangezogen. Sinnvollere Vergleichsmaßstäbe sind die univariaten Prognoseleistungen der jeweils trennschärfsten samplespezifischen Kennzahlen oder aber die univariaten Prognoseleistungen von Kennzahlen, die sich in empirischen Studien häufig als trennscharf herausgestellt haben (siehe Abschnitt 3.3). Wünschenswert wäre ferner, würden die Autoren der verschiedenen Insolvenzprognosestudien die Prognosegüte ihrer eigenen Verfahren der Prognosegüte einfacher unkalibrierter, multivariater Modelle gegenüberstellen.²⁸¹ Als weiterer sinnvoller Gütemaßstab bietet sich ein Vergleich mit den im folgenden Kapitel aufgeführten Ratingmodellen an. Von den vorgestellten 24 Studien enthalten immerhin acht Studien vollständige Dokumentation der getesteten Ratingmodelle.²⁸² Sechs Modelle davon basieren ausschließlich auf Variablen, die *keine* Insiderinformationen erfordern oder zu regionspezifisch sind. Diese sechs Modelle können damit von Modelltestern nachgebaut und auf beliebige Datensätze angewendet werden.

Es scheint naheliegend, daß zumindest ein Teil der Popularität des ALTMANSchen Modells unter Entwicklern von Ratingmodellen, die das Altman-Modell immer noch als Vergleichsratingmodell verwenden, darin begründet liegt, daß das ALTMANSche Modell so einfach zu schlagen ist.

²⁷⁷ Siehe auch DIMITRAS, ZANAKIS, ZOPOUNIDIS (1996, S. 496ff.) für eine Übersicht der 35 am häufigsten verwendeten Kennzahlen, für die 59 Ratingmodelle aus 47 Studien ausgewertet wurden. Vier der fünf ALTMAN-Kennzahlen belegen Ränge unter den Top 10 (!) mit: X_1 : #1, X_2 : #9, X_3 : #4, X_4 : #24, X_5 : #10.

²⁷⁸ Basierend auf Daten von HAYDEN (2003, S. 14, 18) ergeben sich folgende Plazierungen der ALTMAN-Variablen hinsichtlich der univariaten Trenngüte für einen Datensatz österreichischer KMU (Mittelwert der Accuracy-Ratio-Werte für drei unterschiedliche Ausfalldefinitionen): X_1 (HAYDEN Nr. #15): Rang 22/65, X_2 (#58) Rang 11/65, X_3 (#45) Rang 18/65, X_4 (#2) Rang 1/65 (!) (Da die Kennzahl „equity/ liabilities“ in der HAYDEN-Liste nicht enthalten ist, wurde die Kennzahl „equities/assets“ (HAYDEN Nr. 2) verwendet. Diese Kennzahl ist ordinal äquivalent zu „equity/ liabilities“ und sollte deshalb die gleiche Trenngüte aufweisen), X_5 (#42) Rang 38/65. Schlechter sind die Plazierungen der ALTMAN-Variablen $X_{1..5}$ mit den Rängen 31, 63 (vermutlich Datenfehler), 11, 16 und 25 (von 65) auf Basis eines Datensatzes mit deutschen KMU, Quelle: Auswertungen basierend auf HAYDEN (2002, S. 73f.)

²⁷⁹ Zum gegenteiligen Ergebnis kommen FRERICHs, WAHRENBURG (2003, S. 3, 10): “none of the financial ratios used for the Altman Z”-score is chosen in a stepwise selection procedure based on the Deutsche Bundesbank database”, ohne die univariate Trenngüte der einzelnen untersuchten 49 Kennzahlen anzugeben.

²⁸⁰ siehe insbesondere die Ergebnisse der SHUMWAY (2001)-Studie.

²⁸¹ Vorbildlich ist in dieser Hinsicht die Studie von FALKENSTEIN, BORAL, CARTY (2003/2000).

²⁸² Einige dieser Studien enthalten sogar mehrere vollständig dokumentierte Modelle. In den folgenden Aufzählungen wurde jedoch nur ein Modell pro Studie gewertet.

3.5 Schätzgütwerte realer Insolvenzprognoseverfahren

In den folgenden drei Übersichten werden verschiedene Insolvenzprognoseverfahren und die mit ihnen erzielten Prognosegüten vorgestellt. Um eine gute Vergleichbarkeit der Verfahren zu ermöglichen, erfolgt eine getrennte Ausweisung für große Aktiengesellschaften in Tabelle H, internationale mittelständische Unternehmen in Tabelle I und deutsche/ österreichische mittelständische Unternehmen in Tabelle J. Neben den Stichprobengrößen, den Grundgesamtheiten und erreichten Prognosegüten der Verfahren, wird auch angegeben, ob die Verfahren in den entsprechenden Veröffentlichungen so detailliert dokumentiert werden, daß sie komplett nachgebaut werden können. Zusätzlich wird angegeben, ob von den zu beurteilenden Unternehmen Informationen benötigt werden, die sich nicht aus den Jahresabschlußdaten ergeben.

Mit Ausnahme der aus historischen Gründen relevanten Studien von BEAVER (1967) und ALTMAN (1968) werden nur Studien bzw. Verfahren vorgestellt, die auf Stichproben von mindestens 1.000 Unternehmen beruhen und deren Ergebnisse in einer Form veröffentlicht wurden, die eine Ermittlung von Accuracy-Ratio-Werten erlaubten.

Tabelle H: Übersicht Insolvenzprognoseverfahren, große Aktiengesellschaften; *Fortsetzung auf folgenden Seiten*

Studie	Anzahl Unternehmen	Datenbasis	Anmerkungen	Prognosegüte
BEAVER (1966)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR cash flow / total debt: 93,1% (*) $KI_{95\%,AR} = \pm 7 \text{ PP (*)}$ → es wird nur die univariate Prognosefähigkeit von Kennzahlen untersucht
ALTMAN (1968, 2000/1968)	siehe Abschnitt 3.4	siehe Abschnitt 3.4	Def. Ausfall: siehe Abschnitt 3.4	zu AR Z-, Z'- und Z''-Score und $KI_{95\%,AR}$ siehe Abschnitt 3.4; 3 Diskriminanzanalysemodelle mit 4/4/5 Variablen → alle drei Verfahren sind vollständig dokumentiert, Modell 1 benötigt Marktwert des Eigenkapitals
OHLSON (1980)	105 Ausfaller, 2.058 Nichtausfaller (keine separate Validierungsstichprobe)	1970-1976, kapitalmarktnotierte Aktiengesellschaften, USA, Verarbeitendes Gewerbe, Nichtausfaller: COMPUSTAT; Ausfaller: Wall Street Journal Index	17 Fehler-I-II-Kombinationen pro Modell (S. 130), Def. Ausfall: Insolvenz	2 Logit-Modelle, 9 Variablen AR Modell 1: 81,6% (*) AR Modell 2: 86,7% (*) $KI_{95\%,AR} = \pm 5,5 \text{ PP (*)}$ → beide Verfahren sind vollständig dokumentiert
ZMIJEWSKI (1984)	1.600 Nichtausfaller, 81 Ausfaller (davon Validierungsstichprobe: 800 Nichtausfaller, 41 Ausfaller)	1972-1978, alle American oder NY Stock Exchange gelisteten Unternehmen mit Branchen-SIC-Code < 6000 [d.h. ohne Verkehr, Finanzen, Immobilien, Dienstleistungen]; Ausfallinformationen: Capital Changes Reporter, Wall Street Journal Index, COMPUSTAT Research File	Def. Ausfall: Insolvenzantrag	Probit-Modell, 3 Variablen Panel A: MW F I: 17,1%, MW F II: 2,2% (S.71) AR Unweighted Probit Model: 81%-98,5%, $AR_{\alpha\&\beta} = 94\% (*)$ $KI_{95\%,AR} = \pm 9 \text{ PP (*)}$ → die vorgestellten Verfahren („Modellfamilien“) sind vollständig dokumentiert

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
KEENAN, SOBEHART (1999)	siehe Abschnitt 3.3, analog: SOBEHART, KEENAN, STEIN (2000), SOBEHART, STEIN, MIKITYANSKA, LI (2000)	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR return on assets: 53% AR distance to default (“Merton model variant”): 67% AR MOODY’S Public Firm (“nonlinear model”, “MOODY’S model”): 73% , $KI_{95\%,AR} = \pm 3$ PP (*) → die 8+1 verwendeten Variablen werden in SOBEHART ET AL (2000, S. 10) genannt, es werden Marktdaten benötigt, Angaben zur Aggregation fehlen
FALKENSTEIN, BORAL, CARTY (2003/2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR liabilities / assets : 61,9% (public fi.) AR RiskCalc [Can/US]: 76,5% (public fi.) (Knz. S.42) $KI_{95\%,AR} = \pm 2$ PP [public firm] (*) → die zehn verwendeten Variablen werden genannt, Angaben zur Aggregation fehlen [nicht-parametrisches Transformationsverfahren + Probit]
SHUMWAY (2001)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Diskriminanzanalyse- und Logitverfahren, 3-6 Variablen AR net income / total assets: 66,4% (*) AR Modell mit ALTMAN-Variablen: 76,6% (*) AR ZMIJEWSKI: 68,9% (*) AR „Accounting and Market“: 83,7% (*) $KI_{95\%,AR} = \pm 4,5$ PP (*) → alle zwölf Verfahren sind vollständig dokumentiert, einige Verfahren benötigen Marktdaten
CAREY, HRYCAY (2001)	Ausfaller – Unt.-jahre: 1970-1987: 69 - 7.641 1988-1993: 82 - 3.348 1994-1998: 63 - 3.253 (Lernstichprobe: 1970-1987)	Ausfaller: 01/1999 release of MOODY’S Corporate Bond Default Database; Nichtausfaller: von MOODY’S geratete Unternehmen, 06/1999 release of Compustat (Daten von 1970-1998), US-Unternehmen, Nicht-Finanzunt.	Angabe ratingklassenspezifischer Ausfälle (absolut) und Ausfallraten, Def. Ausfall: MOODY’S Ausfalldefinition	Logitmodell, 4 Variablen AR (*) ²⁸³ 1970-87 1988-93 1994-98 1970-98 MOODY’S Ratings: 80,0% 67,4% 70,7% 75,4% Logit-Modell: 86,4% 78,4% 75,0% 82,8% $KI_{95\%,AR}$ (*) ± 7 PP $\pm 6,5$ PP ± 8 PP ± 4 PP → das Verfahren (Logit-Modell) ist vollständig dokumentiert

²⁸³ Die Accuracy Ratios für die MOODY’S-Ratings wurden auf Basis einer 7stufigen Skala ermittelt (siehe ebenda, S. 244, 267), die Accuracy Ratio Werte für das Logit-Modell von CAREY, HRYCAY (2001) auf Basis einer 10stufigen Skala (siehe ebenda, S. 220, 266).

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
KEALHOFER (2003)	(I): „all identified defaults of non-financial companies with public debt ratings from 1979 to 1990“ (S.34) ²⁸⁴ ; 657 Unt. p.a., 1.066 Unt. insgesamt ²⁸⁴ (II): 121 Ausfälle; durchschnittlich 1.347 verschiedene Unt. p.a., 1.579 verschiedene Unt. insgesamt (out-of-sample-Test)	(I): alle Nichtfinanz-Unternehmen mit veröffentlichtem S&P-Rating, 1979-1990 (II): alle Nichtfinanz-Unternehmen mit veröffentlichten Ratings von S&P und/oder (?) MOODY'S, 1990-1999	(I): 10 Kombinationen aus Trefferquoten und ausgeschlossenen Unternehmen pro Verfahren (graf.) (II): (kont.) CAP-Kurve für alle Verfahren (graf. Darstellung), Def. Ausfall: Zahlungsausfall, Restrukturierung (weitgehend identisch mit S&P Definition)	(I) AR S&P-Ratings: 60% (*) AR KMV EDF: 68% (*) KI _{95%,AR} = ± 6,5 PP (*) (II) AR S&P-Ratings (implied): 68% (*) AR MOODY'S Ratings (implied): 68% (*) , AR KMV EDF: 85% (*) KI _{95%,AR} = ± 5,5 PP (*) → unvollständige Nennung der einbezogenen Variablen, Angaben zur Aggregation fehlen
FONS, VISWANATHAN, (2004)	Lernstichprobe: 14.176 Beobachtungen, 371 Ausfälle, (1989-2000): Validierung: 1,378 Unternehmen, 60 Ausfälle (2001) 1.167 Unt. p.a. (Median)	US Nicht-Finanzunternehmen mit MOODY'S-Rating, 1989-2002, Finanzdaten: COMPUSTAT	PD-Profile für alle sechs verwendeten Kennzahlen Def. Ausfall: Zahlungsverzug/ Ausfall Zinsen, Insolvenzantrag, Restrukturierung.	AR Zinsdeckungsquote 68% ²⁸⁵ AR 1999 2000 2001 2002 2003 insg. Bond Implied Ratings 80% 74% 86% 89% 91% 83,3% MDP Implied Ratings 74% 65% 80% 77% 83% 76,0% MOODY'S Ratings 72% 59% 76% 81% 88% 74,3% KI _{95%,AR} = ± 7.5 PP (*) [nur 2001er Daten], KI _{95%,AR} = ± 3.5 PP (*) [Daten von 1989-2001], alle sechs Variablen werden genannt MDP: Aggregation analog FALKENSTEIN, BORAL, CARTY (2003/2000) [nichtlineare Transformation + Probit]

²⁸⁴ Auf Basis von MOODY'S (2005, S. 12f.)-Daten ergibt sich für den Zeitraum von 1979-1990 eine durchschnittliche Ausfallrate von 1,48% p.a. Für S&P werden keine Werte vor 1981 veröffentlicht. Für den Zeitraum 1981-1990 ergibt sich eine durchschnittliche Ausfallrate von 1,31%, siehe S&P (2004, S.7). Legt man eine durchschnittliche jährliche Ausfallrate von 1,4% bei 657 Unternehmen pro Jahr zugrunde, ergeben sich für den Zeitraum 1979-1990 $12 \cdot 1,4\% \cdot 657 = 110$ Ausfälle bei $12 \cdot 657 = 7.884$ Unternehmensjahren.

²⁸⁵ berechnet gemäß ebenda (S. 11): „Indeed, our testing found that a model containing only interest coverage is nearly 90% as accurate as the complete model. A model containing interest coverage and leverage together displays more than 95% of the accuracy of the complete model.“.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
STANDARD & POOR'S (2004)	10.438 Unternehmen, 66.500 Unt.-jahre, davon 1.170 Ausfälle (out-of-time-Test ²⁸⁶)	1981-2003 geographische Herkunft (2000): ²⁸⁷ USA 74%, Europa, Mittlerer Osten, Afrika 19%, Asien 3%, Lateinamerika 4% Aufteilung nach Branchen (2000): Banken 12%, Versicherungen 42% [sic!, siehe obige Fußnote], Industrieunternehmen 46%	Def. Ausfall (S&P (2004, S.7f.)): Zahlungsausfall Finanzverbindlichkeiten, Restrukturierung,	AR S&P Ratings: 83% ²⁸⁸ KI _{95%,AR} = ± 1,5 PP (*) nach Regionen: AR USA 82%, AR Europäische Union 94% → Angaben zur Aggregation fehlen: ²⁸⁹
MOODY'S (2004)	63.500 Unt.-jahre, davon 1.150 Ausfälle (out-of-time-Test)	1983-2003; geographische Herkunft (2000): ²⁹⁰ USA 67%, Europa, Mittlerer Osten, Afrika 19%, Asien 10%, Lateinam.5%, Aufteilung nach Branchen (2000): Banken 29%, Versicherungen 11%, Industrieunternehmen 60% Unternehmensgröße (<i>Median</i> (!) Umsatz) >> 1 Mrld. US\$ ²⁹¹	Def. Ausfall (MOODY'S (2004, S.3)): Zahlungsausfall Finanzverbindlichkeiten, Insolvenz, Restrukturierung	AR MOODY'S Ratings: 85,0% (*) KI _{95%,AR} = ± 1,5 PP (*) nach Regionen: AR Nordamerika: 81,0%, AR Europa: 95,3%, AR Asien/ Pazifik: 84,7% nach Größe (assets): AR „große“ Unternehmen: 84,4%, AR „kleine“ Unternehmen: 74,0% nach Branchen: AR Finanzunternehmen: 92,3%, AR Nicht-Finanzunternehmen: 80,5% zur Information: AR Staaten: 87,0% ²⁹² → Angaben zur Aggregation fehlen

²⁸⁶ Zum Begriff des Out-of-Time siehe SOBEHART, KEENAN, STEIN (2000, S. 1ff.). Getestet wurde die historische Prognosefähigkeit der S&P-Ratings. Zu jedem Zeitpunkt, zu dem Prognosen abgegeben wurden, waren S&P (bestenfalls) die bis zu diesem Zeitpunkt eingetretenen Ausfälle bekannt, so daß auch nur diese in einer Lernstichprobe verwendet worden sein konnten. Falls S&P seine Ratingmethodik regelmäßig überarbeitet, entspricht dieses Vorgehen dem der „walk-forward-Methode“ (siehe oben).

²⁸⁷ Die Angaben basieren auf BASLER AUSSCHUB (2000c, S.33f). Hinsichtlich der Aufteilung der Emittenten nach Branchen (1980-2001) finden sich abweichende Angaben bei S&P (2002, S.14): Finanzinstitute 19,5%, Konsumgüter/Dienstleistungen 13,5%, Luft- und Raumfahrt, Automobilbau, Kapitalgüter, Metall 11%, Versicherungen/Immobilien 10%, Versorger 9%, Freizeit/ Medien 8%, Gesundheit/ Chemie 6%, Energie/ Rohstoffe 5,5%, Telekommunikation 4,5%, Verkehr 4,5%, Technologie 4%, Forst/ Bau 4%;

²⁸⁸ alle Angaben aus S&P (2004b, S. 11)

²⁸⁹ „There are no formulae for combining scores to arrive at a rating conclusion. Bear in mind that ratings represent an art as much as a science. A rating is, in the end, an opinion.“, S&P (2003b, S. 17)

²⁹⁰ Angaben basieren auf BASLER AUSSCHUB (2000c, S.33f). Hinsichtlich der Aufteilung nach Regionen finden sich geringfügig abweichende Angaben bei MOODY'S (2004c) für 2004: Nordamerika 63%, Europa 24%, Asien/ Pazifik 12%.

²⁹¹ siehe CANTOR (2004, S.9) für eine Untergliederung der Median-Unternehmensgrößen nach Regionen und Ratingklassen.

Tabelle I: Übersicht Insolvenzprognoseverfahren, mittelständische Unternehmen (international), *Fortsetzung auf folgender Seite*

Studie	Anzahl Unternehmen	Datenbasis	Anmerkungen	Prognosegüte
VARETTO (1998)	Lernstichprobe: je 1.920 „sound“ und „unsound“ Unt. Validierungsstichprobe: je 449 „sound“ und „unsound“ Unt.	italienische Unternehmen des verarbeitenden Gewerbes, Quelle: Italienische Zentralbank und 50 Geschäftsbanken, 1982-1995	Def. Ausfall: widersprüchliche Verwendung und unvollständige Erklärungen des Begriffs „unsound“	Fehler I. Art =4,7%, Fehler II. Art =6,5% AR Genetischer Algorithmus: 88,9% - 98,8% AR_{α&β} = 96% (*) KI _{95%,AR} = ± 2,5 PP (*) → keine Angaben zu verwendeten Variablen
FALKENSTEIN, BORAL, KOCAGIL (2000)	siehe Abschnitt 3.4	siehe Abschnitt 3.4	Def. Ausfall: siehe Abschnitt 3.4	AR RiskCalc US: 53,7% (US-Unternehmen) AR RiskCalc US: 57,7% (kanadische Unternehmen) AR RiskCalc US: 36,8% (australische Unternehmen) AR RiskCalc Australia: 39,7% (australische Unt.) AR RiskCalc Canada: 58,3% (kanadische Unt.) KI _{95%,AR} = ± 2 PP / ± 2 PP / ± 4,5 PP [US, Austr., Can.] (*) → die verwendeten 8 Variablen werden genannt, Angaben zur Aggregation fehlen
FALKENSTEIN, BORAL, CARTY (2003/2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR liabilities / assets: 44,8% (private fi.) RiskCalc [Can/US]: 54,1% (private fi.) → siehe oben, 8 Variablen KI _{95%,AR} = ± 2 PP (*)
WESTGAARD, WIJST (2001)	70.000 Unt. (1996er Kohorte), 2.000 Ausfälle (1998!) ²⁹³ , (je 50% für Lern- und Validierungsstichprobe)	Vollerhebung aller norwegischen GmbHs, 1995-1999 (ca. 100.000 Unternehmen pro Jahr), mind. 12.500 Euro Bilanzsumme, diverse Bereinigungen, Dun & Bradstreet register of bankruptcies, Norsk Lysningsblad	Def. Ausfall: Insolvenz	Fehler I. Art =5%, Fehler II. Art < 50% ²⁹⁴ AR Logitmodell: mind. 45,0%- 90,0%, AR_{α&β}>74% (*) KI _{95%,AR} = ± 2 PP (*) → das vorgestellten Verfahren ist vollständig dokumentiert; (10 Variablen) es werden norwegenspezifische Variablen (Regionenzugehörigkeiten) verwendet

²⁹² Die Schätzwertwerte für die Untergliederungen nach Regionen, Größenklassen und Branchen wurden MOODY'S (2004c, S.2) entnommen. Aufgrund des moody's-spezifischen Verfahrens der AR-Ermittlung sind die *tatsächlichen* Accuracy-Ratio-Werte etwas größer (siehe Abschnitt 2.3.2).

²⁹³ Angesichts der zur Verfügung stehenden Daten (siehe oben) ist die Beschränkung der Auswertungen auf nur *eine* Kohorte kaum nachvollziehbar. Die hierfür gegebene Begründung legt eine möglicherweise starke Positivverzerrung der Schätzwerteergebnisse nahe: "In the empirical analysis the 1996 accounting data and the 1998 bankruptcy data are used. Several combinations have been tried, but using the 1996-1998 data set appeared to give the best results.", siehe ebenda, S. 344.

²⁹⁴ Im Original (ebenda, S. 347) waren vereinzelte Quantilsangaben zu den per Logitschätzung ermittelten Ausfallwahrscheinlichkeiten für Ausfaller und Nicht-Ausfaller gegeben. Aus diesen Angaben geht hervor, daß 95% der Ausfaller eine Ausfallwahrscheinlichkeit von > 0,6% haben und 50% der Nichtausfaller eine Ausfallwahrscheinlichkeit von < 0,48%.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
KOCAGIL, AKHAVEIN (2001)	siehe Abschnitt 3.4	siehe Abschnitt 3.4	siehe Abschnitt 3.4	AR RiskCalc Japan: 69,4% $KI_{95\%,AR} = \pm 2 \text{ PP} / \pm 2 \text{ PP}$ → die verwendeten 7 Variablen werden genannt, Angaben zur Aggregation fehlen
STEIN, KO-CAGIL, BOHN, AKHAVEIN (2003)	siehe Abschnitt 3.4	siehe Abschnitt 3.4	Def. Ausfall: siehe Abschnitt 3.4	AR KMV Private Firm Model (PFM): 45% (*) AR RiskCalc US/Canada: 52,5% (*) $KI_{95\%,AR} = \pm 1,5 \text{ PP} (*)$ → keine Angaben zu verwendeten Variablen
DWYER, KO-CAGIL, STEIN (2004)	siehe Abschnitt 3.4	siehe Abschnitt 3.4	Def. Ausfall: siehe Abschnitt 3.4	AR EDF RiskCalc V3.1: 57,0% (Knz. S. 34) AR RiskCalc [V1.0]: 49,5% AR PFM: 46,1% $KI_{95\%,AR} = \pm 1,25 \text{ PP} (*)$ → die verwendeten 9-11 Variablen werden genannt, Angaben zur Aggregation fehlen

Tabelle J: Übersicht Insolvenzprognoseverfahren, Mittelständische Unternehmen (Deutschland, Österreich), *Fortsetzung auf folgenden Seiten*

Studie	Anzahl Unternehmen	Datenbasis	Anmerkungen	Prognosegüte
FRITZ, HOSEMANN (2000)	2.580 „gute Konten“, 1.019 „schlechte Konten“, monatliche Kontoführungsdaten (davon 75% Lernstichprobe (2.699 Unt.)), 25% Validierungstichprobe (900 Unt.))	Kontokorrentkonten von Unternehmenskunden der Deutschen Bank AG (?), Umsätze zwischen 2,5 und 25 Mio. Euro, diverse Datenbereinigungen,	Def. Ausfall: erstmalige Wertberichtigung	AR Lineare Diskriminanzanalyse_2: 62% (*) AR k-nächste Nachbarn: 60% (*) AR Genetischer Algorithmus: 62% (*) AR Neuronales Netz_2: 61,5% (*) AR Entscheidungsbaumverfahren: 54,5% (*) $KI_{95\%,AR} = \pm 4,5 \text{ PP} (*)$ → die verwendeten 11-12 aus 98 untersuchten Variablen werden genannt, die Variablen basieren auf monatlichen Werten zur Kontokorrentkontoführung (Kreditlinie, Überziehungstage, Umschlag, ...), Angaben zur Aggregation fehlen
BLOCHWITZ, LIEBIG, NYBERG (2000)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR capital recovery ratio: 53,7% (manufacturing), 44,0% (other enterprises), 47,9% (trade), AR BUNDESBANK Diskriminanzanalyse: 57,4% (S. 14) AR KMV Private Firm Model: 59,7% AR BUNDESBANK Diskriminanzanalyse + Expertensystem: 68,0% ²⁹⁵ $KI_{95\%,AR} = \pm 1,5 \text{ PP} (*)$ → die verwendeten 4-6 Variablen werden genannt (Diskriminanzanalyse), keine Angabe zu Koeffizienten
ESCOTT, GLORMANN, KOCAGIL (2001b)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR Finanzkraft: 56% (*) [Entwicklungsdatensatz], RiskCalc Germany: 59,7% [Validierungsdatensatz] $KI_{95\%,AR} = \pm 2,5 \text{ PP} (*)$ → die verwendeten 9 Variablen werden genannt, Angaben zur Aggregation fehlen

²⁹⁵ In einer Studie der DEUTSCHEN BUNDESBANK (1999, S. 60f), die wegen fehlender Angaben zur Stichprobengröße nicht in obige Übersicht einbezogen wurde, erzielt das Diskriminanzanalyseverfahren der DEUTSCHEN BUNDESBANK eine deutlich bessere Klassifikationsleistung als bei BLOCHWITZ, LIEBIG, NYBERG (2000): Fehler I. Art: 10,4% + 15,8%/2 = 18,3% (Annahme: je 50% der „nicht-klassifizierten“ bestehen aus solventen bzw. insolventen Unternehmen); Fehler II. Art: 12,8% + 19,2%/2 = 22,4% → AR: 59%-84%, **AR_{a&β} = 75%** Dafür kann das nachgeschaltete Expertensystem in dieser Studie die Schätzgüte des Ratings nur marginal verbessern: Fehler I. Art: 16,5% + 6,7%/2 = 19,9%, Fehler II. Art: 15,0% + 5,0%/2 = 17,5% → AR: 63%- 86%, **AR_{a&β} = 78%**.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
SCHWAIGER (2002)	11.610 Unternehmen, Kohorte 1999 (?) (S.437), davon ca. 200 Ausfaller (1,88%), (out-of-time-Test)	Stichprobe: „repräsentativ für den österreichischen Mittelstand“ (S.434), Umsatz zwischen 1 Mio. und 50 Mio. Euro, „repräsentative“ Branchenbesetzung (S.440)	Ausfallquoten und Verteilung der Unternehmen auf 12 Ratingklassen Def. Ausfall: Creditreformindex >500 ²⁹⁶	AR Bonitätsindex CREDITREFORM Österreich (Teilstichprobe): 67,4% (*) KI _{95%,AR} = ± 5 PP (*) → die verwendeten 15 Variablen werden genannt, einige Variablen können nicht aus Jahresabschlüssen entnommen werden („Krediturteil“, „Zahlungsweise der Kunden“, ...), Angaben zur Aggregation fehlen;
LAWRENZ, SCHWAIGER (2002)	2,5 Mio. Unternehmen, davon ca. 49.000-53.000 Ausfaller (2,0%-2,1%) p.a. (out-of-time-Test)	„quasi die Grundgesamtheit [deutscher Unternehmen]“ (S.6), 1998-2000	Ausfallquoten für Kohorten und Pool 1998-2000, Verteilung auf 12 Ratingklassen zum 1.1.2001, Def. Ausfall siehe oben (Creditreformindex >500);	AR Bonitätsindex CREDITREFORM Deutschland (Grundgesamtheit): 50,1% (S. 27), 51,8% (*) (eigene Berechnung) ²⁹⁷ KI _{95%,AR} = ± 0,35 PP (*) [p.a.] → siehe oben
PLATTNER (2002)	3.162 Beobachtungen (Unt.-jahre) durch „geschichtete Ziehung“ (S.42) aus 30.000 Unternehmen; davon 165 Ausfaller (?) ²⁹⁸ , (Lernstichprobe = Validierungsstichprobe)	Portfolio von 30.000 KfW-Kreditnehmern, verschiedene Branchen und Rechtsformen, 1994-1998	Def. Ausfall: Insolvenz	Logitmodell (27 Variablen): FI: 27,0% FII 5,5% AR: 67,5% -94%, AR_{α&β}=86% KI _{95%,AR} = ± 4,5 PP (?) (*) → das Verfahren ist vollständig dokumentiert, erfordert jedoch die Variable „Liquiditätseinschätzung durch Hausbank“;

²⁹⁶ Ein CREDITREFORM-Bonitätsindex ab 500 Punkten deutet auf das Vorliegen „massiver Zahlungsverzüge“ bis „harter Negativmerkmale“ hin. SCHWAIGER (2002, S. 438). Neben Insolvenzen werden „langsame, schleppende Zahlungsweisen“ oder „erhebliche Zielüberschreitungen bis zu 30 Tagen“ als Ausfälle gezählt.

²⁹⁷ Unterstellt man für die einzelnen Kohorten, für die nur die Ausfallquoten gegeben waren, die gleiche *Verteilung* der Unternehmen über die verschiedenen Ratingklassen wie die vom 1.1.2001, so ergeben sich die folgenden Kohorten-Accuracy-Ratios: AR₁₉₉₈ = 52,9%, AR₁₉₉₉ = 50,3%, AR₂₀₀₀ = 51,9%.

²⁹⁸ Dieser Wert wurde basierend auf einer späteren, unveröffentlichten Studie von PLATTNER geschätzt.

<i>Studie</i>	<i>Anzahl Unternehmen</i>	<i>Datenbasis</i>	<i>Anmerkungen</i>	<i>Prognosegüte</i>
KOCAGIL, IMMING, GLORMANN, ESCOTT (2003)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR Cashflow/ Verbindlichkeiten 39% (*) AR Eigenkapitalkennzahl 43% (*) AR RiskCalc Österreich: 54,7% () $KI_{95\%,AR} = \pm 3 \text{ PP (Lernstichprobe) /}$ $\pm 2 \text{ PP (Validierungsstichprobe) (*)}$ → die verwendeten 8 Variablen werden genannt, Angaben zur Aggregation fehlen (Logitmodell auf Basis transformierter Daten)
LEHMANN (2003)	400 Ausfaller, 19.600 Nicht-Ausfaller, 20.000 Unt.jahre (?) (resampling, Testsample: 19.000 Nicht-Ausfaller, 67 Ausfaller)	deutsche mittelständische Unternehmen eines „kommerziellen Kreditinstituts“, verschiedene Branchen, Umsatz < 5 Mio. Euro: 66% der Unternehmen	Def. Ausfall: erstmalige Wertberichtigung,	AR Finanzkennzahlenrating (I): 43,6% (*) AR Zahlungsverhalten („checking account“) (II): 54,2% AR Softfaktorrating (III): 43,8% AR Kombination aus (I) und (II): 58,4% AR Kombination aus (I),(II) und (III): 62,4% → keine Angaben zu den Verfahren $KI_{95\%,AR} = \pm 9 \text{ PP (*)}$ ²⁹⁹
ENGELMANN, HAYDEN, TASCHE (2003)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR ordinary business income / total assets: 57,7% (*) AR bestes Logitmodell: 62,4% ³⁰⁰ $KI_{95\%,AR} = \pm 2,5 \text{ PP(*)}$ ³⁰¹ → alle drei Verfahren (Logitmodelle, 4 Variablen) sind vollständig dokumentiert
HAYDEN (2003)	siehe Abschnitt 3.3	siehe Abschnitt 3.3	Def. Ausfall: siehe Abschnitt 3.3	AR equity / assets: 44,1% (*), AR bestes Logitmodell: 58,9% $KI_{95\%,AR} = \pm 3,5 \text{ PP, } \pm 3 \text{ PP, } \pm 3,5 \text{ PP (Insolvenz, Restrukturierung, Verzug) (*)}$ ³⁰² → alle drei Verfahren (Logitmodelle, 6-9 Variablen) sind vollständig dokumentiert

²⁹⁹ Auf Basis der Resamplinguntersuchungen ermittelt LEHMANN (2003, S. 22) eine Standardabweichung der *AUC* von 2,11%. Dies entspricht – bei Unterstellung normalverteilter *Accuracy Ratios* - einem 95%-Konfidenzintervall für die Accuracy Ratio von $\pm 2,11\% * 2 * 1,96 = \pm 8,3 \text{ PP}$.

³⁰⁰ Auf Basis des gleichen Datensatzes erzielt HAYDEN (2002, S. 75ff.) mit einem Logitmodell mit 12 Variablen eine Accuracy Ratio von 70%.

³⁰¹ Auf Basis der Resamplinguntersuchungen von ENGELMANN, HAYDEN, TASCHE (2003, S. 18), dort als *Bootstrapping* bezeichnet, ergibt sich ein AR-Konfidenzintervall von $\pm 2,49 \text{ PP}$ für Modell 1, dessen Schätzgüte in obiger Tabelle dargestellt wird.

³⁰² Auf Basis der Angaben von HAYDEN (2003, S. 33) ergeben sich folgende 95%-Konfidenzintervalle: $KI_{95\%,AR} = \pm 2,4 \text{ PP, } \pm 4,0 \text{ PP, } \pm 4,9 \text{ PP}$ (Insolvenz, Restrukturierung, Zahlungsverzug). Vermutlich wurden die Konfidenzintervalle des Insolvenzsamples auf Basis der kompletten Stichprobe und nicht der Validierungsstichprobe ermittelt.

Anmerkung zu Tabelle H, Insolvenzprognosestudien großer Aktiengesellschaften

Interessant ist, daß der erhebliche Aufwand, den die kommerziellen Ratingagenturen bei der Erstellung ihrer Ratingurteile betreiben³⁰³ - und die erheblichen Gebühren, die sie hierfür in Rechnung stellen^{304,305} - sich nicht adäquat in der *Schätzgüte* ihrer Prognosen niederschlagen.³⁰⁶ Mit Einjahres-Accuracy-Ratios von ca. 85% erzielen sie zwar überdurchschnittlich gute Ergebnisse. Die meisten der oben zitierten Studien beziehen sich jedoch ausschließlich auf US-amerikanische Nicht-Finanzunternehmen – und gerade hier, also auf Basis *vergleichbarer* (und zum Teil sogar *identischer*) Stichproben, schneiden die Ratingagenturen vergleichsweise besonders schlecht ab.³⁰⁷

Anmerkung zu Tabelle J, Insolvenzprognosestudien deutscher/ österreichischer KMU

Auch hinsichtlich der kommerziellen KMU-Ratingverfahren ist ein ähnlich ungünstiger Befund bezüglich der relativen Prognosegüte zu verzeichnen. In einer aktuellen deutschen Marktstudie³⁰⁸ von 15 kommerziellen Standardlösungen für das Rating von mittelständischen Unternehmen erhielten (nur) die Verfahren CREDITREFORM-Bonitätsindex, MOODY'S RiskCalc Germany und CREDITREFORM-Bilanzrating die Bewertung „sehr gut“. Die Prognosegüte des CREDITREFORM-Bonitätsindex Deutschland ist jedoch – trotz gleicher Grundgesamtheit – schlechter (!) als die Prognosegüte des in Abschnitt 3.2 vorgestellten Benchmarkratings, dessen Insolvenzprognosen lediglich auf der Kenntnis von Rechtsformen und Branchen der Unternehmen basieren.^{309,310} Das multivariate, kennzahlbasierte Verfahren *MOODY'S RiskCalc Germany*³¹¹ erzielt zwar (auf Basis einer anderen Stichprobe) etwas bessere Schätzwertwerte (AR ca. 60%), die Prognosegüte liegt jedoch erstens nur geringfügig über der Prognosegüte der univariaten Erklärungskraft der besten Einzelkennzahl und zweitens ist sie nicht höher als die Schätzgüte von kostenlos verfügbaren Verfahren, die bei vergleichbaren Grundgesamtheiten und auf Basis ähnlich großer Stichproben kalibriert wurden (siehe Tabelle J). Die Schätzgüte des CREDITREFORM-Bilanzrating^{312,313} beträgt ebenfalls nur ca. 60%.

³⁰³ Siehe beispielsweise die über 100 Seiten umfassende Beschreibung der Elemente von Ratingprozessen bei S&P (2003b), während die Beschreibung einfacher statistischer Modelle, wie Diskriminanz- oder Logitanalysen, nur wenige Zeilen erfordert.

³⁰⁴ Nach WHITE (2001, S. 14) betragen die vom zu bewertenden Unternehmen zu tragenden „Listenpreise“ für die Erstellung eines Ratings durch MOODY'S oder S&P 3,25 Basispunkte bei Anleihebeträgen von bis zu 500 Millionen US\$, bei einer Mindestgebühr von 25.000 US\$ und einem Höchstbetrag von 125.000 US\$ (S&P), bzw. 130.000 US\$ (MOODY'S). Für über 500 Mio. US\$ hinausgehende Anleihebeträge berechnen beide Agenturen 2 Basispunkte. S&P kappt den Gesamtbetrag bei 200.000 US\$, verlangt aber eine zusätzliche Gebühr von 25.000 US\$ bei der erstmaligen Erstellung eines Ratings.

³⁰⁵ siehe auch CANTOR, PACKER (1994, S. 4) und TREACY, CAREY (2000/1998, S. 911)

³⁰⁶ Vertreter von Ratingagenturen versuchen die Relevanz dieser Befunde mit Verweis auf die vergleichsweise bessere Leistungsfähigkeit ihrer Ratings bei längeren Prognosehorizonten und die größere Stabilität von Agenturratings im Vergleich zu anderen Prognoseverfahren zu relativieren oder fordern die Berücksichtigung weiterer Komponenten ihrer Ratingurteile ein, insbesondere den „Watchlist/Outlook“-Status von Ratings, siehe CANTOR, MANN (2003, S.25ff.), FONS, VISWANATHAN (2004, S.10) und HAMILTON (2004, S.11)

³⁰⁷ Siehe hierzu vor allem die Studien von CAREY, HRYCAY (2001) und KEALHOFER (2003).

³⁰⁸ Siehe ROMEIKE, WEHRSPORN (2004a und 2004b).

³⁰⁹ In den Jahren 1998-2000 erreichte der Creditreform-Bonitätsindex AR-Werte von 51%, siehe Tabelle J, während das Branchen-Rechtsform-Rating in den Jahren 1998-2001 AR-Werte von 53%-55% erzielte, siehe Tabelle E. Erst in den Jahren 2002 und 2003 fällt die Schätzgüte des Verfahrens auf Werte um 45% ab.

³¹⁰ Gebühren: 400 – 500 Euro Mitgliedsbeitrag pro Jahr zzgl. 16 – 20 Euro pro Bonitätsauskunft, siehe ROMEIKE, WEHRSPORN (2004b, S.23)

³¹¹ Gebühren: „individuelles Preismodell“, Grundpreis ab 20.000US\$, siehe ebenda (S.24)

³¹² Im Gegensatz zum CREDITREFORM-Bonitätsindex erfordert das CREDITREFORM-Bilanzrating das Vorliegen von Jahresabschlüssen.

³¹³ Gebühren: 549,- Euro je Rating siehe ebenda (S.22); Quelle des Accuracy-Ratio-Wert: eigene Berechnungen basierend auf der grafischen Darstellung der CAP-Kurve in ebenda (2004b, S.20)

4 Fazit

Ziel des vorliegenden Diskussionspapiers war es, einen Beitrag zur Verbesserung der Vergleichbarkeit der Schätzgüteregebnisse von Insolvenzprognosestudien zu leisten. Um aus den Erfahrungen *guter*, im Sinne von *mit einer großen Schätzgüte ausgestatteter*, Insolvenzprognoseverfahren lernen zu können, ist es zunächst erforderlich, diese überhaupt erkennen zu können. Angesichts der Vielzahl der in der empirischen Literatur verwendeten Kenngrößen zur Messung von *Prognosequalität*, ist dies keine triviale Aufgabe. In dem Diskussionspapier wird argumentiert, daß ordinale Schätzgütemaße, und hierbei insbesondere die Accuracy Ratio, sich für stichprobenübergreifende Vergleiche am besten eignen. Es wurden verschiedene Verfahren vorgestellt, um den Anwendungsbereich dieser Kenngrößen zu erweitern.

Wesentlich erschwert werden stichprobenübergreifende Vergleiche der Schätzgüteregebnisse unterschiedlicher Verfahren durch den Einfluß der im Abschnitt 3 genannten systematischen und unsystematischen Störfaktoren. Als besonders bedeutend stellte sich dabei der Einfluß der *Unternehmensgröße*, bzw. von hiermit verknüpften Einflußfaktoren wie *Datenqualität* und *Vorselektion von Stichproben* heraus. Während für Insolvenzprognosestudien im Bereich großer, börsennotierter Unternehmen aufgrund von Publizitätspflichten und der Verfügbarkeit hochwertiger externer Bilanzdatenbanken eine unverzerrte Auswahl an Unternehmen sichergestellt werden kann und Jahresabschlußdaten und Ausfallinformationen von hoher (formaler) Qualität vorliegen und zusätzlich Kapitalmarktdaten als weitere Informationsquellen verfügbar sind, ist die Situation bei kleinen Unternehmen weniger vorteilhaft. Fast alle der aufgeführten KMU-Insolvenzprognosestudien basierten auf den Datensätzen von Banken. Neben der ohnehin schon schlechteren formalen Qualität der Jahresabschlüsse von KMU resultieren hieraus – wie in Abschnitt 3 dargelegt und in Anhang V gezeigt wurde – eine Reihe weiterer schwerwiegender Defizite.

Mit Accuracy-Ratio-Werten von 75% bis über 90% zeigen die aufgeführten Studien, daß sich Insolvenzen *bei großen Unternehmen* mit einer sehr hohen Präzision durch kennzahlenbasierte Ratingmodelle vorhersagen lassen, verglichen jedenfalls mit den Schätzgüteregebnissen von Insolvenzprognosen bei kleinen Unternehmen (50%-60%), Privatkunden (über 50%)³¹⁴, dem Branche-Rechtsform-Rating (50%) (siehe Abschnitt 3.2), der univariaten Prognosefähigkeit einzelner Kennzahlen (55%-65%) (siehe Abschnitt 3.3) – aber auch mit den Schätzgüteregebnissen, die auf völlig anderen Anwendungsgebieten erzielt werden (50%-95%).³¹⁵

Einige der in Abschnitt 3.5 aufgeführten, vollständig dokumentierten und damit auch kostenlos nachbaubaren Ratingmodelle erzielen präzisere Insolvenzprognosen als die untersuchten kommerziellen Ratingverfahren.

³¹⁴ Zum Vergleich: auf Basis einer Validierungsstichprobe von 410.000 Datensätzen erzielt das SCHUFA-Scoreverfahren (2001) zur Bonitätsbewertung von *Privatkunden* eine Accuracy-Ratio von 53%, siehe FAHRMEIR, HENKING, HÜLS (2002, S. 27). Entgegen FALKENSTEIN, BORAL, CARTY (2003/2000, S.9) führt die rein quantitative Überlegenheit bei der Verfügbarkeit von Ausfalldaten bei Privatkundenbonitätsanalysen im Vergleich zu Firmenkundenanalysen zu keiner Überlegenheit hinsichtlich der hieraus abgeleiteten Schätzmodelle. Siehe auch die vergleichsweise schlechte Schätzgüte des CREDITREFORM Bonitätsindex für dessen Parametrisierung Millionen Datensätze solventer und zehntausender Datensätze insolventer Unternehmen zur Verfügung stehen.

³¹⁵ So erreichen beispielsweise Regenprognosen eine Accuracy Ratio von 54%-80% (Median: 64%); Eignungstests, beispielsweise zur Prognose von Schulabschlüssen, erzielen AR-Werte von 60%-86%; Röntgenuntersuchungen (Brust) AR-Werte von 96% und CT-Untersuchungen (Gehirn) AR-Werte von 94%, während die Aussagen von Lügendetektoren (sowohl in Feldstudien als auch in Laborexperimenten) Accuracy-Ratio-Werte von durchschnittlich 73% erreichen, siehe SWETS (1988, S. 1288ff. und die dort zitierte Literatur). Prognosen zur Rückfallgefahr von Gewaltstraftätern erzielen AR-Werte von 52% und AIDS-Tests Werte von 84%-94%, siehe SWETS, DAWES, MONAHAN (2000, S.10 und 16) [alle Werte waren im Original in AUC gegeben und wurden in AR umgerechnet]. Hinweis: einige der aufgeführten Beispiele messen nicht die Qualität von *Prognosen*, sondern die Qualität von *Klassifizierungen*.

Literaturverzeichnis

- ALTMAN, E.I. (1968): "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", in Journal of Finance, Bd. 23 (4), S. 589-610, September 1968
- ALTMAN, E.I., HALDEMAN, R.G., NARAYANAN, P. (1977): "ZETA analysis: A new model to identify bankruptcy risk of corporations", in Les Cahiers de Recherche, 1977, S. 1- 45; ebenfalls erschienen in Journal of Banking and Finance, Bd. 1 (1), S. 29- 54, 1977
- ALTMAN, E.I., MARCO, G., VARETTO, F. (1994): "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)", in Journal of Banking and Finance, Bd. 18, S. 505-529, 1994
- ALTMAN, I., SAUNDERS, A. (1998): „Credit Risk Measurement: Developments over the last 20 years“, in Journal of Banking and Finance, Bd. 21, S. 1721-1742, 1998
- ALTMAN, E. I.(2000/1968): „Predicting financial distress of companies: revisiting the Z-score and Zeta ® models“, Working Paper, Stern School of Business, New York University, <http://www.stern.nyu.edu/~ealtman/Zscores.pdf>, (25.4.2005), 07/2000
- ALTMAN, E. I. (2002): "Revisiting Credit Scoring Models in a Basel 2 Environment", Working Paper, Stern School of Business, New York University, <http://www.stern.nyu.edu/fin/workpapers/papers2002/pdf/wpa02041.pdf> (22.2.2005), 05/2002
- ALTMAN, E.I., RIJKEN, H.A. (2004): "How rating agencies achieve rating stability", in Journal of Banking and Finance, Bd. 28, S. 2679-2714, 2004
- AMATO, J.D. FURFINE, C.H. (2004): „Are credit ratings procyclical?“, in Journal of Banking and Finance, Bd. 28, S. 2641-2677, 2004
- AZIZ, M.A., DAR, H.A. (2004): "Predicting Corporate Financial Distress: Whither do We Stand?", <http://www.lut.ac.uk/departments/ec/Researchpapers/2004/JOBF.pdf> (20.1.2005), University of Loughborough Working Paper, 2004
- BALCAEN, S. OOGHE, H. (2004): "35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and their Related Problems", Vlerick Leuven Gent Working Paper Series 2004/15, <http://www.vlerick.be/research/workingpapers/vlgms-wp-2004-15.pdf> (5.1.2005), 2004
- BANGIA, A., DIEBOLD, F. X., SCHUERMAN, T. (2002): "Ratings Migration and the Business Cycle, With Application to Credit Portfolio Stress Testing", in Journal of Banking and Finance, Bd. 26, S. 445-474, 2002
- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2000a): "Summary of Responses Received on the Report 'Credit Risk Modelling: Current Practices and Applications'", Bank for International Settlements, <http://www.bis.org/publ/bcbs71.pdf> (24.8.2004), May 2000
- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2000b): "Range of Practice in Banks' Internal Ratings Systems", Discussion Paper, Bank for International Settlements, <http://www.bis.org/publ/bcbs66.pdf> (28.8.2004), January 2000

- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2000c): “Credit Ratings and Complementary Sources of Credit Quality Information”, Working Paper #3, http://www.bis.org/publ/bcbs_wp3.pdf (30.8.2004), 2000
- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2001): “The internal ratings-based approach: supporting document to the New Basel Capital Accord”, Consultative document, Bank for International Settlements, <http://www.bis.org/publ/bcbsca05.pdf> (24.8.2004), January 2001
- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2004): „Internationale Konvergenz der Kapitalmessung und Eigenkapitalanforderungen, Überarbeitete Rahmenvereinbarung“, Übersetzung der Deutschen Bundesbank, http://www.bundesbank.de/download/-bankenaufsicht/pdf/eigenkapitalempfehlung_de.pdf (7.9.2004), Juni 2004
- BASEL COMMITTEE ON BANKING SUPERVISION (HRSG.) (2005): “Studies on the Validation of Internal Rating Systems”, Working Paper No. 14, http://www.bis.org/publ/bcbs_wp14.pdf (25.5.2005), überarbeitete Version, 05/2005
- BASLER AUSSCHUß: siehe BASEL COMMITTEE ON BANKING SUPERVISION
- BEAVER, W.H. (1966): “Financial Ratios as Predictors of Failure”, Empirical Research in Accounting: Selected Studies, in einer Beilage zu Journal of Accounting Research (4), S. 71- 111, 1966
- BLÖCHLINGER, A., LEIPPOLD, M. (2005/2004): “Economic Benefit of Powerful Credit Scoring”, National Centre of Competence in Research Financial Valuation and Risk Management, University of Zürich, Working Paper No. 216, http://www.nccr-finrisk.unizh.ch/media/pdf/wp/WP216_2.pdf (27.5.2005), 04/2005
- BLOCHWITZ, S., LIEBIG, T., NYBERG, M. (2000): “Benchmarking Deutsche Bundesbank’s Default Risk Model, the KMV® Private Firm Model® and Common Financial Ratios for German Corporations”, Workshop on Applied Banking Research, BASEL COMMITTEE ON BANKING SUPERVISION, <http://www.bis.org/bcbs/oslo/liebigblo.pdf> (16.8.2004), 2000
- BLOCHWITZ, S., HOHL, S., TASCHE, D., WEHN, C.S. (2004): “Validating Default Probabilities on Short Time Series”, http://www.chicagofed.org/publications/capital_and_market_risk_insights/2004/validating_default_probabilities.pdf (1.6.2005), Working Paper, Federal Reserve Bank of Chicago, 05/2004
- BLUME, M.E., LIM, F., MACKINLAY, A.C. (1998): „The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality”, in Journal of Finance, Bd. LIII, No. 4, S. 1389-1413, August 1998
- BOHN, J., ARORA, N., KORABLEV, I. (2005): “Power and Level Validation of the EDF™ Credit Measure in North America”, Moody’s KMV, Working Paper, http://www.moodyskmv.com/research/whitepaper/EDF_Validation_NorthAmerica.pdf (24.5.2005), 03/2005
- BRIER, G.W. (1950): ”Verification of forecasts expressed in terms of probability, Monthly Weather Review 78, S. 1 – 3, 1950
- CANGEMI, B., SERVIGNY, A. DE, FRIEDMAN, C. (2003): “Standard & Poor’s Credit Risk Tracker for Private Firms, Technical Document”, S&P Working Paper, http://www.standardandpoors.co.jp/spf/pdf/rev_CRTTechDocument20031126.pdf (3.3.2005), 2003

- CANTOR, R., PACKER, F. (1994): „The Credit Rating Industry“, Quarterly Review, Federal Reserve Bank of New York, Bd. 19 (2), S. 1-26, 1994
- CANTOR, R., C. MANN (2003): "Measuring the Performance of Corporate Bond Ratings", Special Comment, Report #77916, Moody's Investor's Services, 04/2003
- CANTOR, R. (2004): "Measuring the Quality and Consistency of Corporate Ratings across Regions", Moody's Investors Service, Special Comment, Report # 89168, 11/2004
- CAREY, M. S. HRYCAY, M. (2001): "Parameterizing Credit Risk Models with Rating Data", in Journal of Banking and Finance, Bd. 25 (1), S. 197-270, 2001
- CROSBIE, P., BOHN, J. (2003): "Modeling Default Risk", Modeling Methodology, Moody's KMV, 2003, <http://www.moodyskmv.com/research/whitepaper/ModelingDefaultRisk.pdf> (16.7.2004), 2003
- CROUHY, M., GALAI, D., MARK, R. (2001): "Prototype risk rating system", in Journal of Banking and Finance, Bd. 25, S. 47-95, 2001
- DECISIONEERING (HRSG.) (2000): „Crystal Ball 2000, User Manual®“, Decisioneering, Inc, Denver, Colorado, 2000
- DEUTSCHE BUNDESBANK (HRSG.) (1999): "Zur Bonitätsbeurteilung von Wirtschaftsunternehmen durch die Deutsche Bundesbank", in DEUTSCHE BUNDESBANK, Monatsbericht Januar 1999, S. 51-64, 1999
- DEUTSCHE BUNDESBANK (HRSG.) (2001): "Die neue Baseler Eigenkapitalvereinbarung (Basel II)", in DEUTSCHE BUNDESBANK, Monatsbericht April 2001, S. 15-44, 04/2001
- DEUTSCHE BUNDESBANK (HRSG.) (2003): "Validierungsansätze für interne Ratingsysteme", in Deutsche Bundesbank, Monatsbericht September 2003, S. 61-74, 2003
- DEUTSCHE VEREINIGUNG FÜR FINANZANALYSE, KOMMISSION RATING STANDARDS, ARBEITSKREIS 2 „VALIDIERUNG“ (DVFA, HRSG.) (2004): „DVFA – Validierungsstandards“, in Finanz Betrieb, S. 596-601, 09/2004
- DIMITRAS, A.I., ZANAKIS, S.H., ZOPOUNIDIS, C. (1996): „A survey of business failures with an emphasis on prediction methods and industrial applications, Theory and methodology“, in European Journal of Operational Research, Bd. 90, S. 487-513, 1996
- DVFA: siehe DEUTSCHE VEREINIGUNG FÜR FINANZANALYSE
- DWYER, D.W., STEIN, R. M. (2003): "Inferring the Default Rate in a Population by Comparing Two Incomplete Default Databases", Tech. Report #021216, Moody's KMV, 03/2003
- DWYER, D. W., KOCAGIL, A. E., STEIN, R. M. (2004): „The Moody's KMV EDF™ RiskCalc™ v3.1 model, Next-Generation Technology for Predicting Private Firm Credit Risk“, http://www.moodyskmv.com/research/whitepaper/EDF_RiskCalc_v3_1.pdf, (16.7.2004), Moody's KMV Company, 05/2004
- ELTON, E. J., GRUBER, M. J., AGRAWAL, D., MANN, C. (2004): „Factors affecting the valuation of corporate bonds“, in Journal of Banking and Finance, Bd. 28, S. 2747-2767, 2004

- ENGELMANN, B., HAYDEN, E., TASCHE, D. (2003): „Measuring the Discriminative Power of Rating Systems”, Deutsche Bundesbank, Discussion Paper, Series 2: Banking and Financial supervision, No 01/2003, http://www.bundesbank.de/bankenaufsicht/-bankenaufsicht_diskussionspapiere.en.php (19.7.2004), 2003
- ENGLISH, W. B., NELSON, W. R. (1998): “Bank Risk Rating of Business Loan”, Board of Governors of the Federal Reserve System FEDS, Paper No. 98-51, <http://ssrn.com/abstract=148753> (22.5.2005), 12/1998
- ESCOTT, P., GLORMANN, F., KOCAGIL, A.E. (2001a): „Moody’s RiskCalc™ für nicht börsennotierte Unternehmen: Das Deutsche Modell“, Moody’s Investors Service, Rating Methodology, 06/2001
- ESCOTT, P., GLORMANN, F., KOCAGIL, A.E. (2001b): „RiskCalc™ for Private Companies: The German Model“, Moody’s KMV, Modeling Methodology, <http://www.moodykmv.-com/research/whitepaper/720441.pdf> (18.02.2005), 11/2001
- FAHRMEIR, L., HENKING, A., HÜLS, R. (2002): „Methoden zum Vergleich verschiedener Scoreverfahren am Beispiel der SCHUFA-Scoreverfahren“, Risknews, 11/2002, S. 20-29, 2002
- FALKENSTEIN, E., BORAL, A., KOCAGIL, A. E. (2000): „RiskCalc™ For Private Companies II: More Results and the Australian Model”, Moody’s Investors Service, Rating Methodology, Report # 62265, 12/2000
- FALKENSTEIN, E., BORAL, A., CARTY, L. (2003/2000): „RiskCalc™ For Private Companies“, Modeling Methodology, Moody’s KMV, 2003
- FEELDERS, A.J. (2000): “Credit Scoring and Reject Inference with Mixture Models”, in International Journal of Intelligent Systems in Accounting, Finance and Management, Bd. 9, S. 1-8, 2000
- FISCHER, A. (2004): „Qualitative Merkmale in bankinternen Ratingsystemen: eine empirische Analyse zur Bonitätsbeurteilung von Firmenkunden“, Uhlenbruch Verlag, Bad Soden am Taunus, 2004, zugelassene Dissertation, Universität Münster, 2004
- FONS, J. S. (2002): “Understanding Moody’s Corporate Bond Ratings and Rating Process”, Moody’s Investors Service, Report # 74982, 05/2002
- FONS, J. S., VISWANATHAN, J. (2004): „A User’s Guide to Moody’s Default Predictor Model: an Accounting Ratio Approach“, Moody’s Investors Service, Report # 90127, 12/2004
- FRANKS, J., SERVIGNY, A. DE, DAVYDENKO, S. (2004): “A Comparative Analysis of the Recovery Process and Recovery Rates for Private Companies in the U.K., France, and Germany”, STANDARD AND POOR’S Risk Solution, 06/2004
- FRERICHS, H., WAHRENBURG, M. (2003) :”Evaluating internal credit rating systems depending on bank size”, Working Paper Series: Finance and Accounting, Johann Wolfgang Goethe-Universität Frankfurt Am Main, No. 115, <http://www.wiiv.de/publikationen/-Evaluationinternalcreditrating898.pdf> (25.8.2004), 09/2003
- FRITZ, S., HOSEMAN, D. (2000): “Restructuring the Credit Process: Behaviour Scoring for German Corporations”, in International Journal of Intelligent Systems in Accounting, Finance and Management, Bd. 9, S. 9–21, 2000

- GRICE, J.S., DUGAN, M.T. (2001): "The Limitations of Bankruptcy Prediction Models: Some Cautions for the Researcher", *Review of Quantitative Finance and Accounting*, Bd. 17, S. 151-166, 2001
- GRUNERT, J., NORDEN, L., WEBER, M. (2005): "The role of non-financial factors in internal credit ratings", in *Journal of Banking and Finance*, Bd. 29, S. 509-531, 2005
- GUJARATI, D. (1999/1992): "Essentials of Econometrics", Irwin/McGraw-Hill, 2. Auflage, 1999
- GÜNTHER, T., GRÜNING, M. (2000): „Einsatz von Insolvenzprognoseverfahren bei der Kreditwürdigkeitsprüfung im Firmenkundenbereich“, in „Die Betriebswirtschaft“, Heft 1/2000, S. 39-59, 2000
- GÜNTERBERG, B., WOLTER, H.-J. (2003): „Unternehmensgrößenstatistik 2001/2002 - Daten und Fakten“, IfM-Mat. Nr. 157, Bonn, <http://www.ifm-bonn.org/dienste/dafa.htm> (30.11.2004), 2003
- GUPTON, G.M., STEIN, R. M. (2002): "LossCalcTM: Model for Predicting Loss Given Default (LGD)", Moody's KMV Company, Modeling Methodology, 2002
- GUPTON, G. M., STEIN, R. M. (2005): „LossCalc V2: Dynamic Prediction of LGD Modeling Methodology“, Moody's KMV, Working Paper, http://www.moodyskmv.com/-research/whitepaper/LCv2_DynamicPredictionOfLGD.pdf (25.5.2005), 01/2005
- GÜTTLER, A. (2004): "Using a Bootstrap Approach to Rate the Raters", Working Paper Series: Finance and Accounting, Johann Wolfgang Goethe-Universität Frankfurt, No. 132/2004, <http://opus.zbw-kiel.de/volltexte/2004/2343/pdf/835.pdf> (11.1.2005), erscheint in *Financial Markets and Portfolio Management*, 2004
- HAMERLE, A., RAUHMEIER, R., RÖSCH, D. (2003): "Uses and Misuses of Measures for Credit Rating Accuracy", Working Paper, University of Regensburg, http://defaultrisk.com/-pdf_files/Uses_n_Misuses_o_Measures_4_Cr_Rtng_Accrc.pdf (3.1.2005), 04/2003
- HAMILTON, D.T. (2004): "Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History", Moody's Investors Service, Special Comment, Report # 81068, 02/2004
- HARTUNG, J. (1991): "Statistik: Lehr- und Handbuch der angewandten Statistik", Oldenbourg Verlag, München, Wien, 8. Auflage, 1991
- HAYDEN, E. (2002): „Modeling an Accounting-Based Rating System for Austrian Firms“, Dissertation, Fakultät für Wirtschaftswissenschaften und Informatik, Universität Wien, www.bwl.univie.ac.at/bwl/fwi3/members/hayden/diss.pdf (9.5.2005), 2002
- HAYDEN, E. (2003): "Are Credit Scoring Models Sensitive With Respect to Default Definitions? Evidence from the Austrian Market", Working Paper, University of Vienna <http://www.bwl.univie.ac.at/bwl/fwi3/members/hayden/DefaultDefinitions.pdf> (16.8.2004), 2003
- HULL, J.C. (2003): "Options, futures and other derivatives", 5th edition, Prentice-Hall, Upper Saddle River, 2003

- HUSCHENS, S., HÖSE, S. (2003): „Sind interne Ratingsysteme im Rahmen von Basel II evaluierbar? – Zur Schätzung von Ausfallwahrscheinlichkeiten durch Ausfallquoten“, in Zeitschrift für Betriebswirtschaft zfb, Bd. 73 (2), S. 139-168, 2003
- JANKOWITSCH, R., PICHLER, S., SCHWAIGER, W. S. A. (2003): “Rating Granularity and Basel II Capital Requirements”, TU Wien, Working Paper, <http://www.wu-wien.ac.at/inst/ikw-/hp/download/publ/basel.pdf> (10.2.2005), im Reviewprozeß in Journal of Banking and Finance, 11/2003
- JORDÁO, F., STEIN, R.M. (2003): “What is a more powerful model worth?”, Technical Report, Report #030124, Moody’s KMV Company, 2003
- KEALHOFER, S. (2003): „Quantifying Credit Risk I: Default Prediction“, in Financial Analysts Journal, S. 30–44, Jan./Feb. 2003
- KEENAN, S. C. (1999): “Predicting Default Rates: A Forecasting Model for Moody's Issuer-Based Default Rates”, Moody’s Investors Service, Special Comment, Report # 47729, 08/1999,
- KEENAN, S.C., SOBEHART, J.R. (1999): “Performance Measures for Credit Risk Models”, Moody’s Investors Service, Research Report # 1-10-10-99, 1999
- KOCAGIL, A. E., AKHAVEIN, J. D. (2001): “Moody's RiskCalc™ for Private Companies: Japan”, Moody’s Investors Service, Rating Methodology, Report # 73072, 12/2001
- KOCAGIL, A. E., IMMING, R., GLORMANN, F., ESCOTT, P. (2003): “RiskCalc™ For Private Companies: The Austrian Model”, Moody’s KMV, Modeling Methodology, http://www.moodyskmv.com/research/whitepaper/atmethod_german.pdf (10.05.2005), 11/2003
- KRAFT, H., KROISANDT, G., MÜLLER, M. (2004): „Redesigning Ratings: Assessing the Discriminatory Power of Credit Scores under Censoring“, Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM), Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=302137 (3.3.2005), 09/2004
- KRAHNEN, J.P., WEBER, M. (2001): “Generally accepted rating principles: A primer”, in Journal of Banking and Finance, Bd. 25 (1), S. 3-23, 2001
- KRÄMER, W. (2003): „Die Bewertung und der Vergleich von Kreditausfall-Prognosen“, in Kredit und Kapital, Bd. 36 (3), S. 395-410, 2003
- KRÄMER, W., GÜTTLER, A. (2003): “Comparing the accuracy of default predictions in the rating industry: The case of Moody’s vs. S&P”, Technical Report-Reihe des SFB 475 Nr. 23 (Universität Dortmund), <http://www.wiwi.uni-frankfurt.de/schwerpunkte/finance/wp/332.pdf>, (13.4.2005), 2003
- KÜTING, K., WEBER, C.-P.: (2004/1993): „Die Bilanzanalyse, Lehrbuch zur Beurteilung von Einzel- und Konzernabschlüssen“, 6. Aufl., Schäffer-Poeschel, Stuttgart, 2004
- LAWRENZ, J., SCHWAIGER, W.S.A. (2002): „Bank Deutschland: Aktualisierung der Quantitative Impact Study (QIS2) von Basel II“, in Risknews 01/2002, S. 5-30, 2002

- LEE, W.-C. (1999): "Probabilistic Analysis of Global Performances of Diagnostic Tests: Interpreting the Lorenz Curve-Based Summary Measures", in *Statistics in Medicine*, Bd. 18, S. 455-471, 1999
- LEHMANN, B. (2003): "Is It Worth the While? The Relevance of Qualitative Information in Credit Rating", EFMA 2003 Helsinki Meetings, <http://ssrn.com/abstract=410186> (2.1.2005), 04/2003
- LÖFFLER, G. (2003): "Avoiding the Rating Bounce: Why Rating Agencies are Slow to React to new information", Working Paper Universität Ulm, <http://www.mathematik.uni-ulm.de/dof/papers/ratingbounce.pdf> (1.12.2004), erscheint in: *Journal of Economic Behavior and Organization*, 2003
- LÖFFLER, G. (2004a): "An anatomy of rating through the cycle", in *Journal of Banking and Finance*, Bd. 28, S. 695-720, 2004
- LÖFFLER, G. (2004b): "Ratings versus market-based measures of default risk in portfolio governance", in *Journal of Banking and Finance*, Bd. 28, S. 2715-2746, 2004
- MAHONEY, C. (2002a): "The Bond Rating Process in a Changing Environment", Moody's Investors Service, Special Comment, Report #73741, 01/2002
- MAHONEY, C. (2002b): "The Bond Rating Process: A Progress Report", Moody's Investors Service, Rating Policy, Report #74079, 02/2002
- MATHESON, J. E., WINKLER, R. L. (1976): „Scoring rules for continuous probability distributions“, in *Management Sciences*, Bd. 22, No. 10, 1976
- MCQUOWN, J.A. (1993): "A Comment on Market vs. Accounting-Based Measures of Default Risk", KMV Working Paper, http://www.moodyskmv.com/research/whitepaper/-A_Comment_on_Market_vs_Accounting_Based_Measures_of_Default_Risk.pdf, (16.7.2004), KMV Corporation, 1993
- MILLER, R.M. (1998): "A Nonparametric Test for Credit Rating Refinements", in *Risk Magazine*, 08/1998
- MOODY'S (HRSG.) (2000): "Moody's Investors Service Response to the Consultative Paper Issued by the Basel Committee on Bank Supervision 'A New Capital Adequacy Framework'", Moody's Investors Service, Special Comment, 03/2000
- MOODY'S (HRSG.) (2004a): „Default & Recovery Rates of Corporate Bond Issuers, A Statistical Review of Moody's Ratings Performance, 1920-2003“, Moody's Investors Service, Special Comment, 01/2004
- MOODY'S (HRSG.) (2004b): „Moody's Rating Symbols & Definitions“, Moody's Investors Service, Report #79004, 08/2004
- MOODY'S (HRSG.) (2004c): "The Performance Of Moody's Corporate Bond Ratings: December 2004 Quarterly Update", Moody's Investors Service, 12/2004
- MOODY'S (HRSG.) (2005): "Default and Recovery Rates of Corporate Bond Issuers, 1920-2004", Moody's Investors Service, 01/2005

- MURPHY, A.H., WINKLER, R.L. (1992): "Diagnostic verification of probability forecasts", in International Journal of Forecasting, Bd. 7, S. 435-455, 1992
- NANDA, S., PENDHARKAR, P. (2001): "Linear Models for Minimizing Misclassification Costs in Bankruptcy Prediction", in International Journal of Intelligent Systems in Accounting, Finance and Management, Bd. 10, S. 155-168, 2001
- NORDEN, L., WEBER, M. (2005): „Möglichkeiten und Grenzen der Bewertung von Ratingsystemen durch Markt und Staat“, in Zeitschrift für betriebswirtschaftliche Forschung (ZfbF), Sonderheft 52, S. 31-54, 2005
- OENB: siehe ÖSTERREICHISCHE NATIONALBANK
- OHLSON, J.A. (1980): „Financial Ratios and the Probabilistic Prediction of Bankruptcy“, in Journal of Accounting Research, Bd. 18 (1), S. 109-131, 1980
- ÖSTERREICHISCHE NATIONALBANK (OENB, HRSG.) (2004a): „Ratingmodelle und -validierung“, Leitfadenreihe zum Kreditrisiko, Wien, 2004
- ÖSTERREICHISCHE NATIONALBANK (OENB, HRSG.) (2004b): „Kreditvergabeprozess und Kreditrisikomanagement“, Leitfadenreihe zum Kreditrisiko, Wien, 2004
- ÖSTERREICHISCHE NATIONALBANK (OENB, HRSG.) (2004c): „Neue quantitative Modelle der Bankenaufsicht“, Leitfadenreihe zum Kreditrisiko, Wien, 2004
- PLATTNER, D. (2002): „Warum Firmen Pleite machen, Der Einfluss finanzieller Kennziffern und anderer Faktoren auf die Insolvenzwahrscheinlichkeit kleiner und mittlerer Unternehmen“, in KfW-Beiträge Nr. 28, S. 37-54, August 2002
- RISKMETRICS (1996): „RiskMetrics – Technical Document“, 4th edition, Morgan Guaranty Trust Company, New York, 1996
- ROMEIKE, R., WEHRSPORN, U. (2004a): "Rating-Software im Test", in RATINGaktuell 06/2004, S. 10-19, 2004
- ROMEIKE, F., WEHRSPORN, U. (2004b): „Marktstudie Rating-Software im Test“, <http://www.cre-germany.com/Artikel/Rating-Software-2004.pdf> (15.12.2004), auszugsweise erschienen in RATINGaktuell 06/2004
- RÖSLER, J. (1988): „Die Entwicklung der statistischen Insolvenzdiagnose“, in HAUSCHILD, J. (HRSG.) (1988): „Krisendiagnose durch Bilanzanalyse“, Verlag Dr. Otto Schmidt KG, Köln, S. 102-115, 1988
- SCHMEISER, B. W. (2001): „Some Myths and Common Errors in Simulation Experiments“, in PETERS, B.A., SMITH, J. S., MEDEIROS, D. J., ROHRER, M. W. (HRSG., 2001): „Proceedings of the 2001 Winter Simulation Conference“, Piscataway NJ, IEEE press, S. 39-46, <http://www.informs-sim.org/wsc01papers/006.PDF> (30.5.2005), 2001
- SCHWAIGER, W.S.A. (2002): „Auswirkungen von Basel II auf den österreichischen Mittelstand nach Branchen und Bundesländern“, in Österreichisches Bankarchiv, 06/2002, S.433-446, 2002
- SCOTT, D. W. (1992): "Multivariate Density Estimation: Theory, Practice and Visualization", John Wiley, New York, 1992

- SHANNON, C.E. (2001/1948): “A Mathematical Theory of Communication”, in Bell System Technical Journal, Bd. 27, 1948, S. 379–423, 623–656, reprinted in Mobile Computing and Communications Review, Bd. 5 (I), S. 3 -55, 2001
- SHUMWAY, T. (2001): “Forecasting Bankruptcy More Accurately: A Simple Hazard Model”, in Journal of Business, Bd. 74 (1), S. 101-124, 2001
- SOBEHART, J.R., KEENAN, S.C., STEIN, R.M. (2000): „Benchmarking Quantitative Default Risk Models: A Validation Methodology”, Moody’s Investors Service, Rating Methodology, Report # 53621, 03/2000
- SOBEHART, J. R., STEIN, R. M., MIKITYANSKA, V., LI, L. (2000): “Moody’s Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk”, Moody’s Investors Service, Rating Methodology, Report #53853, 03/2000
- SOMERS, R. H. (1962): “A new asymmetric measure of association for ordinal variables”, American Sociological Review; Dez 1962, Bd. 27 (6), S. 799-811, 1962
- STANDARD AND POOR’S (HRSG.) (2002): „Ratings Performance 2001”, Special Report 02/2002, The McGraw Hills Companies, 2002
- STANDARD AND POOR’S (HRSG.) (2003a): „Ratings Performance 2002, Default, Transition, Recovery, and Spreads“, Special Report 02/2003, The McGraw Hills Companies, 2003
- STANDARD AND POOR’S (HRSG.) (2003b): „Corporate Ratings Criteria“, The McGraw Hills Companies, 2003
- STANDARD AND POOR’S (HRSG.) (2004a): „Ratings Performance 2003“, Special Report 03/2004, The McGraw Hills Companies, 2004
- STANDARD AND POOR’S (HRSG.) (2004b): „S&P Quarterly Default Update & Rating Transitions“, The McGraw Hills Companies, 10/2004
- STANDARD AND POOR’S (HRSG.) (2005): „Annual Global Corporate Default Study: Corporate Defaults Poised to Rise in 2005”, Global Fixed Income Research, The McGraw Hills Companies, 2005
- STATISTISCHES BUNDESAMT (HRSG.) (2004a): „INSOLVENZEN IN DEUTSCHLAND 2003, STRUKTUREN UND ENTWICKLUNGEN“, http://www.destatis.de/presse/deutsch/pk/2004/insolvenzen_2003_i.pdf (04.06.2004), Statistisches Bundesamt — Pressestelle, Wiesbaden, 2004
- STATISTISCHES BUNDESAMT (HRSG.) (2004b): Sonderanfrage Unternehmensinsolvenzen ausgewählter Wirtschaftsbereiche nach Rechtsformen, 1999-2003, Martin Bemann, 20.12.2004
- STATISTISCHES BUNDESAMT (HRSG.) (2004c): „Insolvenzen insgesamt und Insolvenzhäufigkeiten von Unternehmen nach ausgewählten Wirtschaftszweigen, Rechtsformen und Ländern, Deutschland“, <http://www.destatis.de/basis/d/insol/insoltab1.php> (15.2.2005), 2004
- STATISTISCHES LANDESAMT DES FREISTAATES SACHSEN (2004): Sonderanfrage zu branchen- und rechtsformspezifischen Insolvenzquoten in Sachsen, Kamenz, Martin Bemann, 2.11.2004

- STEIN, R.M. (2002): "Benchmarking Default Prediction Models, Pitfalls and Remedies in Model Validation", Moody's KMV, Report #030124, 2002
- STEIN, R.M., KOCAGIL, A.E, BOHN, J., AKHAVEIN, J. (2003): "Systematic And Idiosyncratic Risk In Middle-Market Default Prediction: A Study Of The Performance Of The Risk-Calc™ and PFM™ Models", Moody's Investors Service, Special Comment, Report #77261, 02/2003
- STEIN, R. M. (2005): „The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing“, in Journal of Banking and Finance, Bd. 29, S. 1213-1236, 2005
- STEINER, M., HEINKE, V. G. (1996): "Ratingagenturen an nationalen und internationalen Finanzmärkten", in BÜSCHGEN, H. E., EVERLING, O. (HRSG.) (1996): "Handbuch Rating", Gabler Verlag, Wiesbaden, S. 580 – 627, 1996
- S&P: siehe STANDARD AND POOR'S
- SWETS, J. A. (1973): „The Relative Operating Characteristic in Psychology“, in Science, Bd. 182, S. 990-1.000, 1973
- SWETS, J. A., PICKET, R. M. (1982): „Evaluation of Diagnostic Systems, Methods from Signal Detection Theory“, Academic Press, series in cognition and perception, New York et al., 1982
- SWETS, J. A. (1988): "Measuring the Accuracy of Diagnostic Systems", in Science, Bd. 240, S. 1285- 1293, 1988
- SWETS, J.A., DAWES, R.M., MONAHAN, J. (2000): „Psychological Science Can Improve Diagnostic Decisions“, in Psychological Science in the Public Interest, Bd. 1 (1), S. 1-26, 2000
- TREACY, W. F., CAREY, M. S. (2000/1998): "Credit Risk Rating at Large U.S. Banks", Federal Reserve Bulletin, 11/1998, S. 987-921, in "Credit risk rating systems at large US banks", in Journal of Banking and Finance, Bd. 24 (1-2), S. 167-201, 01/2000
- VARETTO, F. (1998): "Genetic algorithms applications in the analysis of insolvency risk", in Journal of Banking and Finance, Bd. 22, S. 1421-1439, 1998
- VARMA, P., CANTOR, R. (2005): „Determinants of Recovery Rates on Defaulted Bonds and Loans for North American Corporate Issuers: 1983-2003“, in Journal of Fixed Income, Bd. 14 (4), S. 29-44, 2005
- WESTGAARD, S., WIJST, N. VAN DER (2001): „Default probabilities in a corporate bank portfolio: a logistic model approach“, in European Journal of Operational Research, Bd. 135, S. 338-349, 2001
- WHITE, L. J. (2001): "The Credit Rating Industry: An Industrial Organization Analysis", New York University, Center for Law and Business, Research Paper No. 01-001, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=267083 (22.4.2005), 04/2001
- WINKLER, R. L. (1994): „Evaluating Probabilities: Asymmetric Scoring Rules“, in Management Science, Bd. 40, No. 11, S. 1395-1405, 1994

Anhang I: Ober- und Untergrenzen Accuracy Ratio

Ziel: Es soll ermittelt werden, welche Trenngüte, gemessen in Accuracy Ratio, ein ordinales Insolvenzprognoseverfahren mindestens und höchstens besitzen kann, wenn genau ein Punkt der zugehörigen CAP-Kurve (X_0, Y_0) bekannt ist, bzw. wenn für genau einen Trennwert des Verfahrens die zugehörige Kombination aus Fehler I. und II. Art (F_1, F_2) bekannt ist.

X_0 ... Anteil der ausgeschlossenen Unternehmen,

Y_0 ... Trefferquote (=100%-Fehler I. Art)

Technische Nebenbedingungen

Die CAP-Kurve muß im CAP-Diagramm durch die Punkte $(0\%;0\%)$ und $(100\%;100\%)$ verlaufen. Sie muß stetig und schwach-monoton steigend sein, d.h. $CAP(x) \geq 0$. Ihr Anstieg darf maximal $1/PD$ betragen, da um 100% aller Ausfaller zu erfassen, mindestens PD% aller Unternehmen ausgeschlossen werden müssen: $0 \leq CAP(x) \leq 1/PD$.

Inhaltliche Nebenbedingungen

Es wird unterstellt, daß bei den betrachteten Insolvenzprognoseverfahren bessere Ratingklassen stets mit niedrigeren Ausfallwahrscheinlichkeiten einhergehen. Somit werden nur konkave CAP-Kurven betrachtet (technisch: $CAP(x) \geq 0$ und $CAP(x)' \leq 0$).

Im Punkt $(0\%; 0\%)$ muß der Anstieg der CAP-Kurve mindestens 1,0 betragen, ansonsten könnte bei einem schwach monoton *sinkenden* Anstieg der CAP-Kurve der Punkt $(100\%; 100\%)$ nicht erreicht werden.

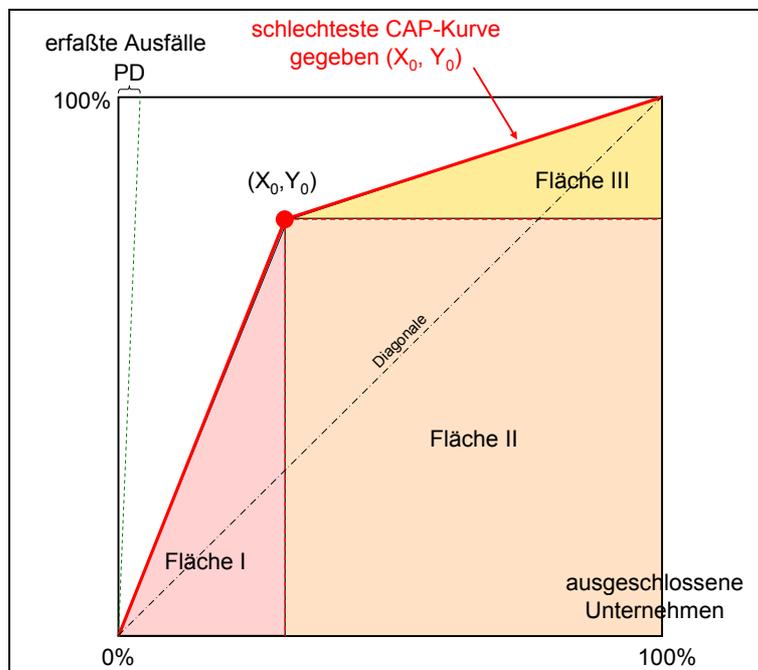


Abbildung 21: Flächenzerlegung für Ermittlung der flächenminimalen AUC_{CAP}

Ermittlung der Untergrenze für Accuracy Ratio

Die „schlechtestmögliche“, durch einen Punkt (X_0, Y_0) im CAP-Diagramm verlaufende CAP-Kurve,³¹⁶ die jedoch alle oben genannten technischen und inhaltlichen Nebenbedingungen erfüllt, ist der Linienzug $(0\%;0\%) - (X_0, Y_0) - (100\%;100\%)$. Diese CAP-Kurve korrespondiert mit einem ordinalen Prognoseverfahren, das nur zwischen den beiden durch X_0 getrennten Gruppen differenzieren kann, innerhalb der Gruppen jedoch keinerlei Trennfähigkeit besitzt – also keine zusätzlichen Informationen zur kategorialen Prognose (X_0, Y_0) beisteuern kann.

Zur Ermittlung von $AUC_{CAP,min}$, dem Flächeninhalt unter der „schlechtesten“ zulässigen CAP-Kurve, wird die Fläche unter der CAP-Kurve in drei Teilflächen zerlegt, deren Inhalt sich jeweils in Abhängigkeit der CAP-Koordinaten X_0 und Y_0 darstellen läßt (vgl. Abbildung 21).

$$\text{F 57) } AUC_{CAP,min} = \frac{X_0 \cdot Y_0}{2} + (1 - X_0) \cdot Y_0 + \frac{(1 - X_0) \cdot (1 - Y_0)}{2}$$

$$\text{F 58) } AUC_{CAP,min} = \frac{X_0 \cdot Y_0}{2} + Y_0 - X_0 \cdot Y_0 + \frac{1 - X_0 - Y_0 + X_0 \cdot Y_0}{2}$$

$$\text{F 59) } \boxed{AUC_{CAP,min} = \frac{1 - X_0 + Y_0}{2}} \text{ oder alternativ}$$

$$\text{F 60) } AUC_{CAP,min} = \frac{1}{2} + \frac{Y_0 - X_0}{2} \text{ mit Fläche unterhalb der Diagonalen} = \frac{1}{2} \text{ und folglich Fläche oberhalb der Diagonalen} = \frac{Y_0 - X_0}{2},$$

Die Differenz $Y_0 - X_0$ wird auch als „vertical distance“ bezeichnet (vertikaler Abstand von $(X_0; Y_0)$ zur Diagonalen). Die maximale vertikale Distanz einer ROC-Kurve stellt ein eigenständiges, in der Literatur verwendetes Gütemaß dar.³¹⁷

Unter Nutzung des in Abschnitt 2.3.2 dargelegten Zusammenhangs zwischen AUC_{CAP} und AR_{CAP} ergibt sich:³¹⁸

$$\text{F 61) } AR_{min} = \frac{2AUC_{CAP,min} - 1}{1 - PD}$$

$$\text{F 62) } AR_{min} = \frac{1 - X_0 + Y_0 - 1}{1 - PD}$$

$$\text{F 63) } \boxed{AR_{min} = \frac{Y_0 - X_0}{1 - PD}} \text{ } Y_0 - X_0 \text{ mißt den Abstand von } (X_0; Y_0) \text{ zur Diagonalen (siehe oben).}$$

In den folgenden Formeln werden die Gütemaße $AUC_{CAP,min}$ und AR_{min} in Abhängigkeit der Fehler I. und II. Art berechnet:

³¹⁶ Die „schlechtestmögliche“ CAP-Kurve ist diejenige der zulässigen CAP-Kurven mit dem kleinsten Flächeninhalt (AUC_{CAP}).

³¹⁷ siehe Lee (1999, S.462).

³¹⁸ Da $AR_{CAP} = AR_{ROC}$ wird bei Accuracy-Ratio-Variablen im folgenden auf die Verwendung der Indizes CAP oder ROC verzichtet. Die Indizes werden aber weiterhin in Verbindung mit AUC-Maßen verwendet.

Es gilt:

$$\text{F 64)} Y_0 = 1 - F_1 \text{ und}$$

$$\text{F 65)} F_2 = \frac{\text{bei } (X_0, Y_0) \text{ ausgeschlossene Nichtausfaller}}{\text{Nichtausfaller}} = \frac{X_0 - (1 - F_1) \cdot PD}{1 - PD} \text{ und damit:}$$

$$\text{F 66)} X_0 = F_2 \cdot (1 - PD) + (1 - F_1) \cdot PD \quad \text{Eingesetzt in Formel F 59 ergibt sich:}$$

$$\text{F 67)} AUC_{CAP, \min} = \frac{1 - (F_2 \cdot (1 - PD) + (1 - F_1) \cdot PD) + (1 - F_1)}{2}$$

$$\text{F 68)} \boxed{AUC_{CAP, \min} = \frac{(1 - F_1 - F_2) \cdot (1 - PD) + 1}{2}} \quad \text{Eingesetzt in Formel F 63 ergibt sich:}$$

$$\text{F 69)} AR_{\min} = \frac{1 - F_1 - F_2 \cdot (1 - PD) - (1 - F_1) \cdot PD}{1 - PD}$$

$$\text{F 70)} \boxed{AR_{\min} = 1 - (F_1 + F_2)} \text{ mit}$$

$$\text{F 71)} AUC_{ROC} = \frac{AR + 1}{2} \text{ und Formel F 70 ergibt sich für die Fläche unter der ROC-Kurve:}$$

$$\text{F 72)} \boxed{AUC_{ROC, \min} = 1 - \frac{F_1 + F_2}{2}}$$

Ermittlung der Obergrenze für Accuracy Ratio

Die „bestmögliche“ durch $(X_0; Y_0)$ verlaufende CAP-Kurve muß folgende Bedingungen erfüllen: ausgehend von $(0\%; 0\%)$ verläuft sie zunächst entlang der gestrichelten PD-Gerade, d.h. die zuerst ausgeschlossenen Unternehmen weisen eine Ausfallwahrscheinlichkeit von 100% auf, anschließend verläuft sie linear durch $(X_0; Y_0)$ bis sie die 100%-Trefferquoten-Gerade schneidet (siehe Abbildung 9).

Beweis: Zu zeigen ist, daß das mittlere Teilstück der flächenmaximalen CAP-Kurve linear sein muß. Hierfür muß gezeigt werden, daß die Kurventeilstücke rechts und links von $(X_0; Y_0)$ linear sein müssen und den gleichen Anstieg haben:

- Für den Anstieg der CAP-Kurve in $(X_0; Y_0)$ a , muß folgende Bedingung erfüllt sein: $\frac{Y_0}{X_0} \geq a \geq \frac{1 - Y_0}{1 - X_0}$ - andernfalls könnten unter den gegebenen technischen und inhaltlichen Nebenbedingungen die Punkte $(0\%; 0\%)$ und $(100\%; 100\%)$ – ausgehend vom Punkt $(X_0; Y_0)$ – nicht „erreicht“ werden.
- Da die Tangente einer (schwach) konkaven Funktion stets oberhalb (oder auf) der Funktion verläuft, ist das rechte Teilstück AUC (=rechtsseitiges Integral der CAP-Kurve) - für ein gegebenes a - dann flächenmaximal, wenn es linear ist. Gleichermäßen läßt sich der linksseitige lineare Verlauf der flächenmaximalen CAP-Kurve begründen.
- Weiterhin müssen beide Geraden (rechts- und linksseitige Teilstücke der CAP-Kurve bei $(X_0; Y_0)$) den *gleichen Anstieg* aufweisen: gegeben ein linksseitiges Geradenteilstück der CAP-Kurve, ist das CAP-Kurvenintegral rechts von $(X_0; Y_0)$ um so größer, je steiler das rechte Geradenteilstück ist. Da das rechte Geradenteilstück nicht steiler als

das linke sein darf (Konkavitätsbedingung), ist das CAP-Integral, unter den gegebenen Nebenbedingungen, dann maximal, wenn das rechte und linke in $(X_0; Y_0)$ mündende Geradenteilstück der CAP-Kurve den gleichen Anstieg aufweisen, d.h. auf *einer* Geraden liegen.

Anstieg und Absolutglied dieser Gerade werden durch einen im folgenden vorgestellten Optimierungsansatz ermittelt. Hierfür wird die Fläche unter der flächenmaximalen CAP-Kurve in vier Komponenten untergliedert (siehe Flächen I..IV). Zusätzlich werden die Hilfsvariablen X_u , Y_u und X_r eingeführt (siehe Abbildung 22).

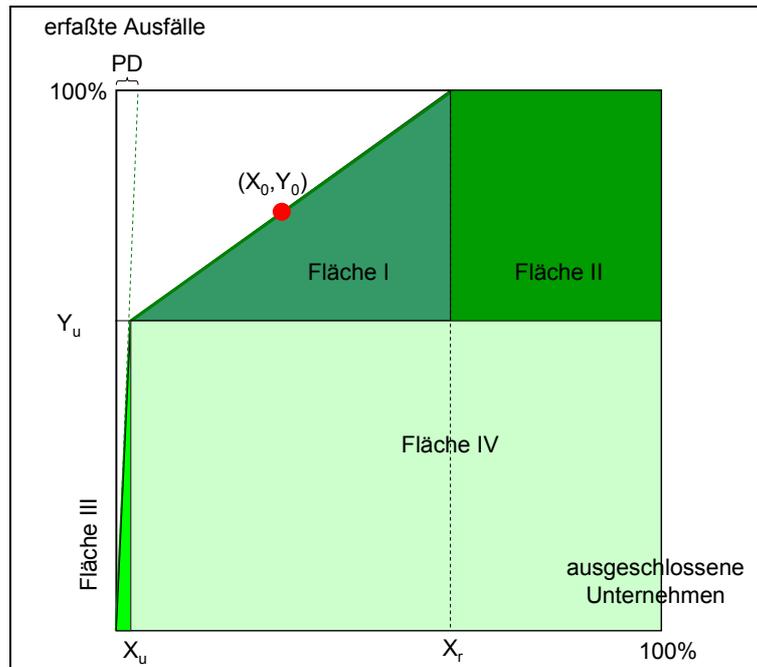


Abbildung 22: Flächenzerlegung für Ermittlung der flächenmaximalen AUC_{CAP}

- X_u ... X-Koordinate des Schnittpunkts der mittleren CAP-Teilstücks mit der PD-Gerade
- Y_u ... Y-Koordinate des Schnittpunkts der mittleren CAP-Teilstücks mit der PD-Gerade,
- X_r ... X-Koordinate des Schnittpunkts der mittleren CAP-Teilstücks mit der 100%-Trefferquotengerade
- a ... Anstieg des linearen CAP-Kurventeilstücks
- b ... Absolutglied des linearen CAP-Kurventeilstücks

Aus der Summation der Teilflächen I..IV folgt:

$$\text{F 73) } AUC_{CAP, \max} = (X_r - X_u) \cdot (1 - Y_u) / 2 + (1 - X_r) \cdot (1 - Y_u) + X_u \cdot Y_u / 2 + (1 - X_u) \cdot Y_u$$

$$\text{F 74) } AUC_{CAP, \max} = \frac{X_r}{2} - \frac{X_u}{2} - \frac{X_r \cdot Y_u}{2} + \frac{X_u \cdot Y_u}{2} + 1 - X_r - Y_u + X_r Y_u + \frac{X_u \cdot Y_u}{2} + Y_u - X_u \cdot Y_u,$$

$$\text{F 75) } AUC_{CAP, \max} = -\frac{X_r}{2} - \frac{X_u}{2} + \frac{X_r \cdot Y_u}{2} + 1$$

Weiterhin gilt (siehe auch Abbildung 22):

$$\text{F 76) } \frac{X_u}{Y_u} = \frac{PD}{100\%}$$

$$\text{F 77) } X_u = Y_u \cdot PD$$

Aus der Geradengleichung des mittleren CAP-Kurventeilstücks folgt:

$$\text{F 78) } aX_u + b = Y_u$$

$$\text{F 79) } aX_0 + b = Y_0$$

F 80) $aX_r + b = 1$ Durch Einsetzen von Formel F 77 in Formel F 78 ergibt sich:

$$\text{F 81) } a \cdot Y_u \cdot PD + b = Y_u$$

$$\text{F 82) } b = Y_u \cdot (1 - a \cdot PD)$$

$$\text{F 83) } Y_u = \frac{b}{1 - a \cdot PD} \quad \text{Aus Formel F 80 folgt:}$$

$$\text{F 84) } X_r = \frac{1 - b}{a} \quad \text{Durch Einsetzen von Formel F 83 in Formel F 78 ergibt sich}$$

$$\text{F 85) } X_u = \frac{b \cdot PD}{1 - a \cdot PD} \quad \text{Aus Formel F 79 folgt:}$$

$$\text{F 86) } a = \frac{Y_0 - b}{X_0} \quad \text{Aus Formel F 83 und Formel F 86 folgt:}$$

$$\text{F 87): } Y_u = \frac{b}{1 - \frac{Y_0 - b}{X_0} \cdot PD} = \frac{X_0 \cdot b}{X_0 - Y_0 \cdot PD + b \cdot PD} \quad \text{Aus Formel F 85 und Formel F 86 folgt:}$$

$$\text{F 88): } X_u = \frac{b \cdot PD}{1 - \frac{Y_0 - b}{X_0} \cdot PD} = \frac{X_0 \cdot b \cdot PD}{X_0 - Y_0 \cdot PD + b \cdot PD} \quad \text{Aus Formel F 84 und Formel F 86 folgt:}$$

$$\text{F 89) } X_r = \frac{1 - b}{\frac{Y_0 - b}{X_0}} = X_0 \cdot \frac{1 - b}{Y_0 - b} \quad \text{Durch Einsetzen von Formeln F 87, F 88 und F 89 in F 75}$$

ergibt sich (Eliminierung der Hilfsvariablen X_r , X_u und Y_u):

$$\text{F 90) } AUC = -\frac{X_0}{2} \cdot \frac{1 - b}{Y_0 - b} - \frac{X_0 \cdot b \cdot PD}{2 \cdot (X_0 - Y_0 \cdot PD + b \cdot PD)} + \frac{X_0}{2} \cdot \frac{1 - b}{Y_0 - b} \cdot \frac{X_0 \cdot b}{X_0 - Y_0 \cdot PD + b \cdot PD} + 1$$

$$\text{F 91) } AUC = \frac{(-X_0 + bX_0) \cdot (X_0 - Y_0 \cdot PD + b \cdot PD) - (X_0 \cdot b \cdot PD) \cdot (Y_0 - b) + bX_0^2 - b^2X_0^2}{2 \cdot (Y_0 - b)(X_0 - Y_0 \cdot PD + b \cdot PD)} + 1$$

$$\text{F 92) } AUC = \frac{b^2 (2 PD X_0 - X_0^2) + b (-PD X_0 + 2 X_0^2 - 2 PD X_0 Y_0) - X_0^2 + PD X_0 Y_0}{2 (Y_0 - b) (b PD + X_0 - PD Y_0)} + 1$$

Anschließend wird dieser Term nach b maximiert.³¹⁹

³¹⁹ Die Ableitungen und Nullstellenermittlungen wurden mit *Mathematica 5* durchgeführt.

$$\begin{aligned}
 \text{F 93) } \frac{\partial \text{AUC}}{\partial b} &= \frac{-\text{PD} X_0 + 2 X_0^2 + 2 b (2 \text{PD} X_0 - X_0^2) - 2 \text{PD} X_0 Y_0}{2 (-b + Y_0) (b \text{PD} + X_0 - \text{PD} Y_0)} \\
 &- \frac{\text{PD} (-X_0^2 + b^2 (2 \text{PD} X_0 - X_0^2) + \text{PD} X_0 Y_0 + b (-\text{PD} X_0 + 2 X_0^2 - 2 \text{PD} X_0 Y_0))}{2 (-b + Y_0) (b \text{PD} + X_0 - \text{PD} Y_0)^2} \\
 &+ \frac{-X_0^2 + b^2 (2 \text{PD} X_0 - X_0^2) + \text{PD} X_0 Y_0 + b (-\text{PD} X_0 + 2 X_0^2 - 2 \text{PD} X_0 Y_0)}{2 (-b + Y_0)^2 (b \text{PD} + X_0 - \text{PD} Y_0)}
 \end{aligned}$$

$$\text{F 94) } \frac{\partial \text{AUC}}{\partial b} = \frac{X_0 ((-1 + b) X_0 + \text{PD} (-b + Y_0)) (X_0 (1 + b - 2 Y_0) - \text{PD} (b - Y_0) (-1 + 2 Y_0))}{2 (X_0 + \text{PD} (b - Y_0))^2 (b - Y_0)^2}$$

$$\text{F 95) } \frac{\partial \text{AUC}}{\partial b} = 0$$

$$\text{F 96) } b_{\text{opt}} = \frac{X_0 - \text{PD} \cdot Y_0 - 2 \cdot Y_0 \cdot X_0 + 2 \cdot \text{PD} \cdot Y_0^2}{-\text{PD} - X_0 + 2 \cdot \text{PD} \cdot Y_0} \quad 320$$

$$\text{F 97) } \text{AUC}_{\text{CAP,max}} = \frac{\left(\frac{X_0 - \text{PD} Y_0 - 2 X_0 Y_0 + 2 \text{PD} Y_0^2}{-\text{PD} - X_0 + 2 \text{PD} Y_0} \right)^2 (2 \text{PD} x - x^2) + \frac{X_0 - \text{PD} Y_0 - 2 X_0 Y_0 + 2 \text{PD} Y_0^2}{-\text{PD} - X_0 + 2 \text{PD} Y_0} (-\text{PD} x + 2 x^2 - 2 \text{PD} x y) - x^2 + \text{PD} x y}{-2 \left(\frac{X_0 - \text{PD} Y_0 - 2 X_0 Y_0 + 2 \text{PD} Y_0^2}{-\text{PD} - X_0 + 2 \text{PD} Y_0} \right)^2 \text{PD} - 2 \frac{X_0 - \text{PD} Y_0 - 2 X_0 Y_0 + 2 \text{PD} Y_0^2}{-\text{PD} - X_0 + 2 \text{PD} Y_0} x + 4 b \text{PD} y + 2 x y - 2 \text{PD} y^2} + 1$$

Durch Vereinfachungen ergibt sich:

$$\text{F 98) } \text{AUC}_{\text{CAP,max}} = 1 - \frac{\text{PD}}{2} - 2X_0 + 2 \cdot X_0 \cdot Y_0 + 2\text{PD} \cdot Y_0 - 2\text{PD} \cdot Y_0^2$$

Aus den Formeln F 61 und F 98 folgt:

$$\text{F 99) } \text{AR} = \frac{2 - \text{PD} - 4X_0 + 4 \cdot X_0 \cdot Y_0 + 4 \cdot \text{PD} \cdot Y_0 - 4\text{PD} \cdot Y_0^2 - 1}{1 - \text{PD}}$$

$$\text{F 100) } \text{AR}_{\text{max}} = \frac{1 - \text{PD} - 4X_0 + 4 \cdot X_0 \cdot Y_0 + 4 \cdot \text{PD} \cdot Y_0 - 4\text{PD} \cdot Y_0^2}{1 - \text{PD}}$$

zur Information:

$$\text{F 101) } X_r = 2X_0 - 2 \cdot \text{PD} \cdot Y_0 + \text{PD}$$

$$\text{F 102) } X_u = 2 \cdot \text{PD} \cdot Y_0 - \text{PD}$$

$$\text{F 103) } Y_u = 2 \cdot Y_0 - 1$$

$$\text{F 104) } a = \frac{Y_0 - 1}{-\text{PD} - X_0 + 2 \cdot \text{PD} \cdot Y_0}$$

$$\text{F 105) } b = Y_0 - \frac{Y_0 - 1}{-\text{PD} - X_0 + 2 \cdot \text{PD} \cdot Y_0} \cdot X_0$$

³²⁰ Das Vorzeichen der zweiten Ableitung von AUC nach b ist negativ. Bei der gefundenen Lösung handelt es sich also um ein Maximum.

Sollen die Gütemaße $AUC_{CAP, \min}$ und AR_{\min} in Abhängigkeit der Fehler I. und II. Art dargestellt werden, ergibt sich aus den Formeln F 64, F 66 und F 98:

$$\text{F 106) } AUC_{CAP, \max} = 1 - \frac{PD}{2} - 2((1 - F_1) \cdot PD + F_2 \cdot (1 - PD)) + 2 \cdot ((1 - F_1) \cdot PD + F_2 \cdot (1 - PD)) \cdot (1 - F_1) + 2PD \cdot (1 - F_1) - 2PD \cdot (1 - F_1)^2$$

$$\text{F 107) } AUC_{CAP, \max} = 1 - \frac{PD}{2} - 2 \cdot PD + 2 \cdot PD \cdot F_1 - 2 \cdot F_2 - 2 \cdot PD \cdot F_2 + (-2 \cdot PD + 2 \cdot PD \cdot F_1 - 2 \cdot F_2 - 2 \cdot PD \cdot F_2) \cdot (-1 + F_1) + 2 \cdot PD - 2PD \cdot F_1 - 2PD \cdot (1 - 2F_1 + F_1^2)$$

$$\text{F 108) } AUC_{CAP, \max} = 1 - \frac{PD}{2} + -2 \cdot PD \cdot F_1 + 2 \cdot PD \cdot F_1^2 - 2 \cdot F_2 \cdot F_1 + 2 \cdot PD \cdot F_2 \cdot F_1 + 2 \cdot PD - 2 \cdot PD \cdot F_1 - 2 \cdot PD + 4 \cdot PD \cdot F_1 - 2 \cdot PD \cdot F_1^2$$

$$\text{F 109) } AUC_{CAP, \max} = 1 - \frac{PD}{2} - 2 \cdot F_2 \cdot F_1 + 2 \cdot PD \cdot F_2 \cdot F_1$$

$$\text{F 110) } \boxed{AUC_{CAP, \max} = 1 - 2 \cdot F_1 \cdot F_2 \cdot (1 - PD) - \frac{PD}{2}}$$

Aus den Formeln F 61 und F 110 folgt:

$$\text{F 111) } AR_{\max} = \frac{2 \cdot \left(1 - 2 \cdot F_1 \cdot F_2 \cdot (1 - PD) - \frac{PD}{2}\right) - 1}{1 - PD}$$

$$\text{F 112) } AR_{\max} = \frac{2 - 4 \cdot F_1 \cdot F_2 \cdot (1 - PD) - PD - 1}{1 - PD}$$

$$\text{F 113) } AR_{\max} = \frac{1 - PD - 4 \cdot F_1 \cdot F_2 \cdot (1 - PD)}{1 - PD}$$

$$\text{F 114) } \boxed{AR_{\max} = 1 - 4 \cdot F_1 \cdot F_2}$$

Aus den Formeln F 71 und F 114 folgt:

$$\text{F 115) } AUC_{ROC, \max} = \frac{1 - 4 \cdot F_1 \cdot F_2 + 1}{2}$$

$$\text{F 116) } \boxed{AUC_{ROC, \max} = 1 - 2 \cdot F_1 \cdot F_2}$$

Ermittlung heuristischer Schätzwerte für die Accuracy Ratio

Sollen aus *einer* Kombination von Fehlern I. und II. Art neben dem *Intervall* möglicher Accuracy-Ratio-Werte (AR-Ober- und Untergrenze) auch einwertige Schätzgrößen ermittelt werden, so bietet sich hierfür der *Mittelwert* aus Ober- und Untergrenze als ein *heuristisches Schätzgütemaß* an:

$$\text{F 117) } AR_{MW} \equiv \frac{AR_{\min} + AR_{\max}}{2}$$

Es handelt sich dabei um einen heuristischen Schätzer, über den sich zunächst nur sagen läßt, daß er stets Accuracy Ratio Werte innerhalb des zulässigen Bereichs liefert (für empirische Untersuchungen der Geeignetheit dieses und anderer Schätzer siehe die folgenden Seiten). Als weitere heuristische Maße zur Abschätzung der Accuracy-Ratios von Verfahrens, für die

lediglich *einzelne* Kombination aus Fehlern I. und II. Art bekannt sind, können die Schätzer AR_α und AR_β – und die aus dem Mittelwert von AR_α und AR_β gebildete Schätzgröße $AR_{\alpha\&\beta}$ – verwendet werden, denen die parametrischen Funktionen für die ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta(x)=1-(1-x)^{1/\beta}$ zugrunde liegen (siehe Anhang IV). Somit ergibt sich folgende zusammenfassende Darstellung:

	CAP-Koordinaten	Fehler I. und II. Art
Untergrenze	$AUC_{CAP,min} = \frac{1 - X_0 + Y_0}{2}$	$AUC_{ROC,min} = 1 - \frac{F_1 + F_2}{2}$
	$AR_{min} = \frac{Y_0 - X_0}{1 - PD}$	$AR_{min} = 1 - (F_1 + F_2)$
Obergrenze	$AUC_{CAP,max} = 1 - \frac{PD}{2} - 2X_0 + 2 \cdot X_0 \cdot Y_0 + 2PD \cdot Y_0 - 2PD \cdot Y_0^2$	$AUC_{ROC,max} = 1 - 2 \cdot F_1 \cdot F_2$
	$AR_{max} = \frac{1 - PD - 4X_0 + 4 \cdot X_0 \cdot Y_0 + 4 \cdot PD \cdot Y_0 - 4PD \cdot Y_0^2}{1 - PD}$	$AR_{max} = 1 - 4 \cdot F_1 \cdot F_2$
AR_{MW}	$AR_{MW} \equiv \frac{AR_{max} + AR_{min}}{2}$	$AR_{MW} \equiv \frac{AR_{max} + AR_{min}}{2}$
AR_α	$AR_\alpha = \frac{\ln(X_0 - Y_0 \cdot PD) - \ln Y_0 - \ln(1 - PD)}{\ln(X_0 - Y_0 \cdot PD) + \ln Y_0 - \ln(1 - PD)}$	$AR_\alpha = \frac{\ln F_2 - \ln(1 - F_1)}{\ln F_2 + \ln(1 - F_1)}$
AR_β	$AR_\beta = \frac{\ln(1 - Y_0) - \ln(1 - PD - X_0 + Y_0 \cdot PD) + \ln(1 - PD)}{\ln(1 - Y_0) + \ln(1 - PD - X_0 + Y_0 \cdot PD) - \ln(1 - PD)}$	$AR_\beta = \frac{\ln F_1 - \ln(1 - F_2)}{\ln F_1 + \ln(1 - F_2)}$
$AR_{\alpha\&\beta}$	$AR_{\alpha\&\beta} \equiv \frac{AR_\alpha + AR_\beta}{2}$	$AR_{\alpha\&\beta} \equiv \frac{AR_\alpha + AR_\beta}{2}$

Tabelle K: Formeln zur Ermittlung exakter Ober- und Untergrenzen und heuristischer Schätzer für die Accuracy Ratio basierend auf CAP-Koordinaten/ Fehler I. und II. Art

Die Eignung der verschiedenen heuristischen Schätzmaße soll im folgenden anhand von empirischen Daten für alle neun Verfahren, die in Abschnitt 3.5 vorgestellt wurden und für welche die für die Ermittlung von ROC-Kurven benötigten Daten verfügbar waren^{321,322}, überprüft werden (siehe Abbildung 23 bis Abbildung 31):

- In den linken Hälften der Abbildungen werden die empirischen ROC-Kurven mit schätzgüteäquivalenten ROC-Kurven der den AR_α - und AR_β -Schätzern zugrundeliegenden Verfahren verglichen. Diese Abbildungen zeigen somit, wie gut die empirischen ROC-Kurven durch kalibrierte ROC_α - bzw. ROC_β -Kurven angenähert werden können.
- In den rechten Hälften der Abbildungen werden für alle gegebenen Fehler-I-II-Kombinationen der empirischen ROC-Kurven die mit den verschiedenen Schätzverfahren ermittelten Accuracy-Ratio-Werte ermittelt. Dargestellt sind auch die tatsächlichen Accuracy Ratios, siehe die fetten, horizontalen Linien. Diese Abbildungen vermitteln somit einen ersten Eindruck über die Güte und Stabilität der verschiedenen Schätzverfahren.

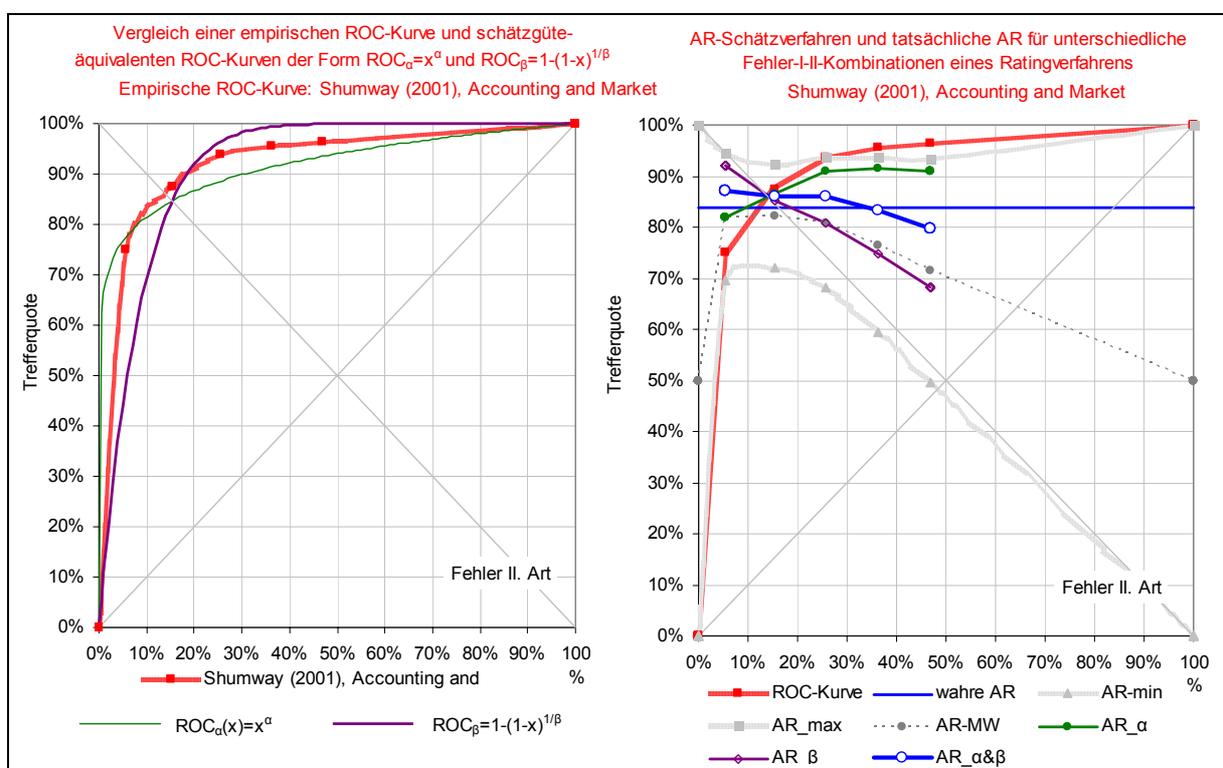


Abbildung 23: links: empirische ROC-Kurve SHUMWAY (2001), *Accounting and Market* und schätzgüteäquivalente ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} , AR_α - und AR_β -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_α - und AR_β -Schätzer ($AR_{\alpha\&\beta}$)

³²¹ Für die Ermittlung der ROC-Kurven wurden *alternativ* benötigt:

- (grafische) Darstellungen der ROC-Kurven,
- Anteile und Ausfallquoten von Unternehmen nach Ratingklassen,
- „ROC-Koordinaten“ (unterschiedliche Kombinationen von Fehlern I. und II. Art (absolut oder in Relation zur Anzahl der Ausfaller bzw. Nicht-Ausfaller)).

³²² Auch wenn die entsprechenden Daten verfügbar waren, erfolgt – wegen des geringen Stichprobenumfangs – keine ROC-Analyse der Ergebnisse von ALTMAN (1968) und BEAVER (1967).

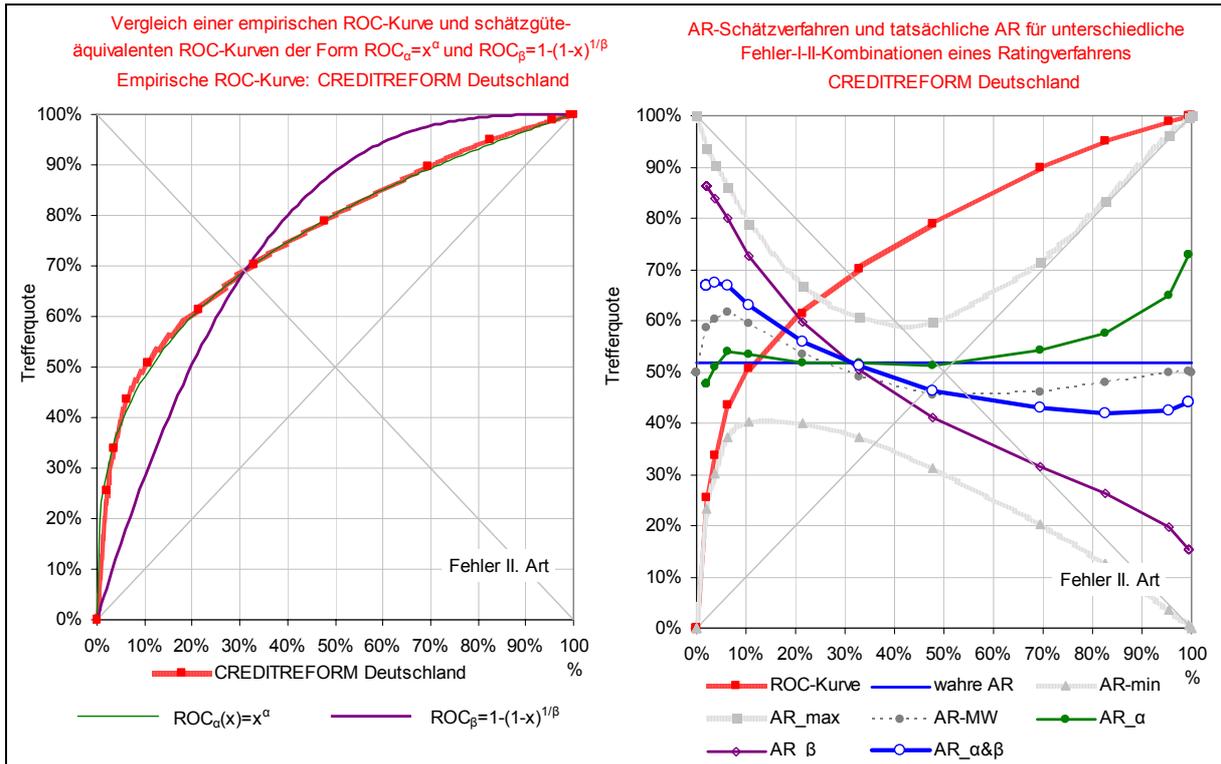


Abbildung 24: links: empirische ROC-Kurve CREDITREFORM-Bonitätsindex Deutschland (Daten aus LAWRENZ, SCHWAIGER (2002)) und schätzgüteäquivalente ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} -, AR_α - und AR_β -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_α - und AR_β -Schätzer ($AR_{\alpha\&\beta}$)

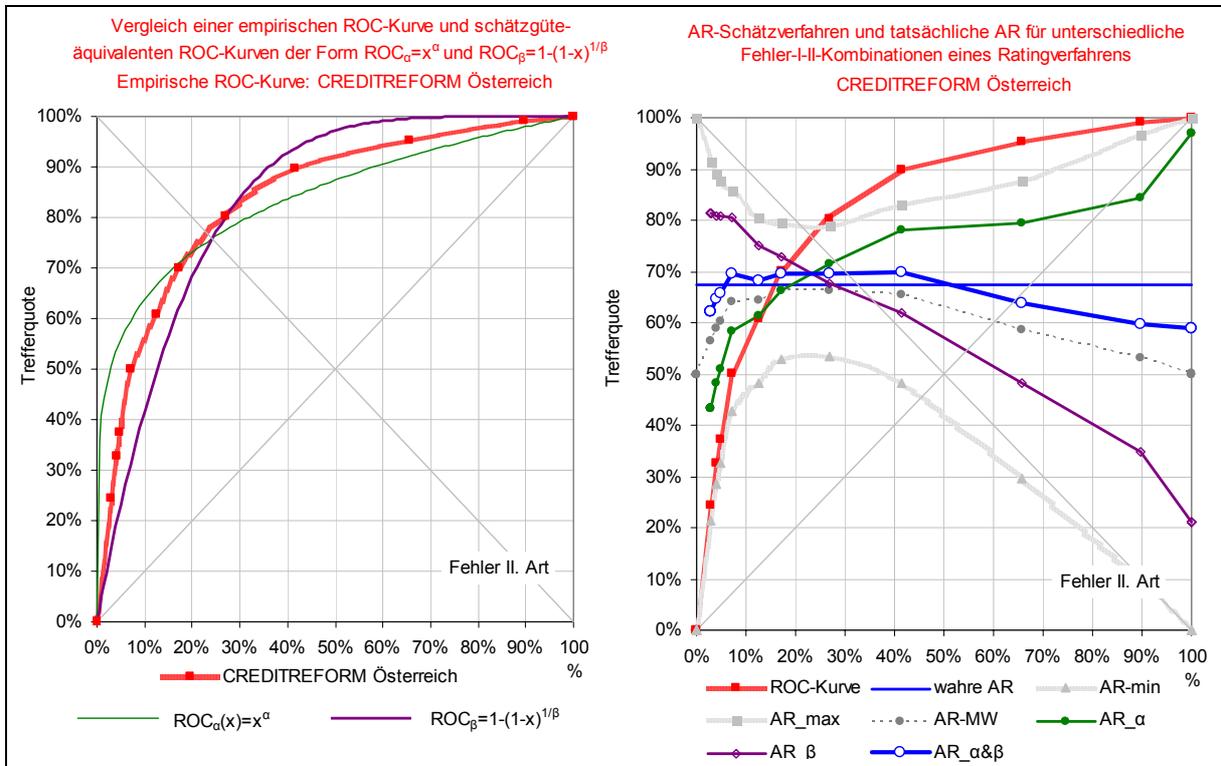


Abbildung 25: links: empirische ROC-Kurve CREDITREFORM-Bonitätsindex Österreich (Daten aus SCHWAIGER (2002)) und schätzgüteäquivalente ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} -, AR_α - und AR_β -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_α - und AR_β -Schätzer ($AR_{\alpha\&\beta}$)

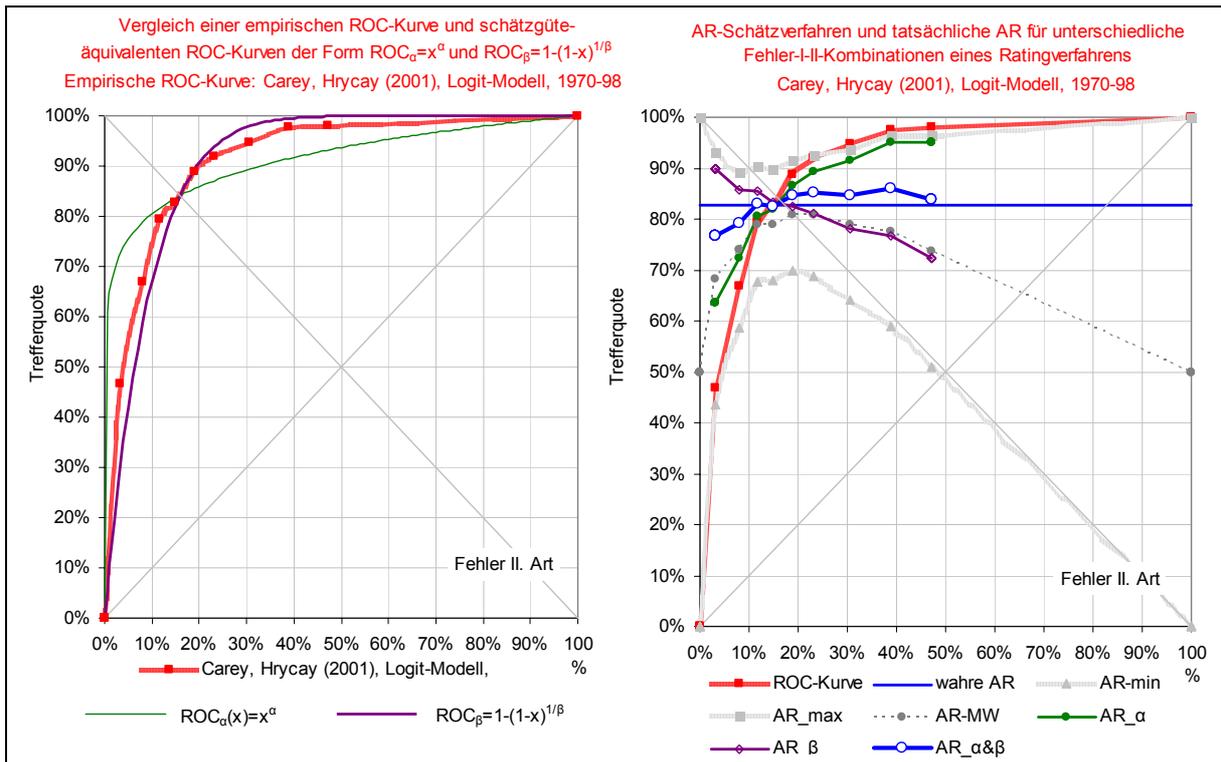


Abbildung 26: links: empirische ROC-Kurve CAREY, HRYCAY (1980), Logit-Modell 1970-1998 und schätzgüteäquivalente ROC-Kurven $ROC_{\alpha}(x)=x^{\alpha}$ und $ROC_{\beta}=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} , AR_{α} - und AR_{β} -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_{α} - und AR_{β} -Schätzer ($AR_{\alpha\&\beta}$)

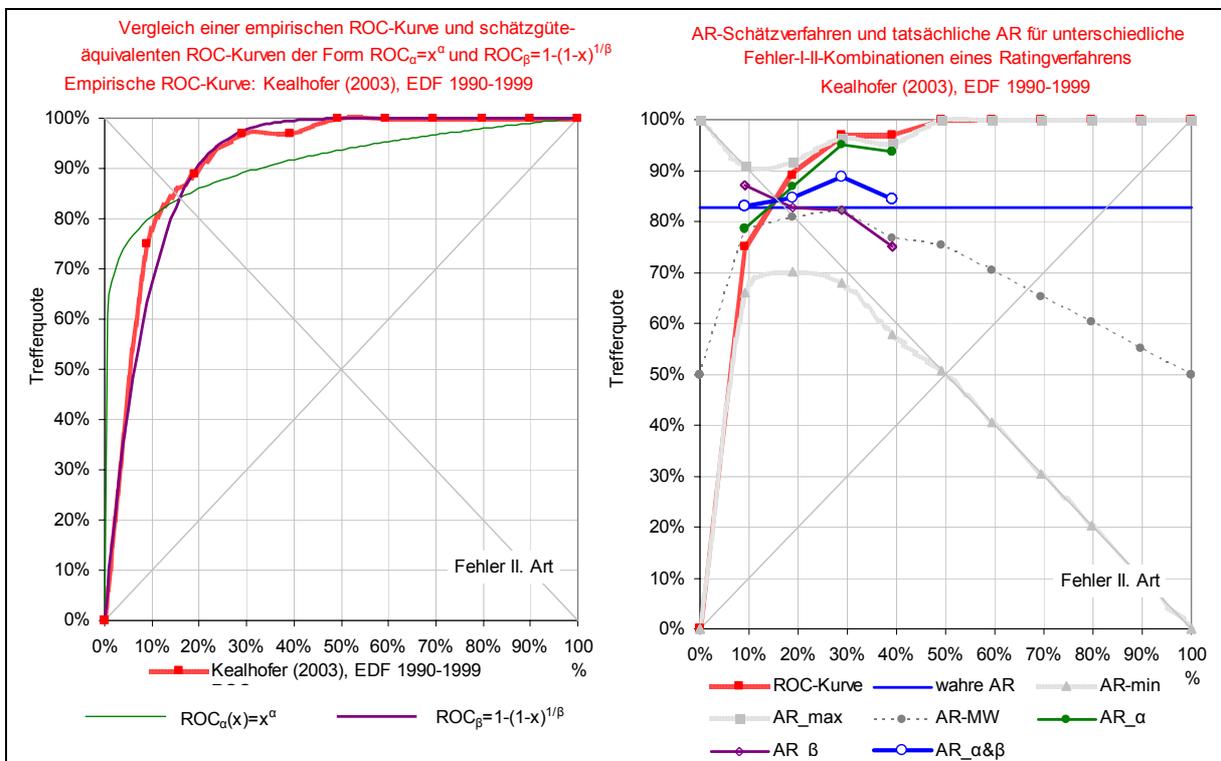


Abbildung 27: links: empirische ROC-Kurve KEALHOEFER (2003), EDF 1990-1999 und schätzgüteäquivalente ROC-Kurven $ROC_{\alpha}(x)=x^{\alpha}$ und $ROC_{\beta}=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} , AR_{α} - und AR_{β} -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_{α} - und AR_{β} -Schätzer ($AR_{\alpha\&\beta}$)

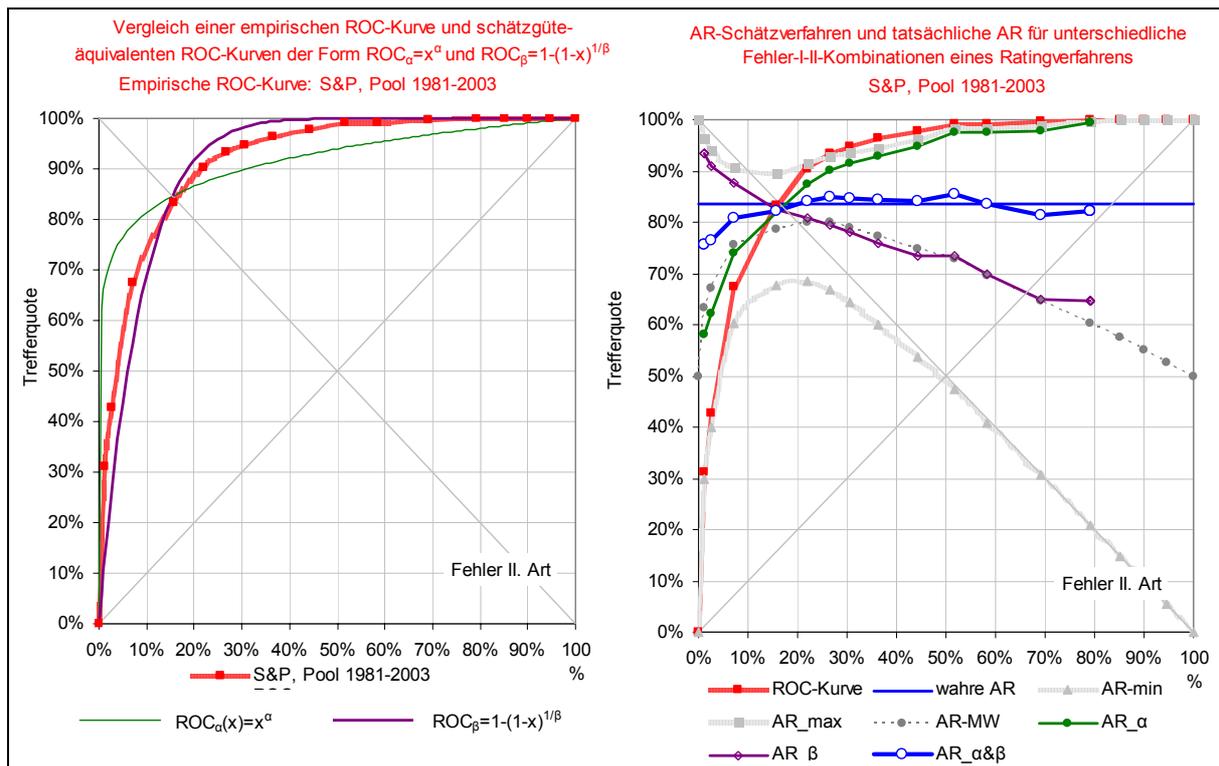


Abbildung 28: links: empirische ROC-Kurve S&P Ratings, Pool 1981-2003 (eigene Auswertungen) und schätzgüteäquivalente ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} , AR_α - und AR_β -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_α - und AR_β -Schätzer ($AR_{\alpha\&\beta}$)

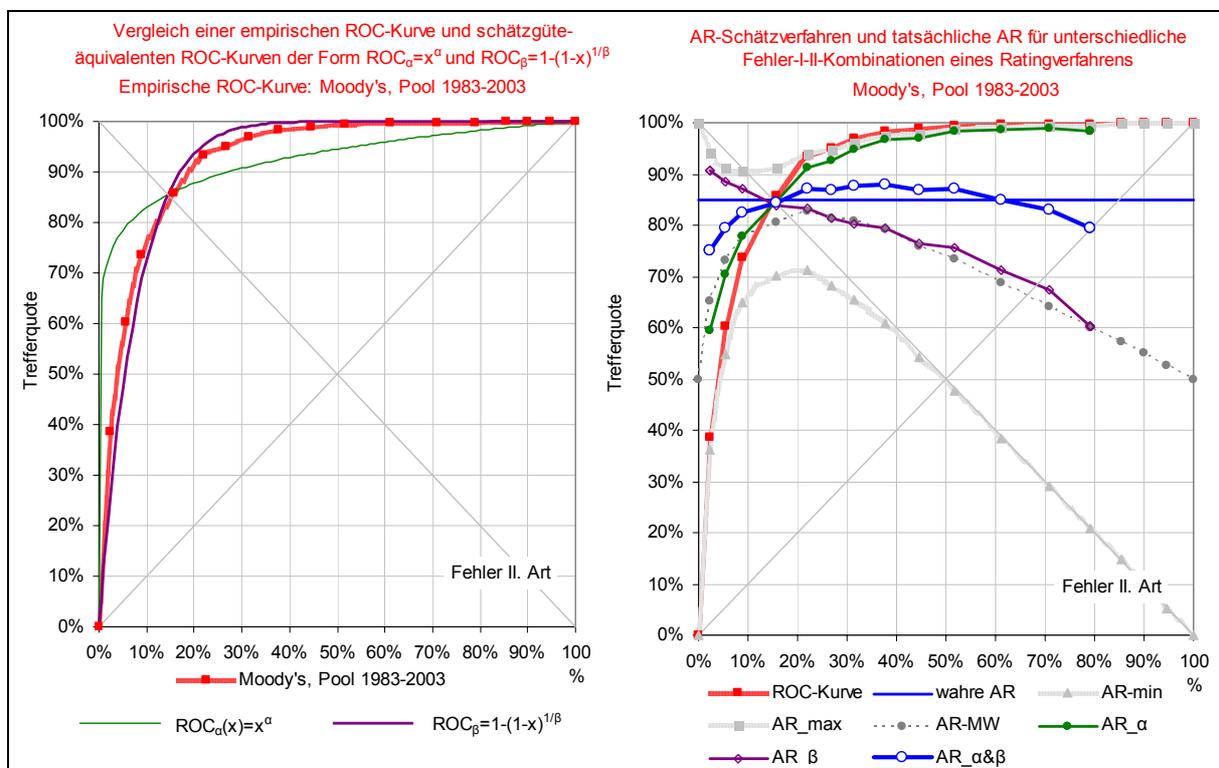


Abbildung 29: links: empirische ROC-Kurve MOODY'S Ratings, Pool 1983-2003 (eigene Auswertungen) und schätzgüteäquivalente ROC-Kurven $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} , AR_α - und AR_β -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_α - und AR_β -Schätzer ($AR_{\alpha\&\beta}$)

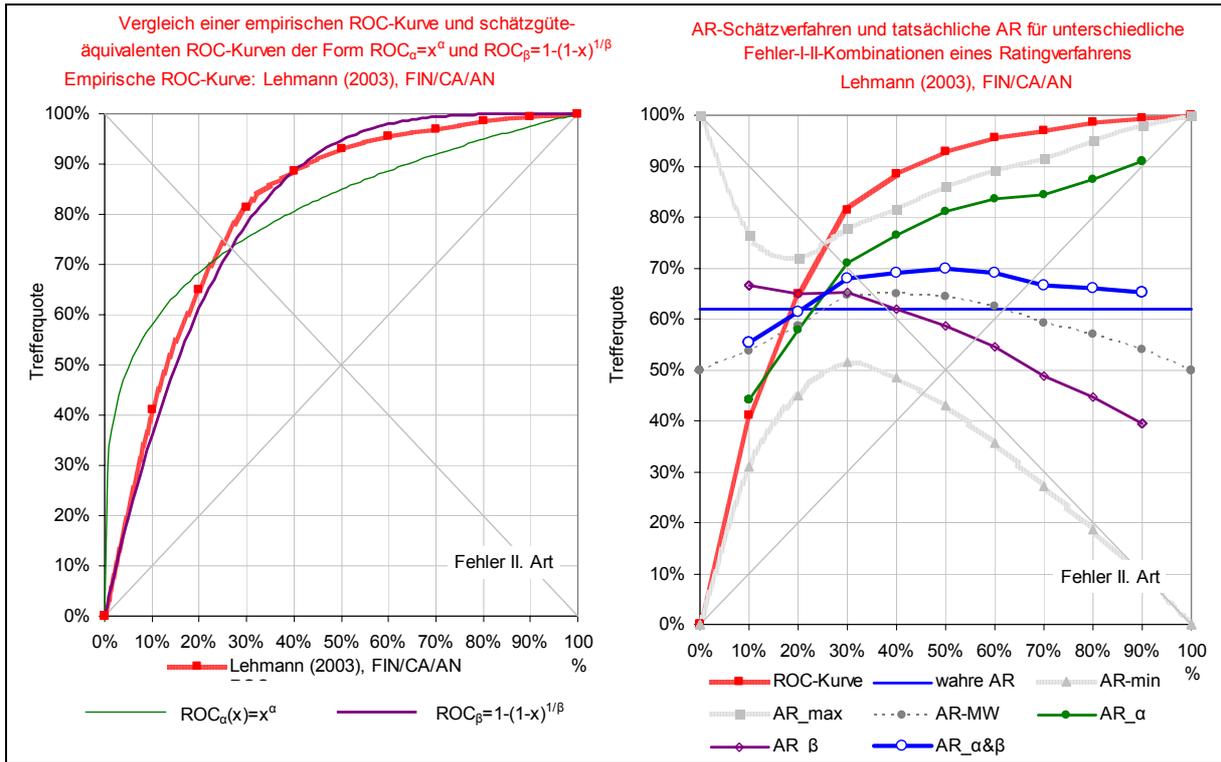


Abbildung 30: links: empirische ROC-Kurve LEHMANN (2003), FIN/CA/AN und schätzgüteäquivalente ROC-Kurven $ROC_{\alpha}(x)=x^{\alpha}$ und $ROC_{\beta}=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} -, AR_{α} - und AR_{β} -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_{α} - und AR_{β} -Schätzer ($AR_{\alpha\&\beta}$)

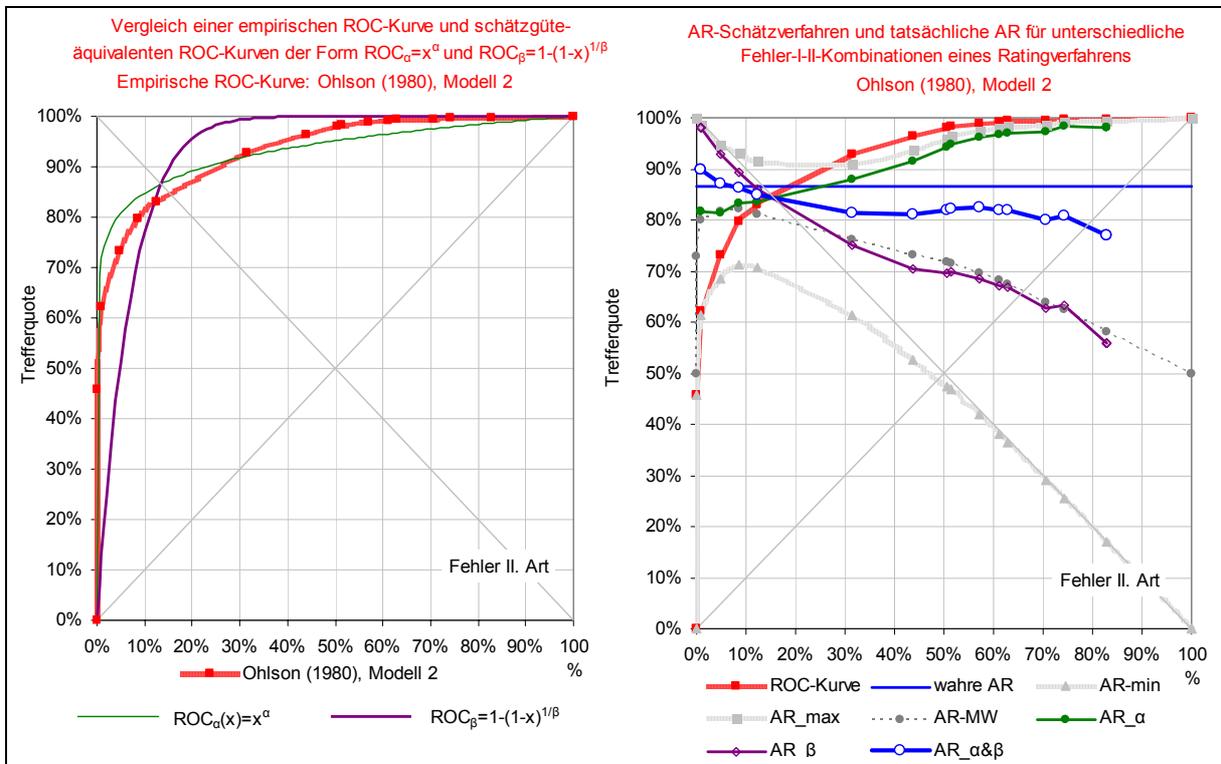


Abbildung 31: links: empirische ROC-Kurve OHLSON (1980), Modell 2 und schätzgüteäquivalente ROC-Kurven $ROC_{\alpha}(x)=x^{\alpha}$ und $ROC_{\beta}=1-(1-x)^{1/\beta}$; rechts: empirische ROC-Kurve und tatsächliche Accuracy Ratio, AR-Ober- und Untergrenzen, AR_{MW} -, AR_{α} - und AR_{β} -Schätzer für verschiedene Fehler-I-II-Kombinationen, Mittelwert aus AR_{α} - und AR_{β} -Schätzer ($AR_{\alpha\&\beta}$)

Die Darstellungen zeigen, daß:

- nur wenige empirische ROC-Kurven sich sehr gut durch die funktionale Form der ROC_α -Kurve, mit $ROC_\alpha(x)=x^\alpha$, approximieren lassen (eine bemerkenswerte Ausnahme stellt die ROC-Kurve des CREDITREFORM Deutschland Bonitätsindex in Abbildung 24 dar), alle anderen ROC_α -Kurven verlaufen jedoch sowohl für sehr kleine als auch für sehr große Fehler II. Art steiler als die entsprechenden empirischen ROC-Kurven, im mittleren Teil hingegen zu flach,
- einige der empirischen ROC-Kurven recht gut durch die funktionale Form der ROC_β -Kurven, mit $ROC_\beta=1-(1-x)^{1/\beta}$, beschrieben werden können (siehe Abbildung 26 bis Abbildung 30) – in allen diesen Fällen verlaufen die ROC_β -Kurven aber sowohl für kleine als auch große Fehler II. Art zu flach und für mittlere Fehler II. Art zu steil, also genau entgegengesetzt zu den Abweichungen der ROC_α -Kurven,
- in den restlichen Fällen (siehe Abbildung 23, Abbildung 25, Abbildung 31) die empirischen ROC-Kurven ungefähr in der Mitte des von der ROC_α - und der ROC_β -Kurve aufspannten Raums verlaufen.

Die Formen der empirischen ROC-Kurven sind somit von einer erheblichen Vielfalt geprägt, einige entsprechen eher dem Typus ROC_α , andere dem Typus ROC_β , andere wiederum einer Mischung beider Typen. Da Kurven des Typs ROC_α nur durch Spiegelung an der Nebendiagonalen in Kurven des Typs ROC_β überführt werden können, gibt es vermutlich keine *einparametrischen* Verfahren, die beide Kurventypen als Spezialfälle beinhalten und damit eine bessere Anpassung an die Fehler-I-II-Kombinationen der empirischen ROC-Kurven ermöglichen können.

Aus oben genannten Gründen weisen die auf einzelnen Fehler-I-II-Kombination basierenden heuristischen Schätzgütemaße AR_α und AR_β sowohl bei sehr großen als auch bei sehr kleinen Fehlern II. Art systematische Verzerrungen auf – jedoch mit umgekehrten Vorzeichen, so daß der beispielsweise der (ungewichtete) Mittelwert beider Größen, $AR_{\alpha\&\beta}$, ein besser geeigneter Schätzer ist.

ROC-Kurven des α -Typs sind durch eine hohe Diskriminanzfähigkeit bei „schlechten“ Unternehmen gekennzeichnet, d.h. sie sind in der Lage, Gruppen von Unternehmen mit sehr hohen Ausfallwahrscheinlichkeiten zu identifizieren. Im Bereich der „guten“ Unternehmen sind sie hingegen durch eine nur geringe Diskriminanzfähigkeit gekennzeichnet (siehe den vergleichsweise steilen, fast linearen Verlauf der ROC-Kurven für große Fehler II. Art). ROC-Kurven des β -Typs sind hingegen durch eine vergleichsweise geringe Diskriminanzfähigkeit bei schlechten Unternehmen (siehe den relativ flachen, fast linearen Verlauf der ROC-Kurve für kleine Fehler II. Art) aber eine sehr gute Diskriminanzfähigkeit bei guten Unternehmen gekennzeichnet (da sie in der Lage sind, relativ große Gruppe von Unternehmen mit sehr niedrigen Ausfallwahrscheinlichkeiten zu identifizieren).

Anmerkung: Stünden mehr als nur *eine* Fehler-I-II-Kombination für die Schätzung der Accuracy Ratio zur Verfügung, ließe sich ein zusätzlicher Parameter (neben α bzw. β) schätzen, der angibt, wie stark die ROC-Kurve entweder dem einem oder dem anderem Extrem (α - bzw. β -Typ) entspricht.

Hinsichtlich der Güte der einzelnen heuristischen Schätzer für die Accuracy-Ratio AR_{MW} , AR_α , AR_β und $AR_{\alpha\&\beta}$ zeigen die empirischen Untersuchungen auf Basis der neun zur Verfügung stehenden ROC-Kurven:

- AR_{MW} : Der aus dem Mittelwert der Accuracy-Ratio-Ober- und –Untergrenzen gebildete Wert konvergiert sowohl für sehr große als auch für sehr kleine Fehler II. Art gegen 50% (da AR_{max} in diesen Fällen gegen 100% und AR_{min} gegen 0% konvergieren) und ist damit, zumindest für alle Ratingverfahren mit einer Accuracy Ratio größer als 50%, tendenziell negativ verzerrt. Auch für Fehler-I-II-Kombinationen mit identischen Fehlern I. und II.

Art (siehe Schnittpunkte der von links oben nach rechts unten verlaufenden Diagonalen mit den ROC-Kurven) ist AR_{MW} tendenziell zu klein (hier: in acht von neun Fällen; allerdings verfehlt AR_{MW} unter diesen Umständen die tatsächlichen Accuracy Ratios nur um 1 bis maximal 5 Prozentpunkte).

- AR_{α} : In allen neun Fällen *unterschätzt* AR_{α} für kleine und *überschätzt* für große Fehler II. Art die tatsächliche Accuracy-Ratio. Im Bereich identischer Fehler I. und II. Art wird die tatsächliche Accuracy Ratio sehr präzise getroffen (nur in einem Fall beträgt die Abweichung mehr als +/- 3 PP).
- AR_{β} : Genau entgegengesetzt zu AR_{α} *überschätzt* AR_{β} für kleine und *unterschätzt* für große Fehler II. Art stets die tatsächliche Accuracy-Ratio. Im Bereich identischer Fehler I. und II. Art wird die tatsächliche Accuracy Ratio sehr präzise getroffen (die Abweichung beträgt nie mehr als +/- 3 PP).
- $AR_{\alpha\&\beta}$: Anders als bei AR_{MW} , AR_{α} und AR_{β} lässt sich beim Mittelwert aus AR_{α} und AR_{β} , $AR_{\alpha\&\beta}$, kein einheitliches Verzerrungsmuster erkennen. Für kleine Fehler II. Art überschätzt $AR_{\alpha\&\beta}$ die tatsächliche Accuracy Ratio in drei Fällen und unterschätzt sie in sechs Fällen. Für sehr große Fehler II. Art überschätzt $AR_{\alpha\&\beta}$ die tatsächliche Accuracy Ratio in einem Fall und unterschätzt sie in fünf Fällen³²³. Dafür trifft $AR_{\alpha\&\beta}$ in einem großen Bereich unterschiedlicher Fehler II. Art die tatsächliche Accuracy Ratio mit einer hohen Präzision.

Fazit: Für Fehler-I-II-Kombinationen mit identischen Fehlern I. und II. Art erzielen alle heuristischen Schätzer, AR_{MW} , AR_{α} , AR_{β} und $AR_{\alpha\&\beta}$, gute Prognosen für die tatsächlichen Accuracy-Ratios. Für asymmetrische Fehler-I-II-Kombinationen weisen die Werte von AR_{MW} , AR_{α} und AR_{β} jedoch systematische Verzerrungen auf. Mit $AR_{\alpha\&\beta}$ hingegen können in einem großen Bereich unterschiedlicher Fehler I. und II. Art die tatsächlichen Accuracy Ratios mit einer hohen Präzision prognostiziert werden. Gegenüber den anderen Schätzern weist $AR_{\alpha\&\beta}$ somit nur Vorteile und keine Nachteile auf und sollte deshalb als einziger Schätzer verwendet werden, wie dies in den Übersichten in den Abschnitten 3.3 bis 3.5 geschieht.

³²³ In den restlichen Fällen stehen keine Fehler-I-II-Kombinationen mit großen Fehlern II. Art zur Verfügung.

Anhang II: Anreizkompatibilität verschiedener Scores

Ziel: Es soll gezeigt werden, daß die Erwartungswerte der bedingten Informationsentropie (CIE) (logarithmische „Strafffunktion“) und des Brierscores (quadratische „Strafffunktion“) dann minimal sind, d.h. ihre bestmöglichen Ausprägungen aufweisen, wenn als prognostizierte Ausfallwahrscheinlichkeiten die erwarteten Ausfallwahrscheinlichkeiten angegeben werden. Ferner wird ein Beispiel für eine nicht anreizkompatible „Strafffunktion“ gegeben.

Entropie

$$\text{F 118) } \text{CIE} = -\frac{1}{n} \sum_{i=1}^n \ln(|\text{PD}_{i,\text{prog}} + \theta_i - 1|)$$

Anmerkung: nicht definiert ist CIE nur für die Fälle, in denen ein Ausfall eintritt, obwohl er mit Sicherheit ausgeschlossen wurde ($\Theta_i=1$ und $\text{PD}_{i,\text{prog}}=0$) oder in denen kein Ausfall eintritt, obwohl er mit Sicherheit prognostiziert wurde ($\Theta_i=0$ und $\text{PD}_{i,\text{prog}}=1$).

$$\text{F 119) } E(\text{CIE}) = -\frac{1}{n} \sum_{i=1}^n \text{PD}_{i,\text{tat}} \cdot \ln \text{PD}_{i,\text{prog}} + (1 - \text{PD}_{i,\text{tat}}) \cdot \ln(1 - \text{PD}_{i,\text{prog}}) \quad \text{für alle } i=1, \dots, n,$$

$$\text{F 120) } \frac{\partial E(\text{CIE})}{\partial \text{PD}_{i,\text{prog}}} = \frac{1}{n} \cdot \left(\frac{-\text{PD}_{i,\text{tat}}}{\text{PD}_{i,\text{prog}}} - \frac{(1 - \text{PD}_{i,\text{tat}})}{1 - \text{PD}_{i,\text{prog}}} \cdot (-1) \right)$$

$$\text{F 121) } \frac{\partial E(\text{CIE})}{\partial \text{PD}_{i,\text{prog}}} = \frac{1}{n} \cdot \frac{-\text{PD}_{i,\text{tat}} \cdot (1 - \text{PD}_{i,\text{prog}}) + (1 - \text{PD}_{i,\text{tat}}) \cdot \text{PD}_{i,\text{prog}}}{\text{PD}_{i,\text{prog}} \cdot (1 - \text{PD}_{i,\text{prog}})},$$

$$\text{F 122) } \frac{\partial E(\text{CIE})}{\partial \text{PD}_{i,\text{prog}}} = \frac{1}{n} \cdot \frac{\text{PD}_{i,\text{prog}} - \text{PD}_{i,\text{tat}}}{\text{PD}_{i,\text{prog}} \cdot (1 - \text{PD}_{i,\text{prog}})},$$

$$\text{F 123) } \frac{\partial E(\text{CIE})}{\partial \text{PD}_{i,\text{prog}}} = 0 \quad \text{für } \text{PD}_{i,\text{prog}} = \text{PD}_{i,\text{tat}}$$

$$\text{F 124) } \frac{\partial \left(\frac{\partial E(\text{CIE})}{\partial \text{PD}_{i,\text{prog}}} \right)}{\partial \text{PD}_{i,\text{prog}}} = \frac{1}{n} \cdot \frac{(\text{PD}_{i,\text{prog}} - \text{PD}_{i,\text{tat}})^2 + \text{PD}_{i,\text{tat}} \cdot (1 - \text{PD}_{i,\text{tat}})}{(\text{PD}_{i,\text{prog}} - 1)^2 \text{PD}_{i,\text{prog}}^2} > 0 \quad \text{für } \text{PD}_{i,\text{prog}} = \text{PD}_{i,\text{tat}},$$

d.h. $E(\text{CIE})$ wird dann minimiert, wenn $\text{PD}_{i,\text{prog}} = \text{PD}_{i,\text{tat}}$ für jedes i ³²⁴

Je kleiner der CIE-Wert, desto besser ist das zugehörige Verfahren. Der kleinste - und damit beste - erzielbare CIE-Wert ist Null und wird dann erreicht, wenn alle Ausfallprognosen entweder 0% oder 100% sind und stets korrekt sind. Auch für nicht perfekt trennfähige Ausfallwahrscheinlichkeiten, wird CIE minimiert, wenn die offenbarten (d.h. prognostizierten) Wahrscheinlichkeiten den wahren subjektiven Wahrscheinlichkeiten entsprechen.

³²⁴ Für die Randlösungen $\text{PD}_{i,\text{prog}}=0\%$ und $\text{PD}_{i,\text{prog}}=100\%$ ist CIE nicht definiert ist (außer für $\text{PD}_{i,\text{tat}}=0$ oder $\text{PD}_{i,\text{tat}}=1$), siehe Anmerkungen oben. Außer wenn $\text{PD}_{i,\text{tat}}=1/0$ ist, strebt CIE gegen unendlich für $\lim \text{PD}_{i,\text{prog}} \rightarrow 1/0$. Bei dem lokalen Optimum bei $\text{PD}_{i,\text{prog}} = \text{PD}_{i,\text{tat}}$ handelt es sich folglich auch um das globale Optimum.

Brierscore

$$\text{F 125) } BS = \frac{1}{n} \sum_{i=1}^n (PD_{i,\text{prog}} - \theta_i)^2$$

$$\text{F 126) } E(BS) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{prog}})^2 + (1 - PD_{i,\text{tat}}) \cdot PD_{i,\text{prog}}^2$$

$$\text{F 127) } E(BS) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} - 2PD_{i,\text{tat}} \cdot PD_{i,\text{prog}} + PD_{i,\text{tat}} \cdot PD_{i,\text{prog}}^2 + PD_{i,\text{prog}}^2 - PD_{i,\text{prog}}^2 \cdot PD_{i,\text{tat}}$$

$$\text{F 128) } E(BS) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} - 2PD_{i,\text{tat}} \cdot PD_{i,\text{prog}} + PD_{i,\text{prog}}^2$$

$$\text{F 129) } E(BS) = \frac{1}{n} \sum_{i=1}^n (PD_{i,\text{prog}} - PD_{i,\text{tat}})^2 + PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{tat}})$$

$$\text{F 130) } \frac{\partial E(BS)}{\partial PD_{i,\text{prog}}} = \frac{2}{n} \cdot (PD_{i,\text{prog}} - PD_{i,\text{tat}}) \quad \text{für alle } i=1, \dots, n$$

$$\text{F 131) } \frac{\partial E(BS)}{\partial PD_{i,\text{prog}}} = 0 \quad \text{für } PD_{i,\text{prog}} = PD_{i,\text{tat}}$$

$$\text{F 132) } E(BS_{\min}) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{tat}})^2 + (1 - PD_{i,\text{tat}}) \cdot PD_{i,\text{tat}}^2$$

$$\text{F 133) } E(BS_{\min}) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{tat}})$$

$$\text{F 134) } \frac{\partial \left(\frac{\partial E(BS)}{\partial PD_{i,\text{prog}}} \right)}{\partial PD_{i,\text{prog}}} = \frac{2}{n} > 0, \quad \text{d.h. } E(BS) \text{ wird minimiert für } PD_{i,\text{prog}} = PD_{i,\text{tat}} \text{ für alle } i$$

Untersuchung der Randlösungen ($PD_{i,\text{prog}}=0$ oder $PD_{i,\text{prog}}=1$ in F 126):

$$\text{F 135) } E(BS, PD_{i,\text{prog}} = 0) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} \geq E(BS_{\min})$$

$$\text{F 136) } E(BS, PD_{i,\text{prog}} = 1) = \frac{1}{n} \sum_{i=1}^n 1 - PD_{i,\text{tat}} \geq E(BS_{\min})$$

Analog CIE: Der Brierscore wird minimiert, für $PD_{i,\text{prog}} = PD_{i,\text{tat}}$ für alle i , d.h. der Brierscore ist *anreizkompatibel* in obigen Sinne.

Score auf Basis eine „Straffunktion“ mit absolutem Betrag als Beispiel für einen nicht anreizkompatiblen Score

$$\text{F 137) } A = \frac{1}{n} \sum_{i=1}^n |PD_{i,\text{prog}} - \theta_i|$$

$$\text{F 138) } E(A) = \frac{1}{n} \sum_{i=1}^n (1 - PD_{i,\text{tat}}) \cdot PD_{i,\text{prog}} + PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{prog}})$$

$$\text{F 139) } E(A) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{prog}} - PD_{i,\text{prog}} \cdot PD_{i,\text{tat}} + PD_{i,\text{tat}} - PD_{i,\text{tat}} \cdot PD_{i,\text{prog}}$$

$$\text{F 140) } E(A) = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{prog}} - 2 \cdot PD_{i,\text{prog}} \cdot PD_{i,\text{tat}} + PD_{i,\text{tat}}$$

$$\text{F 141) } \frac{\partial E(A)}{\partial PD_{i,\text{prog}}} = \frac{1}{n} (1 - 2 \cdot PD_{i,\text{tat}}),$$

$$\text{F 142) } \frac{\partial E(A)}{\partial PD_{i,\text{prog}}} > 0 \text{ für } PD_{i,\text{tat}} < 50\% \text{ und } \frac{\partial E(A)}{\partial PD_{i,\text{prog}}} < 0 \text{ für } PD_{i,\text{tat}} > 50\%$$

$$\text{F 143) } E(A)_{PD_{i,\text{prog}}=0 \wedge PD_{i,\text{tat}} < 50\%} = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}}$$

$$\text{F 144) } E(A)_{PD_{i,\text{prog}}=PD_{i,\text{tat}} \wedge PD_{i,\text{tat}} < 50\%} = \frac{1}{n} \sum_{i=1}^n (1 - PD_{i,\text{tat}}) \cdot PD_{i,\text{tat}} + PD_{i,\text{tat}} \cdot (1 - PD_{i,\text{tat}})$$

$$\text{F 145) } = \frac{1}{n} \sum_{i=1}^n PD_{i,\text{tat}} \cdot 2 \cdot (1 - PD_{i,\text{tat}})$$

$$\text{F 146) } 2 \cdot (1 - PD_{i,\text{tat}}) > 1 \text{ für } PD_{i,\text{tat}} < 50\%$$

Wenn die tatsächlich erwartete Ausfallwahrscheinlichkeit $PD_{i,\text{tat}}$ kleiner als 50% ist, dann ist der Scorewert um so besser (=niedriger), je niedriger die *prognostizierte* Wahrscheinlichkeit $PD_{i,\text{prog}}$ ist. Somit werden Anreize zur Abgabe möglichst niedriger anstelle erwartungstreuer Prognosen gesetzt (vgl. Formel F 143 und F 145).

Bei einer erwartungstreuen Angabe $PD_{i,\text{prog}} = PD_{i,\text{tat}}$ und kleinen erwarteten Ausfallwahrscheinlichkeiten $PD_{i,\text{tat}}$ fällt der Scorewert bei erwartungstreuen Ausfallwahrscheinlichkeiten fast doppelt so hoch (=schlechter) aus, wie bei Angabe von $PD_{i,\text{prog}} = 0\%$.

Anmerkung: Für $PD_{i,\text{tat}} > 50\%$ werden Prognosen $PD_{i,\text{prog}}=100\%$ „belohnt“.

Anhang III: Abhängigkeit verschiedener Schätzgütemaße von der durchschnittlichen Ausfallwahrscheinlichkeit

Ziel: Anhand eines Modells soll formal die Abhängigkeit verschiedener Schätzgütemaße, Brierscore, Skill-Brierscore, Entropie, Skill-Entropie, Accuracy Ratio, von der durchschnittlichen Ausfallwahrscheinlichkeit der Grundgesamtheit untersucht werden.

Per Annahme soll ein Referenzratingsystem unter allen *Umweltbedingungen* (=durchschnittlichen Ausfallraten) in der Lage sein, die zu beurteilenden Unternehmen in zwei Gruppen einteilen zu können, wobei die erste Gruppe nur aus Nicht-Ausfallern bestehen soll und stets einen Anteil $1-a$ an der Gesamtheit aller Unternehmen haben soll. Der Anteil der Unternehmen der zweiten Gruppe soll folglich *umweltunabhängig* a und die Ausfallwahrscheinlichkeit dieser Unternehmen (umweltabhängig) PD/a betragen. Ferner wird unterstellt, daß das Ratingverfahren stets richtig kalibriert ist.

Es werden die Schätzgütemaße dieses Ratingsystems unter verschiedenen *Umweltbedingungen* (im obigen Sinne) untersucht.³²⁵

Brierscore

Bei richtiger Kalibrierung beträgt der Brierscore (siehe Formel F 37):

$$\text{F 147) } BS = \sum_{i=1}^g a_i \cdot PD_i \cdot (1 - PD_i) \quad \text{hier mit } g=2, PD_1=0\%, a_1=1-a, PD_2=PD/a, a_2=a$$

$$\text{F 148) } BS = (1-a) \cdot 0 + a \cdot \frac{PD}{a} \cdot \left(1 - \frac{PD}{a}\right)$$

$$\text{F 149) } BS = PD \cdot \left(1 - \frac{PD}{a}\right)$$

$$\text{F 150) } \frac{\partial BS}{\partial PD} = 1 - \frac{2PD}{a}$$

$$\text{F 151) } \frac{\partial BS}{\partial PD} > 0 \quad \text{für } PD < \frac{a}{2}, \text{ d.h. } \frac{PD}{a} < 50\%$$

Interpretation: Solange die Ausfallwahrscheinlichkeit in der schlechteren der beiden Ratingklassen kleiner als 50% ist, führt jede Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit PD zu einer Erhöhung (= Verschlechterung) des Brierscores. Die Stärke dieser, prinzipiell unerwünschten Abhängigkeit des Brierscores von der durchschnittlichen Ausfallwahrscheinlichkeit, wird mit der folgenden Elastizitätsanalyse quantifiziert.

Die Elastizität des Brierscores bezüglich PD ist wie folgt definiert und gibt (näherungsweise) an, um wieviel Prozent der Brierscore steigt, wenn die durchschnittliche Ausfallwahrscheinlichkeit um ein Prozent (*nicht Prozentpunkt*) steigt:

³²⁵ *Umweltunabhängigkeit* eines Gütemaßes soll im Sinne von CANTOR, MANN (2003, S. 12) verstanden werden und dann vorliegen, wenn ein Gütemaß invariant gegenüber proportionalen Änderungen der ratingklassenspezifischen Ausfallwahrscheinlichkeiten ist (bei einer konstanten Verteilung der Unternehmen über die unterschiedlichen Ratingklassen), siehe hierzu auch DVFA (2004, S. 599). Zur Kritik an diesem Konzept siehe auch die entsprechenden Anmerkungen in Abschnitt 2.3.2. Für eine alternative Definition von Umweltabhängigkeit siehe die Ausführungen von SWETS (1988, S. 1286), siehe auch OENB (2004a, S. 117f.). Gemäß dieser Definition, wäre ein Gütemaß dann umweltunabhängig, wenn es unabhängig vom Ausfallanteil der zu beurteilenden Stichprobe ist, wobei sowohl Ausfaller als auch Nichtausfaller zufällig aus den jeweiligen Grundgesamtheiten gezogen werden.

$$\text{F 152) } \varepsilon_{\text{BS,PD}} = \frac{\left(\frac{\partial \text{BS}}{\text{BS}}\right)}{\left(\frac{\partial \text{PD}}{\text{PD}}\right)} = \frac{\left(\frac{\partial \text{BS}}{\partial \text{PD}}\right)}{\left(\frac{\text{BS}}{\text{PD}}\right)}$$

$$\text{F 153) } \varepsilon_{\text{BS,PD}} = \frac{1 - \frac{2\text{PD}}{a}}{\text{PD} \cdot \left(1 - \frac{\text{PD}}{a}\right)} \quad \text{durch Kürzen von PD im Nenner und Erweiterung mit a}$$

in Nenner und Zähler ergibt sich:

$$\text{F 154) } \varepsilon_{\text{BS,PD}} = \frac{\frac{a - 2\text{PD}}{a}}{\left(\frac{a - \text{PD}}{a}\right)}$$

$$\text{F 155) } \varepsilon_{\text{BS,PD}} = \frac{a - 2\text{PD}}{a - \text{PD}}$$

$$\text{F 156) } \lim_{\text{PD} \rightarrow 0} \varepsilon_{\text{BS,PD}} = \frac{a}{a} = 1$$

Interpretation: Für kleine PD ist $\varepsilon_{\text{BS,PD}} \approx 1$, d.h. für kleine PD steigt der zugehörige Brierscore fast linear mit steigender durchschnittlicher Ausfallwahrscheinlichkeit an. Somit existiert eine starke Abhängigkeit zwischen Brierscore und durchschnittlicher Ausfallwahrscheinlichkeit. Diese starke Abhängigkeit konnte auch empirisch bestätigt werden (siehe Abschnitt 2.4) und hat zur Folge, daß die Schätzgüte von Prognosen, die auf Stichproben mit unterschiedlichen durchschnittlichen Ausfallwahrscheinlichkeiten basieren nicht sinnvoll verglichen werden können.

Statt dessen wird in diesen Fällen die Verwendung von „Skillmaßen“ vorgeschlagen (siehe Abschnitt 2.4), welche die ermittelten Gütwerte – beispielsweise die des Brierscore - in Relation zur Güte eines *naiven Verfahrens* in der gleichen Umgebung betrachten und dadurch die Umweltabhängigkeit der Schätzgütemaße korrigieren sollen. Ein *naives Verfahren* ist ein Verfahren, das für alle zu prognostizierenden Ereignisse stets die gleiche, aber erwartungstreue, Ausfallwahrscheinlichkeit PD prognostiziert.

Skill-Brierscore

Bei richtiger Kalibrierung beträgt der Skill-Brierscore (siehe Formel F 34, F 35 und F 37):

$$\text{F 157) } \text{Skill}_{\text{BS}} = 1 - \frac{\text{BS}}{\text{PD} \cdot (1 - \text{PD})} \quad \text{aus Formel F 149 folgt:}$$

$$\text{F 158) } \text{Skill}_{\text{BS}} = 1 - \frac{\text{PD} \cdot \left(1 - \frac{\text{PD}}{a}\right)}{\text{PD} \cdot (1 - \text{PD})} \quad \text{durch Kürzen des Bruchs mit PD und Erweiterung des ersten Summanden mit (1-PD) folgt:}$$

$$\text{F 159) } \text{Skill}_{\text{BS}} = \frac{1 - \text{PD} - 1 + \frac{\text{PD}}{a}}{1 - \text{PD}}$$

$$\text{F 160) Skill}_{\text{BS}} = \frac{\text{PD} \cdot \left(\frac{1}{a} - 1\right)}{1 - \text{PD}}$$

$$\text{F 161) Skill}_{\text{BS}} = \frac{\frac{1}{a} - 1}{\frac{1}{\text{PD}} - 1}$$

$$\text{F 162) } \frac{\partial \text{Skill}_{\text{BS}}}{\partial \text{PD}} = \frac{\frac{1}{a} - 1}{-\left(\frac{1}{\text{PD}} - 1\right)^2 \cdot -\text{PD}^2} \quad \text{für } 0\% < \text{PD} < 100\% \text{ folgt}$$

$$\text{F 163) } \frac{\partial \text{Skill}_{\text{BS}}}{\partial \text{PD}} = \frac{\frac{1}{a} - 1}{(1 - \text{PD})^2} > 0$$

Interpretation: Jede Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit PD führt zu einer Erhöhung, d.h. *Verbesserung*, des Skill-Brierscores. Dieses Ergebnis ist überraschend, da eine Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit zu einer Erhöhung, d.h. *Verschlechterung*, des zugrundeliegenden Brierscores führt. Ferner ist dieses Ergebnis insofern unbefriedigend, als daß mit dem Skill-Brierscore offensichtlich kein umweltunabhängiges Schätzgütemaß gefunden wurde.

Die Stärke der Abhängigkeit des Skill-Brierscores von der durchschnittlichen Ausfallwahrscheinlichkeit, wird mit der folgenden Elastizitätsanalyse quantifiziert.

$$\text{F 164) } \varepsilon_{\text{SkillBS,PD}} = \frac{\left(\frac{\partial \text{Skill}_{\text{BS}}}{\text{SkS}_{\text{Brier}}}\right)}{\left(\frac{\partial \text{PD}}{\text{PD}}\right)} = \frac{\left(\frac{\partial \text{Skill}_{\text{BS}}}{\partial \text{PD}}\right)}{\left(\frac{\text{Skill}_{\text{BS}}}{\text{PD}}\right)}$$

$$\text{F 165) } \varepsilon_{\text{SkillBS,PD}} = \frac{\frac{\frac{1}{a} - 1}{(1 - \text{PD})^2}}{\frac{\frac{1}{a} - 1}{\text{PD}}} = \text{PD} \cdot \frac{\frac{1}{\text{PD}} - 1}{(1 - \text{PD})^2}$$

$$\text{F 166) } \varepsilon_{\text{SkillBS,PD}} = \frac{1}{1 - \text{PD}} > 0$$

$$\text{F 167) } \lim_{\text{PD} \rightarrow 0} \varepsilon_{\text{SkillBS,PD}} = 1$$

Interpretation: Der Skill-Brierscore ist um so besser, je größer PD ist. Für kleine PD ist $\varepsilon_{\text{Skill},\text{PD}} \approx 1$, d.h. für kleine PD steigt der zugehörige Skill-Brierscore (fast) linear mit steigender PD an.

Entropie

Bei richtiger Kalibrierung beträgt der Entropiewert (siehe Formel F 36):

$$\text{F 168) } \text{CIE}_{\text{kal}} = -\sum_{i=1}^g a_i (\text{PD}_i \cdot \ln \text{PD}_i + (1 - \text{PD}_i) \cdot \ln(1 - \text{PD}_i)) \quad \text{mit } g=2, \text{PD}_1=0\%, a_1=1-a,$$

$\text{PD}_2=\text{PD}/a, a_2=a$ (siehe Annahmen zu Beginn des Abschnitts) ergibt sich:

$$\text{F 169) } \text{CIE} = -\left((1-a) \cdot (0 + 1 \cdot \ln(1)) + a \cdot \left(\frac{\text{PD}}{a} \cdot \ln\left(\frac{\text{PD}}{a}\right) + \left(1 - \frac{\text{PD}}{a}\right) \cdot \ln\left(1 - \frac{\text{PD}}{a}\right) \right) \right)$$

$$\text{F 170) } \text{CIE} = -\left(\text{PD} \cdot \ln\left(\frac{\text{PD}}{a}\right) + (a - \text{PD}) \cdot \ln\left(\frac{a - \text{PD}}{a}\right) \right)$$

$$\text{F 171) } \frac{\partial \text{CIE}}{\partial \text{PD}} = -\left(\ln\left(\frac{\text{PD}}{a}\right) + \text{PD} \cdot \frac{a}{\text{PD}} \cdot \frac{1}{a} - \ln\left(\frac{a - \text{PD}}{a}\right) + (a - \text{PD}) \cdot \frac{a}{a - \text{PD}} \cdot \frac{1}{a} \cdot (-1) \right)$$

$$\text{F 172) } \frac{\partial \text{CIE}}{\partial \text{PD}} = -\left(\ln\left(\frac{\text{PD}}{a}\right) + 1 - \ln\left(\frac{a - \text{PD}}{a}\right) - 1 \right)$$

$$\text{F 173) } \frac{\partial \text{CIE}}{\partial \text{PD}} = \ln\left(\frac{a - \text{PD}}{\text{PD}}\right)$$

Für $\text{PD} < \frac{a}{2}$ (d.h. $\text{PD}/a < 50\%$) ist der Zähler innerhalb der Logarithmusfunktion größer

als der Nenner und damit $\frac{\partial \text{CIE}}{\partial \text{PD}} > 0$ und für $\text{PD} \geq \frac{a}{2}$ gilt $\frac{\partial \text{CIE}}{\partial \text{PD}} \leq 0$.

Interpretation: Solange die Ausfallwahrscheinlichkeit in der schlechteren der beiden Ratingklassen kleiner als 50% ist, führt – wie beim Brierscore – jede Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit zu einer Erhöhung, d.h. *Verschlechterung*, des Entropiewerts.

Die Elastizität des Entropiewerts bezüglich PD berechnet sich wie folgt:

$$\text{F 174) } \varepsilon_{\text{CIE,PD}} = \frac{\left(\frac{\partial \text{CIE}}{\text{CIE}}\right)}{\left(\frac{\partial \text{PD}}{\text{PD}}\right)} = \frac{\left(\frac{\partial \text{CIE}}{\partial \text{PD}}\right)}{\left(\frac{\text{CIE}}{\text{PD}}\right)}$$

$$\text{F 175) } \varepsilon_{\text{CIE,PD}} = \frac{\ln\left(\frac{a - \text{PD}}{\text{PD}}\right)}{-\left(\text{PD} \cdot \ln\left(\frac{\text{PD}}{a}\right) + (a - \text{PD}) \cdot \ln\left(\frac{a - \text{PD}}{a}\right)\right) \cdot \frac{1}{\text{PD}}}$$

$$\text{F 176) } \varepsilon_{\text{CIE,PD}} = -\frac{\ln\left(\frac{a - \text{PD}}{\text{PD}}\right)}{\ln\left(\frac{\text{PD}}{a}\right) + \frac{a - \text{PD}}{\text{PD}} \cdot \ln\left(\frac{a - \text{PD}}{a}\right)}$$

$$\text{F 177) } \varepsilon_{\text{CIE,PD}} = \frac{\ln\left(\frac{\text{PD}}{a - \text{PD}}\right)}{\ln\left(\frac{\text{PD}}{a}\right) + \frac{a - \text{PD}}{\text{PD}} \cdot \ln\left(\frac{a - \text{PD}}{a}\right)}$$

Es gilt $\lim_{\text{PD} \rightarrow 0} \varepsilon_{\text{CIE,PD}} = 1$.³²⁶

Interpretation: Für kleine PD ist $\varepsilon_{\text{CIE,PD}} \approx 1$, d.h. für kleine PD steigt der Entropiewert, ebenso wie der Brierscore, (fast) linear mit steigender PD an.

Skill-Entropie

Aus den Formeln F 29f. und F 177 folgt:

$$\text{F 178) } \text{CIER} = \frac{\text{CIE}_{\text{PD}} - \text{CIE}}{\text{CIE}_{\text{PD}}}$$

$$\text{F 179) } \text{CIER} = \frac{\text{PD} \cdot \ln(\text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD}) - \left(\text{PD} \cdot \ln\left(\frac{\text{PD}}{a}\right) + (a - \text{PD}) \cdot \ln\left(\frac{a - \text{PD}}{a}\right) \right)}{\text{PD} \cdot \ln(\text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD})}$$

$$\text{F 180) } \text{CIER} = \frac{\text{PD} \cdot \ln(\text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD}) - (\text{PD} \cdot \ln \text{PD} - \text{PD} \cdot \ln(a) + (a - \text{PD}) \cdot \ln(a - \text{PD}) - (a - \text{PD}) \cdot \ln(a))}{\text{PD} \cdot \ln(\text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD})}$$

$$\text{F 181) } \text{CIER} = \frac{(1 - \text{PD}) \cdot \ln(1 - \text{PD}) - (a - \text{PD}) \cdot \ln(a - \text{PD}) + a \cdot \ln(a)}{\text{PD} \cdot \ln(\text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD})}$$

mit Zähler_{CIER} = (1 - PD) · ln(1 - PD) - (a - PD) · ln(a - PD) + a · ln(a) und

Nenner_{CIER} = CIE_{PD} = PD · ln(PD) + (1 - PD) · ln(1 - PD) ergibt sich:

$$\text{F 182) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{\frac{\partial \text{Zähler}_{\text{CIER}}}{\partial \text{PD}} \cdot \text{Nenner}_{\text{CIER}} - \text{Zähler}_{\text{CIER}} \cdot \frac{\partial \text{Nenner}_{\text{CIER}}}{\partial \text{PD}}}{\text{Nenner}_{\text{CIER}}^2}$$

$$\text{F 183) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{1}{\text{CIE}_{\text{PD}}^2} \cdot \left(\begin{array}{l} -1 \cdot \ln(1 - \text{PD}) + (1 - \text{PD}) \cdot \frac{1}{1 - \text{PD}} \cdot (-1) \\ + \ln(a - \text{PD}) + (-a + \text{PD}) \cdot \frac{1}{a - \text{PD}} \cdot (-1) \end{array} \right) \cdot \text{Nenner}_{\text{CIER}} - \text{Zähler}_{\text{CIER}} \cdot \ln\left(\frac{\text{PD}}{1 - \text{PD}}\right)$$

$$\text{F 184) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{\ln(a - \text{PD}) \cdot \text{Nenner}_{\text{CIER}} - \ln(\text{PD}) \cdot \text{Zähler}_{\text{CIER}} + \ln(1 - \text{PD}) \cdot (\text{Zähler}_{\text{CIER}} - \text{Nenner}_{\text{CIER}})}{\text{CIE}_{\text{PD}}^2}$$

$$\text{F 185) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{1}{\text{CIE}_{\text{PD}}^2} \cdot \left(\begin{array}{l} \text{PD} \cdot \ln(\text{PD}) \cdot \ln(a - \text{PD}) + (1 - \text{PD}) \cdot \ln(1 - \text{PD}) \cdot \ln(a - \text{PD}) \\ - (1 - \text{PD}) \cdot \ln(1 - \text{PD}) \ln(\text{PD}) + (a - \text{PD}) \cdot \ln(a - \text{PD}) \ln(\text{PD}) - a \cdot \ln(a) \ln(\text{PD}) \\ - (a - \text{PD}) \cdot \ln(a - \text{PD}) \cdot \ln(1 - \text{PD}) + a \cdot \ln(a) \cdot \ln(1 - \text{PD}) - \text{PD} \cdot \ln(\text{PD}) \cdot \ln(1 - \text{PD}) \end{array} \right)$$

$$\text{F 186) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{1}{\text{CIE}_{\text{PD}}^2} \cdot \left(\begin{array}{l} + \ln(\text{PD}) \cdot \ln(1 - \text{PD}) \cdot (-1 + \text{PD} - \text{PD}) \\ + \ln(\text{PD}) \cdot \ln(a) \cdot (-a) \\ + \ln(\text{PD}) \cdot \ln(a - \text{PD}) \cdot (\text{PD} + a - \text{PD}) \\ + \ln(1 - \text{PD}) \cdot \ln(a - \text{PD}) \cdot (1 - \text{PD} - a + \text{PD}) \\ + \ln(a) \cdot \ln(1 - \text{PD}) \cdot a \end{array} \right)$$

³²⁶ Berechnet mit *Mathematica 5*.

$$\text{F 187) } \frac{\partial \text{CIER}}{\partial \text{PD}} = \frac{1}{\text{CIE}_{\text{PD}}^2} \cdot \left(\begin{aligned} & a \cdot \ln(\text{PD}) \cdot \ln(a - \text{PD}) + (1 - a) \cdot \ln(1 - \text{PD}) \cdot \ln(a - \text{PD}) + a \cdot \ln(a) \cdot \ln(1 - \text{PD}) \\ & - \ln(\text{PD}) \cdot \ln(1 - \text{PD}) - a \cdot \ln(\text{PD}) \cdot \ln(a) \end{aligned} \right)$$

Der erste Term innerhalb der Klammer ist betragsmäßig mindestens genauso groß ist wie die Summe der Beträge der beiden negativen Terme 4 und 5:

Beweis: zu zeigen ist, daß $a \cdot \ln(\text{PD}) \cdot \ln(a - \text{PD}) > \ln(\text{PD}) \cdot \ln(1 - \text{PD}) + a \cdot \ln(\text{PD}) \cdot \ln(a)$:

Gemäß der für reelle Exponenten verallgemeinerten BERNOULLISCHEN Ungleichung³²⁷ gilt:

$$\text{F 188) } (1 + x)^r < 1 + x \cdot r \quad \text{für } 0 < r < 1 \text{ und für } x > -1$$

Setzt man $r=a$ mit $0 < a < 1$ und $x = -\text{PD}/a$ mit $-\text{PD}/a > -1$ folgt:

$$\text{F 189) } \left(1 - \frac{\text{PD}}{a}\right)^a < 1 - a \cdot \frac{\text{PD}}{a} \quad \text{und somit}$$

$$\text{F 190) } \left(\frac{a - \text{PD}}{a}\right)^a < 1 - \text{PD} \quad \text{ferner folgt}$$

$$\text{F 191) } a \cdot (\ln(a - \text{PD}) - \ln(a)) < \ln(1 - \text{PD}) \quad \text{und durch Erweiterung mit } \ln(\text{PD}), \ln(\text{PD}) < 0$$

$$\text{F 192) } a \cdot \ln(\text{PD}) \cdot \ln(a - \text{PD}) - a \cdot \ln(\text{PD}) \cdot \ln(a) > \ln(1 - \text{PD}) \cdot \ln(\text{PD}) \quad \text{und somit}$$

$$\text{F 193) } a \cdot \ln(\text{PD}) \cdot \ln(a - \text{PD}) > \ln(\text{PD}) \cdot \ln(1 - \text{PD}) + a \cdot \ln(\text{PD}) \cdot \ln(a) \quad \text{was zu beweisen war.}$$

$$\text{F 194) } \frac{\partial \text{CIER}}{\partial \text{PD}} > 0$$

Ferner gelten:³²⁸

$$\text{F 195) } \lim_{\text{PD} \rightarrow 0} \text{CIER} = 0$$

$$\text{F 196) } \lim_{\text{PD} \rightarrow a} \text{CIER} = 1$$

$$\text{F 197) } \lim_{\text{PD} \rightarrow 0} \frac{\partial \text{CIER}}{\partial \text{PD}} = \infty$$

$$\text{F 198) } \lim_{\text{PD} \rightarrow a} \frac{\partial \text{CIER}}{\partial \text{PD}} = \infty$$

$$\text{F 199) } \lim_{\text{PD} \rightarrow 0} \varepsilon_{\text{CIER,PD}} = 0$$

$$\text{F 200) } \lim_{\text{PD} \rightarrow a} \varepsilon_{\text{CIER,PD}} = \infty$$

Interpretation: Analog zum Skill-Brierscore sind auch bei der Skill-Entropie (CIER) um so größere (=bessere) Werte zu verzeichnen, je größer die durchschnittliche Ausfallrate PD ist. Für kleine PD ist $\varepsilon_{\text{CIER,PD}} \approx 0$, d.h. zumindest für kleine PD ist die Skill-Entropie weniger sensitiv hinsichtlich der durchschnittlichen Ausfallrate als der Skill-Brierscore.

³²⁷ Siehe beispielsweise http://de.wikipedia.org/wiki/Bernoullische_Ungleichung. Der Hinweis für die Anwendbarkeit der Bernoullischen Ungleichung stammt von Jens Eisenschmidt.

³²⁸ Die in den Formeln F 197 bis F 200 dargestellten Gleichungen wurden mit *Mathematica 5* berechnet.

Accuracy Ratio

Für ein Ratingsystem mit zwei Klassen, bei dem in der „schlechtesten“ Klasse 100% der Ausfälle erfasst werden und a% der Unternehmen ausgeschlossen werden ist:

$$\text{F 201) } AR = \frac{a \cdot \frac{(0+1)}{2} + (1-a) \cdot 1 - \frac{1}{2}}{\frac{1}{2} - \frac{PD}{2}}$$

$$\text{F 202) } AR = \frac{a \cdot \frac{1}{2} + 1 - a - \frac{1}{2}}{\frac{1}{2} - \frac{PD}{2}}$$

$$\text{F 203) } AR = \frac{\frac{1}{2} - \frac{1}{2} \cdot a}{\frac{1}{2} - \frac{PD}{2}}$$

$$\text{F 204) } AR = \frac{1-a}{1-PD}$$

Anmerkung: Für $a = PD$ ergibt sich $AR=100\%$ (perfektes Rating), d.h. jedes Unternehmen in der schlechteren Ratingklasse fällt aus und alle Ausfaller befinden sich in der schlechteren Ratingklasse.

$$\text{F 205) } \frac{\partial AR}{\partial PD} = \frac{1-a}{(1-PD)^2} > 0$$

$$\text{F 206) } \varepsilon_{AR,PD} = \frac{\left(\frac{\partial AR}{AR}\right)}{\left(\frac{\partial PD}{PD}\right)} = \frac{\left(\frac{\partial AR}{\partial PD}\right)}{\left(\frac{AR}{PD}\right)} = \frac{\left(\frac{1-a}{(1-PD)^2}\right)}{\left(\frac{1-a}{1-PD}\right)} \cdot \frac{PD}{PD}$$

$$\text{F 207) } \varepsilon_{AR,PD} = \frac{PD}{1-PD}$$

$$\text{F 208) } \lim_{PD \rightarrow 0} \varepsilon_{AR,PD} = \frac{0}{1} = 0$$

Interpretation: Jede Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit PD führt zu einer Erhöhung (= Verbesserung) der Accuracy Ratio. Somit ist auch die Accuracy Ratio ein umweltabhängiges Schätzgütemaß – zumindest dann, wenn *Umweltabhängigkeit* in obigem Sinne definiert wird, nicht jedoch im Sinne von Abschnitt 2.3.2.

Auch wenn die Accuracy Ratio in obigem Sinne nicht völlig umweltunabhängig ist, ist sie zumindest weitgehend *insensitiv* gegenüber Umwelteinflüssen (Änderungen der durchschnittlichen Ausfallrate) ($\varepsilon_{AR,PD} \approx 0$ für kleine PD).

Anmerkung: Für die AR nach MOODY'S Definition, siehe Abschnitt 2.3.2, ergäbe sich anstelle der Formeln F 204 und F 207:

$$\text{F 209) } AR_{\text{Moody's}} = 1 - a$$

$$\text{F 210) } \varepsilon_{AR-\text{Moody's},PD} = 0$$

In Tabelle L und Tabelle M werden die *formalen* Ergebnisse dieses Abschnitts zusammengefaßt. In den darauf folgenden Darstellungen in Abbildung 32 bis Abbildung 36 werden numerische Beispiele in Form von Wertetabellen für die einzelnen Schätzgütemaße gegeben.

	Brierscore	Skill-Brierscore	Accuracy Ratio
Schätzgütemaß	$BS = PD \cdot \left(1 - \frac{PD}{a}\right)$	$Skill_{BS} = \frac{\frac{1}{a} - 1}{\frac{1}{PD} - 1}$	$AR = \frac{1 - a}{1 - PD}$
1. Ableitung nach PD	$\frac{\partial BS}{\partial PD} = 1 - \frac{2PD}{a}$ >0 für PD/a < 50%	$\frac{\partial Skill_{BS}}{\partial PD} = \frac{\frac{1}{a} - 1}{(1 - PD)^2} > 0$	$\frac{\partial AR}{\partial PD} = \frac{1 - a}{(1 - PD)^2} > 0$
Elastizität bzgl. PD	$\varepsilon_{BS,PD} = \frac{a - 2PD}{a - PD}$	$\varepsilon_{SkillBS,PD} = \frac{1}{1 - PD}$	$\varepsilon_{AR,PD} = \frac{PD}{1 - PD}$
Elastizität für kleine PD	$\varepsilon_{BS,PD} \approx 1$	$\varepsilon_{SkillBS,PD} \approx 1$	$\varepsilon_{AR,PD} \approx 0$

Tabelle L: verschiedene Schätzgütemaße, 1. Ableitungen und Sensitivitäten bezüglich PD für ein einfaches Ratingsystem mit 2 Klassen (Brierscore, Skill-Brierscore, Accuracy Ratio)

	Entropie (CIE)	Skill-Entropie (CIER)
Schätzgütemaß	$CIE = -\left(PD \cdot \ln\left(\frac{PD}{a}\right) + (a - PD) \cdot \ln\left(\frac{a - PD}{a}\right)\right)$	$CIER = \frac{(1 - PD) \cdot \ln(1 - PD) - (a - PD) \cdot \ln(a - PD) + a \cdot \ln(a)}{PD \cdot \ln(PD) + (1 - PD) \cdot \ln(1 - PD)}$
1. Ableitung nach PD	$\frac{\partial CIE}{\partial PD} = \ln\left(\frac{a - PD}{PD}\right)$ >0 für PD/a < 50%	$\frac{\partial CIER}{\partial PD} = \frac{1}{CIE_{PD}^2} \cdot \left(\begin{array}{l} a \cdot \ln(PD) \cdot \ln(a - PD) \\ + (1 - a) \cdot \ln(1 - PD) \cdot \ln(a - PD) \\ + a \cdot \ln(a) \cdot \ln(1 - PD) \\ - \ln(PD) \cdot \ln(1 - PD) \\ - a \cdot \ln(PD) \cdot \ln(a) \end{array} \right)$
Elastizität bzgl. PD	$\varepsilon_{CIE,PD} = \frac{\ln\left(\frac{PD}{a - PD}\right)}{\ln\left(\frac{PD}{a}\right) + \frac{a - PD}{PD} \cdot \ln\left(\frac{a - PD}{a}\right)}$	(..) ³²⁹
Elastizität für kleine PD	$\varepsilon_{CIE,PD} \approx 1$	$\varepsilon_{CIER,PD} \approx 0$

Tabelle M: verschiedene Schätzgütemaße, 1. Ableitungen und Sensitivitäten bezüglich PD für ein einfaches Ratingsystem mit 2 Klassen (Entropie, Skill-Entropie)

³²⁹ Die Elastizität von CIER bezüglich PD wurde mit *Mathematica 5* berechnet. Der entsprechende Term ist sehr umfangreich und unübersichtlich; auf eine Wiedergabe wurde deshalb verzichtet. Der Grenzwert dieses Terms für $\lim_{PD \rightarrow 0}$ wurde ebenfalls mit *Mathematica 5* berechnet und ist in der Tabelle dargestellt.

Brierscore		a [Anteil der zweiten Ratingklasse]				
		50%	25%	10%	5%	1%
PD - probability of default	0,50%	0,5%	0,5%	0,5%	0,5%	0,3%
	1,00%	1,0%	1,0%	0,9%	0,8%	0,0%
	1,50%	1,5%	1,4%	1,3%	1,1%	---
	2,00%	1,9%	1,8%	1,6%	1,2%	---
	2,50%	2,4%	2,3%	1,9%	1,3%	---
	3,00%	2,8%	2,6%	2,1%	1,2%	---
	3,50%	3,3%	3,0%	2,3%	1,1%	---
	4,00%	3,7%	3,4%	2,4%	0,8%	---
	4,50%	4,1%	3,7%	2,5%	0,5%	---
	5,00%	4,5%	4,0%	2,5%	0,0%	---
	10,00%	8,0%	6,0%	0,0%	---	---

Abbildung 32: Wertetabelle Brierscore in Abhängigkeit von a und PD

Skill-Brierscore		a [Anteil der zweiten Ratingklasse]				
		50%	25%	10%	5%	1%
PD - probability of default	0,50%	0,5%	1,5%	4,5%	9,5%	49,7%
	1,00%	1,0%	3,0%	9,1%	19,2%	100,0%
	1,50%	1,5%	4,6%	13,7%	28,9%	---
	2,00%	2,0%	6,1%	18,4%	38,8%	---
	2,50%	2,6%	7,7%	23,1%	48,7%	---
	3,00%	3,1%	9,3%	27,8%	58,8%	---
	3,50%	3,6%	10,9%	32,6%	68,9%	---
	4,00%	4,2%	12,5%	37,5%	79,2%	---
	4,50%	4,7%	14,1%	42,4%	89,5%	---
	5,00%	5,3%	15,8%	47,4%	100,0%	---
	10,00%	11,1%	33,3%	100,0%	---	---

Abbildung 33: Wertetabelle Skill-Brierscore in Abhängigkeit von a und PD

Accuracy Ratio		a [Anteil der zweiten Ratingklasse]				
		50%	25%	10%	5%	1%
PD - probability of default	0,50%	50,3%	75,4%	90,5%	95,5%	99,5%
	1,00%	50,5%	75,8%	90,9%	96,0%	100,0%
	1,50%	50,8%	76,1%	91,4%	96,4%	---
	2,00%	51,0%	76,5%	91,8%	96,9%	---
	2,50%	51,3%	76,9%	92,3%	97,4%	---
	3,00%	51,5%	77,3%	92,8%	97,9%	---
	3,50%	51,8%	77,7%	93,3%	98,4%	---
	4,00%	52,1%	78,1%	93,8%	99,0%	---
	4,50%	52,4%	78,5%	94,2%	99,5%	---
	5,00%	52,6%	78,9%	94,7%	100,0%	---
	10,00%	55,6%	83,3%	100,0%	---	---

Abbildung 34: Wertetabelle Accuracy Ratio in Abhängigkeit von a und PD

Entropie (CIE)		a [Anteil der zweiten Ratingklasse]				
		50%	25%	10%	5%	1%
PD - probability of default	0,50%	2,8%	2,5%	2,0%	1,6%	0,7%
	1,00%	4,9%	4,2%	3,3%	2,5%	0,0%
	1,50%	6,7%	5,7%	4,2%	3,1%	---
	2,00%	8,4%	7,0%	5,0%	3,4%	---
	2,50%	9,9%	8,1%	5,6%	3,5%	---
	3,00%	11,3%	9,2%	6,1%	3,4%	---
	3,50%	12,7%	10,1%	6,5%	3,1%	---
	4,00%	13,9%	11,0%	6,7%	2,5%	---
	4,50%	15,1%	11,8%	6,9%	1,6%	---
	5,00%	16,3%	12,5%	6,9%	0,0%	---
	10,00%	25,0%	16,8%	0,0%	---	---

Abbildung 35: Wertetabelle Entropie in Abhängigkeit von a und PD

Skill-Entropie (CIER)		a [Anteil der zweiten Ratingklasse]				
		50%	25%	10%	5%	1%
PD - probability of default	0,50%	11,0%	22,1%	36,9%	48,4%	78,0%
	1,00%	12,5%	25,0%	42,0%	55,3%	100,0%
	1,50%	13,5%	27,1%	45,7%	60,8%	---
	2,00%	14,3%	28,9%	49,0%	65,7%	---
	2,50%	15,1%	30,5%	51,9%	70,4%	---
	3,00%	15,8%	31,9%	54,7%	75,0%	---
	3,50%	16,4%	33,3%	57,3%	79,9%	---
	4,00%	17,0%	34,6%	59,9%	85,1%	---
	4,50%	17,6%	35,8%	62,5%	91,1%	---
	5,00%	18,1%	37,0%	65,1%	100,0%	---
	10,00%	23,0%	48,2%	100,0%	---	---

Abbildung 36: Wertetabelle Skill-Entropie in Abhängigkeit von a und PD

Im Rahmen der getroffenen Modellannahmen konnte formal gezeigt werden, daß eine Erhöhung der durchschnittlichen Ausfallwahrscheinlichkeit:

- zu starken Erhöhungen, d.h. *Verschlechterungen*, des Brierscores und der Entropie (CIE) führen, zu starken/ moderaten Erhöhungen, d.h. *Verbesserungen*, des Skill-Brierscores/ Skill-Entropie führen,
- (fast) keinen Einfluß auf die Accuracy Ratio hat (für kleine PD).

Aufgrund ihrer großen *Umweltabhängigkeit*, d.h. Sensitivität bezüglich der durchschnittlichen Ausfallrate, sind der Brierscore, der Skill-Brierscore und die Entropie (CIE) ungeeignete Gütemaße beim Vergleich der Schätzgüte von Verfahren, wenn diese auf Stichproben mit unterschiedlichen Ausfallwahrscheinlichkeiten beruhen. Nur bedingt geeignet ist die Skill-Entropie; die geringste Umweltabhängigkeit weist die Accuracy Ratio auf.

Anhang IV: Quantifizierung von Informationsverlusten durch Diskretisierung von Ratingskalen

Ziel: Es sollen die Informationsverluste quantifiziert werden, die aus der Diskretisierung kontinuierlicher Ratingscores in diskrete Ratingklassen resultieren.

Beschreibung des weiteren Vorgehens: Die Abschätzung der mit der Diskretisierung einer kontinuierlichen Ratingskala verbundenen Informationsverluste erfolgt mittels zweier Verfahren. Das erste Verfahren basiert auf einer Interpolation der individuellen Ausfallwahrscheinlichkeiten, die anstelle der ratingklassenspezifischen Ausfallwahrscheinlichkeiten für die Schätzgütemessung verwendet werden. Das zweite Verfahren modelliert eine ROC-Kurve durch eine parametrische, formal analysierbare Funktion und prüft für verschiedene Parameter, wieviel Prozent der Fläche unter der ROC-Kurve, AUC_{ROC} , mit einem Ratingsystem mit g Klassen erreicht werden können.

Es zeigt sich, daß beide Verfahren zu weitgehend identischen quantitativen und qualitativen Ergebnissen führen.

Verfahren I: Interpolation der Ausfallwahrscheinlichkeiten

Das Vorgehen gemäß dem ersten Verfahren ist wie folgt:

1. In einem ein Simulationsmodell wird ein Portfolio mit n Unternehmen erzeugt (hier $n=5.000$). Die Unternehmen werden g verschiedenen Ratingklassen (hier $g=17$ mit AAA=1, AA+=2, ..., CCC/C=17) gemäß einer vorgegebenen Häufigkeitsverteilung zugeordnet und nach Ratingklassen aufsteigend sortiert (für Details zur gewählten Parametrisierung siehe unten).
2. Jedes Unternehmen übernimmt die seiner Ratingklasse entsprechende *ratingklassenspezifische* Ausfallwahrscheinlichkeit.
3. Es wird die erwartete Accuracy Ratio dieses Portfolios ermittelt (siehe Formel F 22).
4. Um innerhalb einzelner Ratingklassen variierende Ausfallwahrscheinlichkeiten der Unternehmen zuzulassen, werden die individuellen Ausfallwahrscheinlichkeiten aus Punkt 2 *geglättet*: Bei der Glättung wird jedem Unternehmen i als individuelle Ausfallwahrscheinlichkeit ein gewichteter Mittelwert der (ungeglätteten) Ausfallwahrscheinlichkeiten seiner b „besseren“ und b „schlechteren“ Nachbarn zugewiesen. Aus technischen Gründen werden für Unternehmen am oberen/ unteren Rand zusätzliche „Nachbarn“ mit einer Ausfallwahrscheinlichkeit von PD_min/ PD_max erzeugt.³³⁰

³³⁰ Aus Plausibilitätsgründen sollte PD_min nicht größer sein als die durchschnittliche Ausfallrate der besten Ratingklasse. Für PD_min wurde ein Wert von 0% gewählt. Die genaue Spezifikation von PD_min ist praktisch ohne Bedeutung für die Modellresultate. Ein nennenswerter Einfluß auf die Modellergebnisse, wird jedoch von PD_max ausgeübt. Für PD_max werden im Rahmen der Sensitivitätsanalyse des Modells verschiedene plausible Werte untersucht und deren Wirkung auf die untersuchten Modellergebnisse analysiert. Diese Variable steuert die Heterogenität der Ausfallwahrscheinlichkeiten der Unternehmen in der schlechtesten Ratingstufe. Aus Plausibilitätsgründen sollte diese Variable nicht kleiner sein als die durchschnittliche Ausfallrate der „schlechtesten“ Ratingklasse.

Die Gewichtungsfaktoren nehmen linear mit zunehmendem Abstand der Nachbarn von Unternehmen i ab (Dreieckskernglättung³³¹ mit der Kernbreite $2 \cdot b - 1$):

$$\text{F 211)} \quad PD_{\text{geglättet},i} = \sum_{j=i-b}^{i+b} w_{i,j} \cdot PD_j \quad \text{mit}$$

$$\text{F 212)} \quad w_{i,j} = \frac{1 - \frac{|i-j|}{b}}{\sum_{k=i-b}^{i+b} \left(1 - \frac{|i-k|}{b}\right)}$$

$$W_j \dots \text{Gewichtungsfaktor mit } \sum_{j=i-b}^{i+b} w_{i,j} = 1$$

für $b=1$ ergibt sich beispielsweise $PD_{\text{geglättet},i} = PD_i$ oder für $b=3$ ergibt sich beispiels-

$$\text{weise } PD_{\text{geglättet},i} = \frac{1}{12} \cdot PD_{i-2} + \frac{2}{12} \cdot PD_{i-1} + \frac{3}{12} \cdot PD_i + \frac{2}{12} \cdot PD_{i+1} + \frac{1}{12} \cdot PD_{i+2}$$

5. Durch eine anschließende Kalibrierung mit einem ratingklassenspezifischen Faktor wird sichergestellt, daß die durchschnittlichen *geglätteten* Ausfallwahrscheinlichkeiten je Ratingklasse den durchschnittlichen ungeglätteten Ausfallwahrscheinlichkeiten je Ratingklasse (=ratingklassenspezifischen Ausfallwahrscheinlichkeiten) (siehe Punkt 2) entsprechen:

$$\text{F 213)} \quad PD_{\text{geglättet_kalibriert},i} = PD_{\text{geglättet},i} \cdot \frac{PD_{\text{ungeglättet,Ratingklasse}[i]}}{MW(PD_{\text{geglättet, Ratingklasse}[i]})}$$

6. Es wird die erwartete Accuracy Ratio dieses Portfolios ermittelt und mit dem unter Punkt 3 ermittelten Wert verglichen.

In den Schritten 1 und 2 werden alternativ die historischen Ausfallraten und Häufigkeitsverteilungen der Ratings auf die unterschiedlichen Ratingklassen gemäß S&P (2004) und MOODY'S (2004) für deren 17-stufige Ratingeinteilungen verwendet. Das Modell ließe sich aber auch auf beliebige andere Ratingverfahren, beispielsweise auf solche mit nur 7 Ratingstufen, anwenden.

Die Schritte 4 bis 6 werden für verschiedene Parametrisierungen des Glättungsverfahrens wiederholt und es wird geprüft, welchen Einfluß die letztlich nur willkürlich zu wählenden (aber inhaltlich gut interpretierbaren) Parameter PD_{max} und b des Glättungsverfahrens – innerhalb plausibler Bandbreiten - auf die erhaltenen Ergebnisse haben.

In Tabelle N und Tabelle O werden die Ergebnisse für die S&P- und MOODY'S-Daten wiedergegeben: dargestellt ist jeweils das Verhältnis der Accuracy-Ratio-Werte des Portfolios

³³¹ Die Qualität der Glättungen hängt kaum von der Art der Kernfunktion ab, sondern hauptsächlich vom verwendeten Glättungsparameter, siehe SCOTT (1992, S. 133). Im folgenden wird deshalb nur der Einfluß des Glättungsparameters auf die Modellergebnisse geprüft und nicht die Wahl der Kernfunktion, die aus Gründen der guten Nachvollziehbarkeit und einfachen Berechenbarkeit gewählt wurde.

Die Kernfunktion sollte zwar kontinuierlich, eingipflig und symmetrisch sein (siehe ebenda S. 138) – was von Dreieckskernen erfüllt wird, aber selbst größere Abweichungen von diesen Anforderungen führen nur zu geringfügig „ineffizienteren“ Glättungen, siehe auch SCOTT (1992, S. 140) für eine Darstellung von 8 üblichen Kernfunktionen und zugehörigen Effizienzmaßen, die nur geringfügig variieren. „Therefore the kernel can be chosen for other reasons (ease of computation, differentiability, ..) without undue concern for loss of efficiency.“, SCOTT (1992, S. 139).

mit Unternehmen die gemäß einer diskreten 17-stufigen Ratingskala bewertet wurden (siehe Schritt 3) und des Portfolios mit Unternehmen mit kontinuierlichen Ausfallwahrscheinlichkeiten (siehe Schritt 6).

Werte für $b < 100$ (was einem Glättungskern entspricht, der jeweils nur $\pm 2\%$ aller Unternehmen des Portfolios umfaßt, wobei zu berücksichtigen ist, daß nähere Nachbarn stärker gewichtet werden als weiter entfernte) führen tendenziell dazu, daß außer an den Übergängen zwischen zwei Ratingklassen praktisch keine Glättung der Ausfallwahrscheinlichkeiten erfolgt, da bei kleinen Kernbreiten fast immer nur Unternehmen der gleichen Ratingklasse in die Ermittlung der gewichteten Ausfallwahrscheinlichkeiten einbezogen werden.³³² Sehr große Werte für b , ab ca. $b=300$, führen – wegen der sich in Schritt 5 anschließenden Kalibrierung - zu unplausiblen Sprüngen der Ausfallwahrscheinlichkeiten zwischen Unternehmen verschiedener Ratingklassen. Für $\lim b \rightarrow \infty$ würde sich sogar die Ausgangssituation wieder einstellen, d.h. innerhalb einzelner Ratingklassen gäbe es keinerlei Differenzierung der Ausfallwahrscheinlichkeiten und zwischen Unternehmen benachbarten Ratingklassen käme es zu sprunghaften Veränderungen der Ausfallwahrscheinlichkeiten.

Aufgrund der anschließenden Kalibrierungen ist der Einfluß des Parameters PD_{max} bei gegebenen Werten für b nur gering. Im Fall der S&P-Daten sinkt die relative Güte des 17-stufigen Ratingverfahrens im Vergleich zu einem Ratingverfahren mit kontinuierlichen Ausfallwahrscheinlichkeiten bei einer dramatischen Veränderung von PD_{max} von 30% auf 100% nur marginal um 0,2% Prozentpunkte (PP).³³³ Innerhalb plausiblerer Grenzen für PD_{max} , 40%-75% (siehe die grau schattierten Bereiche in Tabelle N und Tabelle O) sogar nur um 0,1 PP. Im Fall der MOODY's Daten ist der Einfluß von PD_{max} mit 0,6 PP (0,4 PP) etwas größer, was daran liegt, daß die „schlechteste Ratingklasse“ gemäß MOODY's (Caa/C) fast doppelt so viele Unternehmen umfaßt wie die schlechteste S&P-Ratingklasse (CCC/C), so daß eine größere Heterogenität (die von PD_{max} gesteuert wird) der individuellen Ausfallwahrscheinlichkeiten in dieser Klasse, im Fall der MOODY's-Daten einen größeren Einfluß hat.

S&P-Daten		b (Anzahl der bei der Glättung der Ausfallwahrscheinlichkeiten einzubeziehenden besseren und schlechteren Nachbarn („Kernbreite“))					
		50	100	200	300	400	500
max. PD (unkalibr.)	30,0%	99,8%	99,7%	99,5%	99,4%	99,4%	99,4%
	40,0%	99,8%	99,6%	99,5%	99,4%	99,4%	99,4%
	50,0%	99,8%	99,6%	99,4%	99,4%	99,4%	99,4%
	75,0%	99,7%	99,5%	99,4%	99,3%	99,3%	99,3%
	100,0%	99,7%	99,5%	99,3%	99,3%	99,3%	99,3%

Tabelle N: relative Schätzgüte diskrete vs. kontinuierliche Ausfallwahrscheinlichkeiten für verschiedene Glättungsparameter (maximale PD und Anzahl der einzubeziehenden Nachbarn für Glättung der Ausfallwahrscheinlichkeiten), S&P (2004)-Daten (historische Ausfallhäufigkeiten und aktuelle Verteilung der Unternehmen über 17 Ratingstufen), fette Markierung: kleinster Wert; graue Hinterlegung: plausible Wertebereiche

³³² Bei 17 Ratingklassen umfaßt jede Ratingklasse durchschnittlich rund 6% aller Unternehmen.

³³³ Je nach eingestellter Kernbreite, werden im wesentlichen nur die Unternehmen der schlechtesten Ratingklasse (CCC/C) von diesem Parameter betroffen und durch die anschließende Kalibrierung (Schritt 5) wird der tatsächliche Einfluß dieses Parameters automatisch begrenzt. Sehr niedrige Werte dieses Parameters, $PD_{max} < PD_{CCC/C}$, führen zu einem unplausiblen umgekehrt-U-förmigen Verlauf der Ausfallwahrscheinlichkeiten innerhalb der Gruppe der schlechtesten Unternehmen.

MOODY'S-Daten		b (Anzahl der bei der Glättung der Ausfallwahrscheinlichkeiten einzubeziehenden besseren und schlechteren Nachbarn („Kernbreite“))					
		50	100	200	300	400	500
max. PD (unkalibr.)	30,0%	99,8%	99,7%	99,5%	99,5%	99,5%	99,5%
	40,0%	99,8%	99,6%	99,4%	99,4%	99,4%	99,4%
	50,0%	99,7%	99,5%	99,3%	99,2%	99,3%	99,3%
	75,0%	99,6%	99,4%	99,1%	99,0%	99,1%	99,2%
	100,0%	99,5%	99,2%	98,9%	98,9%	99,0%	99,1%

Tabelle O: relative Schätzgüte diskrete vs. kontinuierliche Ausfallwahrscheinlichkeiten für verschiedene Glättungsparameter (maximale PD und Anzahl der einzubeziehenden Nachbarn für Glättung der Ausfallwahrscheinlichkeiten), Moody's (2004)-Daten (historische Ausfallhäufigkeiten und aktuelle Verteilung der Unternehmen über 17 Ratingstufen), fette Markierung: kleinster Wert; graue Hinterlegung: plausibler Wertebereich

Die folgenden Auswertungen basieren auf der eher konservativen Kombination PD_{max}=75% und b=150, welche die Informationsverluste, die durch eine Diskretisierung kontinuierlicher Ratingskalen entsteht eher über- als unterschätzt.

Die aus der Anwendung dieser Parameter resultierenden Verteilungen der geglätteten und kalibrierten Ausfallwahrscheinlichkeiten gemäß den S&P- und MOODY'S-Daten werden in Abbildung 37 und Abbildung 38 dargestellt. In den darauf folgenden Darstellungen werden die aus den geglätteten und kalibrierten Ausfallwahrscheinlichkeiten resultierenden CAP-Kurven und die CAP-Kurven der „diskreten Portfolien“ dargestellt. Ferner wird untersucht, welcher Anteil des erreichten Präzisionszuwachses der „kontinuierlichen Portfolien“ im Vergleich zu den „diskreten Portfolien“ auf die Glättung der 17 einzelnen Ratingklassen zurückzuführen ist. Zur Ermittlung dieser Größe werden - abweichend vom sonstigen Vorgehen - in Schritt 5 die Ausfallwahrscheinlichkeiten aller Unternehmen der jeweils untersuchten Ratingklasse auf die ratingklassenspezifischen Ausfallwahrscheinlichkeiten (siehe Schritt 2) zurückgesetzt und die resultierenden Accuracy Ratios mit den Accuracy-Ratio-Werten der „diskreten Portfolien“ und der „kontinuierlichen Portfolien“ verglichen.

Anmerkung zu Abbildung 39 und Abbildung 40: Die deutlichen Unterschiede zwischen den S&P- und MOODY'S-Daten bezüglich des Anteils der Informationsverluste, die der schlechtesten Ratingstufe CCC/C bzw. Caa/C zuzurechnen sind, liegt vor allem in den deutlich unterschiedlichen Klassengrößen begründet (S&P: ca. 3%, MOODY'S: ca. 6% aller Unternehmen; vgl. auch Abbildung 37 und Abbildung 38). Je breiter eine Ratingklasse ist, d.h. je mehr Unternehmen sie umfaßt, desto größer sind die Informationsverluste die entstehen, wenn alle ratingklasseninternen Unterschiede in den Ausfallraten ignoriert werden und ratingklassenspezifische statt individueller Ausfallwahrscheinlichkeiten unterstellt werden.

Fazit Verfahren I:

Durch die Verwendung diskreter 17-stufiger Ratingskalen statt kontinuierlicher Ratingskalen entstehen Informationsverluste in einer Größenordnung von 0,3%-0,7% (S&P), bzw. 0,4%-1,0% (MOODY'S) (vgl. die schattierten Bereiche in Tabelle N und Tabelle O) d.h. im Durchschnitt 0,6% [falls ein einwertiges Ergebnis gewünscht wird].

Die Untersuchungen zeigen ferner, daß Informationsverluste hauptsächlich in der schlechtesten Ratingklasse auftreten, d.h. eine nennenswerte Verringerung (der ohnehin schon geringen) Informationsverluste kann im wesentlichen nur durch eine stärkere Differenzierung innerhalb der Klassen CCC/C bzw. Caa/C erreicht werden. Eine starke Differenzierung im Bereich der einzelnen Investmentgrade-Klassen (AAA – BBB-, bzw. Aaa – Baa3), beispielsweise durch die Angabe individueller Ausfallwahrscheinlichkeiten, hat hingegen praktisch überhaupt keinen Einfluß auf die gemessene Schätzgüte.

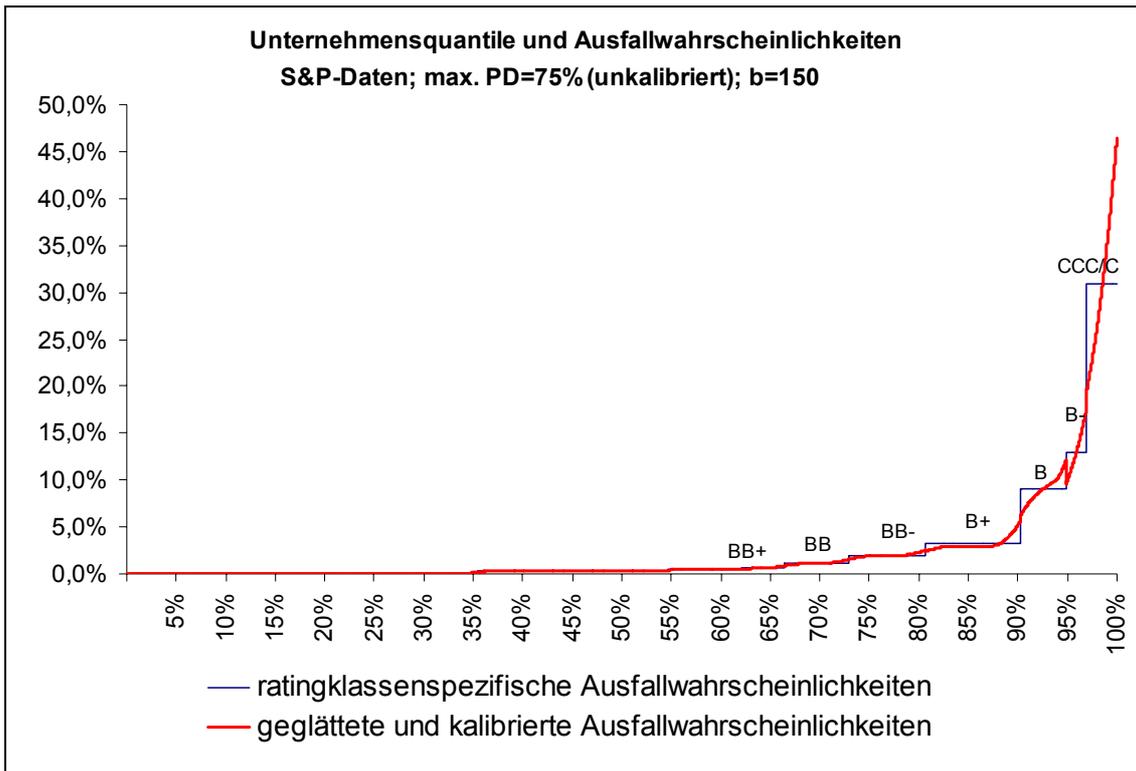


Abbildung 37: Unternehmensquantile, geglättete und kalibrierte vs. ratingklassenspezifische Ausfallwahrscheinlichkeiten, S&P-Daten, max. PD = 75%, b=150

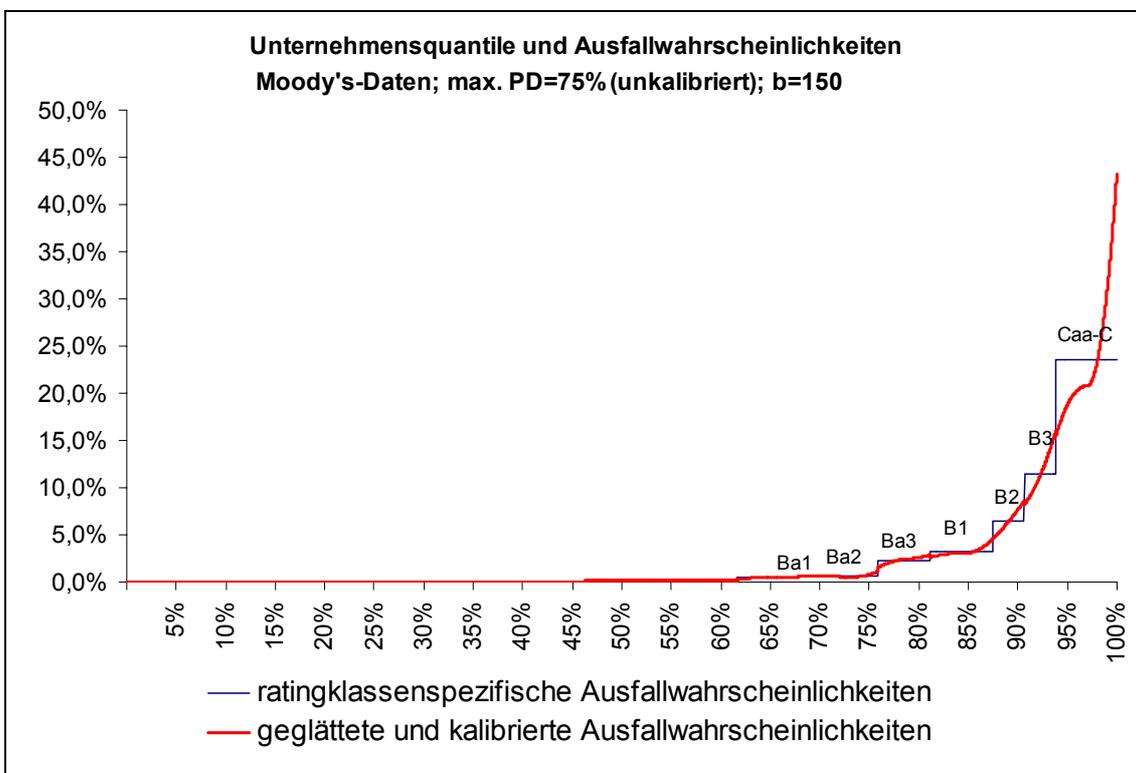


Abbildung 38: Unternehmensquantile, geglättete und kalibrierte vs. ratingklassenspezifische Ausfallwahrscheinlichkeiten, MOODY'S-Daten, max. PD = 75%, b=150

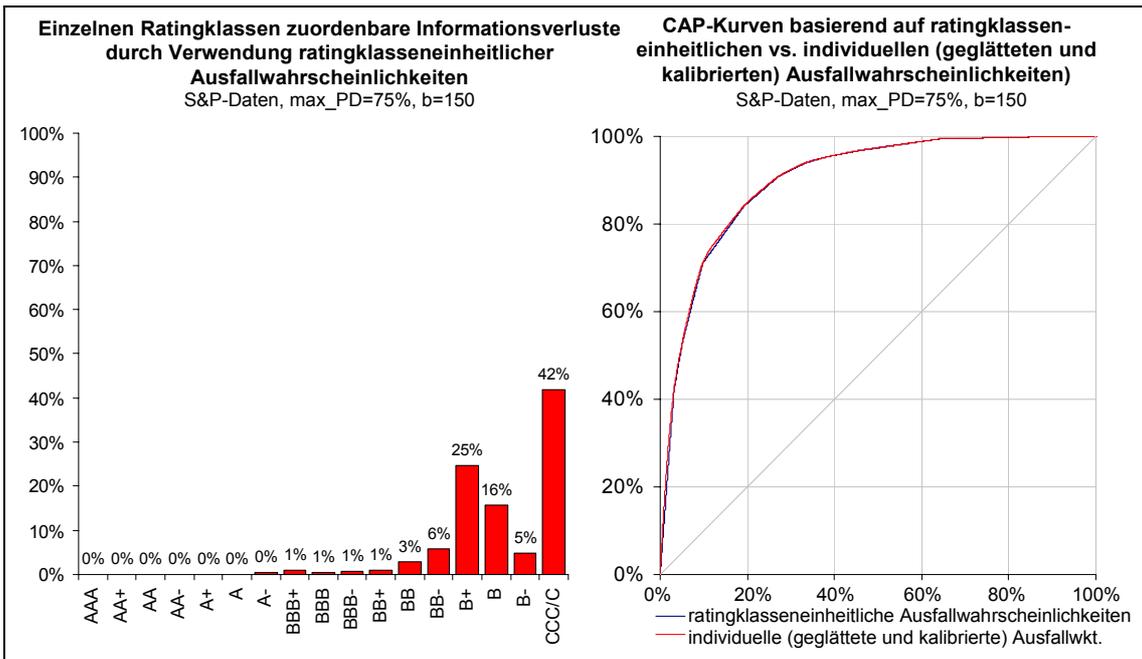


Abbildung 39: einzelnen Ratingklassen zuordenbare Informationsverluste durch Verwendung ratingklasseneinheitlicher (anstatt individueller) Ausfallwahrscheinlichkeiten (links), CAP-Kurven basierend auf ratingklasseneinheitlichen vs. individuellen (geglätteten und kalibrierten) Ausfallwahrscheinlichkeiten), S&P-Daten, max. PD = 75%, b=150 (rechts)

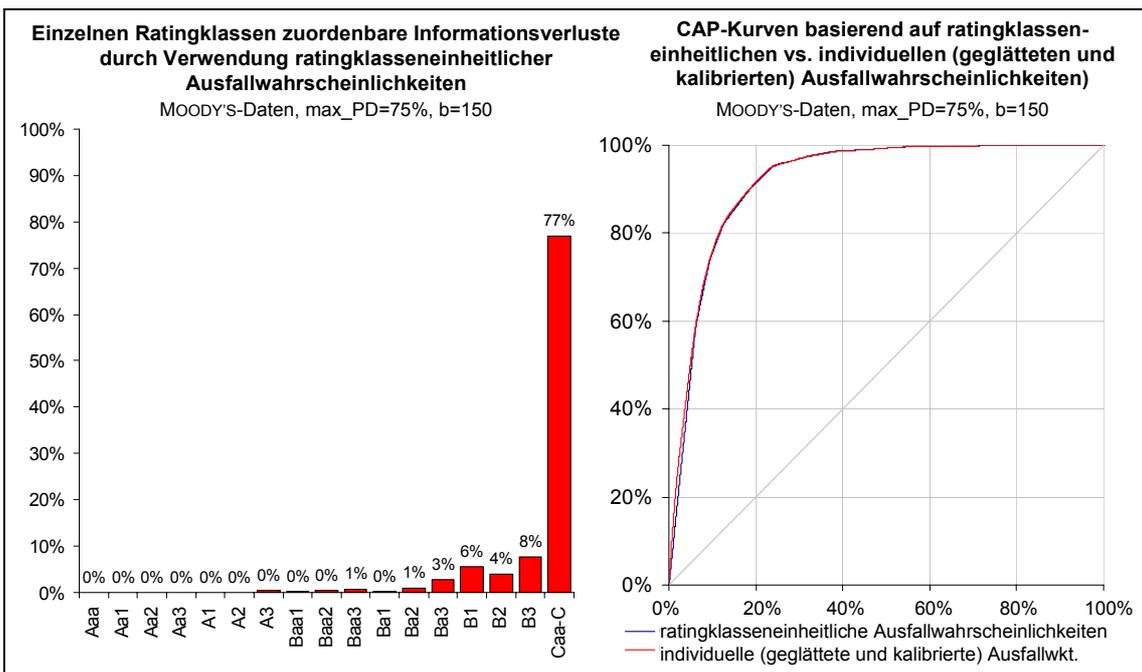


Abbildung 40: einzelnen Ratingklassen zuordenbare Informationsverluste durch Verwendung ratingklasseneinheitlicher (anstatt individueller) Ausfallwahrscheinlichkeiten (links), CAP-Kurven basierend auf ratingklasseneinheitlichen vs. individuellen (geglätteten und kalibrierten) Ausfallwahrscheinlichkeiten), MOODY'S-Daten, max. PD = 75%, b=150 (rechts)

Verfahren II: Lineare Interpolation der ROC-Kurve

Das zweite Verfahren zur Abschätzung der Informationsverluste, die mit einer Diskretisierung kontinuierlicher Ratingskalen verbunden ist, modelliert nicht (explizit) individuelle Ausfallwahrscheinlichkeiten, um die hieraus resultierenden CAP- und ROC-Kurven und die hiermit verbundenen Schätzgütemaße zu ermitteln. Statt dessen modelliert es die ROC-Kurven direkt, d.h. es geht von als bekannt unterstellten, formal analysierbaren ROC-Kurvenfunktionen aus und ermittelt anschließend, wieviel Prozent der Fläche unter der „wahren“ („kontinuierlichen“) ROC-Kurve (AUC_{ROC}) - bzw. wieviel Prozent der Accuracy Ratio - mit einem Ratingssystem mit nur g Klassen erzielt werden können, wobei für g alle Werte zwischen 2 und 30 numerisch untersucht wurden.

Für die formale Darstellung von ROC-Kurven werden zwei *parametrische* Funktionen („Funktionsfamilien“) ROC_α und ROC_β verwendet, wobei α und β frei wählbare Parameter darstellen, durch deren Spezifikation eine Vielzahl möglicher Kurvenverläufe abgebildet werden können. Für ROC_α gilt:

$$F\ 214) \quad ROC_\alpha(x) = x^\alpha \quad \text{mit } ROC_\alpha(x) = 1 - \text{Fehler I. Art und } x = \text{Fehler II. Art}$$

$ROC_\beta(x)$ ergibt sich per Definition aus $ROC_\alpha(x)$ durch Spiegelung an der im ROC-Diagramm von links oben nach rechts unten verlaufenden Diagonalen. Ersetzt man also in Formel F 214 $ROC(x)$ mit $1-x$ und x mit $1-ROC(x)$ und α durch β , ergibt sich:

$$F\ 215) \quad 1 - x = (1 - ROC_\beta(x))^\beta$$

$$F\ 216) \quad ROC_\beta(x) = 1 - (1 - x)^{1/\beta}$$

Empirische ROC-Kurven können gut durch ROC_α - oder ROC_β -Kurven - oder durch Mischungen aus beiden Kurven - beschrieben werden (siehe Anhang I). Für $\alpha=1$ bzw. $\beta=1$ folgen die ROC-Kurven im Fehlalarm-Trefferquoten-Diagramm (ROC-Diagramm) der von links unten nach rechts oben verlaufenden Diagonalen („naives Ratingverfahren“), für $\alpha=0$ bzw. $\beta=0$ verlaufen die entsprechenden ROC-Kurven entlang der Außenseiten des ROC-Diagramms („perfektes Ratingverfahren“). Für Werte für α bzw. β zwischen 0 und 1 werden Kurvenverläufe erzielt, die realen ROC-Kurven ähneln (siehe Anhang I).

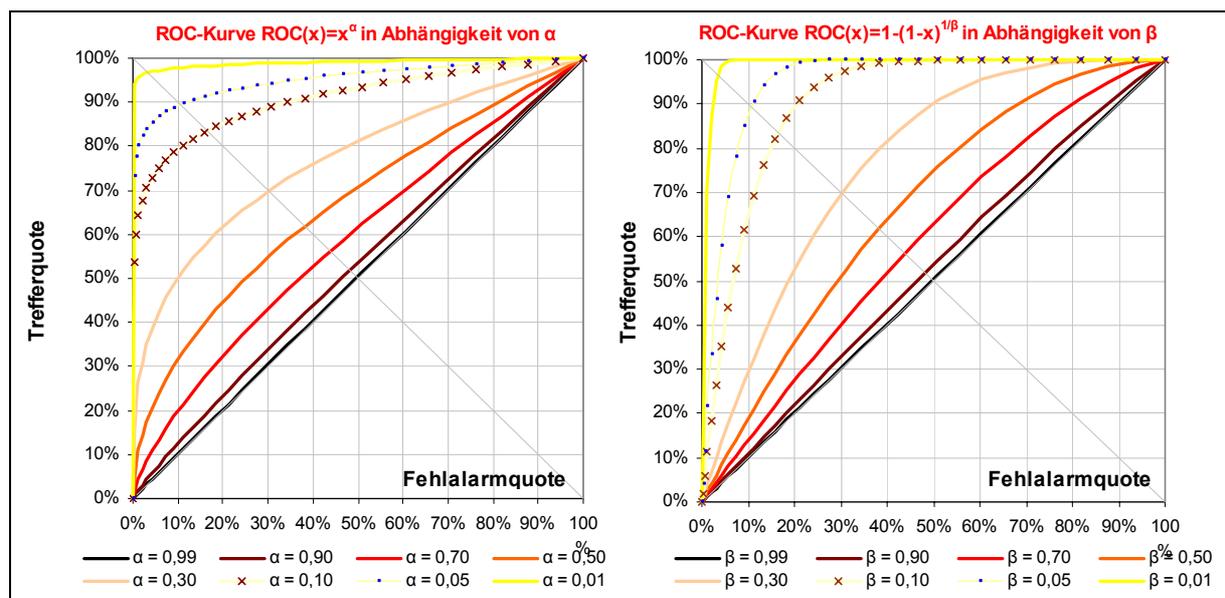


Abbildung 41: ROC-Kurven mit $ROC_\alpha(x)=x^\alpha$ und $ROC_\beta(x)=1-(1-x)^{1/\beta}$ für unterschiedliche α und β

Die Parameter α und β können - gegeben eine Fehler-I-II-Kombination – wie folgt errechnet werden (vgl. Formeln F 214 und F 216)

$$\text{F 217) } \alpha = \frac{\ln(1-F_1)}{\ln F_2} \quad \text{mit } F_1 \dots \text{ Fehler I. Art und } F_2 \dots \text{ Fehler II. Art}$$

$$\text{F 218) } \beta = \frac{\ln(1-x)}{\ln(1-\text{ROC}_\beta(x))}$$

$$\text{F 219) } \beta = \frac{\ln(1-F_2)}{\ln F_1}$$

Die Fläche unter der ROC_α -Kurve und die hierauf basierende Accuracy Ratio betragen:

$$\text{F 220) } \text{AUC}_{\text{ROC}_\alpha} = \int_0^1 \text{ROC}_\alpha(x) dx = \int_0^1 x^\alpha dx$$

$$\text{F 221) } \text{AUC}_{\text{ROC}_\alpha} = \left| \frac{1}{\alpha+1} \cdot x^{\alpha+1} \right|_{x=0}^{x=1}$$

$$\text{F 222) } \text{AUC}_{\text{ROC}_\alpha} = \frac{1}{\alpha+1}$$

$$\text{F 223) } \text{AR}_\alpha = \frac{\text{AUC}_{\text{ROC}_\alpha} - \frac{1}{2}}{\frac{1}{2}} = \frac{1}{\alpha+1} - \frac{1}{2}$$

$$\text{F 224) } \text{AR}_\alpha = \frac{1-\alpha}{1+\alpha} \quad \text{und in Kombination mit Formel F 217:}$$

$$\text{F 225) } \text{AR}_\alpha = \frac{1 - \frac{\ln(1-F_1)}{\ln F_2}}{1 + \frac{\ln(1-F_1)}{\ln F_2}}$$

$$\text{F 226) } \text{AR}_\alpha = \frac{\ln F_2 - \ln(1-F_1)}{\ln F_2 + \ln(1-F_1)}$$

Durch Substitution der ROC-Koordinaten F_1 und F_2 durch die CAP-Koordinaten X_0 und Y_0 gemäß den Formeln F 64 und F 66 ergibt sich ferner:

$$\text{F 227) } \text{AR}_\alpha = \frac{\ln\left(\frac{X_0 - Y_0 \cdot \text{PD}}{1 - \text{PD}}\right) - \ln Y_0}{\ln\left(\frac{X_0 - Y_0 \cdot \text{PD}}{1 - \text{PD}}\right) + \ln Y_0}$$

$$\text{F 228) } \text{AR}_\alpha = \frac{\ln(X_0 - Y_0 \cdot \text{PD}) - \ln(1 - \text{PD}) - \ln Y_0}{\ln(X_0 - Y_0 \cdot \text{PD}) - \ln(1 - \text{PD}) + \ln Y_0}$$

Die Fläche unter der ROC_β -Kurve und die hierauf basierende Accuracy Ratio berechnen sich wie folgt:

$$\text{F 229) } \text{AUC}_{\text{ROC}_\beta} = \int_0^1 \text{ROC}_\beta(x) dx = \int_0^1 1 - (1-x)^{1/\beta} dx$$

$$\text{F 230) } AUC_{\text{ROC}\beta} = \left| x + \frac{\beta}{1+\beta} \cdot (1-x)^{(1+\beta)/\beta} \right|_0^1$$

$$\text{F 231) } AUC_{\text{ROC}\beta} = 1 + \frac{\beta}{1+\beta} \cdot (1-1)^{(1+\beta)/\beta} - \left(0 + \frac{\beta}{1+\beta} \cdot (1-0)^{(1+\beta)/\beta} \right)$$

$$\text{F 232) } AUC_{\text{ROC}\beta} = 1 - \frac{\beta}{1+\beta}$$

$$\text{F 233) } AUC_{\text{ROC}\beta} = \frac{1}{1+\beta} \quad (\text{vgl. Formel F 222})$$

$$\text{F 234) } AR_{\beta} = \frac{\frac{1}{1+\beta} - \frac{1}{2}}{2}$$

$$\text{F 235) } AR_{\beta} = \frac{1-\beta}{1+\beta} \quad (\text{vgl. Formel F 224}) \text{ in Kombination mit Formel F 219}$$

$$\text{F 236) } AR_{\beta} = \frac{1 - \frac{\ln(1-F_2)}{\ln F_1}}{1 + \frac{\ln(1-F_2)}{\ln F_1}}$$

$$\text{F 237) } AR_{\beta} = \frac{\frac{\ln F_1 - \ln(1-F_2)}{\ln F_1}}{\frac{\ln F_1 + \ln(1-F_2)}{\ln F_1}}$$

$$\text{F 238) } AR_{\beta} = \frac{\ln F_1 - \ln(1-F_2)}{\ln F_1 + \ln(1-F_2)} \quad (\text{vgl. Formel F 226})$$

Durch Substitution der ROC-Koordinaten F_1 und F_2 durch die CAP-Koordinaten X_0 und Y_0 gemäß den Formeln F 64 und F 66 ergibt sich ferner:

$$\text{F 239) } AR_{\beta} = \frac{\ln(1-Y_0) - \ln\left(1 - \frac{X_0 - Y_0 \cdot PD}{1-PD}\right)}{\ln(1-Y_0) + \ln\left(1 - \frac{X_0 - Y_0 \cdot PD}{1-PD}\right)}$$

$$\text{F 240) } AR_{\beta} = \frac{\ln(1-Y_0) - \ln(1-PD - X_0 + Y_0 \cdot PD) + \ln(1-PD)}{\ln(1-Y_0) + \ln(1-PD - X_0 + Y_0 \cdot PD) - \ln(1-PD)}$$

Werte für α und β , die zu realistischen Accuracy-Ratio-Werten führen, liegen für sehr trennscharfe Ratingverfahren im Bereich um $\alpha = \beta = 0,1$ ($AR=82\%$), für eher unterdurchschnittliche Ratingverfahren um $\alpha = \beta = 0,3$ ($AR=54\%$) und für vergleichsweise ineffiziente Verfahren um $\alpha = \beta = 0,5$ ($AR=33\%$) (siehe auch Abschnitt 3.5).

Mittels obiger Formeln kann der Flächeninhalt unter den ROC-Kurven, bzw. die zugehörigen Accuracy Ratios, von Ratings mit kontinuierlichen Scores ermittelt werden und mit den entsprechenden Werten von Ratings mit einer endlichen Zahl an Ratingklassen verglichen werden (siehe Formel F 22).

Mathematisch gesehen hängt die erzielbare relative Güte davon ab, wie gut es gelingt, die ROC-Kurve des kontinuierlichen Scores durch eine ROC-Kurve zu approximieren, die aus einer endlichen Anzahl linearer Teilstücke besteht, wobei die Endpunkte der linearen

Teilstücke auf der ROC-Kurve des kontinuierlichen Scores liegen (siehe Abbildung 42).³³⁴ Alle durch lineare Kurventeilstücke approximierten ROC-Kurven müssen mindestens durch die Punkte (0%;0%) und (100%;100%), die zwei „äußeren Stützstellen“, verlaufen. Bei g Ratingklassen stehen zusätzlich $g-1$ auf der ROC-Kurve durch die Wahl entsprechender Cut-Off-Scores „plazierbare“ Punkte zur Bildung der linearen Teilstücke zur Verfügung.

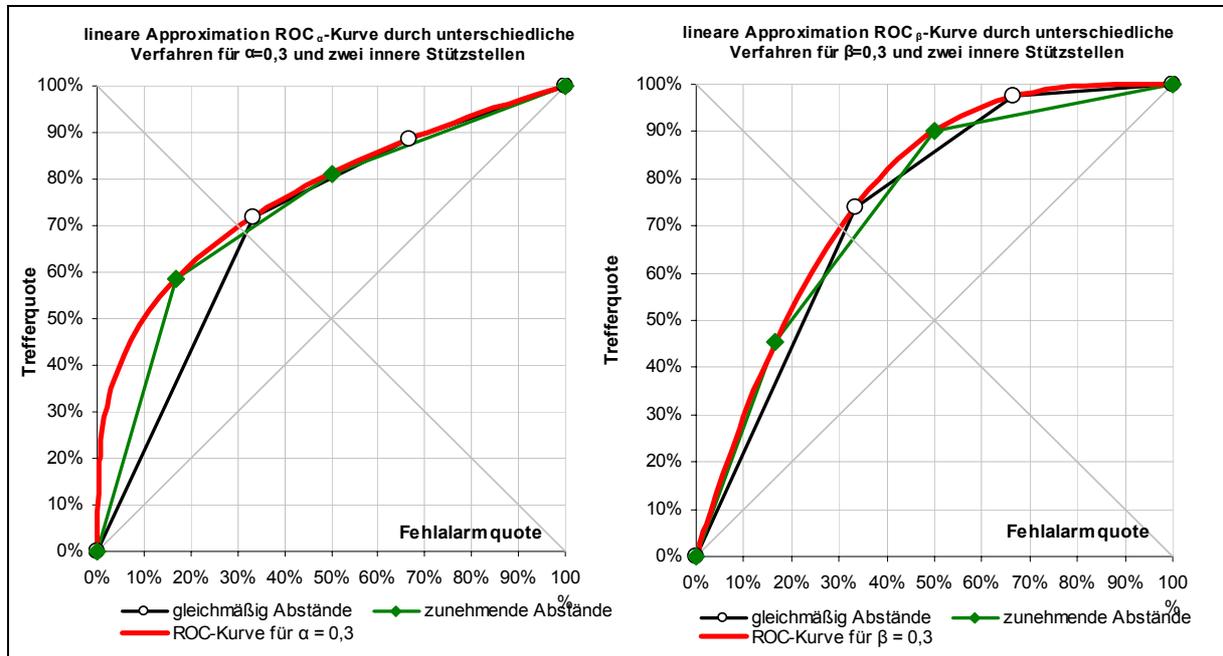


Abbildung 42: lineare Approximationen von ROC_{α} - und ROC_{β} -Kurven durch unterschiedliche Verfahren, $\alpha = \beta = 0,3$, zwei innere Stützstellen

Eine weitere Vorgabe wird jedoch für die Anwendung der linearen Approximation benötigt – die Festlegung der Abszissenwerte (Fehlalarmquoten) für die inneren Stützstellen. In Abbildung 42 sind, sowohl für ROC_{α} - als auch für ROC_{β} -Kurven, zwei Beispiele für unterschiedliche Verfahren zur Festlegung der Abszissenwerte bei einem Ratingsystem mit *drei* Klassen, d.h. *zwei* inneren Stützstellen, dargestellt. Bei dem ersten Verfahren werden die Ratingklassen so abgegrenzt, daß jede Ratingklassen den gleichen Anteil an Fehlalarmen zeugt (hier: jeweils 33,3%). Beim zweiten Verfahren steigt die Klassenbreite linear an, d.h. die zweite Klasse ist doppelt so breit wie die erste Klasse, die dritte Klasse dreimal so breit wie die erste (allgemein: die i -te Klasse erfaßt i mal so viele Fehlalarme wie die erste Klasse), so daß sich hier Klassenbreiten von 16,6%, 33,3% und 50% ergeben.

In Abbildung 43 werden die relativen Accuracy-Ratio-Werte für beide Klassenabgrenzungsverfahren, für verschiedene α - und β -Parameter und für verschiedene Anzahlen an Ratingklassen dargestellt. Besonders hervorgehoben sind die Werte für 2, 7 und 17 Ratingklassen (entsprechend 1, 6 und 16 inneren Stützstellen).

³³⁴ Der lineare Verlauf von Teilstücken einer ROC-Kurve impliziert, daß innerhalb der durch die Teilstücke gekennzeichneten Ratingklassen keine Risikodifferenzierung innerhalb der Unternehmen der jeweiligen Ratingklassen vorgenommen werden kann.

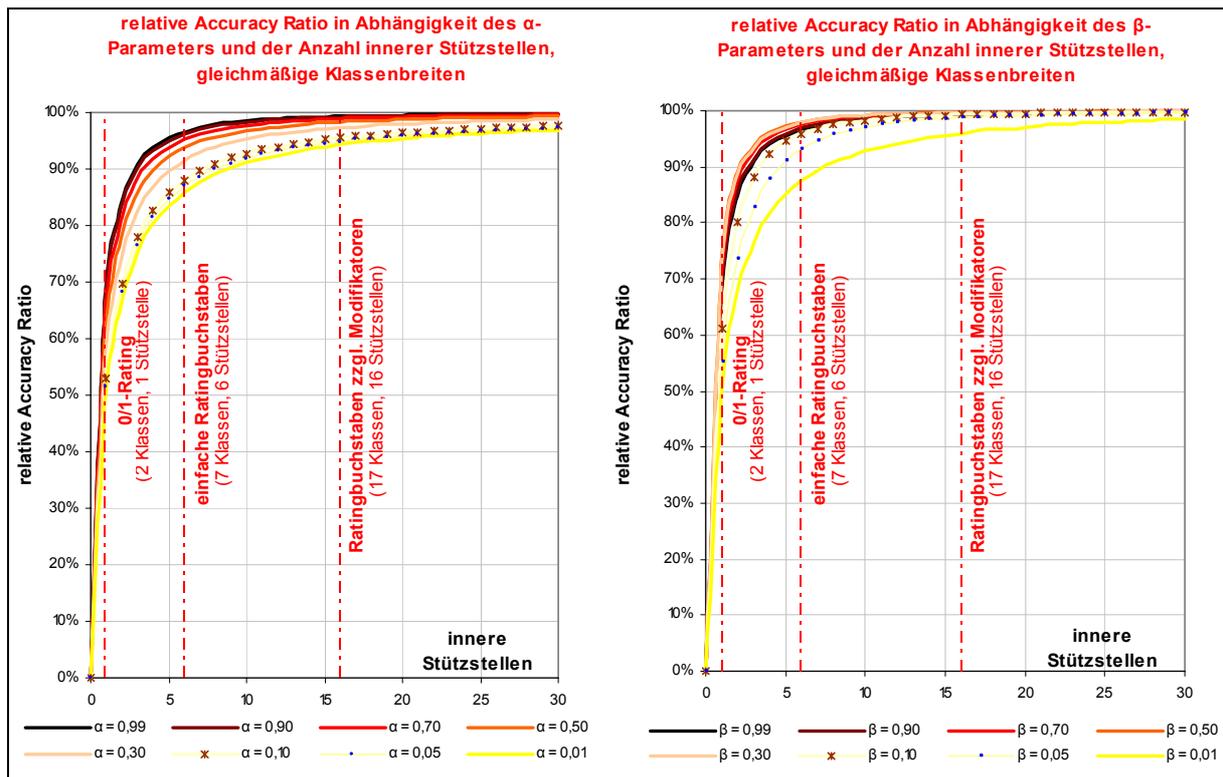


Abbildung 43: relative Accuracy Ratios für verschiedene Parameterwerte der ROC $_{\alpha}$ - (links) und ROC $_{\beta}$ -Kurven (rechts) und 0.30 innere Stützstellen (1..31 diskrete Ratingklasse), gleichmäßige Klassenbreiten

Für 17 Ratingklassen ergeben sich beispielsweise relative Accuracy-Ratio-Werte von 95,5% ($\alpha=0,1$)/ 99,3% ($\beta=0,1$) oder 98,3% ($\alpha=0,5$) / 99,7% ($\beta=0,5$). Die Accuracy Ratio eines sehr schlechten Ratingverfahrens ($\alpha=0,5$, $\beta=0,5$, $AR_{\alpha}=AR_{\beta}=33\%$) würde also auf Basis einer 17-Punkteskala um ca. 0,3% - 1,7% (0,1 - 0,6 Prozentpunkte) zu schlecht, die eines guten Ratingverfahrens ($\alpha=0,1$, $\beta=0,1$, $AR_{\alpha}=AR_{\beta}=82\%$) hingegen um rund 0,7%-4,5% (0,5 - 3,5 Prozentpunkte) zu schlecht ausgewiesen werden – jeweils verglichen mit der Qualität eines auf einem kontinuierlichem Score basierenden Verfahrens.

Obige Werte stellen jedoch zu pessimistische Schätzer für die mit der Diskretisierung von kontinuierlichen Ratingscores verbundenen Informationsverluste dar- insbesondere bei dem α -Verfahren. Bei weiteren Untersuchungen wurde festgestellt, daß die zu verzeichnenden Informationsverluste bei der linearen Interpolation der ROC-Kurve bei gleichmäßiger Klassenaufteilung bei dem α -Verfahren fast ausschließlich auf die in der ersten Ratingklasse zu verzeichnenden Informationsverluste zurückzuführen sind (siehe Abbildung 44 links, vergleiche auch Abbildung 42 links).

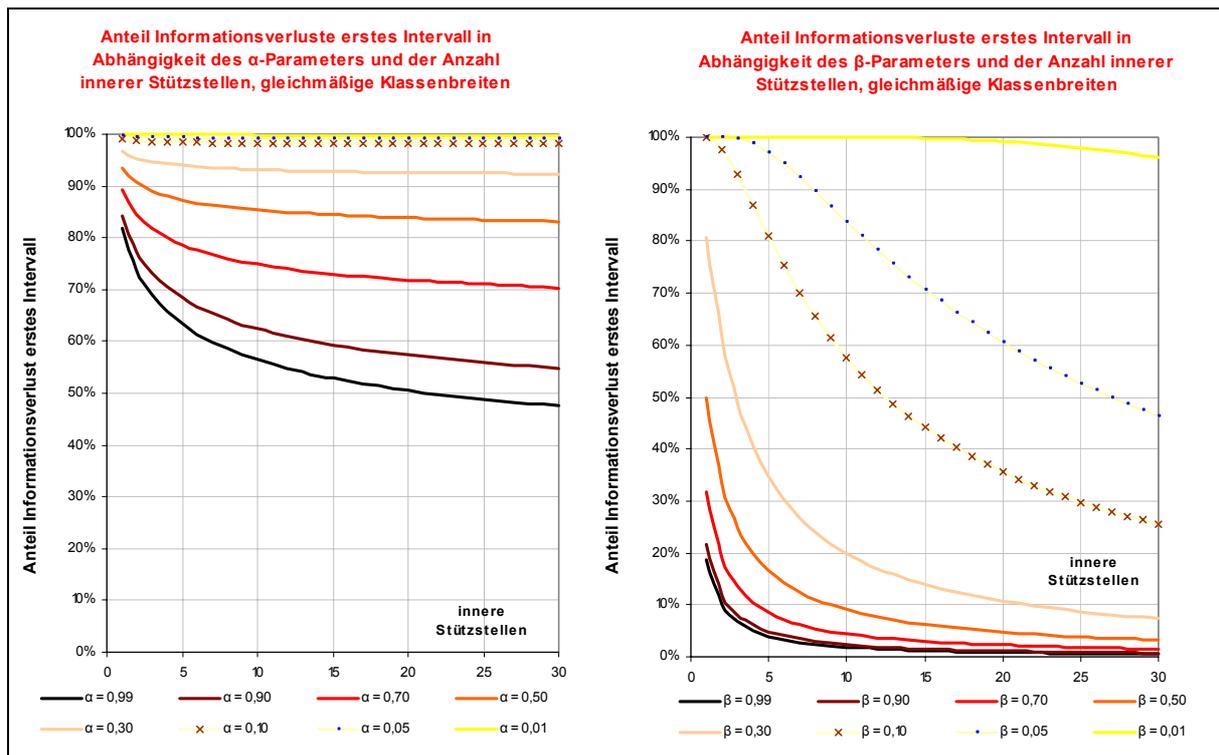


Abbildung 44: Anteil des im ersten Intervall auftretenden Informationsverlusts an den gesamten Informationsverlusten für ROC_{α} - (links) und ROC_{β} -Kurven (rechts) und 0..30 innere Stützstellen (1..31 diskrete Ratingklasse), gleichmäßige Klassenbreiten

Für $\alpha=0,1$ und 30 Ratingklassen treten beispielsweise 98,1% (!) aller Informationsverluste, durch die lineare Approximation in der ersten Ratingklasse auf (für $\beta=0,1$ und 30 Ratingklassen immerhin noch 25,6%) – die restlichen 1,9% an Informationsverlusten verteilen sich auf die übrigen 29 Ratingklassen.

Offensichtlich sind ROC-Kurven des α -Typs im Bereich der ersten Ratingklasse hochgradig nicht-linear (siehe den ausgeprägten konkaven Kurvenverlauf in Abbildung 42, links). Somit ist zu vermuten, daß sich sowohl der Anteil des Informationsverlustes der ersten Ratingklasse an den gesamten Informationsverlusten als auch die Höhe der gesamten Informationsverluste wesentlich verringern lassen sollten, falls die Klassenbreiten der schlechtesten Ratingklassen zugunsten der anderen Ratingklassen reduziert würden. Dies kann beispielsweise mittels des oben erwähnten Verfahrens der proportional zunehmenden Klassenbreiten erreicht werden (siehe Abbildung 45 für den Verlauf der relativen Accuracy Ratios in Abhängigkeit der Qualität des Ratings und der Anzahl an Ratingklassen für das α - und β -Verfahren und Abbildung 46 für den Anteil der ersten Ratingklasse an den insgesamt auftretenden Informationsverlusten).^{335,336}

³³⁵ Ob es sich um ein Verfahren handelt, das die Informationsverluste, gegeben $ROC(x)=x^{\alpha}$, *minimiert*, ist unbekannt und irrelevant. Irrelevant ist es, da die Ratinggeber die relativen Belegungsstärken der einzelnen Ratingklassen ohnehin nur begrenzt, durch die Festlegung von klassenspezifischen Kriterien, steuern können und sie möglicherweise andere Zielstellungen als die Minimierung von Informationsverlusten haben, beispielsweise die Sicherstellung von Mindestbelegungsstärken, um die Aussagefähigkeit statistischer Untersuchungen zu verbessern. Des weiteren ist unwahrscheinlich, daß eine für $ROC(x)=x^{\alpha}$ optimale Klassenaufteilung auch für andere formale Funktionen oder empirische Verteilungen optimal ist. Wichtig für die Wahl des Verfahrens waren nur die stärkere Differenzierung im relevanten vorderen Teil der ROC-Kurve und die einfache Kommunizierbarkeit und Umsetzbarkeit des Verfahrens.

³³⁶ Für weitere Modellierungsmöglichkeiten für formale und numerische Ratingklassenabgrenzungen (“plausible but suboptimal boundaries”, S. 17) siehe JANKOWITSCH, PICHLER, SCHWAIGER (2003, S. 15f. und 21f.).

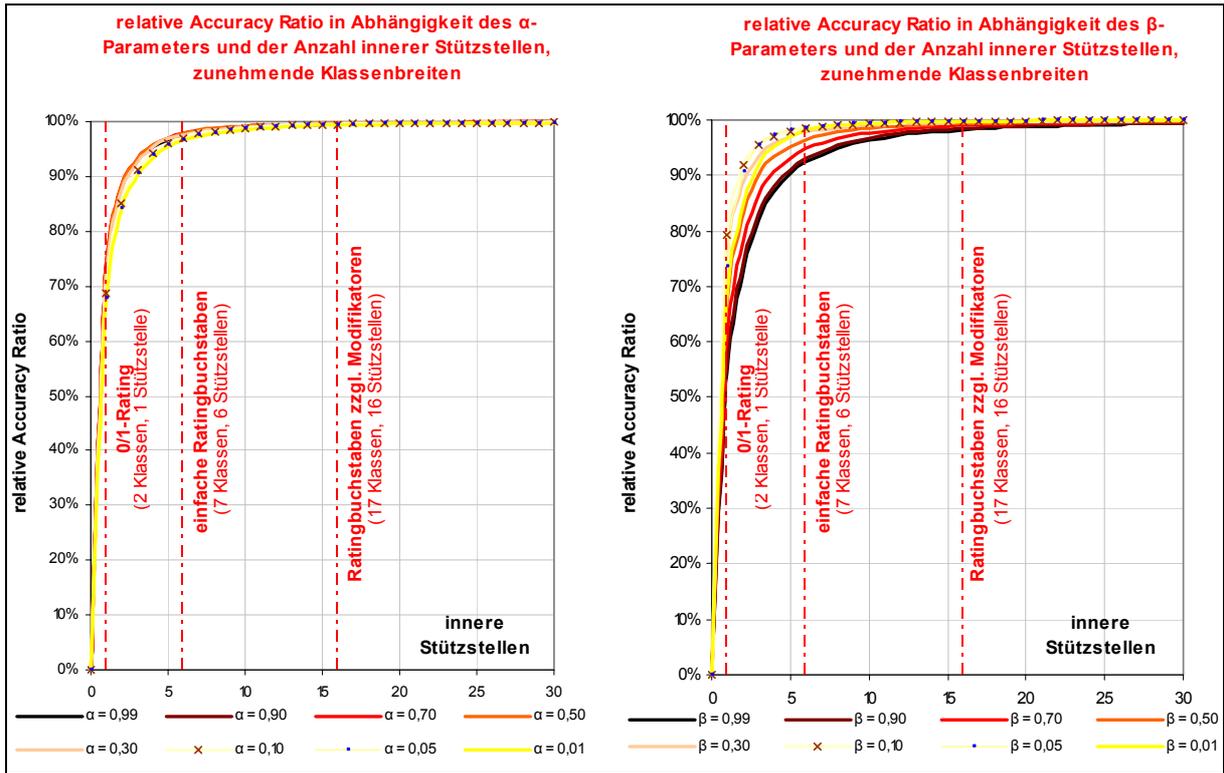


Abbildung 45: relative Accuracy Ratios für verschiedene Parameterwerte der ROC_{α} - (links) und ROC_{β} -Kurven (rechts) und 0..30 innere Stützstellen (1..31 diskrete Ratingklasse), zunehmende Klassenbreiten

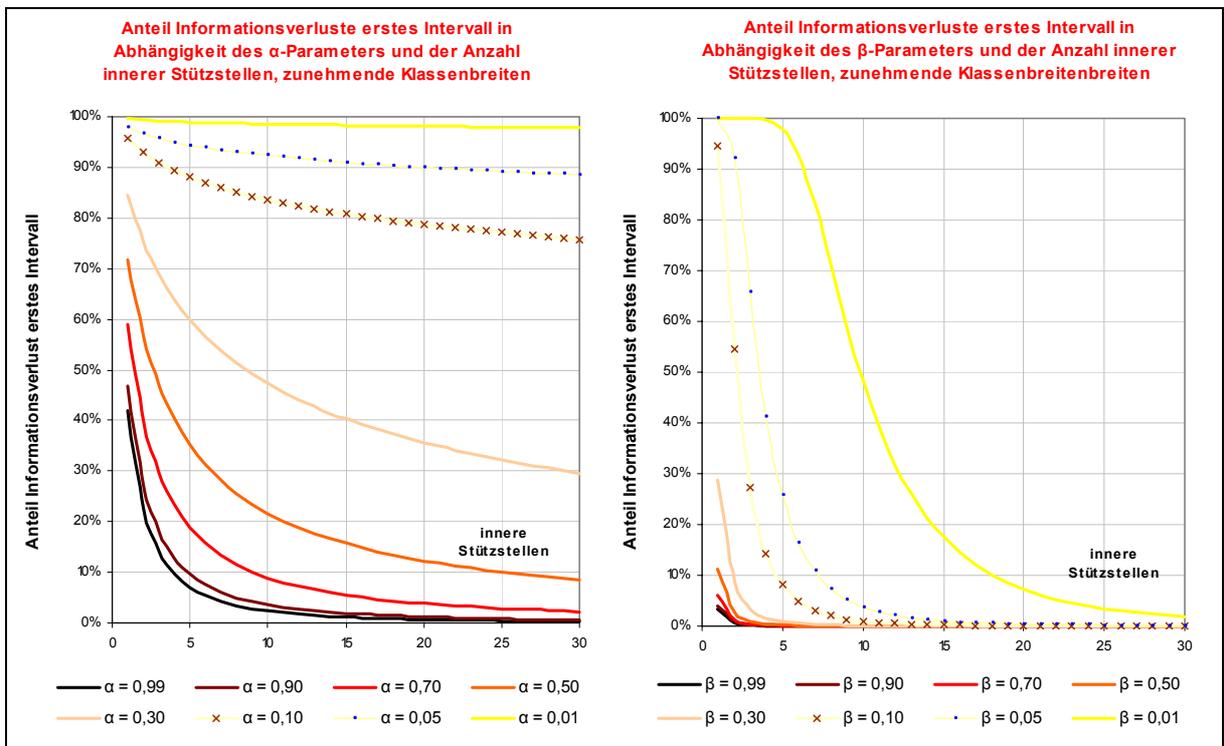


Abbildung 46: Anteil des im ersten Intervall auftretenden Informationsverlusts an den gesamten Informationsverlusten für ROC_{α} - (links) und ROC_{β} -Kurven (rechts) und 0..30 innere Stützstellen (1..31 diskrete Ratingklasse), zunehmende Klassenbreiten

Bei dem Verfahren mit zunehmenden Ratingklassenbreiten und 17 Ratingklassen ergeben sich beispielsweise Accuracy Ratios von 99,5% ($\alpha=0,1$)/ 99,7% ($\beta=0,1$) oder 99,6% ($\alpha=0,5$)/ 99,3% ($\beta=0,1$). Die Accuracy Ratio eines schlechten Ratingverfahrens ($\alpha=0,5$) würde also auf Basis einer 17-Punkteskala nur um ca. 0,4% - 0,7% (ca. 0,2 Prozentpunkte) zu schlecht und die eines guten Ratingverfahrens um rund 0,3%-0,5% (ca. 0,3 Prozentpunkte) zu schlecht ausgewiesen werden – jeweils verglichen mit der Qualität eines auf einem kontinuierlichem Score basierenden Verfahrens.

Fazit Verfahren II

In Tabelle P werden die Ergebnisse dieses Abschnittes zusammengefaßt und die Verhältnisse der Accuracy Ratios von diskreten zu kontinuierlichen Ratings in Abhängigkeit von der Anzahl der Ratingklassen, der Qualität des Ratings, der Verteilung der Ratingklassenbreiten und der funktionalen Gestalt der ROC-Kurven ausgewiesen.

Die für die einzelnen betrachteten Anzahlen an diskreten Ratingklassen für besonders relevant erachteten relativen Accuracy-Ratio-Werte sind schattiert unterlegt (für Erläuterungen siehe unten).

Anzahl Ratingklassen	Qualität des Ratings	gleiche Klassenbreiten		zunehmende Klassenbreiten		Mittelwert α - β -Funktionen	
		α -Funktion	β -Funktion	α -Funktion	β -Funktion	gleiche Klassenbreiten	zunehmende Klassenbreiten
2	gut (AR=82%)	0,529	0,610	0,688	0,794	0,570	0,741
	mittel (AR=54%)	0,580	0,744	0,717	0,757	0,662	0,737
	schlecht (AR=33%)	0,621	0,750	0,732	0,667	0,686	0,699
7	gut (AR=82%)	0,880	0,959	0,971	0,985	0,920	0,978
	mittel (AR=54%)	0,915	0,979	0,977	0,978	0,947	0,977
	schlecht (AR=33%)	0,938	0,980	0,978	0,964	0,959	0,971
10	gut (AR=82%)	0,919	0,980	0,986	0,992	0,950	0,989
	mittel (AR=54%)	0,946	0,990	0,989	0,989	0,968	0,989
	schlecht (AR=33%)	0,963	0,990	0,990	0,982	0,977	0,986
17	gut (AR=82%)	0,955	0,993	0,995	0,997	0,974	0,996
	mittel (AR=54%)	0,973	0,996	0,996	0,996	0,985	0,996
	schlecht (AR=33%)	0,983	0,997	0,996	0,993	0,990	0,995
30	gut (AR=82%)	0,976	0,998	0,998	0,999	0,987	0,999
	mittel (AR=54%)	0,987	0,999	0,999	0,999	0,993	0,999
	schlecht (AR=33%)	0,993	0,999	0,999	0,998	0,996	0,998

Tabelle P: relative Schätzgüte von diskreten Ratings (in Relation zur Schätzgüte eines Ratings mit kontinuierlicher Skala) in Abhängigkeit von der Anzahl der Ratingklassen, der Qualität des Ratings, des Klassenbreitenverfahrens und der funktionalen Gestalt der ROC-Kurven (α - vs. β -Funktion)

Die in Tabelle P zusammengefaßten Ergebnisse zeigen, daß es keinen *eindeutigen* Zusammenhang zwischen der Anzahl der Ratingklassen und den gegenüber einem Ratingmodell mit kontinuierlicher Ratingskala entstehenden Informationsverlusten gibt, sondern daß weitere Einflußgrößen zu berücksichtigen sind. Neben der funktionalen Gestalt der kontinuierlichen ROC-Funktionen, die aber empirisch gut durch eine Mischung aus ROC_{α} - und ROC_{β} -Kurven

beschreibbar ist³³⁷, siehe auch die Untersuchungen in Anhang I, werden die Ergebnisse sehr stark von den unterstellten Aufteilungen der Ratingklassen beeinflusst – für die jedoch, außer für 7 oder 17 Ratingklassen, praktisch keine empirischen Informationen vorliegen.

Im weiteren wird unterstellt, daß die schlechtesten Ratingklassen deutlich unterdurchschnittlich stark belegt sind³³⁸, so daß die Ergebnisse des Verfahrens mit zunehmenden Klassenbreiten tendentiell realitätsnäher sind als die Ergebnisse des Verfahrens mit identischen Klassenbreiten (siehe die entsprechenden Schattierungen in Tabelle P). Die Ergebnisse des Verfahrens mit gleichen Klassenbreiten sind aber den Fällen relevant, in denen die Schätzgüte von Insolvenzprognoseverfahren auf Basis von per Definition gleich großen Gruppen ermittelt werden muß, wobei hier typischerweise jede Gruppe 10% der insgesamt untersuchten Unternehmen umfaßt („Dezile“) (siehe die entsprechenden Hervorhebungen in Tabelle P).³³⁹

Grundsätzlich gilt jedoch, daß sich bei der Abschätzung der für ein *konkretes* Ratingverfahren zu erwartenden Informationsverluste, sich die Verwendung der tatsächlich beobachtbaren Verteilung der Unternehmen auf die einzelnen Ratingklassen (anstelle obiger Verfahren zur Aufteilung der Ratingklassen) empfiehlt.

Als Faustregel kann formuliert werden, daß bei Verwendung eines Ratings mit nur 7 Ratingklassen mit einem Informationsverlust gegenüber einem kontinuierlichen Rating von ca. 3% zu rechnen ist und bei einem Rating mit 17 Klassen mit einem Informationsverlust in Höhe von ungefähr 0,5%.

Auch zeigt sich wiederum, wie bereits bei Verfahren I, daß überproportional hohe Informationsverluste in der schlechtesten Ratingklasse auftreten.

³³⁷ Aus diesem Grund wurden in Tabelle P die Mittelwerte der gemäß des α - und β -Verfahrens bestimmten Werte ermittelt und schattiert hervorgehoben.

³³⁸ Mit Anteilen von 3,1% bzw. 6,2% sind die Ratingklassen CCC/C (S&P) bzw. Caa/C (MOODY'S) auf einer 7er Skala deutlich unterrepräsentiert (durchschnittliche Klassenbreite = $1/7 = 14\%$). Auf der 17er Skala entspricht ihr Anteil zwar ungefähr der durchschnittlichen Klassenbreite ($1/17 = 5,9\%$). Wird jedoch berücksichtigt, daß die Klassen CCC/C bzw. Caa/C in der Modifikatordarstellung fünfmal untergliedert werden (CCC+, CCC, CCC-, CC, C bzw. Caa1, Caa2, Caa3, Ca, C), siehe S&P (2003b, S.8) und MOODY'S (2004b, S.6), so beträgt der durchschnittliche Anteil der diversen CCC/C-Ratingstufen nur 0,6% (S&P) bzw. 1,2% (MOODY'S) und ist damit ebenfalls deutlich unterdurchschnittlich.

In den Ausfallstudien der Agenturen werden bis 2004 die CCC/C-Klassen nur zusammengefaßt ausgewiesen (siehe beispielsweise S&P (2004, S.13f), S&P (2005, S. 18, 31) oder MOODY'S (2004, S. 22, 26)). Die zusammengefaßte Ausweisung erfolgte vermutlich auch aufgrund der schlechten Trennfähigkeit der Agenturratings innerhalb des CCC/C-Bereichs (siehe beispielsweise die Daten in HAMILTON (2004, S. 18): die durchschnittliche Dreijahresausfallrate 1996-2003 [gemäß ebenda, S. 4, wurden alphanumerisch modifizierte Caa-Ratings erst im Juni 1997 eingeführt] für Caa1 (33,4%), ist größer (!) als die für Caa2 (31,3%) und diese wiederum ist größer (!) als die für Caa3 (24,4%). Erst ab 2005 werden – zumindest in der MOODY'S-Ausfallstudie, siehe MOODY'S (2005, S. 18), auch die historischen Ausfallraten für die einzelnen Caa-Teilgruppen Caa1, Caa2, und Caa3 und für Ca/C für den Zeitraum 1998-2004 ausgewiesen. In diesem Zeitraum gilt $PD_{Caa1} < PD_{Caa2} < PD_{Caa3} < PD_{Ca/C}$ und zwar sowohl für die Einjahres- als auch für die Dreijahresausfallraten.

³³⁹ Siehe beispielsweise MCQUOWN (1993, S. 17), KEALHOFER (2003, S. 35), SHUMWAY (2001, S. 118ff.)

Anhang V: Der Einfluß der Vorselektion von Portfolios auf die Schätzgüte von Insolvenzprognosen

Ziel: Es soll die Wirkung einer „positiven Vorselektion“ von Portfolios auf die ermittelte Schätzgüte von Insolvenzprognoseverfahren ermittelt werden, wobei als Selektionskriterium die Insolvenzprognosen selbst zur Anwendung kommen. Hierdurch sollen insbesondere die Konsequenzen auf die ermittelte Qualität der Insolvenzprognosen einer Bank ermittelt werden, die einen bestimmten Prozentsatz der potentiellen Kunden wegen zu schlechter Bonität ablehnt, bzw. in ein Problemkundenportfolio verschiebt, und anschließend die Güte ihrer Insolvenzprognosen nur auf Basis der realisierten Ausfälle der nicht-abgelehnten, bzw. nicht in das Problemkundenportfolio transferierten, Kunden ermittelt. Ebenfalls soll die Wirkung von „negativen Vorselektion“ von Portfolios ermittelt werden.³⁴⁰

Beschreibung des weiteren Vorgehens: Ausgehend von den in Anhang I und IV entwickelten formal analysierbaren ROC-Kurvenfunktionen wird für verschiedene Parameter geprüft, welchen Einfluß Vorselektionen (in obigen Sinne) von Portfolios auf die ermittelte Schätzgüte von Insolvenzprognoseverfahren bei Messung auf Basis der Restportfolios haben. Die Ergebnisse werden anhand von empirischen Daten geprüft.

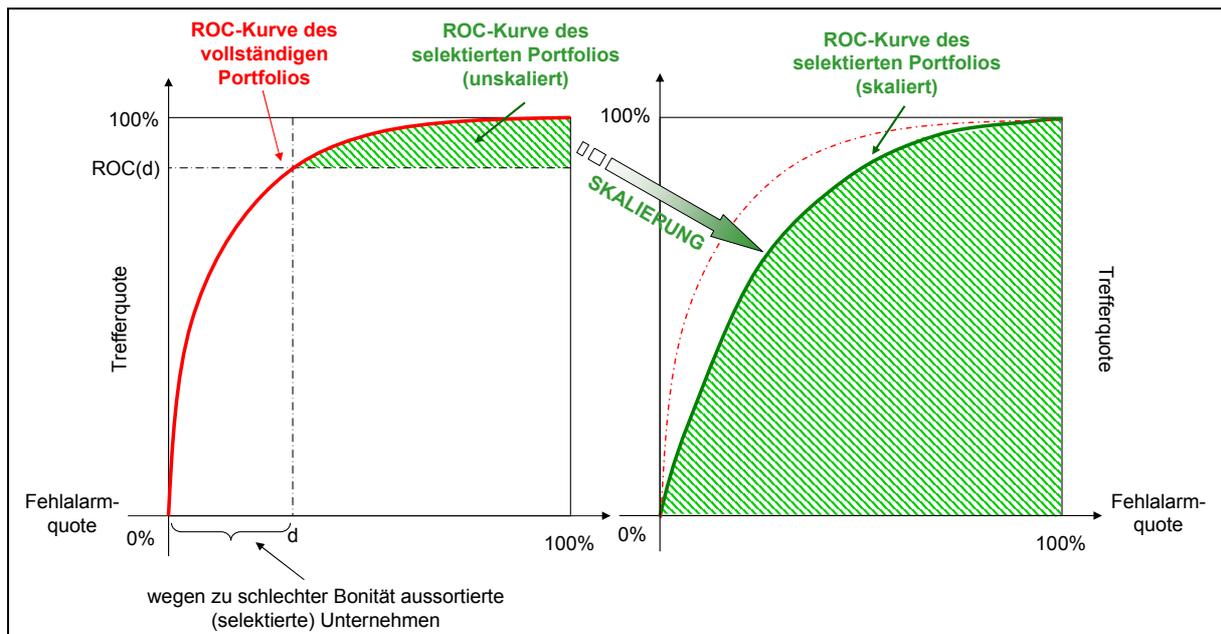


Abbildung 47: Gegenüberstellung der ROC-Kurven eines vollständigen und eines selektierten Portfolios, Skizze

Der formale Ansatz der folgenden Analysen wird in Abbildung 47 skizziert. Die ROC-Kurve eines Portfolios, von dem der Anteil d der Unternehmen mit dem schlechtesten Rating entfernt wurde³⁴¹ ergibt sich durch Skalierung des entsprechenden ROC-Kurvenrumpfstücks (siehe die schraffierte Fläche in der linken Grafik), wobei entlang der Abszisse als Skalie-

³⁴⁰ Für die Relevanz der folgenden Untersuchungen siehe die Ausführungen zu den positiven und negativen Vorselektionen von Bankenportfolios in Abschnitt 3.1.

³⁴¹ Strenggenommen gibt d in obiger Skizze den Anteil der ausgeschlossenen *Nicht-Ausfaller* und nicht den Anteil der ausgeschlossenen *Unternehmen* an. Für kleine durchschnittliche Ausfallwahrscheinlichkeiten PD , d.h. für kleine Ausfaller-Anteile an den Portfolios, wie sie – zumindest bei Betrachtung von Einjahresausfällen – für empirische Studien typisch sind – handelt es sich dabei um eine die folgenden Ergebnisse nur geringfügig verzerrende Annahme. Der exakte Anteil der ausgeschlossenen Unternehmen beträgt $d \cdot (1 - PD) + ROC(d) \cdot PD$.

rungsfaktor $(1-d)^{-1}$ und entlang der Ordinate $(1-ROC(d))^{-1}$ anzuwenden sind. Für numerisch ermittelte Beispiele selektierter Portfolios für verschiedene Werte für d , die auf einer α -ROC-Kurvenfunktion mit $\alpha=0,3$ des vollständigen Portfolios basieren, siehe Abbildung 48.

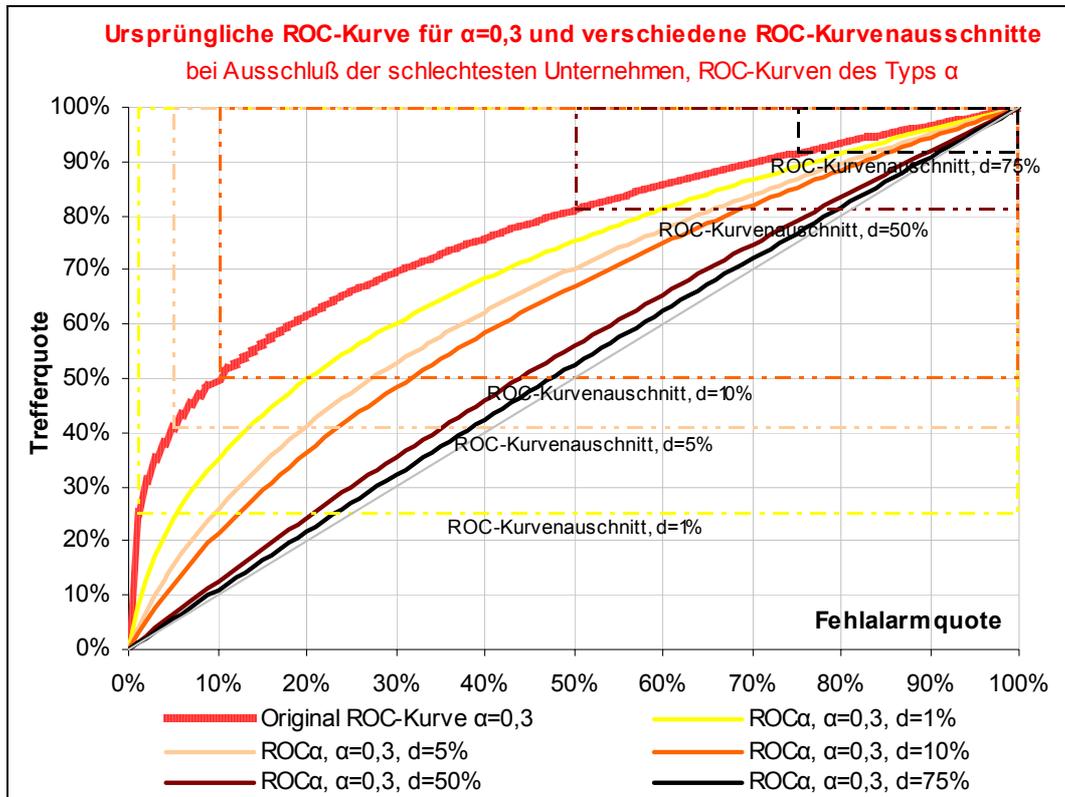


Abbildung 48: Gegenüberstellung der ROC-Kurven eines vollständigen und verschiedener selektierter Portfolios für ein numerisches Beispiel, ROC_{α} -Kurve für $\alpha=0,3$, verschiedene Werte für d

Allgemein gilt für die Fläche unter der skalierten ROC-Kurve des selektierten Portfolios, $AUC_{ROC,d}$, in Abhängigkeit der ROC-Kurvenfunktion des vollständigen Portfolios:

$$F\ 241) \quad AUC_{ROC,d} = \frac{\int_0^1 ROC(x) dx - (1-d) \cdot ROC(d)}{(1-d) \cdot (1-ROC(d))}$$

Im weiteren wird – basierend auf $AUC_{ROC,d}$ – die zugehörige Accuracy Ratio in Abhängigkeit von d errechnet und ins Verhältnis zur Accuracy Ratio des vollständigen Portfolios gesetzt. Für die ROC-Kurven werden die in Anhang I und IV entwickelten formal analysierbaren ROC-Kurvenfunktionen $ROC_{\alpha}(x)=x^{\alpha}$ und $ROC_{\beta}(x)=1-(1-x)^{1/\beta}$ verwendet. Wie in Anhang I dargelegt, handelt es sich bei diesen Funktionen um Extremfälle. Die untersuchten realen ROC-Kurven lassen sich als Mischungen beider Funktionen darstellen, wobei manche reale ROC-Kurven eher dem α -Typ und andere eher dem β -Typ entsprechen. Bei einer durch den Parameter α bzw. β determinierten Prognosegüte eines Verfahrens (siehe die Formeln F 224 und F 235), sind ROC_{α} -Kurven durch eine hohe Diskriminanzfähigkeit im Bereich der Unternehmen mit schlechten Ratings gekennzeichnet und durch eine niedrige Diskriminanzfähigkeit im Bereich der Unternehmen mit guten Ratings (siehe Anhang I). Bei ROC_{β} -Kurven verhält es sich genau umgekehrt.

Somit ist damit zu rechnen, daß die relative Leistungsfähigkeit beider Verfahren bei der Aus-sortierung von Unternehmen mit *schlechten Ratings* stark unterschiedlich betroffen sein wird. Bei ROC_{α} -Kurven ist, verglichen mit ROC_{β} -Kurven, ein deutlich stärkerer Rückgang zu erwarten, da hier gerade die Unternehmen aus dem Sample entfernt werden, bei denen ROC_{α} -

Verfahren „komparative Vorteile“ gegenüber ROC_{β} -Verfahren aufweisen. Bei der Aussortierung von Unternehmen mit guten Ratings ist mit dem umgekehrten Ergebnis zu rechnen.

Für ROC-Kurven des α -Typs gilt:

$$F 242) \int_d^1 ROC_{\alpha}(x) dx = \int_d^1 x^{\alpha} dx = \left[\frac{x^{\alpha+1}}{\alpha+1} \right]_d^1$$

$$F 243) \int_d^1 x^{\alpha} dx = \frac{1-d^{\alpha+1}}{\alpha+1} \quad \text{eingesetzt in Formel F 241 ergibt sich:}$$

$$F 244) AUC_{ROC,d,\alpha} = \frac{\frac{1-d^{\alpha+1}}{\alpha+1} - (1-d) \cdot d^{\alpha}}{(1-d) \cdot (1-d^{\alpha})}$$

$$F 245) AUC_{ROC,d,\alpha} = \frac{1-d^{\alpha+1} - (d^{\alpha} - d^{\alpha+1}) \cdot (\alpha+1)}{(\alpha+1) \cdot (1-d) \cdot (1-d^{\alpha})}$$

$$F 246) AUC_{ROC,d,\alpha} = \frac{1-d^{\alpha} + \alpha d^{\alpha} \cdot (d-1)}{(\alpha+1) \cdot (1-d) \cdot (1-d^{\alpha})}$$

$$F 247) AUC_{ROC,d,\alpha} = \frac{1-d^{\alpha} + \alpha d^{\alpha+1} - \alpha d^{\alpha}}{\alpha - \alpha d - \alpha d^{\alpha} + \alpha d^{\alpha+1} + 1 - d - d^{\alpha} + d^{\alpha+1}}$$

Gemäß Formel F 8 ergibt sich die Accuracy Ratio in Abhängigkeit von AUC wie folgt:

$$F 248) AR_{d,\alpha} = 2 \cdot AUC_{ROC,d,\alpha} - 1 \quad \text{und damit}$$

$$F 249) AR_{d,\alpha} = \frac{2 - 2d^{\alpha} + 2\alpha d^{\alpha+1} - 2\alpha d^{\alpha} - \alpha + \alpha d + \alpha d^{\alpha} - \alpha d^{\alpha+1} - 1 + d + d^{\alpha} - d^{\alpha+1}}{\alpha - \alpha d - \alpha d^{\alpha} + \alpha d^{\alpha+1} + 1 - d - d^{\alpha} + d^{\alpha+1}}$$

$$F 250) AR_{d,\alpha} = \frac{1 - \alpha - d^{\alpha} + \alpha d^{\alpha+1} - \alpha d^{\alpha} + \alpha d + d - d^{\alpha+1}}{(\alpha+1) \cdot (1-d) \cdot (1-d^{\alpha})}$$

Die relative Accuracy Ratio, $AR_{rel,\alpha,d}$, welche die Accuracy Ratio des selektierten Portfolios ins Verhältnis zur Accuracy Ratio des vollständigen Portfolios setzt, ist wie folgt definiert:

$$F 251) AR_{rel,\alpha,d} = \frac{AR_{\alpha,d}}{AR_{\alpha}} \quad \text{und somit ergibt sich in Zusammenhang mit Formel F 224:}$$

$$F 252) AR_{rel,\alpha,d} = \frac{1 - \alpha - d^{\alpha} + \alpha d^{\alpha+1} - \alpha d^{\alpha} + \alpha d + d - d^{\alpha+1}}{(1-\alpha) \cdot (1-d) \cdot (1-d^{\alpha})}$$

Dieser Term läßt sich nicht mehr wesentlich vereinfachen, und die Ableitungen nach d und α sind sehr unübersichtlich. Im folgenden wird er deshalb nur numerisch für verschiedene Kombinationen von d und α untersucht. Die Ergebnisse werden in Abbildung 49 dargestellt.

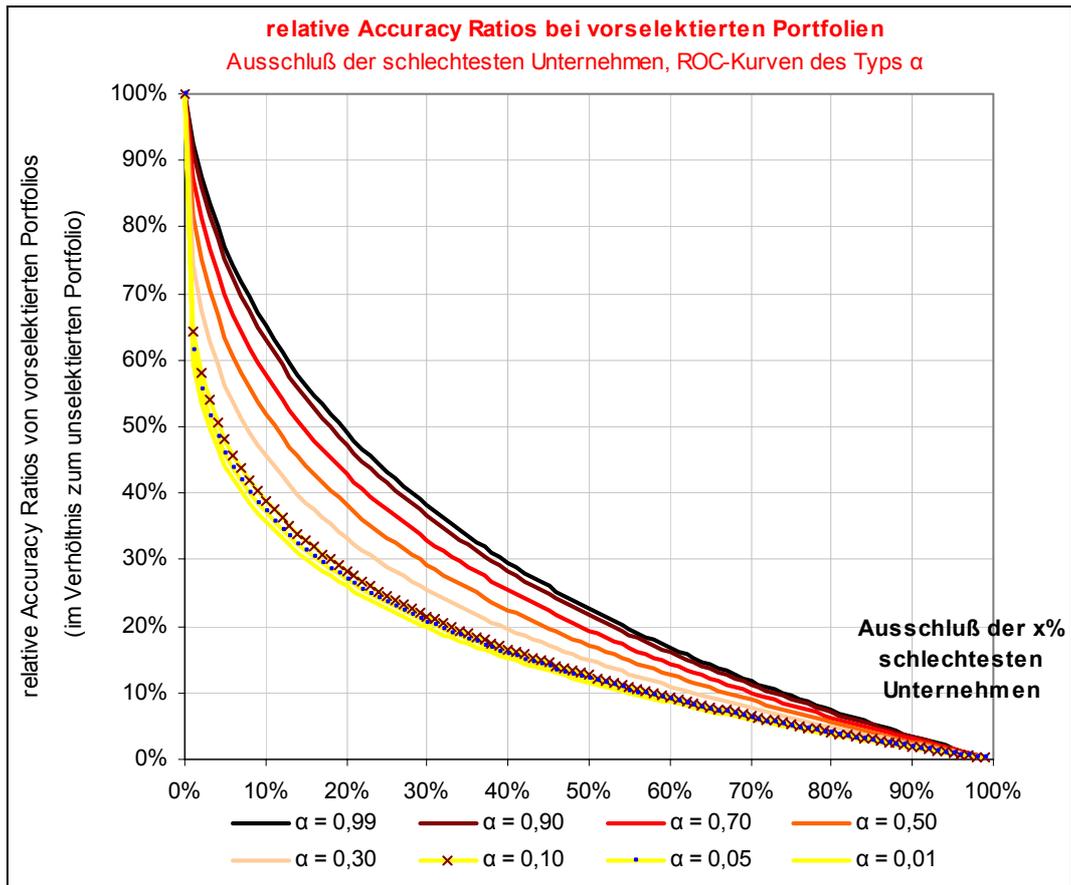


Abbildung 49: Relative Accuracy Ratios vorselektierter Portfolios bei Modellierung der ROC-Kurve des vollständigen Portfolios mittels ROC-Funktionen des α -Typs, Darstellung für verschiedene Werte für α und d (Anteil der ausgeschlossenen „schlechten“ Unternehmen)

Dabei zeigt sich, was zu erwarten war, daß die relativen Accuracy Ratios der selektierten Portfolios um so schlechter (d.h. niedriger) sind, je mehr Unternehmen ausgeschlossen werden, d.h. je größer d ist. Ferner sind die relativen Informationsverluste auch um so größer, je trennschärfer das Verfahren auf Basis des vollständigen Portfolios ist, d.h. je kleiner α ist, wobei sich die Ergebnisse innerhalb des Intervalls mit realistischen α -Werten ($0,1 < \alpha < 0,5$, siehe auch die Anhänge I und IV) nur relativ geringfügig unterscheiden. Entscheidend werden die Ergebnisse hingegen von d , dem Anteil der ausgeschlossenen Unternehmen determiniert. Werden nur 10% der Unternehmen mit den schlechtesten Ratings ausgeschlossen, so reduziert sich die auf Basis des selektierten Portfolios gemessene Schätzgüte um über die Hälfte! Werden 50% der Unternehmen ausgeschlossen, sinkt die gemessene Schätzgüte auf Basis des selektierten Portfolios sogar auf unter ein Sechstel der ursprünglichen Schätzgüte.

Im folgenden werden obige Analysen auch für ROC-Kurven des Typs β durchgeführt und anschließend anhand empirischer Beispiele überprüft.

Für ROC-Kurven des β -Typs gilt:

$$\text{F 253) } \int_d^1 \text{ROC}_\beta(x) dx = \int_d^1 1 - (1-x)^{1/\beta} dx = \left[x + \frac{\beta}{1+\beta} \cdot (1-x)^{(1+\beta)/\beta} \right]_d^1 \quad \text{und damit}$$

$$\text{F 254) } \int_d^1 1 - (1-x)^{1/\beta} dx = 1 - \left(d + \frac{\beta}{1+\beta} \cdot (1-d)^{(1+\beta)/\beta} \right) \quad \text{vereinfacht zu}$$

$$\text{F 255) } \int_d^1 1 - (1-x)^{1/\beta} dx = 1 - d - \frac{(1-d) \cdot \beta}{1+\beta} \cdot (1-d)^{1/\beta} \quad \text{und schließlich}$$

$$\text{F 256) } \int_d^1 1 - (1-x)^{1/\beta} dx = \frac{(1-d) \cdot (1+\beta - \beta \cdot (1-d)^{1/\beta})}{1+\beta}$$

Eingesetzt in Formel F 241 für die Ermittlung von $AUC_{ROC,d,\beta}$ ergibt sich:

$$\text{F 257) } AUC_{ROC,d,\beta} = \frac{\int_d^1 ROC_{\beta}(x) dx - (1-d) \cdot ROC_{\beta}(d)}{(1-d) \cdot (1 - ROC_{\beta}(d))}$$

$$\text{F 258) } AUC_{ROC,d,\beta} = \frac{\frac{(1-d) \cdot (1+\beta - \beta \cdot (1-d)^{1/\beta})}{1+\beta} - (1-d) \cdot (1 - (1-d)^{1/\beta})}{(1-d) \cdot (1 - (1 - (1-d)^{1/\beta}))}$$

$$\text{F 259) } AUC_{ROC,d,\beta} = \frac{\frac{1+\beta - \beta \cdot (1-d)^{1/\beta}}{1+\beta} - 1 + (1-d)^{1/\beta}}{(1-d)^{1/\beta}}$$

$$\text{F 260) } AUC_{ROC,d,\beta} = \frac{1+\beta - \beta \cdot (1-d)^{1/\beta} - (1+\beta) + (1+\beta) \cdot (1-d)^{1/\beta}}{(1+\beta) \cdot (1-d)^{1/\beta}}$$

$$\text{F 261) } AUC_{ROC,d,\beta} = \frac{(1-d)^{1/\beta}}{(1+\beta) \cdot (1-d)^{1/\beta}}$$

$$\text{F 262) } AUC_{ROC,d,\beta} = \frac{1}{1+\beta} = AUC_{ROC,\beta} \quad \text{für alle } d, \text{ vgl. mit Formel F 233!}$$

Die formalen Untersuchungen besagen, daß wenn die ROC-Kurve des vollständigen Portfolios dem β -Typ entsprechen, bei Ausschluß der Unternehmen mit der schlechtesten Bonität *keinerlei* Informationsverluste bei den selektierten Portfolios zu verzeichnen sind – und zwar unabhängig von der Schätzgüte des Verfahrens oder dem Anteil der ausgeschlossenen Unternehmen (auf eine grafische Darstellung der Ergebnisse wird deshalb verzichtet).³⁴² Selbst bei einem Ausschluß von 50%, 90%, 99%, etc. aller Unternehmen bleibt die (erwartete) Schätzgüte auf Basis der Restportfolios unverändert – während bei ROC-Kurven des α -Typs bereits Informationsverluste von über 50% zu verzeichnen sind, wenn nur 10% der Unternehmen ausgeschlossen werden.

Die Ergebnisse zeigen, daß die durch die Aussortierung von Unternehmen mit schlechten Bonitäten entstehenden Schätzgüteeinbußen – bei einer Schätzgütemessung auf Basis des Restportfolios – maßgeblich davon abhängen, ob die Verfahren eher im Bereich von „schlechten“ Unternehmen gut differenzieren können (und damit eher dem α -Typus entsprechen) oder im Bereich von „guten“ Unternehmen (und damit eher dem β -Typus entsprechen).

Im folgenden soll geprüft werden, ob sich die aus obigen formalen Analysen ergebenden erheblichen Unterschiede hinsichtlich der „Selektionssensitivität“ von Verfahren auch für reale

³⁴² Es sind auch Verfahren denkbar, deren Schätzgüte durch den Ausschluß von Unternehmen mit den schlechtesten Ratings *steigt*. Siehe beispielsweise eine ROC-Kurve mit minimaler Diskriminanzfähigkeit bei „schlechten“ Unternehmen und maximaler Diskriminanzfähigkeit bei „guten“ Unternehmen [linearer Verlauf der ROC-Kurve für niedrige Fehler II. Art bis zum Schnittpunkt mit der 100%-Trefferquoten-Gerade]. Wird ein Teil a , $0 < a < 1$, der „schlechten“ Unternehmen ausgeschlossen, so verläuft die ROC-Kurve des selektierten Portfolios weiterhin linear bis zum Schnittpunkt mit der 100%-Trefferquoten-Gerade – allerdings mit einem steileren Anstieg, so daß sich AUC_{ROC} und AR erhöhen.

Verfahren anhand empirischer Daten bestätigen läßt. Dies geschieht mittels drei bereits in Anhang I untersuchten realen Insolvenzprognoseverfahren, siehe die folgende Abbildung 50.

Dargestellt werden die Ergebnisse für die Ratings von S&P und MOODY'S, deren ROC-Kurven eher dem β -Typ entsprechen, und für den CREDITREFORM Bonitätsindex Deutschland, dessen ROC-Kurve fast perfekt dem α -Typ entspricht (siehe Anhang I).

Für die Ermittlung der „Selektionssensitivität“ der einzelnen Ratingverfahren werden zunächst die Accuracy Ratios auf Basis der jeweils vorliegenden ratingklassenspezifischen Angaben hinsichtlich Häufigkeiten und realisierten Ausfallquoten der Unternehmen gemäß Formel F 21 ermittelt. Anschließend werden schrittweise die Anteile der Unternehmen der jeweils verbliebenen schlechtesten Ratingklasse auf Null gesetzt und die Accuracy Ratios der Restportfolios ermittelt. Dieser Schritt wird solange wiederholt, bis die Restportfolios nur noch aus (homogenen) Unternehmen der besten Ratingstufe bestehen.

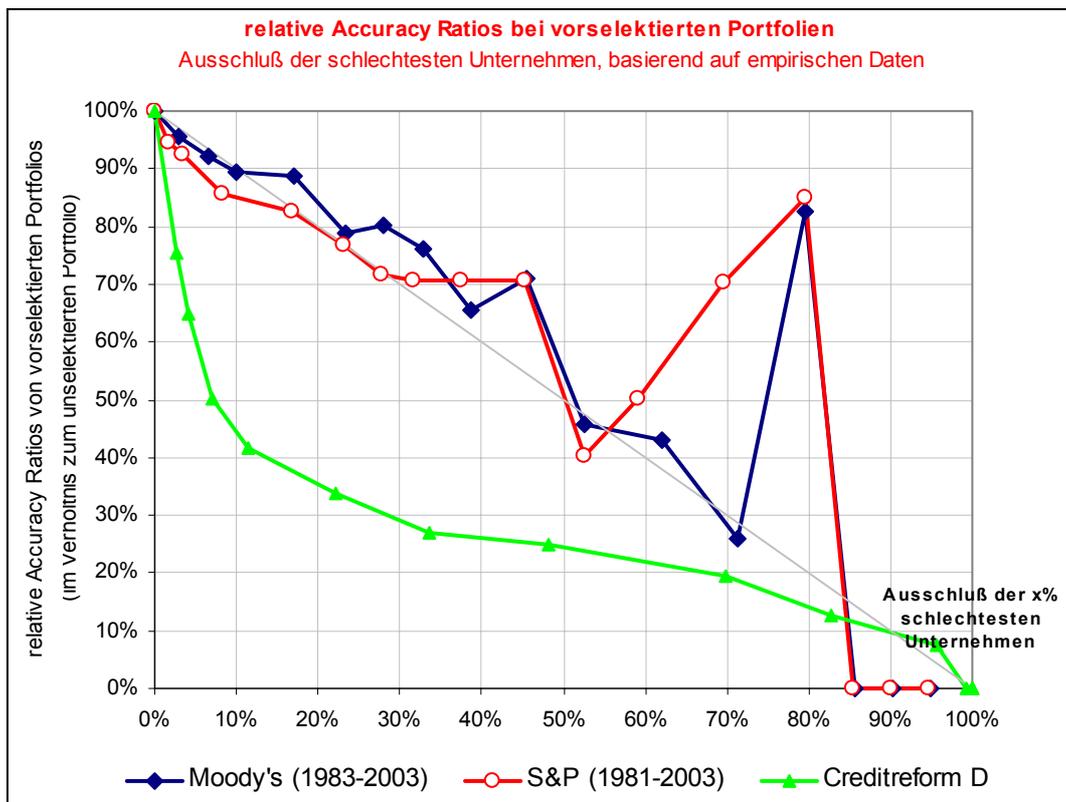


Abbildung 50: Relative Accuracy Ratios bei vorselektierten Portfolios, Ausschluß der schlechtesten Unternehmen, basierend auf empirischen Daten von MOODY'S (1983-2003), S&P (1981-2003) und CREDITREFORM Bonitätsindex Deutschland (1998-2000), Quelle: eigene Auswertungen

Wie zu erwarten war, entspricht der Verlauf der relativen Accuracy Ratios des CREDITREFORM-Bonitätsindex dem für ROC_{α} -Kurven ermittelten Verlauf (vgl. mit Abbildung 49, $\alpha=0,3$). Der Ausschluß von nur 10% der Unternehmen mit der schlechtesten Bonität führt zu einem Rückgang der (ohnehin schon sehr niedrigen) Prognosegüte des CREDITREFORM Bonitätsindex (siehe Abschnitt 3.5) für das Restportfolio um 50%-60%!

Gemäß den MOODY'S- und S&P-Ratings ist für Ausschlüsse von Unternehmen mit den schlechtesten Ratings bis rund 50%³⁴³ des ursprünglichen Portfolios mit einem nahezu proportionalen Rückgang der Prognosegüte auf Basis der Restportfolios zu rechnen (siehe Verlauf der relativen-AR-Kurven entlang der Nebendiagonalen), d.h. der Ausschluß von 10% der

³⁴³ Die Grenze zwischen Investment- und Speculative-Grade-Bereich liegt auf Basis der Pooldaten bei beiden Agenturen bei knapp über 30%. Bei rund 50% beginnt hingegen der Bereich der A/AA/AAA (bzw. A/Aa/Aaa-) Ratings.

schlechtesten Unternehmen führt (nur) zu einem Rückgang der Prognosegüte um rund 10%, der Ausschluß von 20% der Unternehmen zu einem Rückgang der Prognosegüte von 20%, usw. Der hintere Bereich des Kurvenverlaufs ist durch Unstetigkeiten geprägt, die durch statistische Artefakte aufgrund der sehr seltenen Ausfälle (innerhalb eines Prognosezeitraums von einem Jahr) im Investmentgradbereich, speziell im AA- und AAA-Bereich, hervorgerufen werden.³⁴⁴

Der Rückgang der relativen Accuracy-Ratio-Werte auf 0% für hinreichend große Ausschlüsse ist – im Gegensatz zur oben formal hergeleiteten Selektionssensitivität bei reinen ROC_α-Kurven – auch darauf zurückzuführen, daß für die Untersuchungen nur klassierte Daten vorlagen und somit (spätestens) bei Verbleib nur noch einer Ratingklasse die ermittelte Schätzgüte der Verfahren auf Null sinken muß.

Der Vollständigkeit halber werden im folgenden auch die Konsequenzen auf die ermittelte Schätzgüte auf Basis der Restportfolien bei einer Aussortierung der Unternehmen mit den *besten* Ratings ermittelt. Relevant sind die Ausführungen beispielsweise für Investoren, die sich bewußt auf Unternehmen mit geringerer Bonität spezialisieren wollen oder aber auch für Banken, falls es denen nicht gelingt, Unternehmen mit einer besonders guten Bonität als Kunden zu gewinnen.

Verbleiben nur die u% „schlechtesten“ Unternehmen im Portfolio (u wird im folgenden als *Abschneidewert* bezeichnet), so ergibt sich, analog zu den Ausführungen zu Beginn dieses Abschnitts die Fläche unter der skalierten ROC-Kurve des selektierten Portfolios, AUC_{ROC,u}, in Abhängigkeit der ROC-Kurvenfunktion des vollständigen Portfolios zu:

$$\text{F 263) } AUC_{\text{ROC},u} = \frac{\int_0^u \text{ROC}(x) dx}{u \cdot \text{ROC}(u)}$$

Für Kurven des α-Typs gilt dabei:

$$\text{F 264) } \int_0^u \text{ROC}_\alpha(x) dx = \int_0^u x^\alpha dx = \left[\frac{x^{\alpha+1}}{\alpha+1} \right]_0^u \quad \text{und somit}$$

$$\text{F 265) } \int_0^u \text{ROC}_\alpha(x) dx = \frac{u^{\alpha+1}}{\alpha+1}$$

Eingesetzt in Formel F 263 ergibt sich:

$$\text{F 266) } AUC_{\text{ROC},\alpha,u} = \frac{u^{\alpha+1}}{u \cdot u^\alpha} \quad \text{und damit}$$

$$\text{F 267) } AUC_{\alpha,u} = \frac{1}{\alpha+1} = AUC_\alpha \quad \text{für alle } u \text{ (vgl. Formel F 222)!}$$

³⁴⁴ In beiden Fällen werden sehr hohe Schätzgüten auf Basis der Restportfolien erzielt, wenn sämtliche Unternehmen der Ratingklassen BBB+ bis CCC/C (bzw. Baa1-Caa/C) entfernt wurden, da historisch sowohl gemäß den Ratingeinstufungen nach S&P als auch nach MOODY'S Einjahresausfälle nur in der Ratingklasse AA- (bzw. Aa3) auftraten, nicht jedoch in den restlichen Ratingklassen AA, AA+ und AAA (bzw. Aa2, Aa1 und Aaa). Wird im nächsten Schritt die Ratingklasse AA- (Aa3) entfernt, sinkt die Prognosegüte auf Basis des Restportfolios auf Null, da alle verbleibenden Ratingklassen, dann durch die gleichen Ausfallraten (0% bezogen auf einen Einjahreszeitraum) gekennzeichnet sind.

Dies bedeutet, daß bei ROC-Kurven des α -Typs ein Ausschluß von (beliebig vielen) Unternehmen mit den besten Ratings zu *keinem Verlust der Prognosegüte* auf Basis des Restportfolios führt – unabhängig von u und α (auf eine grafische Darstellung wird deshalb verzichtet).

Für Kurven des β -Typs gilt hingegen:

$$\text{F 268) } \int_0^u \text{ROC}_\beta(x) dx = \int_0^u 1 - (1-x)^{1/\beta} dx = \left[x + \frac{\beta}{1+\beta} \cdot (1-x)^{(1+\beta)/\beta} \right]_0^u$$

$$\text{F 269) } \int_0^u \text{ROC}_\beta(x) dx = u + \frac{\beta}{1+\beta} \cdot (1-u)^{(1+\beta)/\beta} - \frac{\beta}{1+\beta}$$

Eingesetzt in Formel F 263 ergibt sich somit:

$$\text{F 270) } \text{AUC}_{\text{ROC},\beta,u} = \frac{u + \frac{\beta}{1+\beta} \cdot (1-u)^{(1+\beta)/\beta} - \frac{\beta}{1+\beta}}{u \cdot (1 - (1-u)^{1/\beta})}$$

$$\text{F 271) } \text{AUC}_{\beta,u} = \frac{u \cdot (1+\beta) + (1-u) \cdot \beta \cdot (1-u)^{1/\beta} - \beta}{(1+\beta) \cdot (u - u \cdot (1-u)^{1/\beta})}$$

$$\text{F 272) } \text{AUC}_{\text{ROC},\beta,u} = \frac{u + u \cdot \beta + \beta \cdot (1-u)^{1/\beta} - u \cdot \beta \cdot (1-u)^{1/\beta} - \beta}{u + u \cdot \beta - u \cdot (1-u)^{1/\beta} - u \cdot \beta \cdot (1-u)^{1/\beta}}$$

Für die *Accuracy Ratio* ergibt sich damit analog zu Formel F 248:

$$\text{F 273) } \text{AR}_{\beta,u} = \frac{2u + 2u \cdot \beta + 2\beta \cdot (1-u)^{1/\beta} - 2u \cdot \beta \cdot (1-u)^{1/\beta} - 2\beta - u - u \cdot \beta + u \cdot (1-u)^{1/\beta} + u \cdot \beta \cdot (1-u)^{1/\beta}}{u + u \cdot \beta - u \cdot (1-u)^{1/\beta} - u \cdot \beta \cdot (1-u)^{1/\beta}}$$

$$\text{F 274) } \text{AR}_{\beta,u} = \frac{u + u \cdot \beta - 2\beta + (1-u)^{1/\beta} \cdot (2\beta - u \cdot \beta + u)}{(1+\beta) \cdot u \cdot (1 - (1-u)^{1/\beta})}$$

Für die *relative Accuracy Ratio* $\text{AR}_{\text{rel},\beta,u}$, ergibt sich gemäß den Formeln F 251 und F 233:

$$\text{F 275) } \text{AR}_{\text{rel},\beta,u} = \frac{\text{AR}_{\beta,u}}{\text{AR}_\beta} = \frac{\text{AR}_{\beta,u}}{\left(\frac{1-\beta}{1+\beta} \right)} \quad \text{und damit}$$

$$\text{F 276) } \text{AR}_{\text{rel},\beta,u} = \frac{u + u \cdot \beta - 2\beta + (1-u)^{1/\beta} \cdot (2\beta - u \cdot \beta + u)}{(1-\beta) \cdot u \cdot (1 - (1-u)^{1/\beta})}$$

Dieser Term wird im folgenden für verschiedene u und β numerisch ermittelt und grafisch dargestellt, siehe die folgende Abbildung 51.

Dabei zeigt sich, daß die relativen Accuracy Ratios der selektierten Portfolios um so schlechter (d.h. niedriger) sind, je mehr gute Unternehmen ausgeschlossen werden, d.h. je kleiner der „Abschneidewert“ u ist.

Ferner sind die relativen Informationsverluste um so *kleiner*, je trennschärfer die Verfahren auf Basis der vollständigen Portfolios sind, d.h. je kleiner β ist. Bei sehr trennscharfen Verfahren ($\beta \geq 0,1$) nimmt die relative Schätzgüte auf Basis der Restportfolios erst für sehr kleine u -Werte, d.h. wenn sehr viele der besten Unternehmen ausgeschlossen werden, dramatisch ab. Im relevanten Bereich (für große u und kleine α bzw. β) reagieren die ermittelten Ergebnisse somit nicht so sensitiv auf Modellierungsdetails (Verwendung von α - oder β -ROC-Kurven) wie im Fall des Ausschlusses von *schlechten* Unternehmen.

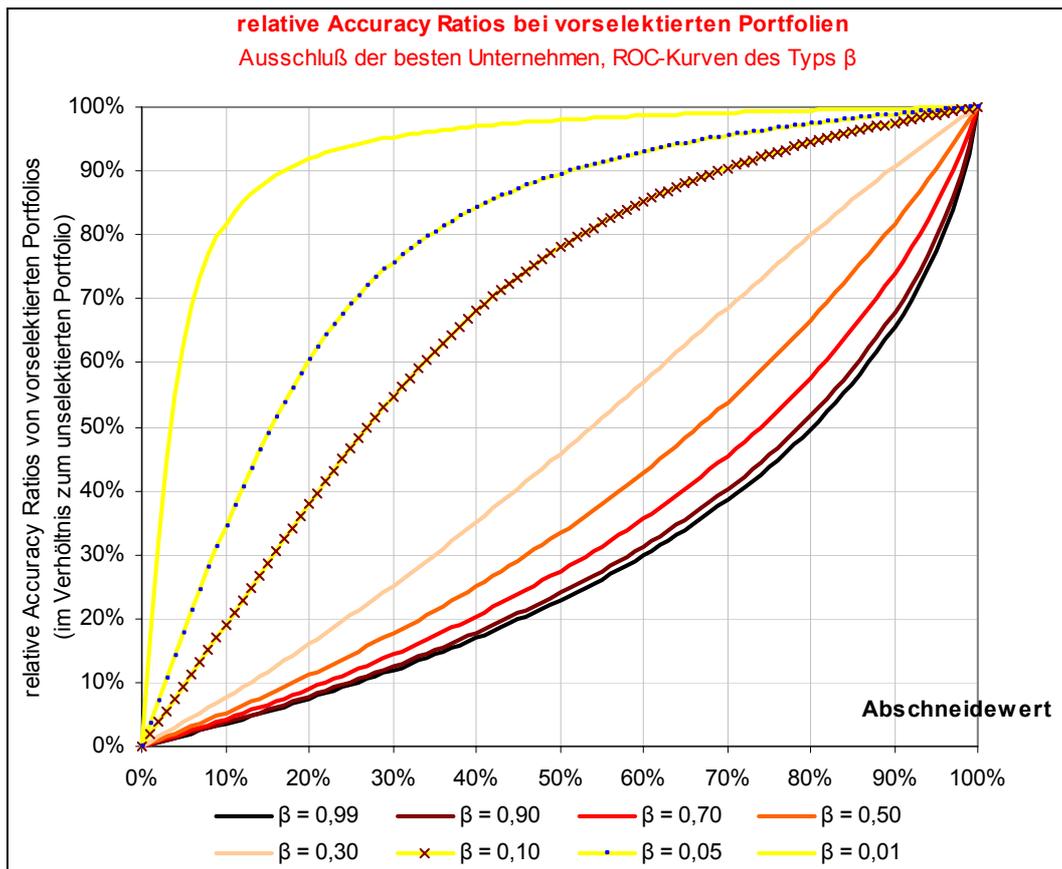


Abbildung 51: Relative Accuracy Ratios vorselektierter Portfolien bei Modellierung der ROC-Kurve des vollständigen Portfolios mittels ROC-Funktionen des β -Typs, Darstellung für verschiedene Werte für β und u (Abschneidewert),

Im folgenden wird wiederum die „Selektionssensitivität“ realer Insolvenzprognoseverfahren anhand empirischer Daten direkt ermittelt, siehe Abbildung 52. Im Gegensatz zu dem oben beschriebenen Vorgehen werden hier jedoch nicht schrittweise die Unternehmen der *schlechtesten*, sondern die Unternehmen der *besten* Ratingklassen entfernt.

Die auf Basis der empirischen Daten gefundenen Ergebnisse lassen sich gut durch die theoretischen Analysen erklären. So verringert sich beim CREDITREFORM Bonitätsindex Deutschland, dessen ROC-Kurve fast perfekt dem α -Typ entspricht, die (sehr niedrige) Prognosekraft auf Basis des Restportfolios praktisch überhaupt nicht, selbst wenn ca. 80% der Unternehmen mit den besten Bonitäten aussortiert werden. Erst bei einer Aussortierung von mehr als 80% der Unternehmen nimmt die Schätzgüte auf Basis des Restportfolios spürbar ab, was aber auch daran liegt, daß für die empirischen Analysen nur *ratingklassenspezifische* Angaben (hinsichtlich Häufigkeiten und Ausfallraten) verwendet werden konnten und nicht die quasi-kontinuierlichen Scores. Verbleibt nur noch eine Ratingklasse, so muß die ermittelte Schätzgüte des Verfahrens auf Basis ratingklassenspezifischer Angaben zwangsläufig auf Null sinken.

Der Verlauf der relativen-Accuracy-Ratio-Kurven gemäß den S&P- und MOODY'S-Ratings, deren ROC-Kurven eher dem β -Typ entsprechen, läßt sich gut durch einen wie in Abbildung 51, für $\beta=0,1$, dargestellten Verlauf beschreiben.

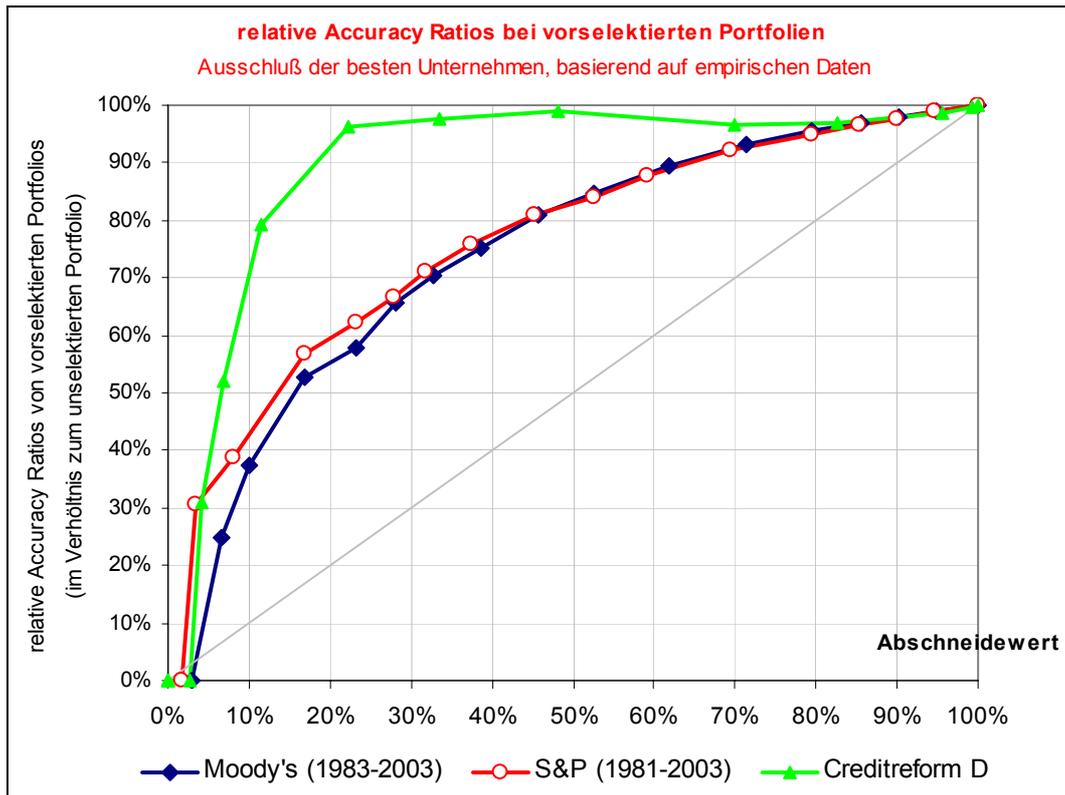


Abbildung 52: Relative Accuracy Ratios bei vorselektierten Portfolios, Ausschluß der besten Unternehmen, basierend auf empirischen Daten von MOODY'S (1983-2003), S&P (1981-2003) und CREDITREFORM Bonitätsindex Deutschland (1998-2000), Quelle: eigene Auswertungen

Insgesamt lassen sowohl die theoretischen als auch die empirischen Ergebnisse darauf schließen, daß eine Aussortierung von Unternehmen mit guten Bonitäten zu weitaus weniger dramatischen und deutlich unterproportionalen Einbußen bei den ermittelten Schätzgüten auf Basis der Restportfolios führt als eine Aussortierung von Unternehmen mit schlechten Bonitäten.

Als Faustregel läßt sich formulieren, daß sich je Prozent ausgeschlossener Unternehmen mit guter Bonität die Schätzgüte des Verfahrens – gemessen auf dem Restportfolio – um ca. ¼ Prozent verringert, d.h. bei Ausschluß von 20% der besten Unternehmen ist mit einem Rückgang der Schätzgüte um (nur) ca. 5% und bei einem Ausschluß von 40% der Unternehmen um (nur) ca. 10% oder weniger zu rechnen.

Dresden Discussion Paper Series in Economics

- 17/03 **Lehmann-Waffenschmidt, Marco / Reina, Livia:** Coalition formation in multilateral negotiations with a potential for logrolling: an experimental analysis of negotiators' cognition processes
- 18/03 **Lehmann-Waffenschmidt, Marco / Böhmer, Robert:** Mentality Matters – Thorstein Veblens ‚Regime of Status‘ und Max Webers ‚Protestantische Ethik‘ aus der Sicht des (radikalen) Konstruktivismus. Eine Anwendung auf die ökonomischen Probleme des deutschen Wiedervereinigungsprozesses
- 19/03 **Eisenschmidt, Jens / Wälde, Klaus:** International Trade, Hedging and the Demand for Forward Contracts
- 20/03 **Broll, Udo / Wong, Kit Pong:** Capital Structure and the Firm under Uncertainty
- 01/04 **Lehmann-Waffenschmidt, Marco:** A Note on Continuously Decomposed Evolving Exchange Economies
- 02/04 **Friedrich, B. Cornelia:** Competition and the Evolution of Market Structure in the E-conomy.
- 03/04 **Berlemann, Michael / Dittrich, Marcus / Markwardt, Gunther:** The Value of Non-Binding Announcements in Public Goods Experiments. Some Theory and Experimental Evidence
- 04/04 **Blum, Ulrich / Schaller, Armin / Veltins, Michael:** The East German Cement Cartel: An Inquiry into Comparable Markets, Industry Structure, and Antitrust Policy
- 05/04 **Schlegel, Christoph:** Analytical and Numerical Solution of a Poisson RBC model
- 06/04 **Lehmann-Waffenschmidt, Marco:** Die ökonomische Botschaft in Goethes „Faust“
- 07/04 **Fuchs, Michaela / Thum, Marcel:** EU Enlargement: Challenges for Germany's New Laender
- 08/04 **Seitz, Helmut:** Implikationen der demographischen Veränderungen für die öffentlichen Haushalte und Verwaltungen
- 09/04 **Sülzle, Kai:** Duopolistic Competition between Independent and Collaborative Business-to-Business Marketplaces
- 10/04 **Broll, Udo / Eckwert, Bernhard:** Transparency in the Interbank Market and the Volume of Bank Intermediated Loans
- 11/04 **Thum, Marcel:** Korruption
- 12/04 **Broll, Udo / Hansen, Sabine / Marjit, Sugata:** Domestic labor, foreign capital and national welfare
- 13/04 **Nyamtseren, Lhamsuren:** Challenges and Opportunities of Small Countries for Integration into the Global Economy, as a Case of Mongolia
- 01/05 **Schubert, Stefan / Broll, Udo:** Dynamic Hedging of Real Wealth Risk
- 02/05 **Günther, Edeltraud / Lehmann-Waffenschmidt, Marco:** Deceleration - Revealed Preference in Society and Win-Win-Strategy for Sustainable Management. Concept and Experimental Evidence
- 03/05 **Sennewald, Ken:** Controlled Stochastic Differential Equations under Poisson Uncertainty and with Unbounded Utility
- 04/05 **Sennewald, Ken / Wälde, Klaus:** "Itô's Lemma" and the Bellman equation: An applied view
- 05/05 **Neumann, Anne / Siliverstovs, Boriss:** Convergence of European Spot Market Prices for Natural Gas?
- 06/05 **Hirschhausen, Christian von / Cullmann, Astrid:** Efficiency Analysis of German Electricity Distribution Utilities
- 07/05 **Seitz, Helmut / Freigang, Dirk / Kempkes, Gerhard:** Demographic Change and Federal Systems: Some Preliminary Results for Germany
- 08/05 **Bemann, Martin:** Verbesserung der Vergleichbarkeit von Schätzgüteregebnissen von Insolvenzprognosestudien

