

Basel II:

Trennschärfemaße zur Validierung von internen Rating-Systemen

Hermann Schulte-Mattler/Ulrich Daun/Thorsten Manns

Ende Juni 2004 veröffentlichte der Baseler Ausschuss für Bankenaufsicht das endgültige Basel-II-Papier. Im Rahmen des Internal Ratings-Based Approach (IRB-Ansatz) wird den Instituten damit erstmalig die Anwendung bankinterner Rating-Systeme zur Berechnung der Eigenkapitalunterlegung gestattet. Wie Ihnen unser Beitrag zeigt, kann die Qualität des verwendeten Rating-Systems durch die Trennschärfe beschrieben und grafisch bzw. numerisch angegeben werden.

Würde eine Bank einen Hellseher zur Rating-Einstufung beschäftigen, hätte sie das „perfekte System“

Das endgültige Papier des Baseler Ausschusses für Bankenaufsicht „International Convergence of Capital Measurement and Capital Standards, A Revised Framework“¹, kurz Basel II genannt, sieht im IRB-Ansatz vor, dass Banken die Eigenkapitalanforderung für einen vergebenen Kredit auf Basis einer schulderspezifischen Prognose der Ausfallwahrscheinlichkeit (Probability of Default, PD) berechnen. Für die Schätzung der Ausfallwahrscheinlichkeit

eines Schuldners hat das Verfahren der logistischen Regression in den letzten Jahren eine zunehmende Verbreitung in der Wissenschaft wie auch in der Praxis erfahren.² So sind im Vergleich zur Diskriminanzanalyse beispielsweise die direkte Modellierung der PD sowie die Robustheit der Schätzergebnisse als Vorteile zu nennen. Jede PD wird eindeutig, zeitlich stabil und auf historischen Daten basierend einer bankinternen Bonitäts- oder Rating-Klasse zugeordnet.

CAP- und ROC-Kurve und Gini-Koeffizient

Die Anwendung des IRB-Ansatzes bedarf der Prüfung und Zustimmung der Bankenaufsicht (in Deutschland also der Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin).³ Der gesamte Prozess zur Überprüfung eines internen Rating-Systems wird „Validierung“ genannt. Im Rahmen der qualitativen Validierung wird das Modelldesign, die Datenqualität und die interne Verwendung des Rating-Systems (Use-Test) in Augenschein genommen. Das Backtesting und das Benchmarking sind Prüfbereiche der quantitativen Validierung. Die BaFin wird also sowohl die Trennschärfe (Info-Box) einzelner Risiko-Parameter als auch den gesamten Rating-Prozess zu beurteilen haben.⁴ In der Praxis spielen dabei in der Regel zwei verwendete Instrumente der Trennschärfeanalyse eine Rolle: das Cumulative Accuracy Profile (CAP) und die Receiver Operating Characteristic (ROC). Die grafischen Informationen der CAP- und ROC-Kurve werden einerseits im Gini-Koeffizient und andererseits im Flächeninhalt unter der ROC-Kurve (Area under the ROC curve, AUROC) verdichtet. Beide Kennzahlen und die Beziehung zwischen ihnen werden ebenfalls analysiert. Da ein Rating-System nur zwischen

solchen Schuldnern differenzieren kann, die sich tatsächlich unterscheiden, ist zudem auf die Portfolioabhängigkeit der vorgestellten Trennschärfemaße hinzuweisen.

Was ist die Trennschärfe?

Die Trennschärfe eines Rating-Systems bezeichnet seine Fähigkeit, ex ante zwischen ausgefallenen und nicht ausgefallenen Kreditnehmern zu unterscheiden. Aus statistischer Sicht ist ein Rating-System umso trennschärfer, je besser es im Voraus ausfallgefährdete Schuldner erkennen kann. Dieses Gütemerkmal ist unmittelbar plausibel, da ein

des Schuldners und desto niedriger ist sein Ausfallrisiko.⁵ Tabelle 1 zeigt, in welche bankinterne Rating-Klasse ein Schuldner in Abhängigkeit der Höhe seiner Eigenkapitalquote eingeordnet wird (Spalte 1 und 2), wobei Rating-Klasse 1 die beste Kategorie bezeichnet. Des Weiteren informiert Tabelle 1 darüber, wie viele Schuldner einer Bonitätsstufe ex-post (beispielsweise ein Jahr nach einem Rating) tatsächlich insolvent wurden (Spalte 3 und 4). Mit diesen Informationen soll die Trennfähigkeit des Rating-Systems grafisch und numerisch beurteilt werden.

Tabelle 1: Anzahl der ausgefallenen Schuldner im Kreditportfolio einer Bank

Bankinterne Rating-Klasse	Zuordnung von EKQ	Anzahl Schuldner		
		ausgefallen	nicht ausgefallen	Summe
(1)	(2)	(3)	(4)	(5)
1	0,8–1,0	2	28	30
2	0,6–0,8	4	14	18
3	0,4–0,6	10	5	15
4	0,2–0,4	14	2	16
5	0,0–0,2	20	1	21
	Summe	50	50	100

Rating-System, das nicht ein Mindestmaß an Trennschärfe aufweist, zur Bonitätseinstufung ungeeignet ist. Zur Veranschaulichung der Grafen und Kennzahlen zur Trennschärfeanalyse kann ein sehr vereinfachtes fiktives Rating-System zu Grunde gelegt werden. Das Kreditportfolio einer Bank soll aus einhundert Krediten an unterschiedliche Schuldner bestehen. Die Kreditsumme soll für jeden Schuldner gleich sein. Die Schuldner werden nur auf Grundlage ihrer jeweiligen Eigenkapitalquote geratet (EKQ = Eigenkapital / Gesamtkapital). Die Systematik des Ratings ist daher denkbar einfach: Je höher die EKQ, desto höher ist die Bonität

1 Vgl. Basel Committee on Banking Supervision (2004), International Convergence of Capital Measurement and Capital Standards, A Revised Framework, Basel, Juni 2004. Die deutsche Übersetzung liegt jetzt auch als Buch vor, Basel II Textausgabe, Köln 2004. Vgl. auch H. Schulte-Mattler/U. von Kenne: Basel II Framework: Meilenstein der Bankenaufsicht, in: Die Bank, Heft 9, 2004, S. 37-40.
 2 Vgl. H. Schulte-Mattler/U. Daun: Basel II: Logistische Regression als das Herz einer Rating-Maschine, in: RATINGaktuell, 3/2004, S. 66-71.
 3 Vgl. H. Schulte-Mattler: Neue Baseler Eigenkapitalübereinkunft (Basel II): Ein Überblick, in: K.-H. Boos/R. Fischer/H. Schulte-Mattler: Kreditwesengesetz. Kommentar zu KWG und Ausführungsvorschriften, München 22004, S. 2254-2310. Vgl. insbesondere für die viel diskutierte Mittelstandsproblematik H. Schulte-Mattler/T. Manns: Basel II: Falscher Alarm für die Kreditkosten des Mittelstandes, in: Die Bank, Heft 6-7, 2004, S. 376-380.
 4 Vgl. Deutsche Bundesbank (2003), Validierungsansätze für interne Rating-Systeme, in: Monatsbericht September 2003, S. 62.
 5 Es ist explizit darauf hinzuweisen, dass die Rating-Systeme der Banken weitaus komplexer sind und nicht nur auf einer Bilanzkennzahl basieren. Da es als Ausgangspunkt für die im Beitrag behandelte Thematik aber lediglich darauf ankommt, dass Schuldner sich in Rating-Klassen befinden, kann auf eine Darstellung der komplexen Rating-Methodik verzichtet werden.

Was zeigt die CAP-Kurve?

Das Ausmaß der Trennschärfe eines Rating-Systems wird von der CAP-Kurve grafisch veranschaulicht. Die Frage ist, ob ausgefallene Schuldner vom Rating-System grundsätzlich schlechter bewertet werden als nicht ausgefallene Kreditnehmer. Wie in Abb. 1 zu erkennen ist, wird auf der horizontalen Achse der CAP-Kurve der kumulative Anteil aller Schuldner („Alarmquote“) und auf der horizontalen Achse der kumulative Anteil aller ausgefallenen Schuldner abgetragen („Trefferquote“). Wenn beispielsweise unter den 52 % aller Schuldner mit den niedrigsten Bonitätsklassen 88 % aller ausgefallenen Schuldner fallen, dann liegt der Punkt P[0,52; 0,88] auf der CAP-Kurve. Ein Rating-System ist umso trennschärfer, je steiler die CAP-Kurve zu Beginn ansteigt.⁶

Wie berechnet man die CAP-Kurve?

Die bankinternen Daten über ausgefallene Schuldner sind in einer bestimmten Weise aufzubereiten, um die CAP-Kurve zeichnen zu können. Zunächst werden alle Kreditnehmer ihrem Rating entsprechend nach sinkender Bonität angeordnet: Im Beispiel ergibt sich die Reihenfolge der bankinternen Rating-Klassen 5, 4, 3, 2 und 1 (siehe Tabelle 2, Spalte 1). Zunächst werden nur die schlechtesten

Schuldner in der Rating-Klasse 5 betrachtet. Sie haben einen Anteil in Höhe von 21 % im Kreditportfolio und von 40 % an den ausgefallenen Schuldnern (Tabelle 2, Spalte 3 und 5). Werden die Schuldner der zweit schlechtesten Rating-Kategorie 4 zusätzlich in die Analyse mit einbezogen, betrachtet man einen Anteil von 37 % im Kreditportfolio und von 68 % aller ausgefallenen Schuldnern. Durch die Hinzunahmen der übrigen Rating-Klassen (1 bis 3) werden sukzessiv die beiden betrachteten Anteilswerte auf 100 % erhöht. Die Anteile der betrachteten Schuldner trägt man auf der x-Achse und den jeweils zugehörigen Anteil der ausgefallenen Schuldnern auf der y-Achse ab. Die lineare Verbindung dieser Wertepaare ergibt die CAP-Kurve.

Wie sieht die CAP-Kurve für extreme Rating-Systeme aus?

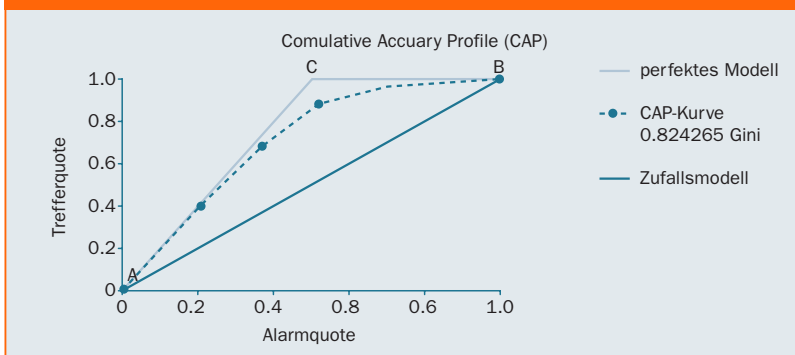
Zwei theoretisch mögliche extreme Verläufe einer CAP-Kurve sind sehr aufschlussreich: das perfekte und das zufällige Rating-System. Im perfekten Fall würde ein Rating-System allen ausgefallenen Schuldnern im Vorhinein die schlechteste Bonitätsklasse zuordnen. Die CAP-Kurve steigt dann zu Beginn linear an und verläuft anschließend horizontal. Ein rein zufälliges Rating-System wäre der andere Grenzfall.

Ein solches System hätte keinerlei Trennfähigkeit. Die CAP-Kurve verläuft in diesem Fall als 45-Grad-Diagonale (Abb. 1). Würde eine Bank einen Hellseher zur Rating-Einstufung beschäftigen, hätte sie das „perfekte System“. Er kennt alle wahren PDs und könnte bereits im Voraus sagen, welche Schuldner in der nächsten Periode insolvent werden. Da der Hellseher sich bei seinen Voraussagen niemals irrt, antizipiert er in unserem Beispiel völlig korrekt, dass 50 % der Schuldner ausfallen werden und zwar unabhängig von ihrer Einstufung mittels der Kennzahl EKQ. Somit begrenzt ein Hellseher die CAP-Kurve „nach oben“. Kein Rating-System kann besser sein als die Weissagungen eines Hellsehers.

Das Zufallsmodell vergibt die Ratings nach dem Zufallsprinzip, so dass die Reihenfolge zwischen den Schuldnern nicht nach bestimmten Kriterien (wie EKQ) gewählt wird. Was bedeutet dies für die CAP-Kurve? Sie verläuft entlang der 45-Grad-Linie: Von zwei zusätzlich betrachteten Schuldnern (entsprechen 2 % des gesamten Portfolios) wird im Schnitt einer zu den Defaultern zählen (entspricht 2 % aller Defaulter). Mit jedem zusätzlichen Schuldner (Erhöhung des Anteils auf der x-Achse) wird im exakt gleichem Ausmaß der Anteil der ausgefallenen Schuldner erhöht (abgetragen auf der y-Achse). Die Stärke eines „echten“ Rating-Systems, das zunächst überproportional viele schlechte Schuldner betrachtet, wird bewusst nicht genutzt. Aus diesem Grund begrenzt das Zufallsmodell die CAP-Kurve „nach unten“.

Abb. 1 zeigt den Verlauf der CAP-Kurve des betrachteten Rating-Mo-

Abb. 1: CAP-Kurve und Gini-Koeffizient



⁶ In Abb. 1 wird bewusst ein flacher Anstieg der CAP-Kurve dargestellt, damit die Ermittlung der erforderlichen Kennzahlen mit dem bloßen Auge erkennbar ist.

Tabelle 2: Aufbereitung der Daten zur Bestimmung der CAP-Kurve

Rating-Klasse geordnet nach aufsteigender Bonität	Anzahl Schuldner ...			
	... an allen Schuldnern (in Prozent)	... an allen Schuldnern (kumuliert)	... an den ausgefallenen Schuldnern (in Prozent)	... an den ausgefallenen Schuldnern (kumuliert)
(1)	(2)	(3)	(4)	(5)
5	21	0	40	0
4	16	21	28	40
3	15	37	20	68
2	18	52	8	88
1	30	70	4	96
		100		100
		Alarmquote		Trefferquote

dells und insbesondere auch die Unterschiede zum perfekten und zufälligen Rating-Modell. In der Praxis sind die Rating-Einstufungen weder perfekt noch zufällig. Die CAP-Kurve verläuft deshalb regelmäßig zwischen diesen beiden Grenzfällen. Der „geknickte“ Verlauf der CAP-Kurve ist der geringen Anzahl der betrachteten Rating-Klassen geschuldet. Wird diese erhöht, würde ihr Verlauf zunehmend „glatter“ erscheinen. Die Informationen der CAP-Kurve zur Trennschärfe eines Rating-Systems lassen sich in einer einzigen Kennzahl zusammenfassen, die nunmehr betrachtet werden soll.

Welche Information gibt der Gini-Koeffizient?

Der Gini-Koeffizient,⁷ auch als Accuracy Ratio, Powerstat oder Somer’s D bezeichnet, gibt an, wie nahe die CAP-Kurve eines Rating-Systems an das perfekte Modell eines Hellsehers kommt oder wie weit entfernt sie vom Zufallsmodell liegt. Zur Berechnung des Giniwertes werden verschiedene Flächen im Diagramm der CAP-Kurve in Beziehung zueinander gesetzt. Als Bezugsgröße für die Bemessung der Trennschärfe wird das Dreieck der Punkte A, B und C herangezogen,

und dann seine Fläche berechnet (also die Fläche die von den CAP-Kurve des perfekten und des zufälligen Modells gebildet wird). Bezeichnet man die Fläche „unterhalb“ der CAP-Kurve des Zufallsmodells bis zur x-Achse mit F_Z und die Fläche „unterhalb“ der CAP-Kurve des perfekten Modells bis zur x-Achse mit F_P (jeweils im Intervall von Null bis Eins), lässt sich die Bezugsfläche des Dreiecks F_{ABC} in Höhe von 0,25 mit der Formel Nr. 1 bestimmen (die numerischen Werte sind die für das verwendete Beispiel, siehe Abb. 1).⁸

Formel Nr. 1

$$F_{ABC} = F_P - F_Z = 0,75 - 0,5 = 0,25$$

Die CAP-Kurve des zu beurteilenden Rating-Systems teilt regelmäßig das Dreieck ABC in zwei Teilflächen ein. Um diese Teilflächen bestimmen zu können, ist zunächst die Fläche F_{CAP} zu bestimmen, die von der CAP-Kurve des Rating-Systems und der x-Achse gebildet wird. Wird von der Fläche F_{CAP} die Fläche F_Z abgezogen, erhält man die untere Teilfläche des Dreiecks ($F_{ABC,U}$), die von der Cap-Kurve des Rating-Systems und der Diagonalen des Zu-

fallmodells gebildet wird. Für das Beispiel ergibt sich die Fläche $F_{ABC,U}$ gemäß Formel 2.

Formel Nr. 2

$$F_{ABC,U} = F_{CAP} - F_Z = 0,7061 - 0,5 = 0,2061$$

Der Gini-Koeffizient (GK) ist als der prozentuale Anteil der unteren Teilfläche ($F_{ABC,U}$) zur Gesamtfläche des Dreiecks ABC (F_{ABC}) definiert, wobei sich für das Beispiel ein Giniwert in Höhe von 0,824 ergibt (Formel Nr. 3).

Formel Nr. 3

$$GK = \frac{F_{ABC,U}}{F_{ABC}} = \frac{0,2061}{0,2500} = 0,824$$

Es ist offensichtlich, dass der Gini-Koeffizient stets im Intervall zwischen Null und Eins liegen muss. Das Zufallsmodell hat einen Giniwert von Null und das perfekte Modell des Hellsehers einen Giniwert von Eins. Niedrige Giniwerte für ein Rating-System deuten auf eine Nähe zum Zufallsmodell hin. Je höher der Giniwert, desto enger schmiegt sich die CAP-Kurve des realen Systems an die des Hellsehers an. Daraus ergibt sich, dass ein Rating-System umso trennschärfer ist, je näher „sein“ Giniwert bei Eins liegt.

Was ist ein üblicher Gini-Wert?

In der angewandten Statistik können für einzelne Bilanzkennzahlen Gini-Werte zwischen 0,2 und 0,4 ermittelt werden. Für komplexere Rating-Systeme, die Informationen

⁷ Der Koeffizient ist benannt nach dem italienischen Mathematiker und Statistiker Corrado Gini (* 1884, † 1965).

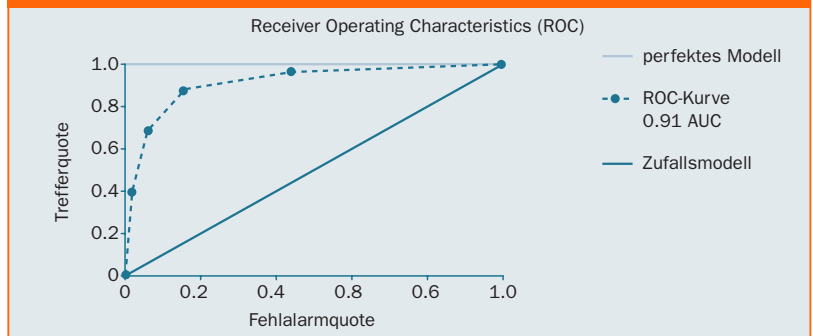
⁸ Sämtliche Flächen lassen sich durch Integration der Funktionen oder als Summe der Integration von Teilstücken der Funktionen (lineare Abschnitte) ermitteln. Schneller geht es mit Programmen wie Mathematica.

aus verschiedenen Quellen beachten,⁹ können Werte im Bereich von etwa 0,6 und höher erreicht werden: Die Ratingagentur Moody's gibt für die deutsche Version ihres Modells RiskCalc™ einen Gini-Koeffizienten in Höhe von 0,597 an.¹⁰ Auch die vom Bank-Verlag angebotene Software BV-Rating Version 3.0 kommt für Unternehmensforderungen auf einen vergleichbaren Wert von 0,57. Im Bereich der Speziallösung „Immobilienfinanzierung“ erreicht BV-Rating 3.0 sogar Werte von 0,7.¹¹

Was veranschaulicht die ROC-Kurve?

Die ROC-Kurve (Receiver-Operator-Characteristic, ROC) ist ein mit der CAP-Kurve verwandtes Konzept. Obwohl die ROC-Kurve ursprünglich aus der Signaltechnik bekannt ist und häufig auch in der Medizin und Biometrie benutzt wird, kann diese Kurve problemlos zur grafischen Darstellung der Trennkraft eines Rating-Systems verwendet werden. Die in ihr enthaltene Information ist identisch mit der in der CAP-Kurve. Die Ausfalldaten des Instituts werden für die ROC-Kurve anders als in der CAP-Kurve aufbereitet (Tabelle 3). Der folgende Gedanke liegt diesem Konzept zu Grunde: Ab welchem Schwellenwert der Rating-Note kann man einen Schuldner a priori als insolvenzgefährdet einstufen? Beispielsweise könnten nur Schuldner der Klasse 5 als Ausfall klassifiziert werden (im Folgenden als Regel A bezeichnet). Genauso gut kann die Regel „Alle Schuldner der Klasse 4 und schlechter gelten als Defaults“ (Regel B) Anwendung finden. Durch die Wahl eines Schwellenwerts und damit einer Entscheidungsregel wird über die „Trefferquote“ und die „Fehlalarmquote“ des Rating-Systems entschieden. Die Trefferquote misst den Anteil aller korrekten Klassifizierungen eines Schuldners als Ausfallkandidat an allen tatsäch-

Abb. 2: ROC-Kurve und AUROC-Kennzahl



lich ausgefallenen Schuldnern.¹² Die Fehlalarmquote hingegen misst die Wahrscheinlichkeit, dass ein tatsächlich solvent gebliebener Schuldner zuvor als potenzieller Ausfallkandidat klassifiziert wurde.¹³

Tabelle 3 zeigt ein einfaches Beispiel anhand der bereits oben betrachteten 100 Schuldner. Bei Anwendung von Regel A werden 20 von 50 ausgefallenen Schuldner korrekt erkannt, so dass die Trefferquote 40 % (20/50) beträgt. Allerdings wird bei einem Schuldner fälschlicherweise der Ausfall prognostiziert, so dass die Fehlalarmquote 2 % (1/50) beträgt. Die Anwendung der „konservativeren“ Regel B erhöht zwar die Trefferquote auf 68 % ((20+14)/50), mit ihr steigt aber auch die Fehlalarmquote auf 6 % ((1+2)/50). Ebenso könnte man Regeln für die Klassen 3, 2 und 1 definieren. Die hiermit verbundenen

Treffer- und Fehlalarmquoten werden in den Spalten 6 und 7 angeführt. An diesen Beispielen ist bereits zu erkennen, dass eine hohe Trefferquote in der Regel mit einer hohen Fehlalarmquote einhergeht, es besteht eine Wechselbeziehung („Trade-off“) zwischen beiden Größen.

9 In das Rating-Ergebnis von BV-Rating gehen neben einer Reihe von Bilanzkennzahlen auch qualitative Faktoren („Softfacts“) und branchenspezifische Einflüsse ein.

10 Vgl. B. Khandani/M. Lozano/L. Carty: Moody's RiskCalc™ For Private Companies: The German Model, London 2001. Die technischen Dokumente von Moody's Investors Service Global Credit Research Rating Methodology können auf der Website von Moody's (www.moody's.com) heruntergeladen werden.

11 Vgl. A. Hamerle/M. Reusch/M. Wadé: Rating gewerblicher Immobilienkreditnehmer nach Basel II, in: Die Bank, Heft 3, 2003, S. 198-204.

12 Dies ist das Gegenstück zum so genannten Fehler erster Art (alpha-Fehler), bei dem einem ausgefallenen Schuldner irrtümlich eine hohe Bonität zugeschrieben wurde.

13 Dieser Sachverhalt wird auch als „Fehler zweiter Art“ (Beta-Fehler) bezeichnet. Offenbar lässt sich die Bank in diesem Fall ein lohnendes Geschäft entgehen.

Tabelle 3: Aufbereitung der Daten zur Bestimmung der ROC-Kurve

Rating-klasse	Zuordnung von EKQ	Anzahl Schuldner			Koordinaten der ROC-Kurve	
		ausgefallen	nicht ausgefallen	Summe	Fehlalarmquote (in Prozent)	Trefferquote (in Prozent)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0,8–1,0	2	28	30	100	100
2	0,6–0,8	4	14	18	44	96
3	0,4–0,6	10	5	15	16	88
4	0,2–0,4	14	2	16	6	68
5	0–0,2	20	1	21	2	40
	Summe	50	50	100		

Die ROC-Kurve visualisiert diesen Trade-off, indem sie sämtliche Kombinationen von Treffer- und Fehlalarmquote abbildet, die sich durch die Variation der Klassifizierungsregeln erreichen lassen. Somit ist man bei der Beurteilung der Trennschärfe von der tatsächlichen Wahl des Schwellenwertes unabhängig. Ein Rating-System ist umso trennschärfer, je steiler die ROC-Kurve zu Beginn ansteigt. Im Fall eines perfekten Systems bestünde die „ROC-Kurve“ nur aus der Verbindung der Punkte [0;0], [0;1] und [1;1].¹⁴ Bei einem rein zufälligen Rating-Verfahren verläuft die ROC-Kurve genau auf der Diagonalen. Dies liegt auch hier daran, dass man gute und schlechte Schuldner a priori nur proportional klassifizieren kann.

Was wird mit AUROC bezeichnet?

Auch für die ROC-Kurve lässt sich eine aggregierte Kennzahl als Trennschärfemaß angeben, die man AUROC nennt. Die AUROC-Kennzahl ergibt sich als Fläche unterhalb der ROC-Kurve und der x-Achse (F_{ROC}) und liegt stets zwischen Null und Eins. Je näher AUROC für ein Rating-System an dem Wert Eins liegt, desto trennschärfer ist es. Für das Beispiel ergibt sich die AUROC-Kennzahl in Höhe von 0,91 gemäß Formel Nr. 4:

Formel Nr. 4

$$AUROC = F_{ROC} = 0,91$$

In welcher Beziehung stehen GK und AUROC?

Zwischen dem Gini-Koeffizient (GK) und der Fläche unter der ROC-Kurve (AUROC) besteht ein einfacher linearer Zusammenhang (Formel Nr. 5).¹⁵

Formel Nr. 5

$$GK = 2 \cdot AUROC - 1$$

Die beiden Trennschärfemaße sind äquivalent, da durch eine einfache lineare Transformation die eine Kennzahl in die jeweils andere überführt werden kann. Mit dieser Gleichung lässt sich auch schnell das Ergebnis für AUROC in unserem Beispiel überprüfen (Formel Nr. 6).

Formel Nr. 6

$$AUROC = \frac{GK + 1}{2} = \frac{0,824 + 1}{2} = 0,91$$

Für das Zufallsmodell ist der AUROC-Wert 0,5. Diese Berechnung kann zum einen auf Basis des Zusammenhangs zwischen AUROC und Gini erfolgen, zum anderen

durch die Berechnung des Flächeninhalts unter der ROC-Kurve des Zufallsmodell (45°-Linie) mittels der Pythagorasformel für Dreiecksflächen.

Welchen Einfluss hat die Struktur des Kreditportfolios auf die Trennschärfemaße?

Die Interpretation der Trennkraft eines Rating-Systems erfor-

¹⁴ Der Koordinatenpunkt [0;0] kommt zu Stande, wenn man keinen Schuldner als Ausfall klassifiziert. In diesem Fall macht man keinen Fehler, landet aber auch keinen Treffer. Dieser Punkt wird daher auch bei der ROC-Kurve eines realen Systems berücksichtigt.

¹⁵ Vgl. für einen Beweis dieser Formel B. Engelmann/E. Hayden/D. Tasche: Measuring the Discriminative Power of Rating Systems, Discussion Paper Nr. 01/2003 der Deutschen Bundesbank, Anhang A, 2003, S. 23.

Info-Box 1: Trennschärfe versus Kalibrierung:

Von dem Begriff der Trennschärfe eines Rating-Systems ist der Begriff der Kalibrierung des Rating-Systems abzugrenzen. Während die Trennschärfe eines Rating-Systems die Fähigkeit beschreibt, im Voraus ausfallgefährdete Kreditnehmer zu identifizieren, wird unter der Kalibrierung im Allgemeinen die Zuweisung von Ausfallwahrscheinlichkeiten (Probability of Default, PD) zu den verschiedenen Rating-Klassen verstanden. Unter Berücksichtigung dieser Definitionen kann ein Rating-System somit durchaus trennscharf, nicht aber gut kalibriert sein. Dies wäre dann der Fall, wenn die Kreditnehmer als ausfallgefährdet identifiziert werden und die tendenziell guten (schlechten) Rating-Klassen einen geringen (hohen) Anteil der später ausfallenden Kunden aufweisen, jedoch die mit Hilfe des Rating-Systems geschätzten Ausfallwahrscheinlichkeiten signifikant von den tatsächlichen Ausfallraten abweichen. Umgekehrt kann eine perfekte Kalibrierung vorliegen, ohne dass ein Minimum an Trennschärfe gegeben ist. Dies wäre der Fall, wenn ein Rating-System darauf verzichtet, Schuldner in verschiedene Bonitätsstufen einzuteilen, das heißt, es erfolgt keine Trennung zwischen „guten“ und „schlechten“ Risiken. Stattdessen wird jedem Schuldner die durchschnittliche PD des kommenden Jahres zugewiesen. Im Rückblick würde dann die Portfolioausfallrate exakt mit der PD-Prognose übereinstimmen, die Kalibrierung wäre perfekt. Da aber keine individuelle Risiko-Einstufung vorgenommen wurde, ist keine Trennschärfe vorhanden. Im Rahmen der Umsetzung des Basel-II-Frameworks ist die Untersuchung der Kalibrierung eines Rating-Systems deshalb vonnöten, weil sich die Eigenkapitalunterlegung nach der „Baseler IRB-Formel“ aus den geschätzten Ausfallwahrscheinlichkeiten ergibt.

Literaturempfehlungen

Deutsche Bundesbank (2003): Validierungsansätze für interne Rating-Systeme, wie Anm. 4. Eine gut lesbare Einführung in die Problematik der Überprüfung zur internen Validierung.

B. Engelmann/E. Hayden/D. Tasche (2003): Measuring the Discriminative Power of Rating Systems, wie Anm. 15. Ein Übersichtsartikel zum Thema CAP- und ROC-Kurven, der auch auf die mathematischen Aspekte des Themas eingeht, ohne dabei den Praxisbezug zu verlieren.

K. Füsler/K. Rödel (2003): Prüfung von Ratingverfahren, in: Die Wirtschaftsprüfung, 56. Jg. Heft 23, 2003, S. 1289–1301. Dieser einführende Artikel geht zusätzlich auf Aspekte der Kalibrierung von Rating-Systemen ein.

R. Rauhmeier (2004): Validierung und Performancemessung bankinterner Rating-Systeme, Bad Soden/ Ts 2004. Dieses Buch bietet eine umfassende Darstellung des Themas. Der Autor arbeitet insbesondere die Portfolioabhängigkeit von Trennschärfemaßen heraus.

J. Sobehart/S. Keenan/R. Stein (2000): Benchmarking Quantitative Default Risk Models: A Validation Methodology, Moody's Investors Service, 2000. Die Autoren konzentrieren sich in ihrem Artikel auf die Berechnung der ROC-Kurven und ihre Interpretation.

H. Schulte-Mattler (2004): Neue Baseler Eigenkapitalübereinkunft (Basel II), wie Anm. 3. Der Autor stellt zusammenhängend und vertiefend das gesamte Basel-II-Regelungswerk vor.

dert Vorsicht, da die Werte von AUROC- und Gini-Koeffizient in hohem Maße portfolioabhängig sind. Ein einfaches Beispiel kann dies verdeutlichen: Das Portfolio von Bank A entspricht dem bisher betrachteten Beispiel (Tabelle 1), für das ein AUROC-Wert in Höhe von 0,91 ermittelt wurde. Bank B besitzt ebenfalls ein Portfolio von 100 Schuldnern, von denen 50 ausgefallen sind. Diese Schuldner unterscheiden sich aber nicht durch ihre Eigenkapitalquote, sie beträgt für alle 0,5. Bank B wendet das Rating-System auf Basis der EKQ an, so dass alle B-Kunden der mittleren Klasse 3 zugeordnet werden. Berechnet man die ROC-Kurve auf Basis dieses Portfolios, so erhält man die des zufälligen Modells, der zugehörige AUROC-Wert beträgt 0,5. Da keine Unterschiede bezüglich der Kennzahl EKQ vorliegen, kann keine Trennung vorgenommen werden. Somit würde das hier vorgestellte Rating für das Kreditportfolio von Bank A eine hohe Trennschärfe ausweisen, nicht aber für das von Bank B. Ursache dafür ist einzig und allein die unterschiedliche Schuldnerstruktur. Bedeutet dies, dass verschiedene Rating-Systeme nicht bezüglich ihrer Trennkraft verglichen werden können? Nein! Der Vergleich muss jedoch auf Grundlage des selben Kreditportfolios geschehen. Außerordentlich gute Giniwerte sind häufig das Resultat heterogener Portfolios. Die angewandte Statistik zieht hieraus den Schluss, dass man bei der Entwicklung von Rating-Funktionen die vorhandenen Daten in eine Entwicklungs- und eine Validierungsstichprobe unterteilt. Das Rating-System wird auf Basis der Entwicklungsstichprobe konstruiert und mittels der Validierungsstichprobe insbesondere auf seine Portfolioabhängigkeit hin überprüft.

Welches Fazit lässt sich ziehen?

Auf die Validierung von Rating-Systemen bereiten sich sowohl Banken als auch die Bankenaufsicht intensiv vor. Der Baseler Ausschuss hat eigens für diesen Problemkreis eine eigene Arbeitsgruppe für Validierungsfragen unter Federführung der Research Task Force eingesetzt. Auch in den Fachgremien der BaFin beschäftigt man sich mit diesem Thema. Die in der Praxis häufig verwendeten Instrumente der Trennschärfenanalyse sind sehr einfach konzipiert und daher leicht zu verstehen (CAP-Kurve und Gini-Koeffizient sowie ROC-Kurve und AUROC-Koeffizient). Die Eignung einzelner Trennschärfemaße zur Validierung sind in der Wissenschaft jedoch heftig umstritten. So wird die bankaufsichtliche Überprüfung von Rating-Systemen sich wohl nicht auf die Anwendung nur eines einzelnen Verfahrens stützen, sondern sich als Gesamtbild aus verschiedenen qualitativen und quantitativen Methoden ergeben. In den vor uns liegenden Jahren wird sich die Qualität und Quantität der verfügbaren Daten entscheidend verbessern. Die daraus und aus der intensiven wissenschaftlichen Arbeit gewonnenen Erkenntnisse werden die bankaufsichtlichen Standards für die Validierung künftig prägen. ■

Prof. Dr. Hermann Schulte-Mattler lehrt Finanzwirtschaft und Controlling an der FH Dortmund (Dortmund University of Applied Sciences). **Ulrich Daun** ist Mitarbeiter im Gemeinschaftsprojekt Internes Rating, das der Bundesverband deutscher Banken zusammen mit dem Bank-Verlag aufgelegt hat. Sein Aufgabengebiet umfasst die im Zusammenhang mit dem IRB-Ansatz notwendigen statistischen Analysen. **Thorsten Manns** ist Doktorand an der University of Plymouth (UK) und Vorstand des Finance Risk Lab e.V. (www.FRL.de).