

Methoden zum Vergleich verschiedener Scoreverfahren am Beispiel der SCHUFA-Scoreverfahren

Ein Beitrag von Ludwig Fahrmeir, Andreas Henking und Ralf Hüls

Einleitung

Zur Bewertung von Scoreverfahren existiert eine Fülle von Methoden. Buchstäblich jeder Anbieter und jeder Anwender von Scoreverfahren verwendet unterschiedliche Methoden bzw. äquivalente Methoden unter verschiedenen Bezeichnungen. Um dieser Vielzahl von Methoden und Begriffen Herr zu werden, initiierte die SCHUFA eine Kooperation zur Methodenrecherche mit dem STABLAB (Statistisches Beratungslabor, Institut für Statistik, LMU München). Der vorliegende Artikel ist ein Ergebnis dieser Kooperation.

Im Dezember 1996 führte die SCHUFA ihren "Kreditbüro-Score" ASS ein, um der kreditgebenden Wirtschaft Breitband-Scores zur Unterstützung der gesamten Einschätzung von Kreditrisiken auf Basis der SCHUFA-Daten zur Verfügung zu stellen. Im Juli 1997 folgte ein Scoreverfahren für den Telekommunika-

tionssektor. Da die bestehenden Scoreverfahren nach drei bis vier Jahren im Produktionseinsatz verbesserungswürdig erschienen, wurde im August 2000 beschlossen, neue Scoreverfahren zu entwickeln und im SCHUFA-System zu implementieren. Die SCHUFA-Scoreverfahren mit den SCHUFA-Datensätzen sind bestens geeignet, die unterschiedlichen Methoden zur Bewertung und zum Vergleich von Scoreverfahren praktisch aufzuzeigen.

Wir besprechen zwei Darstellungsformen zur Analyse der Trennschärfe von Scoreverfahren. Diese sind die ROC (Receiver Operating Characteristic) und das CAP (Cumulative Accuracy Profile). Bei der ROC werden die Defaults, das heißt Eintrag eines SCHUFA-Negativmerkmals, einer Stichprobe gegen die Non Defaults abgetragen. Bei den CAPs hingegen werden die Defaults gegen die Gesamtstichprobe abgetragen. Dies macht formal einen wesentlichen Unterschied. Der qualitative Verlauf der Darstellungen ist jedoch bei den beiden Ansätzen ähnlich.

Häufig werden Methoden wie ROC oder CAP auch Power Curve genannt (Blochwitz et al. (2000), Sobehart et al. (2000)).

Die Bewertungsmethoden sind Schätzverfahren und folglich mit statistischer Unsicherheit belegt. Daher ist es sinnvoll, die Genauigkeit der Bewertungsmethoden zu untersuchen. Es werden Konfidenzbänder für ROCs und Konfidenzintervalle für die Fläche unter den ROCs eingeführt.

Auch in anderen Bereichen werden ROCs verwendet, zum Beispiel in der medizinischen Diagnostik. Dies ermöglicht, Erkenntnisse zu den ROCs aus der medizinischen Diagnostik auf die Bewertung von Scoreverfahren für die Kreditwürdigkeitseinschätzung zu übertragen.

Die Methoden werden anhand der SCHUFA-Scoreverfahren erprobt und diskutiert.

SCHUFA-Scoreverfahren

Als moderner Dienstleister ermöglicht und beschleunigt die SCHUFA HOLDING AG die Kreditgewährung. Ihre Produkte unterstützen die Bonitätsprüfung ihrer Vertragspartner wie Banken, Sparkassen und kreditgewährenden Handels- und Dienstleistungsunternehmen. Weiterhin helfen Dienstleistungen wie Scoreverfahren, das Kreditrisiko im Privatkundengeschäft mit wissenschaftlich fundierten Methoden zu kontrollieren. Damit entspricht die SCHUFA den veränderten Anforderungen des Kreditmarktes, der im

Hinblick auf Basel II fundierte Ratingtools fordert.

Im untersuchten Praxis-Beispiel wird das SCHUFA-Scoreverfahren für die Kreditwirtschaft aus dem Jahr 1996 mit dem SCHUFA-Scoreverfahren für Banken von 2001 verglichen. Während dem Scoreverfahren für die Kreditwirtschaft von 1996 noch alle kreditgebenden Institutsgruppen wie Groß- und Privatbanken, Sparkassen, genossenschaftliche Institute oder Teilzahlungsbanken zugrunde gelegt wurden, erfolgte in 2001 für die Banken der Einstieg in drei spezialisierte Scoreverfahren. Aus dem Analyseprozess bei der SCHUFA liegen große Stichproben über Datensätze mit zwei Scores aus dem Verfahren für die Kreditwirtschaft und aus dem spezialisierten Verfahren für Banken vor. Zu den jeweiligen Beobachtungen wurden die tatsächlichen Zahlungsverläufe nach einer Reifezeit ermittelt.

Für die Entwicklung und Validierung der neuen Scoreverfahren analysierte die SCHUFA insgesamt etwa 6,6 Millionen anonymisierte Datensätze. Für das vorliegende Beispiel wurde die Validierungsstichprobe des Bankenbereichs mit ca. 410.000 Datensätzen verwendet. Beiden Scoreverfahren liegt als statistische Methodik die logistische Regression zugrunde.

Der Score s kann Werte in $\{0, \dots, S\}$ annehmen, wobei kleine Scorewerte negative Einschätzungen und große Scorewerte positive Einschätzungen darstellen. Das bedeutet, dass das Kreditausfallrisiko, gemessen in Wahrscheinlichkeiten, für kleine Scorewerte größer als für große Scorewerte ist. Bei den SCHUFA-Scoreverfahren ist $S = 1.000$.

Receiver Operating Characteristic (ROC)

Um die Konstruktion und Eigenschaften der ROC zu beschreiben, werden einige Bezeichnungen eingeführt.

X : Score eines Non Default

Y : Score eines Default

Z : Score eines beliebigen Individuums der Grundgesamtheit

F bezeichnet die Verteilungsfunktion von X und G die von Y .

Für die Stichprobe werden die Scores von n Non Default-Individuen und von m Default-Individuen beobachtet. Die Kreditwürdigkeiten der einzelnen Individuen seien unabhängig voneinander.

Zur Beurteilung der Qualität statistischer Verfahren sind der Fehler 1. Art und der Fehler 2. Art wichtige Indikatoren. Übertragen auf Creditscoreverfahren bedeuten diese (siehe auch Sobehart et al. (2000)):

Fehler 1. Art: Einschätzung von hoher Kreditwürdigkeit, obwohl ein Default vorliegt.

Fehler 2. Art: Einschätzung von niedriger Kreditwürdigkeit, obwohl ein Non Default vorliegt.

Für einen Scorewert c beziffert $1-G(c)$ den Fehler 1. Art und $F(c)$ den Fehler 2. Art.

Konstruktion der ROC

Die ROC erhält man, indem man für alle Scorewerte $c = 1, \dots, S$ die Punkte

$(F(c), G(c))$ in einem Koordinatensystem, dem Einheitsquadrat, einträgt und verbindet (Abb. 1). Jeder Punkt der ROC stellt damit den Fehler 1. und 2. Art für einen bestimmten Score dar, wobei sich der Fehler 1. Art über die Transformation $1 - \text{Fehler 1. Art} = G(c)$ errechnet.

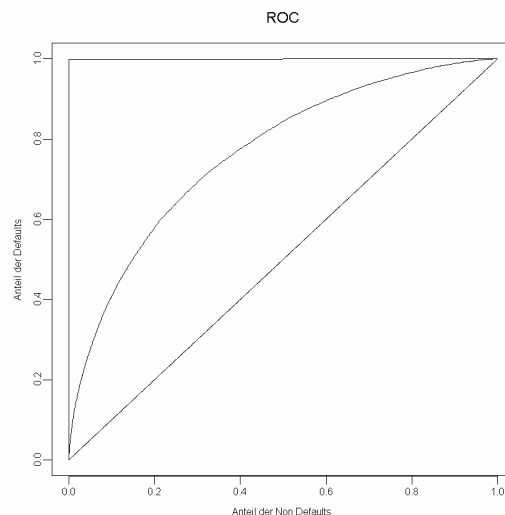


Abbildung 1: Die ROC des SCHUFA-Scoreverfahrens 2001 und die ideale ROC

Da die beiden Verteilungsfunktionen im Allgemeinen nicht bekannt sind, ersetzt man F und G durch empirische Verteilungsfunktionen. Damit ist die ROC, dargestellt mit den empirischen Verteilungsfunktionen, ein Schätzer.

Eigenschaften der ROC

Liefert ein Scoreverfahren keine Information, liegt die (theoretische) ROC auf der Winkelhalbierenden. Üblicherweise liegt die Kurve oberhalb der Winkelhalbierenden. Es ist jedoch prinzipiell möglich, dass die ROC die Winkelhalbierende schneidet und damit abschnittsweise unter ihr verläuft. Das sollte bei ausreichend großen Stichproben jedoch nicht vorkommen und ist ein starker Indikator

dafür, dass das Scoreverfahren einer Überprüfung unterzogen werden sollte.

ROCs sind monoton steigend, da sowohl F als auch G bzw. ihre empirischen Pendanten monotone Funktionen sind. Für kleine Scorewerte sollte G schneller anwachsen als F , für große Scorewerte vice versa. Dadurch ergibt sich die typische bauchige Form der ROCs. Bei einem idealen Scoreverfahren verläuft die ROC vom Ursprung (0,0) auf der Default-Koordinatenachse zum Punkt (0,1) und dann am oberen Rand des Einheitsquadrats zum Punkt (1,1) (Abb. 1).

ROCs verschiedener Scoreverfahren können sich schneiden und dabei die gleiche Fläche umschließen. Das bedeutet, dass ein Scoreverfahren nicht immer gleichmäßig besser oder schlechter als ein anderes Scoreverfahren sein muss.

Die Fläche unter der ROC kann als die Wahrscheinlichkeit interpretiert werden, dass für zwei zufällig ausgewählter Individuen, jeweils eines aus der Non Default und eines aus der Default Stichprobe, der Non Default richtigerweise einen höheren Scorewert als der Default aufweist (Hanley, McNeil (1982), Lee (1999)). Beträgt diese Wahrscheinlichkeit 0,5, so liefert das Scoreverfahren offenbar keine Information. Je näher die Wahrscheinlichkeit an der 1 ist, desto trennschärfer ist das Scoreverfahren. Beide Interpretationen dieser Wahrscheinlichkeit bestätigen die oben beschriebenen nicht informative und ideale ROC.

Konfidenzbänder für die ROCs der SCHUFA-Scoreverfahren

Die ROC wird aus Daten geschätzt und unterliegt damit bereits durch die Stichprobenauswahl zufälligen Einflüssen. Um diese Zufälligkeit zu quantifizieren und zu visualisieren, kann man Konfidenzbänder um die ROC konstruieren. Simultane Konfidenzbänder geben an, in welchem Bereich sich die gesamte Kurve mit mindestens einer vorgegebenen Wahrscheinlichkeit befindet. Punktweise Konfidenzbänder geben an, in welchem Bereich sich die einzelnen Punkte der ROC mit einer vorgegebenen Wahrscheinlichkeit befinden.

Die verschiedenen Konstruktionsprinzipien für Konfidenzbänder bei ROCs haben folgendes gemeinsam: Zunächst bestimmt man zu einer vorgegebenen Überdeckungswahrscheinlichkeit γ Konfidenzbänder um die empirischen Verteilungsfunktionen \hat{F} und \hat{G} . Dadurch erhält man für jeden Scorewert ein Konfidenzrechteck um den korrespondierenden Punkt der ROC. Alle Rechtecke zusammen ergeben das Konfidenzband für die ROC.

Wir gehen von gleichen Überdeckungswahrscheinlichkeiten γ für \hat{F} und \hat{G} aus. Die Überdeckungswahrscheinlichkeit für die ROC bzw. deren einzelne Punkte ist dann wegen der Unabhängigkeit von X und Y

$$\gamma^2 = 1 - \alpha.$$

Die Untergrenzen der Konfidenzbänder werden mit u und die Obergrenzen mit o indiziert.

Simultane Konfidenzbänder

Campbell (1994) schlägt vor, simultane Konfidenzbänder mit Hilfe der Kolmogorov-Verteilung zu konstruieren. Mit der Kolmogorov-Statistik modelliert man den maximalen Abstand zwischen theoretischer und empirischer Verteilungsfunktion, also zwischen F und \hat{F} bzw. G und \hat{G} . Die zugehörige Kolmogorov-Verteilung findet man bei Büning, Trenkler (1994) tabelliert.

Mit vorgegebener Überdeckungswahrscheinlichkeit γ erhält man aus der Kolmogorov-Verteilung d und f , so dass

$$\hat{F}_u(c) = \max\{0, \hat{F}(c) - d\} \quad \text{und} \\ \hat{F}_o(c) = \min\{\hat{F}(c) + d, 1\} \quad \text{für alle } c \in \{1, \dots, S\}$$

das Konfidenzband für \hat{F} beschreibt und

$$\hat{G}_u(c) = \max\{0, \hat{G}(c) - f\} \quad \text{und} \\ \hat{G}_o(c) = \min\{1, \hat{G}(c) + f\} \quad \text{für alle } c \in \{1, \dots, S\}$$

das Konfidenzband für \hat{G} .

Für eine Überdeckungswahrscheinlichkeit von $\gamma = 0,99$ erhält man approximativ für $n > 40$ und $m > 40$:

$$d = \frac{1,63}{\sqrt{n}} \quad \text{und} \quad f = \frac{1,63}{\sqrt{m}}.$$

Die Punkte $(\hat{F}_u(c), \hat{G}_o(c))$, $(\hat{F}_o(c), \hat{G}_o(c))$, $(\hat{F}_u(c), \hat{G}_u(c))$ und $(\hat{F}_o(c), \hat{G}_u(c))$ bilden ein Konfidenzrechteck um den Punkt $(\hat{F}(c), \hat{G}(c))$.

Durch Verbindung der Punkte $(\hat{F}_u(c), \hat{G}_o(c))$ wird das obere einhüllende Konfidenzband gebildet, das untere

durch Verbindung der Punkte $(\hat{F}_o(c), \hat{G}_u(c))$.

Die Konfidenzbänder wurden für die ROCs der SCHUFA-Scoreverfahren von 2001 und 1996 auf Basis der Validierungsstichproben gebildet (Abb. 2). Die ROC des SCHUFA-Scoreverfahrens 2001 verläuft oberhalb der ROC des SCHUFA-Scoreverfahrens 1996. Die Konfidenzbänder haben eine Überdeckungswahrscheinlichkeit von 0,98. Dank des großen Umfangs der SCHUFA-Validierungsstichprobe sind die Konfidenzbänder sehr schmal. Man erkennt, dass die Konfidenzbänder, außer in den Randregionen, nicht überlappen (Abb. 2). Dies ist eine hinreichende Bedingung, um auf die statistisch signifikante Verbesserung durch das SCHUFA-Scoreverfahren 2001 zu schließen.

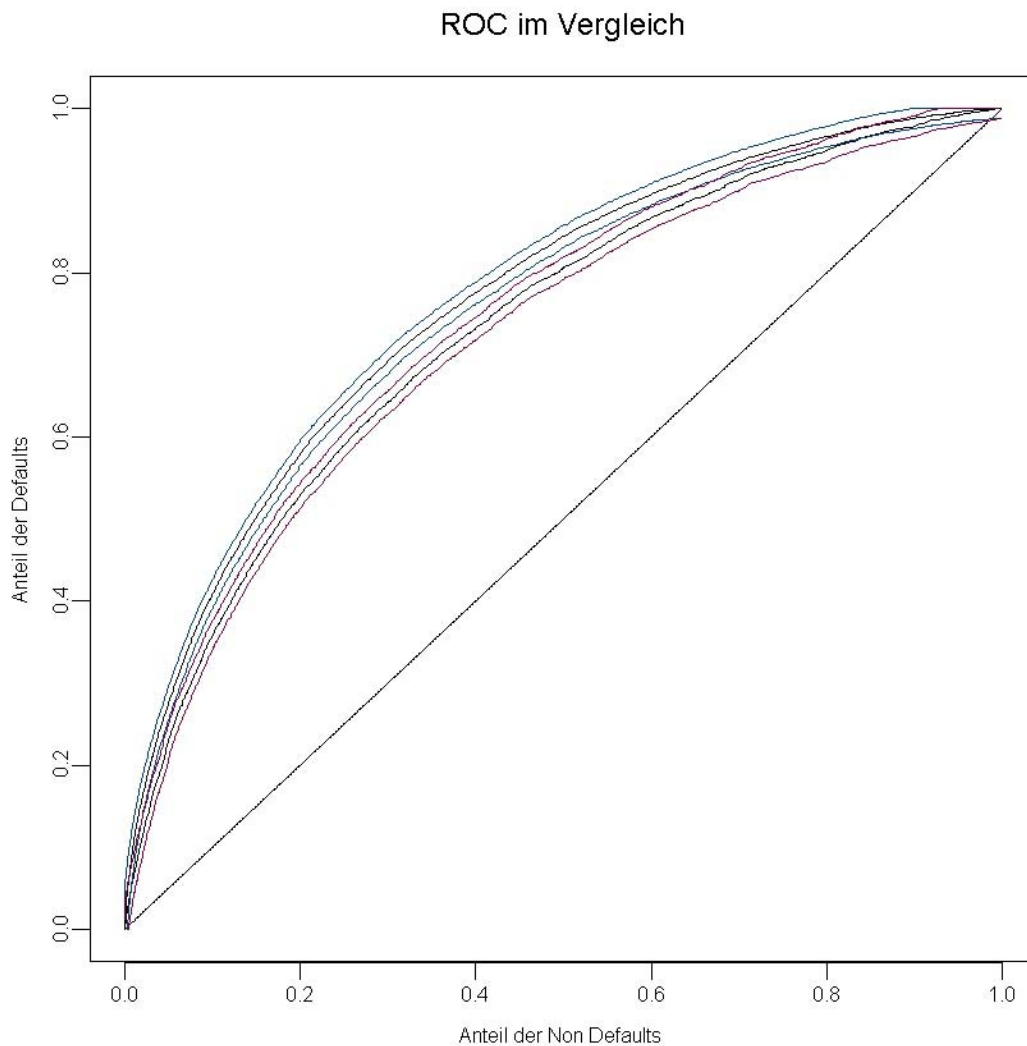


Abbildung 2: Die ROCs der SCHUFA-Scoreverfahren 2001 und 1996 mit simultanen Konfidenzbändern

Punktweise Konfidenzbänder

Punktweise Konfidenzbänder werden konstruiert, indem man für jeden Punkt der ROC ein Konfidenzintervall zum vorgegebenen Niveau bildet. Dies geht nach dem Konstruktionsprinzip für Konfidenzintervalle einer Wahrscheinlichkeit, sowohl für die Werte der Verteilungsfunktion der Defaults als auch der Non Defaults. Man erhält wieder Konfidenzrech-

tecke, die man analog zu den simultanen Konfidenzbändern zur Bildung eines einhüllenden Konfidenzbandes verwendet.

Für eine ausführliche Darstellung zur Konstruktion von Konfidenzintervallen für Anteilswerte sei auf Standardwerke der Statistik verwiesen (Fahrmeir et al. (2001)). Konstruktionsprinzipien für punktweise Konfidenzbänder der ROCs finden sich bei Schäfer (1994).

Die punktweisen Konfidenzbänder belegen, dass sich die ROC des SCHUFA-Scoreverfahrens 2001 für die meisten Scorewerte statistisch signifikant von der ROC des SCHUFA-Scoreverfahrens 1996 unterscheidet. Lediglich für Scorewerte

größer als 808 überlappen die punktwweisen Konfidenzbänder. Dies erklärt sich durch geringe Zellenbesetzung bei den Defaults für die hohen Scorewerte.

CAP

Bei den CAPs (Cumulative Accuracy Profiles, Sobehart et al. (2000)) wird die empirische Verteilungsfunktion aus der gesamten Validierungsstichprobe bestimmt, die gegen die empirische Verteilungsfunktion der Defaults abgetragen wird (Abb. 3).

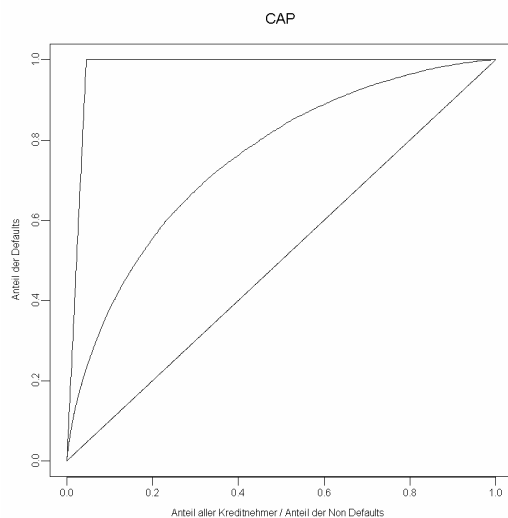


Abbildung 3: Das CAP des SCHUFA-Scoreverfahrens 2001 und ideales CAP

Das CAP des trivialen Scoreverfahren ist, wie bei der ROC, die Winkelhalbierende. Die Neigung des ersten Teilstücks der idealen CAP entspricht genau dem Anteil der Defaults (Abb. 3).

Vergleich von ROC und CAP

Bei den ROCs sind die beiden betrachteten Teilpopulationen disjunkt bzw. X und Y sind stochastisch (bedingt) unabhängig, bei den CAPs hingegen nicht. Dies erschwert die Interpretation und Untersuchung statistischer Eigenschaften bei den CAPs.

Die ROCs sind bauchiger als die korrespondierenden CAPs. Daher ist die Fläche zwischen der Winkelhalbierenden und der ROC größer als bei den CAPs, was eine optische Aufwertung der Trennschärfe zur Folge hat.

Je größer der Anteil der Defaults in der Grundgesamtheit, desto größer ist die Abhängigkeit zwischen Y und Z . Das bedeutet, dass die CAPs im Vergleich zur korrespondierenden ROC flacher verlaufen und sich damit stärker von der ROC unterscheiden. Anders ausgedrückt: in Abhängigkeit von dem Anteil der Defaults ändert sich das CAP bei gleichbleibenden Fehlern 1. und 2. Art. Es kann zur Verwirrung beitragen, wenn die Bewertung eines Scoreverfahrens abhängig von dem Anteil der Defaults in der Validierungsstichprobe ist. Dies ist ein gewichtiges Argument für die Verwendung der ROC.

Man erkennt zwischen der ROC und dem CAP für das SCHUFA-Scoreverfahren von 2001 einen Unterschied, der jedoch auf Grund des geringen Anteils der Defaults (4,5%) klein ist (Abb. 4).

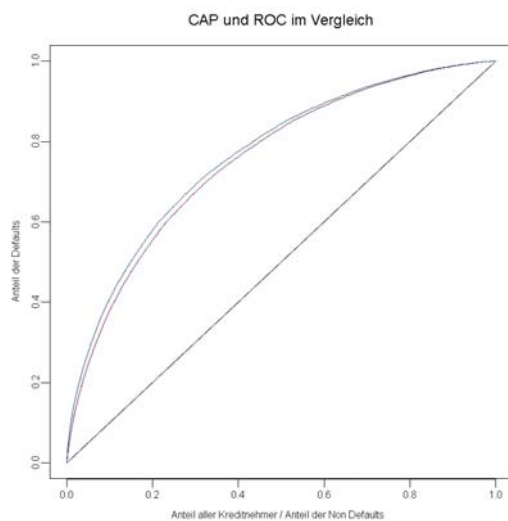


Abbildung 4: CAP (rot) und ROC (blau) des SCHUFA-Scoreverfahrens von 2001 im Vergleich

Messung der Trennschärfe

Es gibt verschiedene eindimensionale Maße zur Trennschärfemessung. Wir betrachten hier nur Maße, die direkt aus den oben beschriebenen Kurven abgeleitet werden. Darüber hinaus gibt es noch einige weitere Maße, zum Beispiel: Coefficient of Concordance (Fritz, Hosemann (2000)), Pietra-Index (Lee (1999)), Entropie (Sobehart et al. (2000)) und Kullback-Leibler-Distanz.

Die Reduzierung der ROC oder des CAP auf ein eindimensionales Maß hat einen Informationsverlust zur Folge. Daher sollten diese Maße nie ohne die zugehörige Kurve betrachtet werden (Zweig, Campbell (1993)).

Gini-Koeffizient

Der Gini-Koeffizient, auch Accuracy Ratio (Sobehart et al. (2000) und Power Statistic, wird üblicherweise als die Fläche zwischen der Winkelhalbierenden und der Kurve (ROC oder CAP) im Verhältnis zur Fläche zwischen der idealen Kurve und der Winkelhalbierenden definiert (Blochwitz et al. (2000), Lee (1999), Sobehart et al. (2000)). Dadurch erhält man ein normiertes Maß in $[0,1]$. Hat der Gini-Koeffizient den Wert 0, liegt ein triviales Scoreverfahren vor. Je größer der Gini-Koeffizient ist, desto trennschärfer ist das Verfahren.

Der Gini-Koeffizient hat für CAPs und ROCs genau den gleichen Wert, wie über den funktionalen Zusammenhang von ROC und CAP nachgewiesen werden kann. Das SCHUFA-Scoreverfahren 1996 hat einen Gini-Koeffizienten von 0,47, das SCHUFA-Scoreverfahren 2001 hat einen Gini-Koeffizienten von 0,53. Dies zeigt eine deutliche Verbesserung durch das SCHUFA-Scoreverfahren 2001 an, wie bereits durch die Konfidenzbänder belegt wurde.

Die den SCHUFA-Scoreverfahren zugrunde liegenden Daten entstammen der SCHUFA-Datenbank. Demzufolge sind Daten wie Einkommen und Familienstatus nicht in den SCHUFA-Scoreverfahren berücksichtigt. Der Gini-Koeffizient von 0,53 des SCHUFA-Scoreverfahrens 2001 ist daher für einen „Kreditbüro-Score“, als gut einzustufen.

Fläche unter der ROC

Naheliegender ist es, die Fläche unter der ROC als Gütemaß zu wählen. Dies ist die Fläche zwischen den Koordinatenachsen und der ROC bzw. die Fläche zwischen der ROC und der Winkelhalbierenden plus 0,5. Mehrere Autoren verwenden dieses Maß (Hanley, McNeil (1982), Lee (1999), Zweig, Campbell (1993)). Bei Lee (1999) bekommt es das Kürzel AUC (area under the curve), welches wir im folgenden benutzen.

Auch AUC ist auf $[0,1]$ normiert. Nimmt AUC den Wert 1 an, liegt eine ideale ROC vor. Beim Wert von 0,5 liegt ein triviales Scoreverfahren vor. Bei Werten kleiner als 0,5 liefert das Scoreverfahren einen negativen Informationsgewinn. AUC steht in linearem Zusammenhang mit dem Gini-Koeffizienten für ROCs: $Gini = 2(AUC - 0,5)$.

AUC kann als die Wahrscheinlichkeit, dass ein Default-Individuum im Vergleich zu einem Non Default-Individuum richtig, nämlich riskanter, eingeschätzt wurde, interpretiert werden. Diese Wahrscheinlichkeit und damit die AUC kann über die Wilcoxon Statistik geschätzt werden (Hanley, McNeil (1982), zur Wilcoxon-Statistik Büning, Trenkler (1994)). Für die Wilcoxon Statistik lässt sich die Varianz (Hanley, McNeil (1982)) der AUC schätzen und ein Konfidenzintervall für die AUC angeben.

Das SCHUFA-Scoreverfahren 1996 hat eine AUC von 0,733, das SCHUFA-Scoreverfahren 2001 eine AUC von 0,765. Das Konfidenzintervall für die AUC des SCHUFA-Scoreverfahrens von 1996 auf dem 98%-Niveau ist $[0,729; 0,737]$, das Konfidenzintervall für die AUC des SCHUFA-Scoreverfahrens von

2001 ist $[0,761; 0,769]$. Die beiden Konfidenzintervalle sind disjunkt, daher kann geschlossen werden, dass die beiden Flächen statistisch signifikant verschieden sind. Das SCHUFA-Scoreverfahren von 2001 stellt also eine Verbesserung dar, wie dies auch schon die Konfidenzbänder für die ROCs belegen.

Schlussbemerkungen

In vorliegendem Bericht haben wir zwei Verfahren, die ROC und das CAP, zur geometrischen Darstellung der Güte von Scoreverfahren vorgestellt und verglichen. Aufgrund statistischer Eigenschaften eignet sich die ROC besser zur Interpretation und Bewertung der Trennschärfe eines Scoreverfahrens. Die Fläche unter der ROC lässt sich direkt als Wahrscheinlichkeit darstellen, was die Interpretation wesentlich vereinfacht.

Simultane und punktweise Konfidenzbänder gestatten es, die durch Stichproben bedingte Unsicherheit in der Darstellung der ROC zu quantifizieren. Es ist möglich, statistisch gesicherte Aussagen zu treffen, ob ein Scoreverfahren eine vorgegebene Güte einhält oder sogar besser ist. So konnte gezeigt werden, dass das SCHUFA-Scoreverfahren von 2001 eine statistisch signifikante Verbesserung gegenüber dem SCHUFA-Scoreverfahren von 1996 darstellt. Dies belegen auch der Gini-Koeffizient und die AUC, für die Konfidenzintervalle gebildet werden können.

Kontakt:

Prof. Dr. Ludwig Fahrmeir

Ludwig Maximilians Universität München
Institut für Statistik
Ludwigstr. 33
80799 München
email: fahrmeir@stat.uni-muenchen.de

Dr. Andreas Henking

RiskSIM
Zeheterstr. 9a
83607 Holzkirchen
Tel.: ++49.8024.476607
email: andreas.henking@risk-sim.de

Ralf Hüls

SCHUFA Holding AG
Bismarckplatz
44866 Bochum
Tel.: ++49.2327.911428
email: ralf.huels@schufa.de

Literatur:

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“, KMV.

Büning, H., Trenkler, G. (1994): „Nichtparametrische statistische Methoden“, Walter de Gruyter.

Campbell, G. (1994): „Advances in Statistical Methodology for the Evalua-

tion of Diagnostic and Laboratory Tests“, Statist. Med 13.

Fahrmeir, L., Künstler, R., Pigeot, I., Tutz, G. (2001): „Statistik. Der Weg zur Datenanalyse“, Springer Verlag.

Fritz, S. Hosemann, D. (2000): „Restructuring the Credit Process: Behaviour Scoring for German Corporates“, Int. J. Intell. Sys. Acc. Fin. Mgmt. 9.

Hanley, J., McNeil, B. (1982): „The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve“, Radiology 143.

Jensen, K., Müller, H., Schäfer, H. (2000): „Regional confidence bands for ROC curves“, Statist. Med. 19.

Lee, W. (1999): „Probabilistic Analysis of Global Performances of Diagnostic Tests: Interpreting the Lorenz Curve-Based Summary Measures“, Statist. Med. 18.

Schäfer, H. (1994): „Efficient Confidence Bounds for ROC Curves“, Statist. Med. 13.

Sobehart, J., Keenan, S., Stein, R. (2000): „Benchmarking Quantitative Default Risk Models: A Validation Methodology“, Moody's Investors Services.

Zweig, M., Campbell, G. (1993): „Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine“, Clin. Chem. 39.