

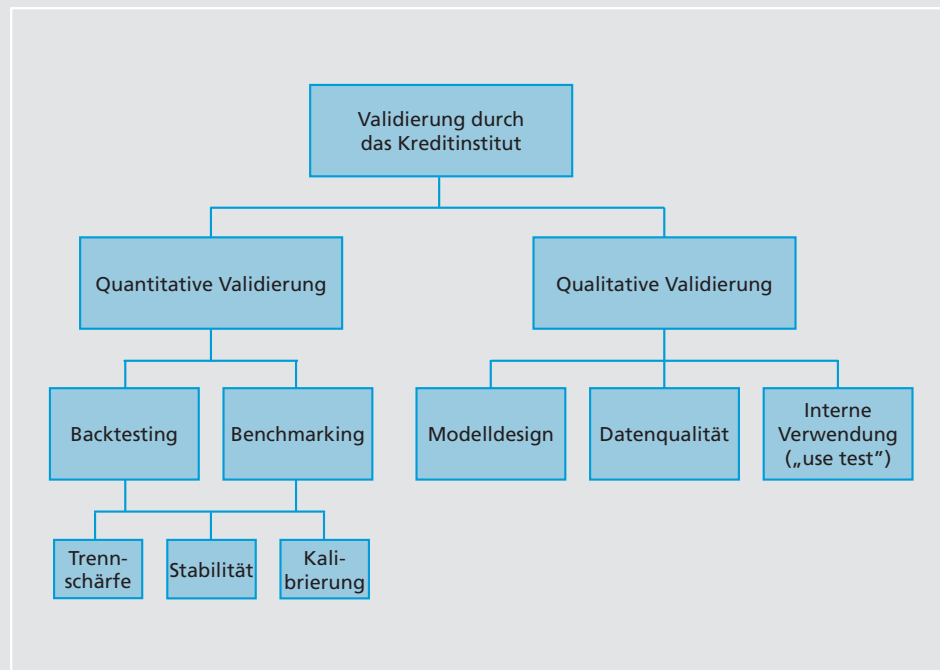
## Validierungsansätze für interne Ratingsysteme

Der neue internationale Eigenkapitalstandard für Kreditinstitute (Basel II) gibt Banken die Möglichkeit, ihre eigenen Ratingverfahren zur Bestimmung der für die Kapitalunterlegung relevanten Risikogewichte zu verwenden. Im Gegenzug sind die Banken verpflichtet, ihre Ratingsysteme regelmäßig zu überprüfen (Validierung). Regulatorische Standards für die Validierung sollen einheitliche Rahmenbedingungen für die aufsichtliche Zertifizierung und die laufende Überwachung der eingesetzten Ratingsysteme gewährleisten.

Die Validierung stellt für Banken und Bankenaufseher eine große Herausforderung dar. Die zur quantitativen Validierung verwendeten statistischen Verfahren eignen sich zwar als Indikatoren für mögliche Fehlentwicklungen. Aus ihnen lässt sich jedoch in der Regel kein strenges Kriterium für die Eignung eines Ratingverfahrens ableiten. Daher werden qualitative Kriterien bei der Validierung eine wichtige Rolle spielen.

Nicht zuletzt auf Grund der steigenden Verfügbarkeit verlässlicher Daten sind in den nächsten Jahren methodische Weiterentwicklungen der dargestellten Verfahren zu erwarten. Insbesondere dürften die zukünftigen Diskussionen von Forschung und Bankpraxis weitere Erkenntnisse über die eingesetzten Methoden zur Schätzung der Risikoparameter liefern.

## Aspekte der Validierung



Deutsche Bundesbank

Ratingverfahren dienen der Ermittlung des Kreditrisikos einzelner Kreditnehmer. Dazu werden mit Hilfe unterschiedlicher Methoden den betreffenden Kreditnehmern Risikokennzahlen zugewiesen, die den Grad ihrer Kreditwürdigkeit anzeigen.

Auf Grund der bevorstehenden aufsichtlichen Anerkennung bankinterner Ratingsysteme in den beiden IRB-Ansätzen („Internal Ratings Based Approach“) ist deren quantitative und qualitative Validierung ein aktuell viel diskutierter Problembereich. Unter Validierung versteht man dabei den gesamten Prozess der Überprüfung eines internen Ratingsystems, von der Validierung seiner Trennschärfe bis zur prozessorientierten Validierung („use test“). Einen Überblick über die wesentlichen Teilaspekte des Validierungsprozesses für

Ratingverfahren gibt das oben stehende Schaubild.

Mit der Validierung von Ratingsystemen eng verbunden ist die Validierung weiterer Risikoparameter, die aus den Ratingeinstufungen abgeleitet werden und in den IRB-Ansätzen der neuen Baseler Mindestanforderungen (Basel II) die Höhe des vorzuhaltenden Eigenkapitals einer Bank maßgeblich bestimmen. Dieser Aufsatz soll näher in die Validierungsthematik einführen, ohne dabei ein aufsichtliches Präjudiz für oder gegen bestimmte Verfahren zu schaffen. Im Sinne einer „Best-Practice“-Auswahl fließen Antworten aus einer Umfrage unter deutschen Banken ein, die im Frühjahr 2003 durchgeführt wurde.

## Quantitative Aspekte der Validierung

Die Ausgestaltung der quantitativen, aber auch der qualitativen Validierung hängt in hohem Maße von der Art des verwendeten Ratingsystems ab. Man unterscheidet hier grundsätzlich zwischen modellgestützten Systemen und solchen, die auf einem Expertenurteil beruhen.

### *Modell- gestützte Ratingsysteme*

Modellgestützte Systeme, wie die Diskriminanzanalyse oder verschiedene Arten der Regressionsanalyse, werden typischerweise auf Basis historischer Ausfalldaten entwickelt. Liegen solche Daten nicht in ausreichendem Umfang vor, wird häufig ein so genanntes „Shadow-Rating“ durchgeführt, das die Bonitätseinstufung externer Ratingagenturen nachbildet. Allen modellgestützten Verfahren ist gemein, dass sie – mit Hilfe statistischer Methoden – eine Reihe von Risikofaktoren (z. B. Kreditsumme, Eigenkapital oder Berufsgruppe) in eine Risikokennzahl (Bonitätsscore) abbilden.

### *Expertenurteil*

Liegen insgesamt wenig statistisch verwertbare Informationen vor oder sind die Kreditbeziehungen materiell bedeutsam oder komplex, so wird die Bank eher auf eine Expertenmeinung vertrauen. In der Regel wird auch in einem solchen Ratingsystem die Ratingeinstufung in einem standardisierten Verfahren erfolgen. Der Unterschied zu modellgestützten Verfahren liegt hier vor allem in der fehlenden statistischen Modellierung des Bonitätsscores.

### *Mischsysteme*

Am häufigsten findet man in der Praxis Mischformen beider Klassen von Ratingsystemen vor. In solchen gemischten Systemen

kann der zuständige Kreditspezialist die modellgestützte Einstufung korrigieren, falls ihm Informationen vorliegen, die vom modellgestützten Ratingsystem nicht oder nur unzureichend berücksichtigt werden.

Alle Ratingsysteme – modellgestützte oder auf Expertenurteil beruhende – lassen sich im Prinzip quantitativ validieren. Eine quantitative Validierung setzt allerdings voraus, dass eine ausreichende Anzahl von Kreditausfällen vorliegt. Typischerweise ist diese Voraussetzung im Mengengeschäft gegeben, also für Kredite an kleine und mittlere Unternehmen oder für solche an Privatpersonen. Die maßgeblichen Kriterien bei der quantitativen Validierung eines Ratingsystems sind seine Trennschärfe, seine Stabilität und seine Kalibrierung.

### **Trennschärfe und Stabilität**

Die Trennschärfe eines Ratingsystems bezeichnet seine Fähigkeit, ex ante zwischen ausgefallenen und nicht ausgefallenen Kreditnehmern zu diskriminieren. Zur Überprüfung der Trennschärfe bieten sich eine Reihe statistischer Trennschärfemaße an, von denen einige im Anhang näher erläutert werden. Allerdings ist die absolute Höhe der Trennschärfe eines Ratingsystems nur bedingt aussagekräftig. Nur unter Berücksichtigung statistischer Unschärfen kann etwa ein direkter Vergleich verschiedener Ratingsysteme durchgeführt werden. Ein solcher Vergleich muss dann auf demselben Datensatz erfolgen.

Die Trennschärfe sollte zudem nicht nur im Entwicklungsdatensatz überprüft werden,

*Kriterien der  
quantitativen  
Validierung*

*Trennschärfe  
eines  
Ratingsystems*

## Gütekriterien für Ratingsysteme

### **Trennschärfe:**

Die Trennschärfe von Ratingsystemen beschreibt die Fähigkeit, im Voraus ausfallgefährdete Kreditnehmer zu erkennen. Ein maximal trennscharfes Ratingsystem könnte demnach bereits im Vorfeld alle später ausfallenden Kreditnehmer exakt identifizieren. In der Praxis gibt es solche Ratingsysteme jedoch nicht. Eine hohe Trennschärfe wird attestiert, wenn die tendenziell guten Ratingklassen insgesamt einen geringen Anteil der später ausfallenden und einen hohen Anteil der später nicht ausfallenden Kreditnehmer aufnehmen, während es sich in den tendenziell schlechten Klassen genau umgekehrt verhält.

### **Stabilität:**

Ein stabiles Ratingsystem zeichnet sich dadurch aus, dass es die Ursache-Wirkungs-Beziehung zwischen den Risikofaktoren und der Bonität adäquat modelliert. Es vermeidet Scheinabhängigkeiten auf Grund empirischer Korrelationen. Im Gegensatz zu stabilen Systemen weisen instabile Systeme häufig eine im Zeitablauf stark nachlassende Prognosegüte auf.

### **Genauigkeit der Kalibrierung:**

Kalibrierung bezeichnet üblicherweise die Zuweisung von Ausfallwahrscheinlichkeiten zu den Ratingklassen. Ein Ratingsystem ist gut kalibriert, wenn die geschätzten Ausfallwahrscheinlichkeiten nur wenig von den realisierten Ausfallraten abweichen. In einem weiteren Sinn zählt man auch die Zuweisung zusätzlicher Risikoparameter, wie etwa die Verlustquote und die Kredithöhe zum Zeitpunkt des Ausfalls, zur Kalibrierung des Ratingsystems.

Deutsche Bundesbank

sondern auch in einem davon unabhängigen (Out-of-Sample-Validierung). Andernfalls besteht die Gefahr, durch Überanpassung an die Daten des Entwicklungsbestandes die Trennschärfe zu überschätzen. Häufig wird dann das Ratingsystem auf einem vom Entwicklungsbestand unabhängigen, aber strukturell ähnlichen Datenbestand eine relativ niedrige Trennschärfe aufweisen. Das Ratingssystem würde also über eine geringe Stabilität verfügen.

Eine Möglichkeit, die Stabilität eines modellgestützten Ratingsystems zu prüfen, kann darin liegen, die statistische Signifikanz der verwendeten Risikofaktoren zu bewerten. Zusätzlich sollten bestehende Korrelationseffekte geprüft werden. Hohe oder instabile Korrelationen können die Stabilität des Ratingsystems negativ beeinflussen.

*Stabilität eines  
Ratingsystems*

## Kalibrierung

Die Höhe der Eigenkapitalanforderungen einer Bank nach Basel II wird in den beiden IRB-Ansätzen durch die bankinternen Schätzungen der Risikoparameter für jede Kreditforderung bestimmt. Diese werden aus den bankinternen Bonitätseinstufungen abgeleitet. Zu nennen sind hier die Ausfallwahrscheinlichkeit des Kreditnehmers (Probability of Default: PD) sowie, für den fortgeschrittenen IRB-Ansatz, die erwartete Verlustquote (Loss Given Default: LGD) und die Kredithöhe zum Zeitpunkt des Ausfalls (Exposure At Default: EAD). Man spricht in diesem Zusammenhang auch von der Kalibrierung des Ratingsystems. Da die Risikoparameter von der Bank selbst bestimmt werden können, ist

*Die Risiko-  
parameter  
nach Basel II*

aus aufsichtlicher Sicht die Güte der Kalibrierung ein entscheidendes Kriterium für die Beurteilung von Ratingsystemen.

Dabei stellen neben der Ausfallwahrscheinlichkeit auch die Verlustquote und die Kredithöhe bei Ausfall Zufallsgrößen dar, da sie zum Zeitpunkt der Bonitätseinstufung des Kreditnehmers nicht vollständig bekannt sind. Insbesondere hängen sie von der Werthaltigkeit der Sicherheiten beziehungsweise von dem bis zum Ausfall in Anspruch genommenen Kreditbetrag ab. Im Gegensatz zur Ausfallwahrscheinlichkeit müssen diese Parameter jedoch nur im fortgeschrittenen IRB-Ansatz von der Bank selbst geschätzt werden, während sie im Basis-IRB-Ansatz aufsichtlich vorgegeben werden.

*Verfahren zur  
Ermittlung der  
Ausfallwahrscheinlichkeiten*

Es gibt mehrere bewährte statistische Verfahren, um aus einem Ratingverfahren Ausfallwahrscheinlichkeiten abzuleiten. Dabei lassen sich direkte und indirekte Verfahren unterscheiden. Bei den direkten Verfahren, wie Logit-, Probit- und Hazardratenmodellen, kann der ermittelte Bonitätsscore selbst als Ausfallwahrscheinlichkeit des Kreditnehmers interpretiert werden. Die Ausfallwahrscheinlichkeit einer Ratingklasse wird dann in der Regel als Mittelwert der Ausfallwahrscheinlichkeiten der einzelnen Schuldner pro Klasse berechnet.

Sofern der Bonitätsscore nicht als Ausfallwahrscheinlichkeit verstanden werden kann (wie bei der Diskriminanzanalyse), wird auf indirekte Verfahren zurückgegriffen. Eine einfache Möglichkeit besteht darin, die Ausfallwahrscheinlichkeit für jede Ratingklasse aus

historischen Ausfallraten zu schätzen. Eine weitere Methode ist die Schätzung der Scoreverteilungen der ausgefallenen Kreditnehmer einerseits und der nicht ausgefallenen Kreditnehmer andererseits. Anschließend kann mit Hilfe der Bayesschen Formel jedem Kreditnehmer eine spezifische Ausfallwahrscheinlichkeit zugeordnet werden.

In der Praxis werden die Schätzungen der Ausfallwahrscheinlichkeiten einer Bank von den später tatsächlich beobachteten Ausfallraten abweichen. Die entscheidende Frage ist, ob die Abweichungen rein zufällig sind oder ob sie systematisch erfolgen. Sowohl aus bankaufsichtlicher wie auch aus bankinterner Sicht ist eine systematische Unterschätzung der Ausfallwahrscheinlichkeiten kritisch zu beurteilen, da die Eigenmittelanforderung einer Bank in diesem Fall im Vergleich zu ihren Risiken zu niedrig ausfällt.

In der Wissenschaft werden verschiedene statistische Verfahren zur Überprüfung der Schätzqualität von Ausfallwahrscheinlichkeiten diskutiert. Die meisten dieser Verfahren beruhen auf einem so genannten „Backtesting“. Allerdings weisen diese Verfahren in der Bankpraxis Mängel auf, die eine rein mechanische Anwendung nicht ratsam erscheinen lassen. Dies soll anhand des Binomialtests veranschaulicht werden, dessen technische Details im Anhang beschrieben werden.

Der Binomialtest hat bereits im Rahmen des Backtestings von Marktrisikomodellen Einzug in die Aufsichtspraxis erhalten. Auch bei der Beurteilung von Ausfallwahrscheinlichkeiten

*Binomialtest*

lässt sich unter vereinfachenden Annahmen ein statistischer Test konstruieren, der auf der Binomialverteilung beruht. Dabei wird vorausgesetzt, dass die Ausfälle pro Ratingklasse statistisch unabhängig sind. Unter der Hypothese, dass die geschätzten Ausfallwahrscheinlichkeiten der Ratingklassen korrekt sind, wäre die tatsächlich nach einem Jahr zu beobachtende Zahl von Ausfällen pro Ratingklasse binomialverteilt. Ergeben sich starke Abweichungen der Ausfallrate von der geschätzten Ausfallwahrscheinlichkeit der Ratingklasse, so müsste die Hypothese der korrekten Schätzung abgelehnt werden. Das Ratingmodell wäre also schlecht kalibriert.

Ein Schwachpunkt dieses Tests ist die Annahme, dass die Ausfälle der Kreditnehmer unabhängige Ereignisse darstellen. In der Praxis sind die Ausfälle jedoch auf Grund konjunktureller Einflüsse mehr oder weniger stark korreliert. Rein theoretisch wäre eine Lösung dieses Problems denkbar, wenn die Ausfallkorrelationen bekannt wären. Die Bestimmung von Ausfallkorrelationen ist jedoch schwierig. Somit ist selbst ein modifizierter Binomialtest allenfalls als Indikator für eine gute oder schlechte Kalibrierung geeignet.

Ein anderer Ansatz bei der statistischen Validierung von Ausfallwahrscheinlichkeiten ist die Verwendung von Referenzportfolios. In der Bankpraxis ist zum Beispiel der Bezug auf externe Daten von Ratingagenturen und anderen kommerziellen Anbietern als Benchmark weit verbreitet. Systematische Abweichungen der bankinternen Schätzungen von den Schätzungen für das Referenzportfolio

müssten überprüft werden. Das Benchmarking kann eine wertvolle Ergänzung des Validierungsprozesses darstellen. Die Nützlichkeit dieses Ansatzes hängt aber sehr von der Wahl eines geeigneten Referenzportfolios ab. Auch die Auswahl eines Benchmark-Ratings ist im Allgemeinen keine leichte Aufgabe.

Neben der Schätzung der Ausfallwahrscheinlichkeit sieht Basel II in dem fortgeschrittenen IRB-Ansatz auch die bankinterne Schätzung der Verlustquote und der Kredithöhe bei Ausfall vor. Eine quantitative Validierung der Verlustquoten besteht darin, die bankinternen Schätzungen zu verifizieren. Bei Bankkrediten wird die Verlustquote hauptsächlich durch die Verwertung der Kreditsicherheiten bestimmt. Fällt ein Kredit aus, so wird dem Kreditinstitut die Höhe des tatsächlich eingetretenen Verlustes erst nach dem Abwicklungszeitraum bekannt. Der Abwicklungszeitraum kann – abhängig von den Kreditmerkmalen und insbesondere von der Besicherung – sehr unterschiedlich ausfallen. Er beträgt in der Regel 18 Monate bis drei Jahre, in Ausnahmefällen sogar mehr als zehn Jahre.

Zur Bestimmung des tatsächlichen Verlustes sind sämtliche Zahlungsströme zu berücksichtigen, die während der Abwicklung anfallen, und gegebenenfalls einzelnen Sicherheiten zuzuordnen. Die Zahlungsströme setzen sich zusammen aus an die Bank geleisteten Zahlungen und solchen, die die Bank selbst zu leisten hat. Erstere bestimmen sich vor allem aus Teilzahlungen des Kreditnehmers oder Erlösen aus der Verwertung von Sicherheiten. Letztere bestehen zum Beispiel aus Kosten für Anwälte, Gerichtskosten sowie auflaufen-

*Messung der  
Verlustquoten*

*Verwendung  
von Referenz-  
portfolios und  
externen  
Datenquellen*

den Zinsen und Refinanzierungskosten während der Abwicklung. Auf Grund der Dauer des Abwicklungsverfahrens müssen die anfallenden Zahlungsströme diskontiert werden, um den tatsächlichen ökonomischen Verlust bei Ausfall zu bestimmen.

Zu den Verlustquoten von börsengehandelten Unternehmensanleihen kann bereits auf eine Reihe statistischer Untersuchungen zurückgegriffen werden. Einheitliche Datenbanken für Verluste aus Buchkrediten befinden sich dagegen erst im Aufbau. Allerdings ist anzunehmen, dass auch für Buchkredite die Verlustquoten deutlich branchenspezifisch und stark mit den Ausfallquoten korreliert sind. Die Verlustdatenbank muss die Verluste vollständig erfassen und auch die ausgefallenen Kredite enthalten, die in ihren Blankoanteilen nicht zu Verlusten geführt haben. Die ausschließliche Erfassung von Krediten, die tatsächlich zu Verlusten geführt haben, würde zu einer Überschätzung der Verlustquoten führen. Häufig tritt auch der Fall ein, dass mehrere Kredite durch ein und dieselbe Sicherheit besichert sind (z.B. Globalgrundschuld). In der Regel wird ein Kreditinstitut versuchen, für jede Sicherheitenklasse eine eigene Erlösquote zu schätzen. Bei Globalsicherheiten muss die Sicherheit auf die einzelnen Kredite verteilt werden.

*Höhe der  
Kreditforderung  
bei Ausfall*

Wie bei der Verlustquote beruht die Validierung der Kredithöhe bei Ausfall auf einer Überprüfung der bankinternen Schätzungen. Für bilanzielle Forderungen sehen die Baseler Mindestanforderungen vor, dass die Schätzwerte nicht geringer sein dürfen als der aktuell in Anspruch genommene Kreditbetrag,

wobei die Effekte des bilanziellen Nettings berücksichtigt werden dürfen. Für derivative Geschäfte wird der Kreditäquivalenzbetrag aus den Wiederbeschaffungskosten zuzüglich eines Aufschlags für mögliche zukünftige Verpflichtungen ermittelt. Somit konzentrieren sich die ergänzenden bankaufsichtlichen Anforderungen an die bankinternen Schätzungen der Kredithöhe bei Ausfall auf außerbilanzielle Geschäfte. Zentrales Problem ist dabei die Ermittlung des in Anspruch genommenen Anteils von zugesagten Beträgen zum Zeitpunkt des Ausfalls. Untersuchungen zeigen, dass es signifikante Zusammenhänge zwischen der Kredithöhe bei Ausfall und der Restlaufzeit des Kredits sowie zwischen der Kredithöhe bei Ausfall und der Bonität des Schuldners gibt. Die Kredithöhe bei Ausfall ist durch zusätzliche Inanspruchnahme von Kreditzusagen tendenziell umso höher, je länger die Restlaufzeit des Kredits ist. Dies ist plausibel, denn je länger die Restlaufzeit eines Kredits ist, desto größer ist die Wahrscheinlichkeit, dass sich die Bonität des Schuldners verschlechtert und sich seine Zugriffsmöglichkeiten auf alternative Finanzierungsquellen verringern. Andere Untersuchungsergebnisse besagen, dass die Ausnutzung der Kreditzusagen im Ausfallzeitpunkt tendenziell umso niedriger ist, je schlechter die Bonität eines Schuldners zum Zeitpunkt der Einräumung des Kreditrahmens war. Als Erklärung dafür wird angeführt, dass eine Bank bei Kreditnehmern mit niedriger Bonität Klauseln in den Kreditvertrag aufnimmt, die ihnen die Inanspruchnahme des zugesagten Kreditrahmens bei einer weiteren Bonitätsverschlechterung erschweren.

Die Schätzungen lassen sich erheblich vereinfachen, falls Abhängigkeiten von der Bonität und der Restlaufzeit nicht berücksichtigt werden müssen. Dies birgt allerdings die Gefahr in sich, dass durch Vernachlässigung dieser Abhängigkeiten die Schätzwerte für die Kreditanspruchnahme systematisch verzerrt sind.

### Qualitative Aspekte der Validierung

*Qualitative Validierung ergänzt quantitative Validierung*

Die quantitativen Validierungstechniken müssen durch qualitative – also nichtstatistische – Verfahren ergänzt werden. Nicht zuletzt dient die qualitative Validierung dazu, die Anwendbarkeit quantitativer Techniken sicherzustellen. In diesen Fällen wird die qualitative Validierung vor der quantitativen durchgeführt werden müssen. Die qualitativen Analysen überprüfen vor allem drei Aspekte: das Design der Ratingmodelle, die Datenqualität für Ratingentwicklung und -einsatz sowie die interne Verwendung des Ratingsystems im Kreditvergabeprozess („use test“).

*Modelldesign*

Die Überprüfung des Modelldesigns spielt vor allem bei modellgestützten Systemen, aber nicht nur bei diesen, eine wichtige Rolle. Dies trifft insbesondere dann zu, wenn eine quantitative Validierung auf Grund der Datenlage nur eingeschränkt möglich ist. In jedem Fall muss der Prozess der Ratingvergabe transparent und gut dokumentiert sein. Der Einfluss der Risikofaktoren sollte deutlich aufgeschlüsselt werden und ökonomisch plausibel sein. Bei modellgestützten Systemen ist darüber hinaus der Nachweis der statistischen Fundierung entscheidend.

Generell sollte eine Bank auf die Integrität ihrer Daten sowie deren konsistente Erhebung achten. Nur eine solide Datenbasis mit hinreichend großer Datenhistorie erlaubt die Entwicklung eines hochwertigen Rating-systems sowie zuverlässige Schätzungen der aufsichtlich geforderten Risikoparameter. Liegen dem Kreditinstitut selbst nur wenig Ausfallinformationen vor, so kann es – wie bereits erwähnt – gegebenenfalls auf externe Datenquellen zurückgreifen.

*Datenqualität und -verfügbarkeit*

Maßgeblich bei der qualitativen Validierung interner Ratingsysteme ist darüber hinaus die tatsächliche Verwendung der Ratingergebnisse im bankinternen Risikomanagement und Reporting. Diese Art der qualitativen Validierung prüft die Ausgestaltung der bankinternen Prozesse und wird daher auch als „prozessorientierte Validierung“ bezeichnet. Beispiele für ein Kreditrisikomanagement mit Hilfe von Ratingsystemen sind unter anderem ratingbasierte Kreditentscheidungen und Kreditvergabekompetenzen, eine nach Bonitätsstufen ausgerichtete Kreditrisikostategie sowie entsprechend strukturierte Limitsysteme. Bei all diesen Anwendungen stellt ein Kreditinstitut bei wichtigen geschäftspolitischen Entscheidungen auf die Risikobewertung durch die bankinternen Bonitätsstufen ab.

*„use test“*

Aus aufsichtlicher Sicht spiegelt die Art und Weise, wie die Bank ihr Ratingsystem für interne Entscheidungsprozesse nutzt, das Vertrauen wider, das sie ihrem eigenen System entgegenbringt. Wenn bankinterne Ratingsysteme intern nicht oder nur für einzelne, isolierte Zwecke verwendet werden, kann dies als bankinterne Einschätzung der (man-



gelnden) Qualität der Ratingsysteme interpretiert werden. Ein Ratingsystem ohne ausreichende Einbindung in die bankinternen Kreditprozesse wird daher nicht anerkanntsfähig sein.

Auch die Quantifizierung des Risikos, ausgedrückt in Ausfallwahrscheinlichkeiten sowie Erlös- beziehungsweise Verlustquoten, sollte bankintern genutzt werden. Das wichtigste Beispiel hierfür ist die Ermittlung von Standardrisikokosten als Teil der Deckungsbeitragsrechnung. Außerdem ist die Bestimmung der Risikovorsorge auf Basis von Standardrisikokosten als Indikator für die interne Nutzung der Ratingverfahren vorstellbar.

#### *Unabhängigkeit*

Die Baseler Mindestanforderungen an bankinterne Ratingsysteme verlangen außerdem, dass Ratingentscheidungen nicht von anderen Geschäftsbereichen, die direkt oder indirekt von der Kreditentscheidung profitieren, beeinflusst werden. Besonders wichtig ist die unabhängige Ratingvergabe beim Einsatz von Expertenurteilen. In diesen Fällen sollte die endgültige Ratingkompetenz im Bereich „Marktfolge“ und keinesfalls im Bereich „Markt“ liegen. Dies gilt umso mehr, wenn die Vertriebsmitarbeiter volumenabhängig vergütet werden. Zu den qualitativen Kriterien gehört deshalb, dass die Ratingvorentcheidung, die potenziell durch den Kundenbetreuer geschehen kann, durch einen unabhängigen Dritten überprüft und bestätigt wird.

#### *Andere Faktoren*

Weitere wichtige Punkte des Validierungsverfahrens sind die Qualifizierung der Mitarbeiter und die Akzeptanz der Ratingverfahren

bei ihren Nutzern. Diese müssen über ein gutes Verständnis für das Ratingverfahren verfügen und es im täglichen Geschäft tatsächlich einsetzen.

### **Anwendungsperspektiven eines zentralen Kreditregisters für Validierungszwecke**

---

Sofern ein Institut die Zulassung von internen Ratingsystemen gemäß Basel II beantragt, muss es eine angemessene Validierung seines Systems nachweisen. Aufgabe der Bankenaufsicht ist es, die Ratingsysteme zu zertifizieren und die Einhaltung dieser Mindestanforderungen durch die Bank laufend zu überwachen. Im Rahmen dieses Prozesses müssen auch die bankinternen Validierungsverfahren beurteilt werden. Hierbei können zentrale Kreditregister eine wichtige Rolle spielen. Im Baseler Ausschuss für Bankenaufsicht werden daher zurzeit deren Anwendungsperspektiven erörtert.

Voraussetzung für die Verwendung eines zentralen Kreditregisters für bankaufsichtliche Zwecke ist vor allem die Verfügbarkeit von Informationen über Kreditausfälle, über bankinterne Ratingeinstufungen sowie über die Besicherung der Kredite. Zum Teil sind diese Informationen bereits jetzt in zentralen Kreditregistern einiger Länder vorhanden. Ein zentrales Kreditregister bietet gegenüber der Alternative von Einzelanfragen auf Grund einheitlicher Kreditnehmernummern den Vorteil einer direkten Vergleichsmöglichkeit der Ratingeinstufung verschiedener Banken für ein und denselben Kreditnehmer (Benchmarking).

*Erforderliche  
Datenbasis*

king). Die als Vergleich gewählte Stichprobe wäre flexibel bestimmbar. Zudem würde das Meldewesen die Banken in ihrer Gesamtheit abdecken.

*Anwendung für  
Backtesting*

Ein weiteres Anwendungsgebiet von Kreditregistern im Zusammenhang mit Validierungsfragen ist das Backtesting. Wie bereits dargestellt, bezeichnet das Backtesting den Vergleich der Schätzungen für die Ausfallwahrscheinlichkeiten mit den tatsächlich aufgetretenen Kreditausfällen. Damit könnte im Prinzip die bankinterne quantitative Validierung überprüft werden.

Grundsätzlich können zentrale Kreditregister also eine unterstützende Rolle bei der aufsichtlichen Zertifizierung von Ratingsystemen und deren Überwachung spielen. Je nach Umfang der dafür notwendigen Untersuchungen erfordert dies Änderungen der zentralen Kreditregister in ihrer bestehenden Form. Hierüber sollte unter sorgfältigen Kosten-Nutzen-Erwägungen entschieden werden. Der wesentliche Anwendungszweck von zentralen Kreditregistern wird voraussichtlich das Benchmarking von Schätzungen unterschiedlicher Kreditinstitute sein. Demgegenüber sind der Verwendung von Kreditregistern für das Backtesting durch die hohen Anforderungen an den Detailliertheitsgrad der hierfür benötigten Kreditinformationen Grenzen gesetzt.

## Ausblick

---

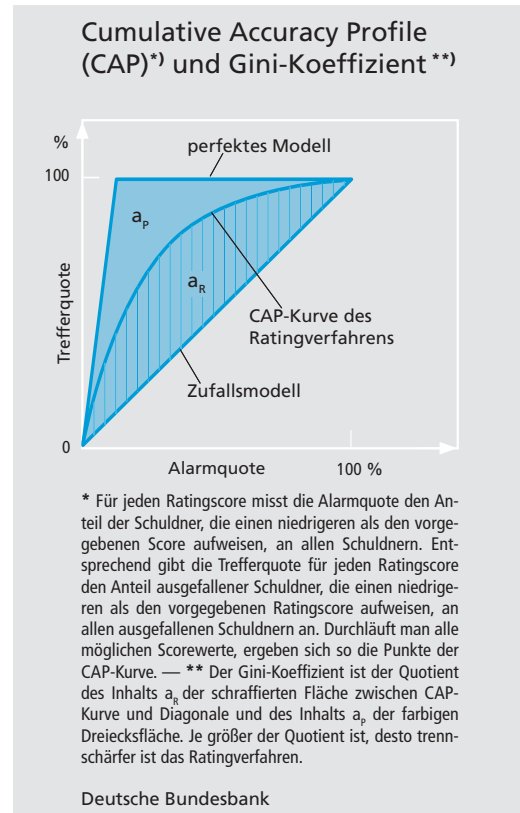
Banken und Bankenaufsicht bereiten sich gegenwärtig intensiv auf die Validierung von Ratingsystemen vor. Um die Entwicklung der Validierungsansätze weiter voranzutreiben, wurde unter der Federführung der Research Task Force des Baseler Ausschusses für Bankenaufsicht eine Arbeitsgruppe für Validierungsfragen eingesetzt. Die in den letzten Jahren stark angestiegene Zahl von Veröffentlichungen zu diesem Thema zeigt, dass sich auch die Wissenschaft dieser Fragen angenommen hat. Die Eignung einzelner Verfahren ist aber noch umstritten. Fest steht, dass die Beurteilung der bankinternen Ratingssysteme nicht mittels einer einzelnen Validierungsmethode erfolgen kann, sondern sich als Gesamtbild aus verschiedenen quantitativen und qualitativen Methoden ergeben wird. Die gegenwärtige Diskussion trägt zu einer Weiterentwicklung der Validierungstechniken bei. Außerdem wird sich die Qualität und Quantität der verfügbaren Daten in den kommenden Jahren entscheidend verbessern. Die daraus gewonnenen Erkenntnisse werden in die aufsichtlichen Validierungsstandards einfließen.

## Anhang

### Statistische Trennschärfemaße

#### Cumulative Accuracy Profile (CAP)

Die CAP-Kurve liefert eine graphische Veranschaulichung der Trennschärfe eines Ratingverfahrens. Hierzu wird für den Datenbestand, auf dem die Trennschärfe des Ratingmodells untersucht werden soll, der Bonitätsindikator (Scorewert) für jeden Kreditnehmer ermittelt. Dieser kann eine stetige Größe sein, etwa das Ergebnis einer Diskriminanzanalyse oder einer Logit-Regression. Er kann auch eine ganze Zahl sein, die die Ratingklasse repräsentiert, in die der Kreditnehmer eingestuft wurde. In der folgenden Analyse wird angenommen, dass ein hoher Scorewert eine gute Bonität widerspiegelt. Im ersten Schritt werden die Kreditnehmer nach aufsteigenden Scorewerten geordnet. Die CAP-Kurve wird dann bestimmt, indem man auf der horizontalen Achse den kumulativen Anteil aller Schuldner („Alarmquote“) aufträgt und auf der vertikalen Achse den kumulativen Anteil aller ausgefallenen Schuldner („Trefferquote“). Dies wird in dem nebenstehenden Schaubild illustriert. Wenn beispielsweise unter den 30 % aller Schuldner mit den niedrigsten Ratingscores 70 % aller ausgefallenen Schuldner enthalten sind, dann liegt der Punkt (0,3;0,7) auf der CAP-Kurve. Ein Ratingverfahren ist umso trennschärfer, je steiler die CAP-Kurve zu Beginn ansteigt. Im Idealfall würde das Ratingverfahren allen ausgefallenen Schuldnern die niedrigsten Scorewerte zuordnen. Die CAP-Kurve würde dann zu Beginn linear ansteigen und anschließend horizontal verlaufen. Der andere Grenzfall wäre eine rein zufällige Ratingeinstufung. Ein solches Ratingverfahren hätte keinerlei Trennfähigkeit. Die zu erwartende CAP-Kurve wäre in diesem Fall identisch mit der Diagonalen. In der Realität sind Ratingeinstufungen weder perfekt noch zufällig. Die entsprechende CAP-Kurve



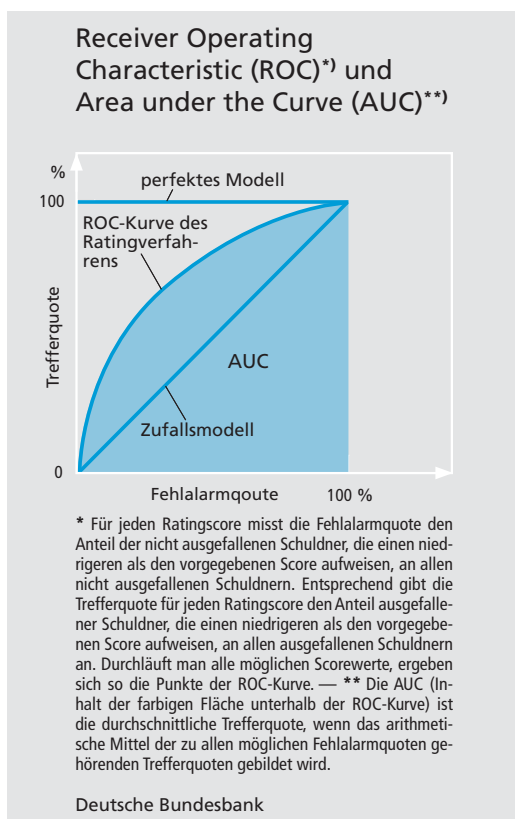
verläuft daher zwischen diesen beiden Grenzfällen. Die Trennschärfe eines Ratingverfahrens lässt sich mit Hilfe der CAP-Kurve in einer einzigen Zahl, dem so genannten „Gini-Koeffizienten“<sup>1)</sup> (GK) zusammenfassen. Im oben stehenden Schaubild wird die Fläche zwischen dem perfekten Rating und dem zufälligen Rating mit  $a_p$  bezeichnet, die Fläche zwischen dem untersuchten Rating und dem zufälligen Rating mit  $a_R$ . Der Gini-Koeffizient ist dann definiert als das Verhältnis von  $a_R$  zu  $a_p$ , das heißt

$$GK = \frac{a_R}{a_p}$$

Der Gini-Koeffizient liegt stets zwischen minus Eins und Eins. Ein Ratingverfahren ist umso trennschärfer, je näher er bei Eins liegt.

Gini-Koeffizient (GK)

1 Der Gini-Koeffizient wird in der englischsprachigen Literatur häufig als „Accuracy Ratio“ bezeichnet.



Receiver Operating Characteristic (ROC)

Ein mit der CAP-Kurve verwandtes Konzept ist die ROC-Kurve. Dabei wird für die ausgefallenen Kreditnehmer einerseits und für die nicht ausgefallenen Kreditnehmer andererseits jeweils die empirische Score-Verteilung ermittelt. Das Resultat könnte etwa so aussehen wie im Schaubild auf Seite 73 dargestellt. Im nächsten Schritt gibt man sich einen Scorewert  $C$  vor. Anhand dieses Scorewerts  $C$  kann man eine einfache Entscheidungsregel definieren, um potenzielle Ausfallkandidaten zu identifizieren. Alle Kreditnehmer mit einem Scorewert größer als  $C$  werden als kreditwürdig eingestuft, solche mit kleinerem Scorewert als nicht kreditwürdig. Ein gutes Ratingverfahren ist dadurch gekennzeichnet, dass es eine möglichst hohe Trefferquote (korrekte Klassifizierung eines Kreditnehmers als Ausfallkandidat) bei einer gleichzeitig möglichst niedrigen Fehlarmlarmquote (inkorrekte Klassifizierung eines Kreditnehmers guter Bonität als Ausfallkandidat) aufweist.

Um unabhängig von der Wahl des Schwellenwerts  $C$  die Trennschärfe eines Ratingverfahrens analysieren zu können, wird für jedes  $C$ , das zwischen dem maximalen und dem minimalen Scorewert liegt, sowohl die Fehlarmlarmquote als auch die Trefferquote berechnet. Die so ermittelten Punktpaare ergeben die ROC-Kurve (vgl. nebenstehendes Schaubild). Ein Ratingverfahren ist umso trennschärfer, je steiler die ROC-Kurve zu Beginn ansteigt. Im Fall eines perfekten Ratingverfahrens bestünde die ROC-Kurve nur aus der Verbindung der Punkte  $(0;0)$ ,  $(0;1)$  und  $(1;1)$ . Bei einem rein zufälligen Ratingverfahren verläuft die ROC-Kurve im nebenstehenden Schaubild genau auf der Diagonalen.

Ähnlich wie für die CAP-Kurve lässt sich auch für die ROC-Kurve eine aggregierte Kennzahl angeben. Diese ergibt sich aus der Fläche unterhalb der ROC-Kurve und wird als AUC bezeichnet. Die Kennzahl AUC liegt stets zwischen Null und Eins. Je näher AUC an dem Wert Eins liegt, desto trennschärfer ist ein Ratingverfahren. Im nächsten Abschnitt wird auf den Zusammenhang zwischen AUC und GK sowie auf die statistischen Eigenschaften von AUC und GK eingegangen. Ein wichtiges Resultat ist die Äquivalenz von AUC und GK. Man kann durch eine einfache lineare Transformation die eine Kennzahl in die andere überführen.

„Area under the Curve“ (AUC)

Ein weiteres in der Praxis gebräuchliches Trennschärfemaß ist der minimale Klassifizierungsfehler. Seine Berechnung lässt sich anhand des Schaubilds auf Seite 73 verdeutlichen. Als Klassifizierungsfehler bezeichnet man den Mittelwert der relativen Häufigkeiten von ausgefallenen Kreditnehmern und nicht ausgefallenen Kreditnehmern, die bei einem Schwellenwert  $C$  falsch klassifiziert worden sind. Der Anteil der ausgefallenen Kreditnehmer, die unter Beachtung des Schwellenwerts  $C$  als kreditwürdig eingestuft worden sind, entspricht der

Minimaler Klassifizierungsfehler

Fläche unter der Kurve der Score-Verteilung der ausgefallenen Schuldner, die sich rechts von C befindet. Analog entspricht der Anteil der nicht ausgefallenen Schuldner, die irrtümlich als nicht kreditwürdig klassifiziert worden sind, der Fläche unter der Score-Verteilung der nicht ausgefallenen Schuldner, die sich links von C befindet. Die Hälfte der Summe dieser beiden Flächen entspricht dem Klassifizierungsfehler. Berechnet man den Klassifizierungsfehler für jeden Wert von C, der zwischen dem minimalen und dem maximalen Scorewert liegt, und bestimmt das Minimum, so erhält man den minimalen Klassifizierungsfehler. Dieser ist umso kleiner, je trennschärfer das Ratingsystem ist. Alternativ kann der minimale Klassifizierungsfehler mit Hilfe der Kolmogoroff-Smirnoff-Statistik bestimmt werden, die die maximale Differenz der beiden Score-Verteilungsfunktionen misst.

#### Statistische Eigenschaften von GK und AUC

Zwischen den beiden Trennschärfemaßen Gini-Koeffizient (GK) und Fläche unter der ROC-Kurve (AUC) besteht der einfache lineare Zusammenhang

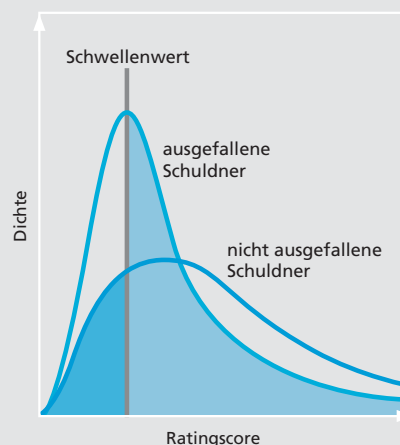
$$GK = 2 \cdot AUC - 1.$$

Im Weiteren werden hauptsächlich die statistischen Eigenschaften von AUC dargestellt, da diese anschaulicher zu interpretieren sind. Über die obige Relation bekommt man dann die äquivalenten Eigenschaften für GK.

Bildet man sämtliche Kombinationen von Paaren aus jeweils einem ausgefallenen Schuldner und einem nicht ausgefallenen Schuldner, so kann man die Mann-Whitney-Statistik definieren als

$$U(a, b, c) = \frac{1}{N_D \cdot N_{ND}} \sum_{(D, ND)} u_{D, ND},$$

#### Wahrscheinlichkeitsdichten der Ratingscores \*) und Klassifizierungsfehler \*\*)



\* Für die beiden Verteilungen der Ratingscores auf den Populationen der ausgefallenen und der nicht ausgefallenen Schuldner. — \*\* Bei gegebenem Schwellenwert ergibt sich der Klassifizierungsfehler als die Hälfte der Summe der Inhalte der beiden farbigen Flächen.

Deutsche Bundesbank

wobei hier  $N_D$  die Anzahl der ausgefallenen Schuldner und  $N_{ND}$  die Anzahl der solventen Schuldner bezeichnet. Der Ausdruck  $u_{D,ND}$  ist definiert als

$$u_{D,ND} = \begin{cases} a, & \text{falls } S_D < S_{ND} \\ b, & \text{falls } S_D = S_{ND} \\ c, & \text{falls } S_D > S_{ND} \end{cases}$$

Hier ist  $S_D$  der Ratingscore des ausgefallenen Kreditnehmers und  $S_{ND}$  der Ratingscore des solventen Kreditnehmers. Für das Trennschärfemaß AUC lässt sich die Beziehung

$$AUC = U(1, 0.5, 0)$$

beweisen. Berücksichtigt man die Definition von U, so erhält man

$$AUC = P(S_D < S_{ND}) + 0.5 P(S_D = S_{ND}).$$

Diese Gleichung lässt sich anschaulich interpretieren. Zieht man zufällig einen Schuldner aus der

Gesamtheit der ausgefallenen Kreditnehmer sowie einen Schuldner aus der Gesamtheit der solventen Kreditnehmer, so wird man vermuten, dass derjenige Kreditnehmer mit dem höheren Ratingscore der solvente Kreditnehmer ist. Haben beide Kreditnehmer denselben Ratingscore, so wird man das Los entscheiden lassen. Die Wahrscheinlichkeit, dass man durch diese Entscheidungsregel den solventen Kreditnehmer identifiziert, beträgt  $P(S_D < S_{ND}) + 0.5 P(S_D = S_{ND})$ . Diese Wahrscheinlichkeit ist exakt gleich der Fläche unter der ROC-Kurve.

*Konfidenzintervalle und Tests für AUC und GK*

Der Zusammenhang zwischen der Fläche unter der ROC-Kurve und der Mann-Whitney-Statistik lässt sich verwenden, um auf relativ einfache Weise Konfidenzintervalle für AUC zu berechnen. Er erlaubt ferner einen Test auf Unterschiedlichkeit zwischen den AUC-Werten zweier Ratingverfahren, die auf demselben Datensatz validiert werden. In beiden Fällen wird ausgenutzt, dass die Mann-Whitney-Statistik beziehungsweise die normierte Differenz zweier Mann-Whitney-Statistiken asymptotisch normalverteilt ist. Die zugehörigen Varianzen lassen sich aus den empirischen Daten leicht berechnen.<sup>2)</sup>

### Mathematische Beschreibung des Binomialtests

Nachfolgend wird die Funktionsweise des Binomialtests beschrieben. Der Binomialtest lässt sich auf eine einzelne Ratingklasse anwenden. Hierbei wird angenommen, dass sämtliche  $K$  Schuldner einer Ratingklasse dieselbe Ausfallwahrscheinlichkeit  $PD$  besitzen. Als Verteilung der Ausfälle innerhalb der Ratingklasse ergibt sich die Binomialverteilung, wenn man annimmt, dass die Ausfallereig-

nisse statistisch unabhängig sind. Jedem Schuldner wird eine Indikatorvariable  $I_i$  zugeordnet, wobei  $I_i$  den Wert Eins annimmt, wenn der Schuldner ausfällt, ansonsten ist er gleich Null. Die Anzahl der Ausfälle  $D_K$  ergibt sich zu

$$D_K = \sum_{i=1}^K I_i.$$

Die Null-Hypothese, dass die tatsächliche Ausfallwahrscheinlichkeit höchstens den Wert  $PD$  hat, kann nun zu einem Konfidenzniveau  $\alpha$  abgelehnt werden, wenn die tatsächliche Ausfallrate eine kritische Schwelle  $d_{K,\alpha}$  überschreitet, welche bestimmt wird aus

$$P[D_K \geq d_{K,\alpha}] \leq \alpha.$$

Unter Verwendung der Zähldichte der Binomialverteilung errechnet sich  $d_{K,\alpha}$  zu

$$d_{K,\alpha} = \min \left\{ d : \sum_{i=d}^K \binom{K}{i} PD^i (1 - PD)^{K-i} \leq \alpha \right\}.$$

Die Wahrscheinlichkeit, dass unter der Annahme der Binomialverteilung die kritische Schwelle  $d_{K,\alpha}$  überschritten wird, ist also höchstens  $\alpha$ . Zur Bestimmung von  $d_{K,\alpha}$  wird unterstellt, dass alle Ausfälle in einer Ratingklasse unabhängig sind. Dies ist in der Realität nicht der Fall, weil Ausfallraten im Konjunkturzyklus schwanken und somit Ausfälle miteinander korreliert sind. Die Folge ist, dass der Binomialtest  $d_{K,\alpha}$  im Allgemeinen unterschätzt. Der Binomialtest stellt somit einen konservativen Indikator für die Güte der Kalibrierung der Ausfallwahrscheinlichkeit einer Ratingklasse dar.

---

<sup>2</sup> Auf die genaue Wiedergabe der entsprechenden Formeln wird hier verzichtet. Die Ausdrücke sind nicht sehr übersichtlich. Dies stellt für die Anwender dieser Methoden jedoch keine Einschränkung dar, weil die Methoden in den gängigen Statistik-Softwarepaketen implementiert sind.