

VALIDIERUNG VON RATINGVERFAHREN

Ronny Parchert

Beim nachfolgenden Beitrag handelt es sich um einen modifizierten Vorab-Auszug aus dem Artikel „Validierung von Risikomanagementsystemen“, welcher im Frühjahr 2005 im Handbuch „Basel II“ (Hrsg. Becker/Gaulke/Wolf) im SchäfferPoeschelVerlag erscheint.

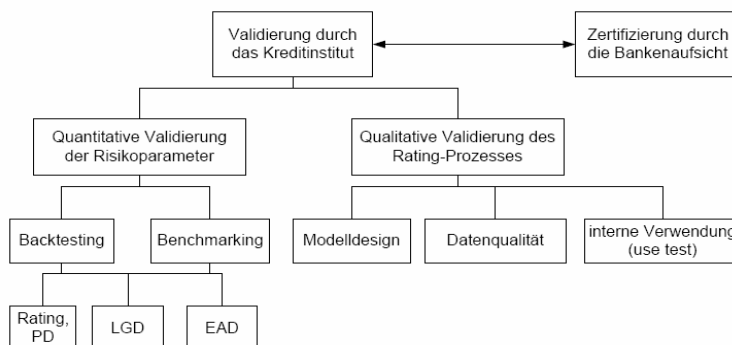
Inhalt

≡ Definition Validierung	1
≡ Qualitative Überprüfung	2
≡ Quantitative Validierungsansätze.....	3
≡ Fazit	7

≡ Definition Validierung

Mit Basel II kommt dem Überprüfungsprozess der internen Risikomanagementverfahren und insbesondere der im Rahmen eines IRB-Ansatzes verwendeten Ratingsysteme eine besondere Bedeutung zu. Im Rahmen der zweiten Säule der Basler Rahmenvereinbarung wird deutlich gefordert, dass die nationale Bankenaufsicht bei Umsetzung ihres Überprüfungsverfahrens verstärkt die internen Risikomanagementsysteme beurteilen und validieren soll. Dabei übernimmt die Aufsicht nicht die vollständige Validierung sondern beurteilt insbesondere die Ergebnisse der internen Validierung.

Unter dem Begriff der Validierung wird nachfolgend der gesamte Prozess der Überprüfung des Ratingssystems und der Verfahren zur Quantifizierung von Risikoparametern verstanden. Die nachfolgende Abbildung zeigt, in welche Komponenten der Validierungsprozess aufgeteilt werden kann.



☰ Qualitative Überprüfung

Die qualitative Validierung von Risikomanagementsystemen und -quantifizierungsverfahren umfasst alle nicht statistischen Verfahren zur Beurteilung der Prognosegüte und Funktionalität der Verfahren. Hierunter sind in der Regel alle prozessorientierten Überprüfungen zu subsumieren.

Die wichtigste Komponente ist die Beurteilung des Maßes der Integration der intern ermittelten Risikoparameter in die täglichen Geschäftsabläufe. Im Kontext von Basel II geht es vor allem um die Integration der Ergebnisse des bankinternen Ratingsystems. Wichtige Anhaltspunkte liefern den deutschen Kreditinstituten hierbei die Vorgaben der MaK. Die Intensität der einzelnen Prozesse sollte nach den Mindestanforderungen am Risikogehalt der Geschäfte, welcher sich in der Definition von Basel II aus der Kategorisierung durch das Ratingssystem ergibt, ausgerichtet sein. Dies betrifft neben der Ausgestaltung der einzelnen Prozesse auch die Definition der Übergangskriterien zwischen Normal-, Intensiv- und Problemkreditbetreuung. Gleichzeitig sollten die Ergebnisse Eingang in die Kompetenzordnung, das Pricing und die Limitierung finden.

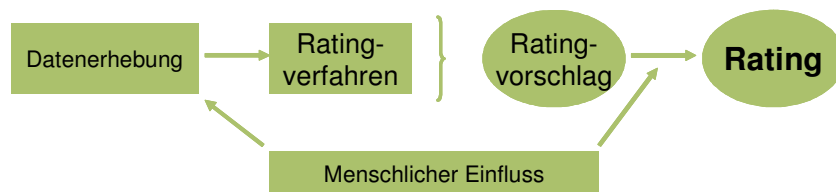
Ein weiterer Aspekt der qualitativen Überprüfung ist das bankinterne Verständnis für das Ratingsystem und die Sicherstellung der Verantwortung der Geschäftsleitung. Es ist sicherzustellen, dass alle im Gesamtprozess des Risikomanagement involvierten Mitarbeiter entsprechende Kenntnis über die verwendeten Verfahren besitzen und im Sinne der Risikostrategie die tägliche Anwendung der Verfahren vornehmen. Die Verantwortung der Gesamtgeschäftsleitung ist in zahlreichen aufsichtsrechtlichen Vorschriften oder internationalen Prinzipien für die Risikosteuerung verankert. Die Geschäftsleitung muss neben der Funktionalität des Systems vor allem die adäquate Beurteilung aller Risiken im Sinne der Risikostrategie mit den eingesetzten Verfahren sicherstellen. Dabei ist zu beachten, dass unter Berücksichtigung ökonomischer Zwänge und der Risikotragfähigkeit des Instituts durchaus der Einsatz pragmatischer Lösungen für Geschäfte, deren Risikogehalt nur einen geringen Anteil am Gesamtbankrisiko haben, angemessen sein kann. Beispielsweise müssen für die Kreditvergabe in bestimmten Nischenmärkten keine umfangreichen statistischen Ratingverfahren entwickelt werden, wenn die Risikoeinschätzung näherungsweise auch mit einfachen, eher subjektiven Kriterien erfolgen kann.

Neben der Objektivität und der Stabilität der eingesetzten Verfahren ist die Integration der einzelnen Risikomessverfahren für verschiedene Risikoarten in die Gesamtbanksteuerung ein wesentlicher Aspekt für die Qualität und das interne Vertrauen in die einzelnen Verfahren. So sind beispielsweise die Ergebnisse von Kreditverlustparameterschätzungen in die Kalkulation von risikoadjustierten Margen zu berücksichtigen. In den Baseler Ausführungen sind diese Anforderungen in Bezug auf die Ratingverfahren als „Use Test“ der bankinternen Ratingsysteme aufgeführt.

Die einzelnen Risiken müssen im Rahmen der Kapitalallokation sinnvoll mit Risikodeckungsmassen bedacht werden, wobei auch Effekte, welche erst auf Gesamtbankebene entstehen, wie beispielsweise Konzentrationen und Korrelationen, zu berücksichtigen sind. Gleichzeitig ist es eine zwingende Anforderung für die interne Steuerung und das externe Berichtswesen (Meldewesen, Eigenkapitalunterlegung) die gleichen Risikoparameter zu verwenden.

Die dargestellten prozessorientierten Aspekte sind durch komponentenorientierte Validierungen der einzelnen Verfahren zu ergänzen. Grundsätzlich hat sich in der Praxis nicht zuletzt mit den MaIR der Ansatz der prozessorientierten Überprüfung durchgesetzt. Ausgangspunkt ist hierbei die Unterteilung des Ratingprozesses in seine wesentlichen Komponenten. Die Prüfungshandlungen werden dann an den definierten Komponenten orientiert durchgeführt.

Aufgabe der komponentenorientierten Validierung ist es, die drei Elemente – Datenerhebung und –erfassung, quantitatives Verfahren und menschlicher Einfluss – des Ratingprozesses auf ihre Tauglichkeit und Funktionsfähigkeit zu analysieren.



Im Fokus sollte neben der quantitativen Verfahrensbeurteilung, welche zentrales Thema im nachfolgenden Teil dieses Beitrages ist, vor allem die Qualität der verwendeten Daten stehen. Die Qualität der Daten ist die Grundvoraussetzung für eine aussagefähige Risikoquantifizierung. Gleichzeitig wird ein besonderes Augenmerk auf den an vielen Stellen eines Gesamtrisikomanagementsystems möglichen menschlichen Einfluss gelegt. Am Beispiel eines Ratings ist zu untersuchen und zu validieren, ob die Gesamtaussagekraft und Trennschärfe des Verfahrens durch eventuelle Überschreibungen von Experten nachhaltig verschlechtert werden. Ein Instrument hierzu können die nachfolgend dargestellten quantitativen Verfahren sein.

≡ Quantitative Validierungsansätze

Die besondere Schwierigkeit bei der quantitativen Validierung von Kreditrisikomessverfahren liegt in Abgrenzung zu den Marktpreisrisiken hauptsächlich in der geringeren zeitlichen Häufigkeit des Risikoevents und dem längeren Betrachtungszeitraum von 1 Jahr.

Grundsätzlich kann eine auf mathematisch-statistischen Verfahren basierende Validierung nur dann erfolgen, wenn eine ausreichende Anzahl von Ausfällen vorliegt. Dies ist in vielen Kreditinstituten für einen Teil der Forderungsklassen regelmäßig nicht der Fall.

Die Validierung von Kreditrisikomessverfahren setzt regelmäßig auch im Kontext zu Basel II bei der Überprüfung der einzelnen Risikoparameter - beispielsweise Ausfallwahrscheinlichkeiten (PD); Verlustquoten (LGD) und Exposures (EaD) ¹ - an und konzentrieren sich auf die Validierung des Ratingsystems. Die maßgeblichen Kriterien bei der Beurteilung dieser sind dabei seine Trennschärfe, eine ausreichende Kalibrierung und die Stabilität der Schätzungen.

Die Kalibrierung des Systems beschreibt die Zuweisung von Ausfallwahrscheinlichkeiten zu den einzelnen Ratingklassen. Eine gute Kalibrierung liegt dann vor, wenn die geschätzten Ausfallwahrscheinlichkeiten nahe bei den tatsächlich realisierten PD's liegen. Die Kalibrierung ist hierbei in der Regel monoton, d.h. mit steigender Bonität zwischen den Klassen sinken die PD's.

In der Praxis ist regelmäßig zu beobachten, dass die beobachteten von den prognostizierten Ausfallwahrscheinlichkeiten abweichen. Die eingesetzten Verfahren müssen nunmehr die Frage beantworten, ob diese Abweichungen systematisch oder rein zufällig sind. In Anlehnung an Marktrisikomodelle ist der Binomialtest weit verbreitet. Bei der Beurteilung der Güte von Ausfallwahrscheinlichkeiten lässt sich unter Annahmen, welche die Realität stark vereinfachen, ein Test konstruieren, der auf der Binomialverteilung beruht. Beim Binomialtest wird jede Ratingklasse für sich betrachtet, wobei angenommen wird, dass jeder Schuldner dieser Ratingklasse die Ausfallwahrscheinlichkeit der Ratingklasse besitzt. Die tatsächlich nach einem Jahr beobachtete Ausfallrate der Ratingklasse wäre unter Annahme einer korrekt bestimmten PD binomialverteilt. Ergeben sich für eine definierte Irrtumswahrscheinlichkeit² jedoch starke Abweichungen der Ausfallrate von den geschätzten Ausfällen, müssten die geschätzten Ausfallwahrscheinlichkeiten abgelehnt werden.

Der größte Schwachpunkt des Binomialtests liegt in der Annahme, dass die Ausfälle der Kreditnehmer unabhängige Ereignisse darstellen. Die Annahmen sind jedoch in der Bankpraxis nicht erfüllt, da Ausfälle von Kreditnehmern keine unabhängigen Ereignisse darstellen. Rein theoretisch wäre eine Lösung dieses Problems denkbar, wenn die Ausfallkorrelationen bekannt wären. Die Bestimmung von Ausfallkorrelationen ist jedoch noch deutlich schwieriger als die Bestimmung von Ausfallwahrscheinlichkeiten an sich. Der Binomialtest sollte dabei als Indikator verstanden werden und niemals Alleinkriterium für die Ablehnung eines Ratingmodells sein.³

¹ Definitionen vgl. Basel II (2004) – in der internen Steuerung kommen weitere Parameter, wie beispielsweise Korrelationsannahmen hinzu.

² Darstellbar als 1-Konfidenzniveau.

³ Vgl. Bundesbank Monatsbericht September 2003.

Für die Beurteilung der Trennschärfe von Ratingverfahren existieren ebenfalls mehrere praxisübliche Ansätze. Hierbei geht es darum, zu beurteilen, ob das Ratingverfahren in der Lage ist, eine geeignete Rangfolge der Kreditnehmer bezüglich Ihrer Bonität herzustellen. Weit verbreitete statistische Maße für die Beurteilung sind „Cumulative Accuracy Profile“ (CAP) und die „Receiver Operating Characteristic“ (ROC). Die Verfahren werden nachfolgend an einem Beispiel erläutert.

In einem Ratingverfahren mit 5 Ratingklassen, wobei die Ratingklasse 5 die Kreditnehmer mit der besten und die Ratingklasse 1 diejenigen mit der schlechtesten Bonität beinhalten soll, werden folgende Verteilungen angenommen.

Score	Jahr 0	Jahr 1	
	Gesamt	Solvent	Insolvent
1	20	10	10
2	50	44	6
3	110	105	5
4	50	47	3
5	20	19	1
Gesamt	250	225	25

Beim oftmals als „Ginitest“ bezeichneten Verfahren liefert die CAP-Kurve eine grafische Veranschaulichung der Trennschärfe. Hierzu werden auf der Horizontalen die kumulativen prozentualen Anteile aller Schuldner mit steigender Ratingklasse (Bonität) abgetragen und auf der Vertikalen der kumulative prozentuale Anteil der insolventen Schuldner. Im Beispiel ergibt sich folgende Grafik (mittlere Kurve).

In der Abbildung sind zwei weitere Kurven abgetragen. Die „obere“ Kurve stellt den Verlauf eines optimalen Modells dar, in welchem die 25 insolventen Kreditnehmer bei idealer Trennschärfe mit 20 auf die Ratingklasse 1 und mit 5 Kreditnehmern auf die Klasse 2 verteilt wären. Die untere Kurve stellt den Verlauf in einem zufälligen Modell dar, hier steigt der kumulierte relative Anteil der Ausfälle linear an.

Zur Ermittlung der Trennschärfe des Systems wird nun das Accuracy Ratio (AR) berechnet. Diese Kennzahl ist Ausdruck für das Verhältnis der Flächen unter den Kurven des optimalen und des verwendeten Modells. Üblicherweise ist AR definiert als Quotient der Fläche zwischen dem eigenen Modell und dem zufälligen und der Fläche zwischen dem optimalen und dem zufälligen Modell⁴. AR liegt demnach immer zwischen den Werten 0 und 1 und das Verfahren ist umso trennschärfer je näher AR bei 1 liegt. Im Beispiel ergibt sich ein AR von 0,47. In der Praxis wird oftmals auch das Verhältnis der Gesamtflächen unter dem Ratingmodell und dem optimalen Modell, der so genannte Gini-

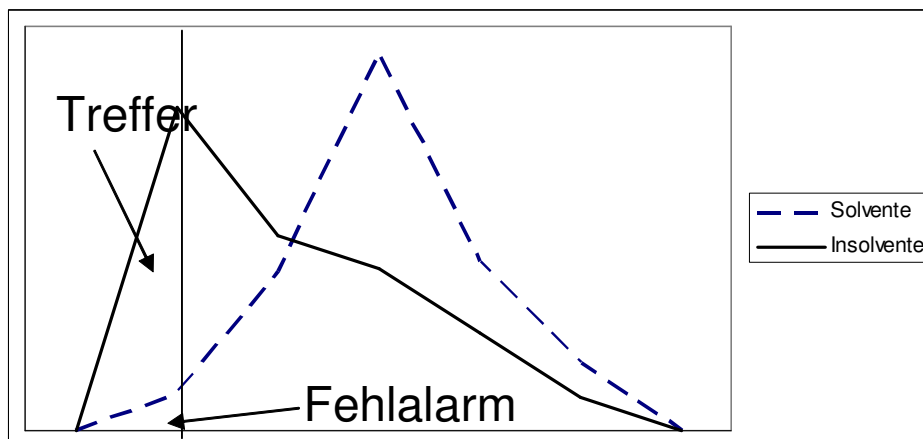
⁴ AR wird in der Praxis teilweise auch als Quotient der Gesamtflächen unter der Kurve des eigenen und der des optimalen Modells definiert.

Koeffizient berechnet. Dieser sollte üblicher Weise bei Werten von über 70% liegen. Im gewählten Beispiel ergibt sich ein Gini von 75,5%.

Ein anderer eng mit der CAP-Kurve verwandter Ansatz ist die ROC-Kurve. In diesem Fall wird für die insolventen Kreditnehmer und für die solventen Kreditnehmer separat die empirische Score-Verteilung ermittelt.

Score	Häufig Solvent	Häufig Insolv
1	0,044	0,400
2	0,196	0,240
3	0,467	0,200
4	0,209	0,120
5	0,084	0,040

Nunmehr ist für eine bestimmte Scoregrenze beurteilbar, wie viele Ausfalltreffer das System erzeugt hat. Dies sind alle ausgefallenen Kreditnehmer mit einem kleineren (schlechteren) Score. Gleichzeitig kann bestimmt werden, wie oft ein Fehlarmed ausgelöst wurde. Dies ist genau dann der Fall, wenn ein solventer Kreditnehmer einen kleineren Scorewert zugewiesen bekommen hat.



Die Grafik zeigt am gewählten Beispiel die Flächen für die Treffer und Fehlarmed bei einer Scoregrenze von 2. Für eine von der Scoregrenze unabhängige Analyse der Trennschärfe des Systems wird diese Erhebung der Treffer- und Fehlalarmquote für jeden Scorewert durchgeführt. Die ermittelten Punktepaare je Scorewert werden zur graphischen Veranschaulichung derart dargestellt, dass auf der horizontalen Achse die Fehlalarmquote und auf der vertikalen Achse die Trefferquote aufgetragen wird. Dies geschieht ebenfalls für ein optimales und ein zufälliges Ratingmodell. Für ein perfektes Ratingmodell bestünde die Kurve nur aus der Verbindung des Achsenursprungs mit den Punkten 0;1 und 1;1.

In Anlehnung an die CAP-Kurve wird nun eine aggregierte Kennzahl ermittelt, welche zur Beurteilung der Trennschärfe des Systems dient. Diese „Area under the curve“ (AUC) beschreibt die Fläche unter der Kurve für das Ratingmodell. Der Wert für AUC liegt dabei ebenfalls zwischen 0 und 1. Im Beispiel ergibt sich ein Wert von 72,7 %. Für ein Rating-

modell mit stetiger Ratingskala lässt sich zwischen AR und AUC der Zusammenhang $AR=2*AUC-1$ zeigen.⁵

Neben den Ausfallwahrscheinlichkeiten sind die Verlustparameter LGD und EaD, hier vor allem die Credit Conversion Factors (CCF) zu validieren. Die Validierung der LGD ist deutlich schwerer als bei PD's. Die Hauptursache hierin liegt vor allem in der Komplexität der Bestimmung von Verlustquoten, welche durch den langen Zeitraum von Verwertungen, zusätzlichen Kosten und Teilzahlungen bestimmt wird. Aus diesen Gründen kann im klassischen Kreditgeschäft nicht direkt zum Ausfallzeitpunkt eine Verlustquote bestimmt werden. Diese Probleme führen zu einer fehlenden Standardisierung der Erhebungen und vorrangig qualitativen Validierungsansätzen.

≡ Fazit

Die Darstellungen haben gezeigt, dass die Validierung von Ratingverfahren eine ebenso komplexe Aufgabe ist, wie deren Entwicklung und Aufbau. Es wurde deutlich, dass es nicht das „eine richtige“ System zur Beurteilung gibt, vielmehr muss die Bewertung immer als Kombination der Ergebnisse vieler Ansätze der qualitativen und quantitativen Validierung sein.

Vor allem die beschriebenen qualitativen Aspekte zeigen die besondere Anforderung an die Aufsicht im Rahmen des Supervisory Review Process im Rahmen der Säule 2 von Basel II. Die Aufsicht muss sich in der Prüfung hier auf die Individualität der einzelnen Kreditinstitute einlassen.

Im Zusammenhang mit der Anwendung der IRB-Ansätze wird zu beobachten sein, wie hoch die Anforderungen seitens der Aufsicht bezüglich der Validierung an Institute sein werden, welche auf gepoolte Daten zurückgreifen werden. Hier stellt sich die Frage, ob bestimmte, vor allem quantitative Validierungen an den Poolverwalter ausgelagert werden können und dieser diese Aufgaben zentral übernimmt und ob es ausreichend ist, wenn die einzelnen Institute qualitativ die Anwendung der Verfahren und die Aussagekraft für ihr Portfolio (als Teil des Pools) unterlegen.

Gleichzeitig bleibt zu beachten, dass nicht nur für die Risikoklassifizierungsverfahren und die PD-Schätzungen geeignete Validierungsverfahren zu entwickeln sind, sondern auch die internen Schätzungen der weiteren Verlustparameter (EaD, LGD) im Rahmen des fortgeschrittenen IRB-Ansatzes entsprechend zu überprüfen sind.

⁵ Vgl. hierzu Bundesbank (2003).
© 1 PLUS i GmbH