

## **Evaluierung quantitativer Ratingverfahren**

Thorsten Poddig  
Universität Bremen  
Hochschulring 4  
28359 Bremen  
poddig@uni-bremen.de

Andreas Oelerich  
Universität Bremen  
Hochschulring 4  
28359 Bremen  
andreas.oelerich@uni-bremen.de

### **Zusammenfassung**

Der Beitrag stellt ein Simulationsystem vor, welches die Prognosefähigkeiten eines quantitativen Ratingverfahrens zu evaluieren erlaubt. Das Simulationsystem wird dabei auf die logistische Regression angewendet und verschiedene Modellierungstechniken vergleichend diskutiert. Die üblicherweise im Zusammenhang mit Insolvenzprognosen diskutierten Trefferquoten der solventen bzw. insolventen Unternehmen sind kein Indiz für ein leistungsstarkes Ratingverfahren. Die detaillierte Ratingprognose ist teilweise deutlich schlechter als die einfache dichotome Prognosequalität. Das Verfahren mit der höchsten dichotomen Trefferquote muss nicht die höchste polytome Trefferquote haben.

**Keywords:** quantitative Ratingverfahren, Logistische Regression, Bootstrap, Kreditrisiko, Evaluierung, Prognosequalität

## **1 Einleitung**

Ratingverfahren bilden die Grundlage der zukünftigen Kreditrisikomanagement-Prozesse in Kreditinstituten. Hierfür gibt es zwei Gründe. Zum einen schreibt die neue Eigenkapitalvereinbarung deren Verwendung vor und zum anderen sind Banken aus Wettbewerbsgründen angehalten ihr Risikomanagement nachhaltig zu optimieren. Das Eigenkapital wird dabei mehr an den ökonomischen Risiken (Bonitätsrisiken) gekoppelt. Kreditratings stellen Beurteilungen dieses Risikos dar. Viele Kreditinstitute leiten, auf Basis eines quantitativen Verfahrens, einen Ratingvorschlag ab, um zu einem abschließenden Urteil zu gelangen. Die Verwendung der Ratingverfahren ist mit der jeweiligen Prognosequalität des unterliegenden, quantitativen Verfahrens eng verknüpft. Die Trefferquote kann (bei quantitativen Ratingverfahren) als ein bedeutendes Maß für die Qualität eines Modells bei dichotomen Zielva-

riablen gelten und wird auch in empirischen Studien als Beurteilungsmaßstab herangezogen.<sup>39</sup>

Für praktische Belange, insbesondere im Hinblick auf Kreditrisikomodelle, ist allerdings nicht die Trefferquote in Bezug auf die Solvenz bzw. Insolvenz als umfassender Beurteilungsmaßstab entscheidend. Vielmehr ist die Prognose der tatsächlichen Ratingklasse von Bedeutung, da die Ratingklasse als Grundlage der wichtigen Zuordnung der Ausfallwahrscheinlichkeit eines Unternehmens dient und für die Berechnung des Bonitätsgewichts entscheidend ist.

Das Rating eines Unternehmens kann beispielsweise auf Basis des logistischen Regressionsmodells, mit einer geeigneten Einteilung des Einheitsintervalls, erfolgen. Man muss also nicht nur die Trennkraft zwischen solventen und insolventen Unternehmen als Qualitätsmaßstab heranziehen, sondern auch die Fähigkeit, die konkrete Ratingklasse zu treffen.

Diese Prognosekraft ist in empirischen Studien nicht evaluierbar, da die a priori Ratingklasse unbekannt ist und somit nicht modelliert werden kann. Man kennt nur die Solvenz bzw. Insolvenz eines Unternehmens. Solche detaillierten Qualitätsbewertungen der Ratingverfahren können nur durch Simulationsstudien bestimmt werden, in denen die Unternehmensdatenbank und der Insolvenzgenerierungsprozess bekannt sind. Unter diesen Bedingungen kann die polytome Prognosequalität bezüglich aller Ratingklassen bestimmt werden. Es ist nicht garantiert, dass Verfahren mit hohen Trefferquoten bezüglich der Solvenz/Insolvenz-Klassifizierung auch bessere Trefferquoten bezüglich der detaillierten Ratingskala haben. Das beste Verfahren zur dichotomen Prognose muss aber nicht zwangsläufig zugleich das beste Verfahren zur polytomen Prognose sein.

Der Beitrag stellt die Umsetzung eines solchen Simulationssystems mit SAS und die erzielten Ergebnisse vor. Der Beitrag strukturiert sich wie folgt. Zunächst wird das Rahmenwerk nach Basel II kurz vorgestellt und dabei auf Veränderungen bezüglich der Behandlung des Kreditrisikos fokussiert. Im Anschluss wird ein Simulationssystem erörtert, welches die Analyse der Ratingprognosen ermöglicht. Die Ergebnisse werden schließlich im vierten Abschnitt dargestellt. Der Beitrag schließt mit einer Zusammenfassung.

---

<sup>39</sup> Vgl. beispielsweise die Untersuchungen von Anders und Szczesny (1998) und Leker und Schewe (1998).

## 2 Basel II

Im Zusammenhang mit der Diskussion der neuen Eigenkapitalanforderungen für Banken ist das Thema Rating verstärkt in den Mittelpunkt des Interesses gerückt. Die aktuelle Eigenkapitalvereinbarung aus dem Jahre 1988 stellt auf das gesamte Eigenkapital einer Bank ab, das für die Begrenzung des Insolvenzrisikos und den daraus resultierenden Folgen für die Einleger eines Kreditinstitutes entscheidend ist.<sup>40</sup> In Erweiterung dieses bestehenden Systems soll durch die neue Regelung die Sicherheit und Solidität des globalen Finanzmarktes dadurch vergrößert werden, dass interne Kontrollsysteme, die Überprüfung durch die Aufsicht und die Marktdisziplin eine höhere Bedeutung erhalten. Hierbei sollen einerseits die Kapitalanforderungen an Banken stärker als bisher vom ökonomischen Risiko einzelner Kreditengagements abhängig sein.<sup>41</sup> Andererseits soll die Implementierung von neuen/modernen Entwicklungen an den Finanzmärkten sowie im Risikomanagement der Institute gefördert werden, welches durch die Zulassung unterschiedlich komplexer Methoden erreicht wird.<sup>42</sup> Entscheidend für ein Kreditinstitut ist es, die eingegangenen Risiken zu steuern und dauerhaft tragen zu können. Die internen Risikosteuerungssysteme haben daher einen hohen Stellenwert. Der Ausschuss will darauf hinwirken, dass sie weiter verbessert und ständig überprüft werden.<sup>43</sup> Darüber hinaus ist eine Ergänzung dieser neuen Eigenkapitalanforderungen und des aufsichtsrechtlichen Überprüfungsprozesses durch die Marktdisziplin vorgesehen. Der neue Baseler Eigenkapitalakkord besteht daher aus drei sich gegenseitig ergänzenden Säulen, um die Stabilität des nationalen und des internationalen Bankensystems besser abzusichern.

Die bestehende Vereinbarung sieht vor, dass das Eigenkapital mindestens acht Prozent der Summe der risikogewichteten Aktiva entsprechen muss.<sup>44</sup> Dies wird auch durch die neue Eigenkapitalvereinbarung nicht grundsätzlich modifiziert, dennoch werden zukünftig zwei wesentliche Änderungen in der Berechnung der Eigenkapitalquote vorgesehen. Die Veränderungen betreffen

---

<sup>40</sup> Vgl. Becker und Peppmeier (2000), S. 40f und Eilenberger (1996), S. 165ff.

<sup>41</sup> Vgl. Heinke (2001), S. 178, Guennemann und Szczesny (2001), S. 18 und Heinke (2002), S. 3.

<sup>42</sup> Vgl. Berndt (2001), S. 170.

<sup>43</sup> Vgl. hierzu die vier zentralen Grundsätze der aufsichtsrechtlichen Überprüfung in BIS (2003), S. 162ff.

<sup>44</sup> Vgl. Becker und Peppmeier (2000), S. 48.

die Definition der gewichteten Risikoaktiva. Das bisher als implizit durch Markt- und Kreditrisiken abgedeckte operationelle Risiko soll zukünftig explizit als dritte Risikokomponente berücksichtigt werden.<sup>45</sup>

Bei der Messung des Kreditrisikos wird ein Standardansatz, ein auf internen Ratings basierender Ansatz (IRB-Basisansatz) und ein fortgeschrittener IRB-Ansatz zugelassen. Kern der Änderung gemäß Basel II betrifft die Neuerung bei der Bestimmung des Risikogewichts, das bei der Berechnungsvorschrift des Eigenkapitalbedarfs eines Kreditengagements von Bedeutung ist.<sup>46</sup>

$$\text{Eigenkapitalanforderung für Kreditrisiko} = 0.08 * \text{Bemessungsgrundlage} * \text{Risikogewicht}$$

Das Risikogewicht bestimmte sich früher allein aus einer gesetzlich vorgeschriebenen Einteilung in sechs Bonitätsklassen. Unternehmen (Nicht-Banken) waren einheitlich mit dem Risikogewicht von 100% versehen, wodurch das tatsächliche Risiko nicht ausreichend differenziert wurde. Diese Risikogewichte werden durch die neue Eigenkapitalvereinbarung an die tatsächlichen Risiken eines Kreditnehmers, welche durch das erteilte Rating des Kreditnehmers symbolisiert werden, angeglichen. Durch die geplante Neuregelung ist das Risikogewicht per Definition vom individuellen Rating des Kreditnehmers abhängig, da zu jeder Ratingklasse ein Gewichtungsfaktor definiert wird (vgl. Tabelle 1).

**Tabelle 1:** Neue Risikogewichte einzelner Ratingklassen beim Standardansatz.<sup>47</sup>

Rating	AAA bis AA-	A+ bis A-	BBB+ bis BB-	unter B-	nicht beurteilt
Gewichtungsfaktor	20%	50%	100%	150%	100%

Die Tabelle 1 verdeutlicht auch die Bedeutung der Ratings ganz deutlich: Ein Kreditengagement mit einem Unternehmen mit Rating AA- erfordert für Kreditinstitute hiernach (c.p. aller anderen Einflussparameter) demnach nur ein Fünftel der Eigenkapitalunterlegung gegenüber einem Unternehmen mit

<sup>45</sup> Vgl. BIS (2003), S. 6ff, S. 38ff und S. 120ff sowie Poddig und Kunze (2003), S. 699.

<sup>46</sup> Der Solvabilitätskoeffizient beträgt dabei 8%. Vgl. Becker und Peppmeier (2000), S. 49 und Grunert u.a. (2002), S. 1051.

<sup>47</sup> Vgl. Eigermann (2002), S.43 und Wilkens u.a. (2001), S. 188.

Rating BBB+.<sup>48</sup> Banken können bzw. müssen dieses zukünftig bei der Preisgestaltung mit berücksichtigen.

Neben dem Standardansatz steht es Banken offen (anstelle der externen Ratings), eigene Ratings zu verwenden. „Danach können Banken bei Erfüllung bestimmter Mindestanforderungen die aufsichtrechtliche Risikogewichtung ihrer Kreditnehmer auf Grundlage von mittleren Ausfallwahrscheinlichkeiten (Probability of Default; PD) der einzelnen Risikoklasse selbst bestimmen.“<sup>49</sup> Hier schätzen Banken zu jeder Risikoklasse intern die Parameter, die zur Berechnung des Risikogewichts notwendig sind. Dieses sind beim Basisansatz die PD und beim fortgeschrittenen Ansatz neben der PD auch die LGD (loss given default), EAD (exposure at default) sowie Vertragsparameter wie Restlaufzeit.<sup>50</sup> Der Kern der Änderungen sieht zusammenfassend eine individuelle Risikobeurteilung und damit auch die Bewertung eines einzelnen Kreditengagements vor. Die Bonitätseinstufungen und damit die Ratingverfahren haben dabei eine hohe Bedeutung im Kreditrisikomanagement. Wohl auch aus diesem Grund findet seit der Publikation des neuen Konsultationspapiers eine rege Forschung auf dem Gebiet der quantitativen Ratingverfahren statt. Neue, vielversprechende Methoden werden traditionellen Ratingverfahren gegenübergestellt und anhand der Trefferquoten verglichen. Zu diesen modernen Verfahren zählen u.a. neuronale Netze, genetische Algorithmen, Fuzzy Logic oder auch Support Vector Machines. Das folgende Simulationssystem ermöglicht die Evaluierung verschiedener quantitativer Ratingverfahren. Es eignet sich vor allem dazu, die generelle Prognosefähigkeit derartiger Verfahren, im Hinblick auf eine detaillierte Ratingskala zu bestimmen.

### 3 Simulationssystem

Empirische Studien basieren auf realen Daten, die i.A. in Form einer Tabelle zur Verfügung stehen. Diese Tabelle enthält alle Informationen, die zu den Unternehmen bekannt sind und zur Insolvenzerklärung möglicherweise einen Beitrag leisten. Die Anzahl der Unternehmen wird im Folgenden durch  $N$  symbolisiert. Unternehmensindividuelle (Daten-)Vektoren (im Folgenden als  $\mathbf{x}_k$  mit  $k=1, \dots, N$  bezeichnet) bilden diese Tabelle. Die Tabelle

---

<sup>48</sup> Vgl. beispielsweise Tholen (2000), S. 547f.

<sup>49</sup> Grunert u.a. (2002), S. 1049.

<sup>50</sup> Vgl. Wilkens u.a. (2001), S. 189 und Sironi und Zazzara (2003), S. 108.

$\mathbf{X}=(\mathbf{x}_1, \dots, \mathbf{x}_N)^T$  besteht dann aus  $N$  Zeilen und  $q$  Spalten, wenn  $q$  die Anzahl der Faktoren bezeichnet, die zu jedem Unternehmen erfasst werden. Diese Faktoren sind Jahresabschlussdaten, -kennzahlen sowie andere quantitative und qualitative, potenzielle insolvenzklärende Informationen. Wird exemplarisch nur die Branche, die Rechtsform, die Eigenkapitalquote und die Mitarbeiteranzahl erfasst, haben diese Vektoren beispielsweise die Form  $\mathbf{x}^*=(\text{Dienstleistung, GmbH, 0.3, 50})$ .

In vielen empirischen Studien zur Insolvenzforschung werden die quantitativen Einflussfaktoren standardisiert, d.h. der Mittelwert wird auf Null und die empirische Varianz auf Eins projiziert.<sup>51</sup> Der Vorteil dieses Vorgehens ist die einfachere Interpretation der Modellparameter, da alle (quantitativen) Einflüsse auf mehr oder weniger dergleichen Skala beobachtet werden. Dieses in der Praxis übliche Standardisieren ermöglicht eine sehr einfache Generierung quantitativer Unternehmensdaten: Man kann standard-normalverteilte Zufallszahlen verwenden.

Die Generierung qualitativer Informationen stellt sich als problemlos dar; es sind nur Gruppierungsvariablen. Einzig die Anzahl der Faktoren und die jeweilige Anzahl der Faktorstufen sind vorzugeben. Aus diesen Vorgaben resultieren zu jedem Faktor dessen Faktorstufen und damit auch alle Faktorstufenkombinationen. Es ist dann noch die Anzahl der Unternehmen zu jeder dieser Kombination vorzugeben. Diese Art der qualitativen Faktorsimulation ist weit verbreitet und soll auch in dieser Arbeit verwendet werden.<sup>52</sup>

Insgesamt kann durch das beschriebene Vorgehen eine beliebige Unternehmensdatenbank (die Matrix  $\mathbf{X}$ ) konstruiert werden. Es sind als Variablen vorzugeben: Die Anzahl der qualitativen Faktoren, die Anzahl der Faktorstufen pro qualitativen Faktor, die Anzahl der Unternehmen zu jeder resultierenden Faktorstufenkombination und die Anzahl der metrischen (quantitativen) Faktoren. Die Datenbank ist im Rahmen dieser Untersuchung mit SAS/BASE aufgebaut.

Als zweiter, wichtiger Schritt ist die Vorgabe des Insolvenzkennzeichens notwendig. Aus Sicht der Statistik ist dies die abhängige Variable. Die Konstruktion dieses Kennzeichens erfordert eine funktionale Verknüpfung der Unternehmensdaten mit der Insolvenzwahrscheinlichkeit, um hieraus eine dichotome Kodierung bestimmen zu können. Diese dichotome Kodierung ist

---

<sup>51</sup> Vgl. beispielsweise das Vorgehen von Moodys (2001), S. 8 und Petersmeier (2003), S. 229f.

<sup>52</sup> Vgl. u.a. Langer (1998), Siemer (2002) und Oelerich und Poddig (2004).

bernoulliverteilt:  $Y_k \sim B(\pi_k)$ . Aus diesem Grund bietet sich die logistische Regressionsgleichung zur Datengenerierung an, da dieses Modell das natürliche Modellierungswerkzeug für dichotome (zu modellierende) Variablen ist.

Das logistische Regressionsmodell, mit der Modellgleichung<sup>53</sup>

$$\ln\left(\frac{\pi_k}{1-\pi_k}\right) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

stellt die notwendige Verknüpfung zwischen den Unternehmensdaten und der Wahrscheinlichkeit für die Insolvenz her. Die Insolvenzwahrscheinlichkeit  $\pi_k$  ist der (auf  $\mathbf{x}_k$  bedingte) Erwartungswert der zu Grunde liegenden Bernoulliverteilung, der durch die sogenannte Linkfunktion ( $f(z)=\ln(z/(1-z))$ ) transformiert wird. Dieser transformierte Erwartungswert wird durch den linearen Prediktor erklärt.

Neben der Unternehmensdatenbank wird daher zusätzlich der Modellparametervektor des logistischen Regressionsmodells benötigt, um die Wahrscheinlichkeit zu spezifizieren. Auf Basis der resultierenden Insolvenzwahrscheinlichkeiten erfolgen zum Abschluss noch zwei Aufgaben: Zum einen wird mit diesen Insolvenzwahrscheinlichkeiten das genannte Bernoullixperiment durchgeführt, um die dichotome Kodierung zu erhalten und zum anderen wird die a priori Ratingklasse vorgegeben.

Die konkret künstlich generierte Unternehmensdatenbank beinhaltet zwei qualitative Faktoren, die jeweils fünf Kategorien haben. Dies könnten beispielsweise Rechtsformen 'Einzelkaufmann', 'OHG', 'KG', 'GmbH', 'AG' und Branchenkodierungen, wie etwa 'Dienstleistung', 'Baugewerbe', 'Handel', 'Landwirtschaft' und 'Industrie' sein. Neben diesen qualitativen Faktoren sind auch quantitative Einflussgrößen erklärungsrelevant. Hier werden zehn Faktoren integriert, die exemplarisch Jahresabschlussdaten oder hieraus konstruierte Kennzahlen sein können. Die gewählte Anzahl der Einflussgrößen orientiert sich an Ergebnissen zahlreicher empirischer Studien, die zum Teil auf mehr als 20 potenziellen Einflussgrößen basieren und in denen unterschiedlich viele Einflussfaktoren als signifikant identifiziert werden.<sup>54</sup> Die

---

<sup>53</sup> Vgl. zum logistischen Regressionsmodell insbesondere Hosmer und Lemeshow (2000).

<sup>54</sup> Vgl. die Untersuchung von Hayden (2002); die Autoren nutzen für ihr Rating österreichischer Unternehmen 21 Kennzahlen, von denen insgesamt neun als signifikant selektiert werden (vgl. S. 46ff.). Für ihr Rating deutscher Unternehmen

Wahl der Modellkomplexität ist in Grenzen willkürlich, reflektiert aber bisherige Erkenntnisse und antizipiert die Integration qualitativer Einflussfaktoren.<sup>55</sup>

Die Tabelle 2 zeigt die Parameter der Faktorstufen der beiden qualitativen Einflussfaktoren, die mit A und B bezeichnet sind. Dabei sind die fünf Modellparameter pro Faktorstufe in beiden Hauptfaktoren mit den Konstanten  $\beta_i=3-i$  vorgegeben. Diese Effekte sind mit den in Tabelle 2 angegebenen Faktoren multipliziert worden. Die Parameter sind sowohl mit der Referenzkodierung als auch mit der Effektkodierung konform, da jeweils eine Faktorstufe mit dem Einfluss Null berücksichtigt wird bzw. die jeweiligen Parameter sich zu Null summieren. Eine Wechselwirkung der beiden qualitativen Faktoren ist nicht vorhanden, da alle Modellparameter der Faktorstufenkombinationen ( $\eta_{ij}$ ) als Null vorgegeben sind. Des Weiteren sind die quantitativen Einflussfaktoren in der Tabelle 2 aufgeführt. Alle Einflussfaktoren haben einen unterschiedlich starken Einfluss auf die Wahrscheinlichkeit, die Wechselwirkung ist effektlos.

**Tabelle 2:** Modellparameter im Simulationssystem.  
(Die hier angegebenen Modellparameter bilden zusammen den verwendeten Modellparametervektor.)

Parameter	A	B	AB	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Wert	-1	-0.5	0	-1.5	-1	-0.75	0.25	0.5	0.4	0.1	0.4	0.9	-1.25

Durch die künstliche Unternehmensdatenbank und den Modellparametervektor ist die Insolvenzwahrscheinlichkeit eines Unternehmens spezifiziert und kann berechnet werden. Auf Basis dieser Wahrscheinlichkeit wird ein Bernoulliexperiment zu jedem Unternehmen durchgeführt, d.h. die Zufallsvariable  $Y_k$  wird durch eine Realisierung auf Basis der Vorgaben ersetzt.

Mit dieser künstlich generierten (zu modellierenden) Zielvariable und den vorgegebenen Designvektoren  $\mathbf{x}_k$  ist ein Datensatz konstruiert, zu dem eine

---

sind es 21/12 Kennzahlen (vgl. S. 67ff). Bei Leker und Schewe (1998) ist das Verhältnis 31/5 (vgl. S. 884). Bei Anders und Szczesny (1998) enthält die resultierende Modellgleichung nur zwei quantitative Einflüsse (vgl. S. 905).

<sup>55</sup> Die Bedeutung qualitativer Faktoren stellen Grunert u.a. (2002) heraus.



klassische „Fallstudie“ durchgeführt werden kann. Es wird auf Basis eines künstlichen Datensatzes der Parametervektor ( $\beta$ ) geschätzt. Aus diesem Parameter resultiert, in Verbindung mit dem individuellen Designvektor eines Unternehmens, die prognostizierte Insolvenzwahrscheinlichkeit und damit die Ratingprognose. Es liegen demnach zu jedem Unternehmen die a priori bekannte und die prognostizierte Ratingklasse vor, sodass eine Evaluierung des Verfahrens hinsichtlich der Deskription einer Ratingklasse erfolgen kann.

Der Datengenerierungsprozess erfolgt dabei nach dem folgenden Muster:

1. Vorgabe des Stichprobenumfanges.
2. Vorgabe des Modellparametervektors, der Designvektoren und Ermittlung der resultierenden Insolvenzwahrscheinlichkeit  $\pi_k$  und der Ratingklasse zu jedem Wirtschaftssubjekt.
3. Durchführung der Bernoulliexperimente zur Generierung der künstlichen Zielvariable zu jedem Wirtschaftssubjekt.
4. Berechnung des Schätzers des Modellparametervektors, der prognostizierten Insolvenzwahrscheinlichkeit sowie der resultierenden Ratingklasse für alle Wirtschaftssubjekte.

Wegen der Vielfalt der potenziell möglichen Klassendefinitionen können nicht alle evaluiert werden. Es ist daher eine Vorauswahl zu treffen. In dieser Untersuchung stellt die Ausgangsbasis ein auf neun Klassen basierendes Ratingsystem dar (vgl. Tab. 3). Die nebenstehende Tabelle zeigt die Zuordnung der Wahrscheinlichkeiten zu einer Klasse.<sup>56</sup>

**Tabelle 3:** Die verwendete Einteilung des Wahrscheinlichkeitsintervalls auf neun Klassen.

Klassen	Untergrenze	Obergrenze
1	0,0000	0,0002
2	0,0002	0,0005
3	0,0005	0,0011

---

<sup>56</sup> Die Masterskala (wie in Rolfes (2000), S. 12, dargestellt) entspricht in diesem Ratingsystem den Klassen eins bis acht; die neunte Klasse ist notwendig, um jede mögliche Wahrscheinlichkeit eindeutig einer Klasse zuordnen zu können.

4	0,0011	0,0040
5	0,0040	0,0133
6	0,0133	0,0770
7	0,0770	0,1699
8	0,1699	0,2000
9	0,2000	1,0000

## 4 Simulationsergebnisse

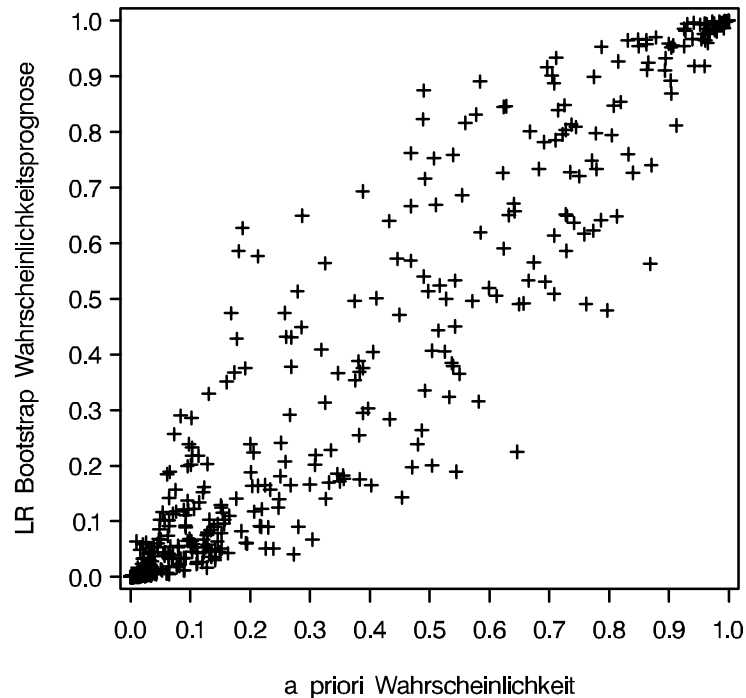
Die künstliche Datenbank wird im Rahmen dieser Untersuchung mit dem klassischen logistischen Regressionsmodell (mit und ohne Selektionsalgorithmus) und dem Bootstrap-Verfahren als klassisches resampling Verfahren ausgewertet.<sup>57</sup>

Beim verwendeten Bootstrap-Verfahren wird durch Ziehen mit Zurücklegen aus der bestehenden in-sample Datenbank eine neue, gleichgroße Datenbank bestimmt. Es kann dabei vorkommen, dass ein Unternehmen gar nicht oder mehrmals selektiert wird. Auf Basis des resultierenden Datensatzes wird zu allen Unternehmen die Wahrscheinlichkeitsprognose bestimmt. Dieses Vorgehen wird 200 Mal wiederholt. Der Mittelwert der empirischen Verteilung der Wahrscheinlichkeitsprognose dient als Schätzer der Ausfallwahrscheinlichkeit. Alle resultierenden Modellgleichungen werden auf die out-of sample Datenbank angewendet und der resultierende Mittelwert der empirischen Wahrscheinlichkeitsverteilung (für jedes Unternehmen) wird als prognostizierte Wahrscheinlichkeit verwendet.

Dabei ist vor allem die Güte der Wahrscheinlichkeitsprognose in der out-of sample Stichprobe von Interesse. Die Abbildung 1 zeigt exemplarisch die Prognosen auf Basis des Bootstrap-resampling Ansatzes. Die Abweichungen von den vorgegebenen Wahrscheinlichkeiten und den prognostizierten ist deutlich zu erkennen, da perfekte Prognosen durch die Identität gegeben sind. Die entsprechenden Abbildungen der klassischen logistischen Regression (ohne Bootstrap) sind dieser sehr ähnlich.

---

<sup>57</sup> Eine detaillierte Darstellung des logistischen Regressionsmodells findet man beispielsweise in Hosmer und Lemeshow (2000) und zum Bootstrap-Verfahren vgl. insbesondere Efron und Tibshirani (1993).



**Abbildung 1:** Qualität der Wahrscheinlichkeitsprognose auf Basis des Bootstrap Verfahrens im logistischen Regressionsmodell. Die Ordinate zeigt die Bootstrap-Wahrscheinlichkeitsprognose und die Abszisse die a-priori Wahrscheinlichkeit der künstlichen Datenbank in der out-of sample Stichprobe bei 500 Unternehmen.

Der Vergleich der Trefferquoten liefert einen praxisrelevanten Vergleich der Modellierungstechniken. Insbesondere der out-of sample Vergleich ist für die Praxis bedeutend. Die dichotomen Trefferquoten des klassischen logistischen Regressionsmodells und des Bootstrap Verfahrens sind in den folgenden Tabellen dargestellt.

Das Bootstrapping liefert in-sample erstaunlich gute Prognosefähigkeiten, die Trefferquoten sind in-sample über 90%. Die Verwendung des Waldmodells, also die Modellgenerierung auf Basis des forward-Selektionsalgorithmus auf Basis des Waldtests, liefert die schlechteste Prognosegüte in der in-

sample Stichprobe. Das Vollmodell, die Integration alle Einflussvariablen, ist gegenüber dem Waldmodell in-sample zu bevorzugen. Allerdings lässt sich diese qualitative Abstufung der gewählten Modellierungstechniken out-of sample nicht bestätigen. Das Waldmodell hat in dieser Stichprobe die größte Prognosegüte bei den beiden kleinen Stichprobenumfängen. Das Bootstrap-Verfahren schneidet hier am schlechtesten ab, mit Ausnahme beim größten Stichprobenumfang.

**Tabelle 4:** Gesamttrefferquoten der dichotomen Insolvenzprognosen aller Modelle bei unterschiedlichen Stichprobenumfängen.

Modell	In sample			Out-of sample		
	Stichprobenumfang			Stichprobenumfang		
	125	250	500	125	250	500
Vollmodell	92.00	90.00	87.60	74.40	76.80	81.80
Waldmodell	87.20	88.40	86.20	76.80	77.60	81.80
Bootstrap	100	94.40	90.60	68.80	76.00	82.20

Für eine detaillierte Untersuchung sind die tatsächlichen Trefferquoten beider Gruppen von Interesse, also die Prognosefähigkeit solventer bzw. insolventer Unternehmen. Die Insolvenzprognose ist dabei von besonderem Interesse, da insolvente Unternehmen möglichst frühzeitig erkannt und ein Kreditengagement vermieden werden sollte. Die Ergebnisse dieser detaillierten (dichotomen) Prognosegüte sind in Tabelle 5 dargestellt.

Ins Auge fällt die Verbesserung der Insolvenzprognosen beim kleinsten Stichprobenumfang des Bootstrap-Verfahrens gegenüber dem klassischen Wald- und Vollmodell. Die relativ geringe Gesamt-Prognosegüte des Bootstrap-Verfahrens begründet sich somit durch eine sehr schwache Solvenzprognose. Mit steigendem Stichprobenumfang relativiert sich die Insolvenzprognosegüte des Bootstrap-Verfahrens gegenüber den klassischen Ansätzen.

**Tabelle 5:** Detaillierte Prognosegüte aller Modellierungstechniken bei unterschiedlichen Stichprobenumfängen.

Modell	Gruppe	In sample			Out-of sample		
		Stichprobenumfang			Stichprobenumfang		
		125	250	500	125	250	500
Vollmodell	solvent	90.80	89.35	87.32	78.65	71.89	84.14
	insolvent	94.74	91.36	88.24	63.89	90.77	76.19
Waldmodell	solvent	87.36	89.35	85.59	73.03	72.97	82.72
	insolvent	86.84	86.42	87.58	86.11	90.77	79.59
Bootstrap	solvent	100	92.31	89.05	59.55	70.27	83.85
	insolvent	100	98.77	94.12	91.67	92.31	78.23

Die detaillierten Ratingklassen können in diesem Simulationssystem ebenfalls evaluiert werden. Die Konstruktion der hier verwendeten Ratingdefinition ist im Vorfeld bereits diskutiert, sodass an dieser Stelle nur die Ergebnisse präsentiert werden sollen. Die Prognosegüte der detaillierten Ratingklassen ist zu jedem Verfahren als schwach einzustufen. In der Literatur wird eine Trefferquote von 75% als ‚gut‘ bezeichnet.<sup>58</sup> Bezogen auf diesen Anspruch ist keines der hier betrachteten Verfahren zur Ratingprognose als ‚gut‘ einzustufen. Die höchste beobachtete Trefferquote in diesem Kontext beträgt (out-of sample) 62.80% (Waldmodell). Das Vollmodell prognostiziert beim kleinsten Stichprobenumfang nicht einmal jedes dritte Unternehmen richtig. Beim Waldmodell sind es immerhin bei allen Stichprobenumfängen über 50%. Auffällig ist der Vergleich zwischen dem Waldmodell und dem Bootstrap bei 250 Unternehmen. Die Prognosegüte des Bootstrap ist knapp 10%-Punkte höher. Die dichotome Prognosekraft

<sup>58</sup> Vgl. Leker und Schewe (1998), S. 887.

(Tab. 4) weist allerdings ein umgekehrtes Verhältnis aus: Hier ist die Prognosegüte des Waldmodells um knapp 1.5%-Punkte höher als beim Bootstrap-Verfahren. Beim Stichprobenumfang von 500 Unternehmen ist diese genau umgekehrt: Das Waldmodell ist bei der dichotomen Prognosegüte schwächer und bei der polytomen Prognosegüte besser als das Bootstrap-Verfahren.

**Tabelle 6:** Gesamttrefferquoten der polytomen Ratingprognosen aller Modelle bei unterschiedlichen Stichprobenumfängen.

Modell	In sample			Out-of sample		
	Stichprobenumfang			Stichprobenumfang		
	125	250	500	125	250	500
Vollmodell	36.80	52.00	48.80	29.60	48.40	45.80
Waldmodell	55.20	49.20	64.40	57.60	52.80	62.60
Bootstrap	56.00	54.00	51.80	52.00	61.60	50.80

Die Untersuchung zeigt daher, dass ein Verfahren, in einem gegebenen Modell, mit der höchsten dichotomen Trefferquote, keinesfalls auch die höchste polytome Trefferquote aufweisen muss. Dieses Phänomen ist auch unter anderen Datenkonstellationen zu beobachten und stellt keineswegs ein einmaliges Ergebnis dar. Für die praktische Ratingbildung auf Basis eines ökonometrischen Modells ist die alleinige Trefferquote der dichotomen Prognosefähigkeit, wie sie in der gängigen Forschung angegeben und als Qualitätsmaßstab interpretiert wird, kein Indiz für eine optimale Ratingprognose.

Auffällig ist auch die Fluktuation der Ergebnisse. Diese lassen sich durch die kleinen Stichproben begründen. Die logistische Regression schätzt nur asymptotisch konsistente Parameter, sodass mit verzerrten Wahrscheinlichkeitsprognosen bei kleinen Stichprobenumfängen zu rechnen ist. Darüber hinaus können Ausreißer negativ auf die Prognoseergebnisse wirken. Es

bietet sich daher der Einsatz von Verfahren, die robust gegenüber diesen Konstellationen sind an.

Für praktische Anwendungen dieses quantitativen Ratinggenerierungsprozesses bleibt damit die Frage unbeantwortet, welches Verfahren in realen Anwendungen zum Einsatz kommen sollte. Der Einsatz von Monte Carlo Simulationen kann die Prognosequalität eines Ratingverfahrens aufzeigen. Dieses kann erste Ansätze liefern, welches Verfahren unter gegebenen Bedingungen zu präferieren ist. Schließlich kann die Monte-Carlo Technik an einem vorliegenden Datensatz angewendet werden, sofern eine Wahrscheinlichkeit für die Bernoulli-Experimente bekannt ist. Der Original in-sample Datensatz einer empirischen Studie enthält auch die Abhängigkeiten, die in den Jahresabschlussdaten enthalten sind. Man kann daher unter den tatsächlichen Bedingungen eines Kreditinstitutes das optimale Verfahren identifizieren.

Dieses Vorgehen erfordert auf der einen Seite einen immensen Rechenaufwand, eröffnet aber auf der anderen Seite die Identifikation desjenigen Verfahrens, welches unter realen Bedingungen zu optimalen Prognosen führt.

## **5 Zusammenfassung**

Dieser Beitrag diskutiert Prognosefähigkeiten ausgewählter, quantitativer Ratingverfahren, die durch die neue Eigenkapitalvereinbarung im Kreditrisikomanagement zum Einsatz kommen können. Weit verbreitet ist die Nutzung des logistischen Regressionsmodells. Dieses Modell wird zur Generierung einer künstlichen Datenbank verwendet und definiert zu jedem Unternehmen eine a priori Wahrscheinlichkeit. Auf Basis dieser Wahrscheinlichkeit wird die a priori Ratingklasse bestimmt. Die Anwendung verschiedener Modellierungstechniken (Vollmodell, Waldmodell und Bootstrap-Verfahren) wird an einem künstlichen Datensatz untersucht und die Prognosen mit den bekannten Ratingklassen verglichen. Die Diskussion der Ergebnisse zeigt schließlich auf, dass Verfahren mit der höchsten dichotomen Trefferquote nicht die höchste polytome Trefferquote haben müssen. Die dichotome Trefferquote ist daher kein Indiz für ein leistungsstarkes Ratingverfahren.

## **Literatur**

- [1] Anders, U. und Szczesny, A. (1998). Prognose von Insolvenz-wahrscheinlichkeiten mit Hilfe logistischer neuronaler Netzwerke. *Zeitschrift für Betriebswirtschaftliche Forschung*, 50, 892 – 915.
- [2] Becker, H.P. und Peppmeier, A. (2000). *Bankbetriebslehre*. 4. Auflage, Friedrich Kiehl Verlag, Ludwigshafen.
- [3] Berndt, H. (2001). Kreditwesen - Umfrage zu Basel II. *Zeitschrift für das gesamte Kreditwesen*, 4, 170 - 17.
- [4] Efron, B. und Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [5] BIS (2003). Bank for International Settlement. Consultative Document: The New Basel Capital Accord.
- [6] Eigermann, J. (2002). *Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale*. 2. Auflage. Verlag Wissenschaft & Praxis, Sternenfels.
- [7] Eilenberger, G. (1996). *Bankbetriebswirtschaftslehre*. 6. Auflage, Oldenbourg Verlag, München.
- [8] Grunert, J., Kleff, V., Norden, L. und Weber, M. (2002). Mittelstand und Basel II: Der Einfluss der neuen Eigenkapitalvereinbarung für Banken auf die Kalkulation von Kreditzinsen. *Zeitschrift für Betriebswirtschaft*, 72, 1045 – 1064.
- [9] Guennemann, D.-C. und Szczesny, A. (2001). Behandlung von Kreditrisiken durch die Bankenaufsicht: Fragen und Antworten in Zeiten des Umbruchs. In: Szczesny, A. (Hrsg.): *Kreditrisikomessung und Kreditrisikomanagement*, Nomos Verlagsgesellschaft, 11 – 34.
- [10] Hayden, E. (2002) *Modelling an Accounting-Based Rating System for Austrian Firms*, Fakultät für Wirtschaftswissenschaften und Informatik der Universität Wien, Dissertation.
- [11] Heinke, E. (2001). Basel II und seine Bedeutung für die mittelständ. Wirtschaft. *Zeitschrift für das gesamte Kreditwesen*, 4, 174-178.



- [12] Heinke, E. (2002). Das Baseler 3-Säulen-Konzept und die Rolle der dezentralen Bankenaufsicht. In: Tietmeyer, H. und Rolfes, B. (Hrsg.): Basel II - Das neue Aufsichtsrecht und seine Folgen. Gabler Verlag, 1 – 12.
- [13] Hosmer, D.W. und Lemeshow, S. (2000). Applied Logistic Regression. John Wiley & Sons, New York.
- [14] Langer, F. (1998): Berücksichtigung von Kovariablen im nicht-parametrischen gemischten Modell, Institut für Mathematische Stochastik, Universität Göttingen, Dissertation.
- [15] Leker, J. und Schewe, G. (1998). Beurteilung des Kreditausfallrisikos im Firmenkundengeschäft der Banken. Zeitschrift für Betriebswirtschaftliche Forschung, 50, 877 – 891
- [16] Moody's Investors Service (2001) Moody's RiskCalc™ für nicht börsennotierte Unternehmen: Das Deutsche Modell.
- [17] Oelerich, A. und Poddig, Th. (2004): Modified Wald Statistics for Generalized Linear Models. Journal of the German Statistical Society, 1, 23 – 34.
- [18] Poddig, Th. und Kunze, B. (2003). Riskmanagementsysteme bei Banken vor dem Hintergrund der staatlichen Regulierung des Finanzsektors. Finanzbetrieb, 1, 693 – 702.
- [19] Rolfes, B. und Emse, C. (2000): Rating basierte Ansätze zur Bemessung der Eigenkapitalunterlegung von Kreditrisiken. ecfs – Forschungsbericht, 3.
- [20] SAS Institute Inc. (1995). Logistic Regression Examples Using the SAS System, Version 6, First Edition. Cary, NC: SAS Institute Inc.
- [21] Siemer, A. (2002): Die statistische Auswertung von ordinalen Daten bei zwei Zeitpunkten und zwei Stichproben, Mathematisch-naturwissenschaftliche Fachbereich der Universität Göttingen, Dissertation.
- [22] Sironi, A. und Zazzara, C. (2003): The Basel Committee proposals for a new capital accord: implication for Italian banks. Review of Financial Economics, 12, 99 – 126.
- [23] Tholen, M. (2000): Basel II: Was erwartet den Mittelstand? Zeitschrift für das gesamte Kreditwesen, 10, 547 - 549.

*A. Oelerich, T. Poddig*

- [24] Wilkens, M., Entrop, O. und Völker, J. (2001): Strukturen und Methoden von Basel II – Grundlegende Veränderungen der Bankenaufsicht. *Zeitschrift für das gesamte Kreditwesen*, 4, 187 – 193.