



WESTFÄLISCHE
WILHELMUS-UNIVERSITÄT
MÜNSTER

Validierung eines Ratingsystems am Beispiel eines mittelständischen Unternehmens

Diplomarbeit

vorgelegt von
Ulrich Konrad Frye

Betreuer: Privatdozent Dr. Volkert Paulsen
Institut für Mathematische Statistik
Fachbereich 10 - Mathematik und Informatik
Westfälische-Wilhelms-Universität Münster

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnisverzeichnis	VII
Einleitung	1
1 Statistische Grundlagen	5
1.1 Grundlagen	6
1.2 Einführung von Ratingklassen	8
1.3 Monotonie	10
2 Trennschärfe	15
2.1 Cumulative Accuracy Profile	15
2.2 Reciever Operating Characteristic	21
2.3 Weitere Trennschärfemaße	26
2.3.1 Bayes Fehler	26
2.3.2 Pietra-Index	28
2.3.3 Bedingter Entropie Quotient	30
3 Kalibrierung von Ratingsystemen	33
3.1 Einfaktor-Modell	33
3.2 Verteilung einer Ausfallquote	36
3.3 Test einer Ausfallwahrscheinlichkeit	38
4 Datenanalyse	49
4.1 Trennschärfeberechnung	51
4.2 Test einer Ausfallwahrscheinlichkeit	56
Fazit	59
Anhang	61
Literaturverzeichnis	IX

Abbildungsverzeichnis

1	Komponenten der Validierung	2
1.1	Beispiel für bedingte Dichten	7
1.2	bedingte Ausfallwahrscheinlichkeit als Funktion der Score Werte	8
2.1	CAP - Kurve für Ratingsysteme	16
2.2	Bedingte Dichten bei perfektem Ratingsystem	17
2.3	CAP - Kurve für Ratingsysteme	19
2.4	ROC - Kurve für Ratingsysteme	22
2.5	Ableitung der ROC - Kurve aus Abbildung 2.4	23
2.6	ROC - Kurve bei Identität der bedingten Vtlg.Fkt.	29
2.7	Informationsentropie als Funktion der Ausfallwahrscheinlichkeit	30
4.1	Transaktionsmaske mySAP ERP	50
4.2	CAP - Kurve des Ratingsystems der Fa. KRONE	52
4.3	ROC - Kurve des Ratingsystems der Fa. KRONE	53

Tabellenverzeichnis

4.1	Score-Werte s und ihre Einteilung in Ratingklassen	49
4.2	Datenerhebung	50
4.3	Trennung in Defaulter und Non-Defaulter	50
4.4	kum. Wahrscheinlichkeiten, Angaben in %	51
4.5	Funktionswerte $CAP(u)$	51
4.6	Anteil in den Ratingklasse, Angaben in %	52
4.7	Funktionswerte $ROC(u)$	53
4.8	Ableitung von $ROC(u)$	54
4.9	Fehlerwahrscheinlichkeit der Ratinklassen, in %	55
4.10	Abstand der Verteilungsfunktionen	55
4.11	Entropie der bedingten empirischen Ausfallwahrscheinlichkeit	56
4.12	prognostizierte Ausfallwahrscheinlichkeit der CreditReform	61
4.13	theoretische Ausfälle des CreditReform Ratings	61
4.14	kum. Wahrscheinlichkeiten, Angaben in %	62
4.15	Anteil in den Ratingklasse, Angaben in %	62
4.16	Kritische Werte des Kolmogorov-Smirnoff Tests	63

Einleitung

Die Geschichte des Ratings reicht bis ins 19. Jahrhundert zurück. Mit der Ausdehnung des Eisenbahnnetzes über den US-amerikanischen Kontinent, wurden enorme Kredite benötigt, die einzelne Banken alleine nicht mehr aufbringen konnten. Aus diesem Grund gaben die Industrieunternehmen zusätzliche Anleihen aus, um finanzielle Mittel von Investoren zu beschaffen. Da die Anleger ohne Informationen über das Risiko nicht bereit waren ihr Geld an die Eisenbahngesellschaften zu geben, veröffentlichte Henry Varnum Poor 1868 das „Manual of the Railroads of the United States“. In diesem Schreiben wurden Informationen über die Eisenbahngesellschaften zusammengefasst. Somit war dieses die erste Bonitätsbeurteilung eines Unternehmens. Seitdem sind Bonitätsbewertungen unumgänglich in der Finanzindustrie.

Mit der Umsetzung von Basel II in das europäisches Recht 2007 wurde die Bedeutung des Ratings in Deutschland erhöht. Basel II ist ein Empfehlungsschreiben des Baseler Ausschusses für Bankenaufsicht. Mit der Umsetzung von Basel II wird die Eigenkapitalanforderung der Banken an das Risiko des Portfeuille gekoppelt. Das bedeutet, dass die Bank bei einem höheren Ausfallrisiko des Kredits diesen mit mehr Eigenkapital hinterlegen muss. Das wiederum führt zu einer Verteuerung des geliehenen Geldes für die Bank. Die höheren Kosten werden an den Antragssteller durch z.B. höhere Zinsen weitergegeben.

Aber auch für andere Industriezweige ist die Bonitätsbewertung von Bedeutung. So prüfen viele Unternehmen zuerst die Bonität von Kunden, bevor sie mit ihnen Geschäftsbeziehungen eingehen. Im Falle einer schlechten Bewertung werden entweder besondere Zahlungsbedingungen gefordert oder die Geschäftsbziehung sogar ganz abgelehnt. Die Bewertung der Bonität wird mit Hilfe von Ratingsystemen ermittelt.

Unter dem Begriff der *Validierung* wird der gesamte Prozess der Überprüfung von Ratingsystemen verstanden. Grundsätzlich wird zwischen der qualitativen und quantitativen Validierung unterschieden. In Abbildung 1 ist die Struktur von Vali-

dierungsprozessen aufgezeigt.

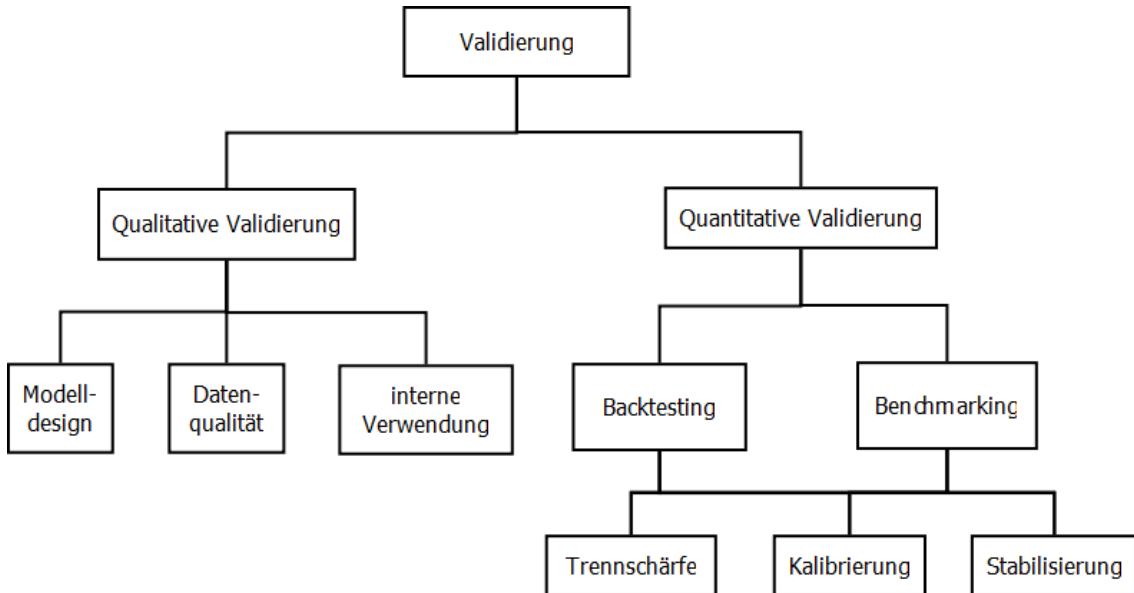


Abbildung 1: Komponenten der Validierung

Die qualitative Validierung ist der quantitativen Validierung in der Regel vorangestellt. Dabei besteht diese aus den Bereichen Modelldesign, Datenqualität und interne Verwendung.

Die Hauptaufgabe des Modelldesigns ist die Dokumentation des Ratingsystems. Wichtige Kriterien sind hierbei der Umfang und die Vollständigkeit der Dokumentation. Das bedeutet, dass der Ratingprozess bzw. die Architektur des Ratingmodells transparent gemacht werden sollen. Weiterhin soll dadurch die Auswahl sowie der Einfluss von Risikoparametern ökonomisch begründet werden.

Der Datenqualität kommt bei der Entwicklung eines Ratingsystems eine eminent Bedeutung zu und ist damit schon ein erstes Gütemaß für das Modell. Bei der Validierung der Datenqualität wird untersucht, ob die Daten vollständig sind. Ebenfalls muss geprüft werden, ob die zur Entwicklung oder Validierung verwendete Stichprobe repräsentativ ist. Nur wenn dieses erfüllt ist, lässt sich das ermittelte Ratingergebnis nachvollziehen.

Bei der Validierung der internen Verwendung wird lediglich überprüft, inwieweit das Ratingsystem in das Risikomanagement eingebunden ist. Die interne Verwendung ist jedoch ein Indiz für das Vertrauen in das Ratingmodell.

Die zweite Säule der Validierung besteht aus der quantitativen Validierung des Ratingmodells. Wichtig für diese Validierungsmethode ist ein ausreichender Stichprobenumfang. Die quantitative Validierung unterteilt sich in das statistische Backtesting sowie das Benchmarking.

Das Backtesting wird anhand von firmeninternen Daten durchgeführt, wobei beim Benchmarking das Ratingmodell anhand von Vergleichsdaten überprüft wird. Dieses ermöglicht den Vergleich verschiedener Ratingmodelle auf einer identischen Datenbasis. Bei beiden quantitativen Validierungskomponenten wird die Trennschärfe, die Kalibrierung der Ausfallwahrscheinlichkeiten sowie die Stabilität des Ratingmodells untersucht.

Die Fähigkeit *ex ante* zwischen ausgefallenen (Defaulter) und solventen (Non-Defaulter) Debitoren zu unterscheiden, wird Trennschärfe genannt. Je besser das Ratingmodell ausfallgefährdete Debitoren erkennen kann, desto höher ist die Trennschärfe. Die Trennschärfe dient somit als Gütemaß des Ratingmodells.

Die Kalibrierung der Ausfallwahrscheinlichkeit untersucht die Abweichung der prognostizierten von der tatsächlichen Ausfallwahrscheinlichkeit. Dabei dürfen sich diese für ein adäquates Risikomanagement nicht signifikant unterscheiden.

Die Beständigkeit des Ratingmodells über einen längeren Zeitraum sowie sich ändernde Rahmenbedingungen, nennt man Stabilität des Ratingmodells. Die Stabilität wird validiert, indem die vorangegangenen Validierungskomponenten mit Hilfe verschiedener Szenarien untersucht werden.

Die vorliegende Arbeit widmet sich ausschließlich dem statistischen Backtesting. Im ersten Kapitel werden die statistischen Grundlagen geklärt sowie ein Grundgerüst für die weiteren Validierungstechniken entwickelt. Das zweite Kapitel beschäftigt sich mit der Messung der Trennschärfe. Dabei werden ausgewählte Beispiele der Trennschärfemessung vorgestellt.

Der Kalibrierung des Ratingsystems widmet sich das dritte Kapitel. Anhand des Ein-Faktor Modells wird dabei die prognostizierte Ausfallwahrscheinlichkeit gegen die wahre Ausfallwahrscheinlichkeit getestet. Im vierten Kapitel werden die vorgestellten Methoden der Validierung auf die Datenerhebung der Fahrzeugwerk Bernard KRONE GmbH angewandt und somit das Ratingsystem der Fa. KRONE überprüft. Abschließend wird ein Fazit über die gewonnenen Ergebnisse gezogen.

Ich möchte ganz herzlich Herrn Privatdozent Dr. Volkert Paulsen für die Betreuung dieser Arbeit danken. Ein besonderer Dank gilt auch der Fahrzeugwerk Bernard

KRONE GmbH, insbesondere Herrn Hubert Börger für die Durchführung der Arbeit im Rahmen meines Betriebspraktikums. Für die Durchsicht ein Dank an Felix Brinkmann.

Gemäß § 21 (6) der Diplomprüfungsordnung für den Studiengang Mathematik der Westfälischen Wilhelms-Universität Münster vom 15. Juli 1998 versichere ich, dass ich die vorliegende Diplomarbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel benutzt habe.

Münster, den 28. April 2010

Ulrich K. Frye

Kapitel 1

Statistische Grundlagen

In diesem Kapitel wird ein Modell vorgestellt, welches als Gerüst der gängigen Validierungstechnik dient. Zunächst werden die statistischen Grundlagen erklärt.

Dazu sei $(\Omega, \sigma(\Omega), P)$ ein Wahrscheinlichkeitsraum, welcher unspezifiziert bleibt. Ω kann dabei als Menge aller Debitoren aufgefasst werden.

Jedem Debitor werden nun zwei Zufallsvariablen S und X zugeordnet.

$S : \Omega \rightarrow \mathbb{R}$ gibt dabei den sogenannten Score -entlang einer stetigen Skala- des Debitors an. Der Score -Realisation der Zufallsvariablen S - fasst die aus der Buchführung gewonnenen Daten bzgl. eines Debitors in einem einzelnen Wert zusammen und ist somit eine Kennzahl für die Kreditwürdigkeit bzw. Bonität des Debitors. Dabei gilt, dass ein hoher Score auch eine hohe Bonität impliziert bzw. ein niedriger Score eine geringe Bonität. In diesem Modell wird S auch als Score-Variable bezeichnet.

Die Zufallsvariable $X : \Omega \rightarrow \{0, 1\}$ soll den Status des Debitors -ausgefallen bzw. nicht ausgefallen- am Ende einer Beobachtungsperiode angeben. Die Zufallsvariable ordnet dem Debitor hierzu im Falle eines Ausfalls den Status 1 bzw. im Falle eines solventen Debitors den Wert 0 zu. Formal für den i -ten Debitor:

$$X(i) = X_i = \begin{cases} 1, & i \text{ ist ausgefallen} \\ 0, & i \text{ ist nicht ausgefallen} \end{cases}$$

Der Status des Debitors ist zu Anfang der Beobachtungsperiode nicht bekannt. Die Zufallsvariable X ist somit latent. Ziel ist es, mit Hilfe des Scores s , den zukünftigen Zustand des Debitors vorherzusagen.

1.1 Grundlagen

Zunächst sollen einige Bezeichnungsweisen für den weiteren Verlauf eingeführt werden.

Die Menge Ω der Debitoren lässt sich am Ende der Beobachtungsperiode in die disjunkten Mengen der Defaulter D und der Non-Defaulter N aufteilen. Die Aufteilung wird dabei kanonisch durch die Realisation der Zufallsvariablen X gegeben. Die Randverteilung von X ist offensichtlich Binomial-verteilt mit

$$p \stackrel{\text{def}}{=} P(X = 1) = 1 - P(X = 0) \quad (1.1)$$

Dabei sei mit p die totale Ausfallwahrscheinlichkeit -also die Wahrscheinlichkeit, dass am Ende einer Periode ein Debitor ausgefallen ist- gemeint. Mit $f_D : \mathbb{R} \rightarrow [0, 1]$ ist die bedingte Dichte bzgl. des Lebesgue Maßes von S -gegeben $X = 1$ bezeichnet. Die Verteilungsfunktion ist entsprechend gegeben durch

$$F_D(s) = P(S \leq s \mid X = 1) = \int_{-\infty}^s f_D(u) du.$$

Des Weiteren wird mit $f_N : \mathbb{R} \rightarrow [0, 1]$ die bedingte Dichte von S -gegeben $X=0$ bezeichnet. Die Verteilungsfunktion von S -gegeben $X=0$ ist gegeben durch

$$F_N(s) = P(S \leq s \mid X = 0) = \int_{-\infty}^s f_N(u) du.$$

Die unbedingte Dichte der Verteilung S sei $f : \mathbb{R} \rightarrow [0, 1]$. Die Verteilungsfunktion von S wird festgelegt durch

$$F(s) = P(S \leq s) = \int_{-\infty}^s f(u) du \quad (1.2)$$

In der Abbildung (1.1) ist ein Beispiel für die bedingten Dichten f_D und f_N gegeben.

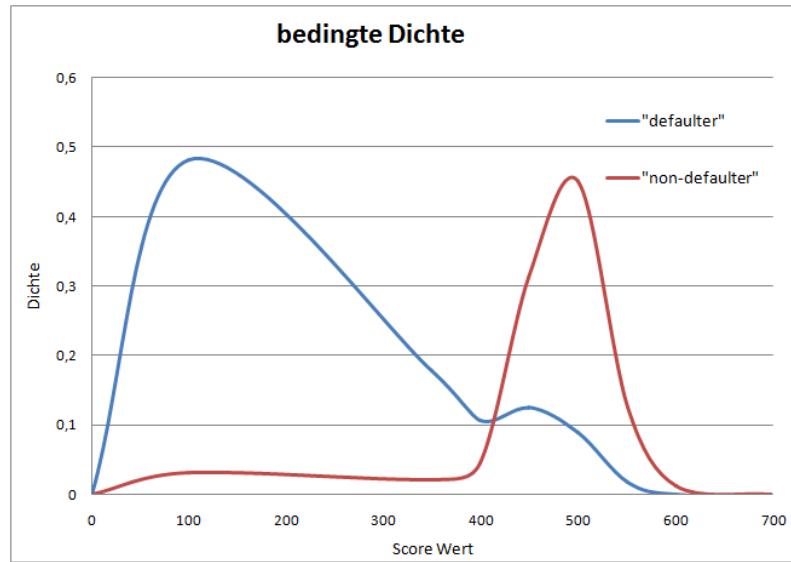


Abbildung 1.1: Beispiel für bedingte Dichten

Vermöge dem Satz von der totalen Wahrscheinlichkeit, kann die unbedingte Dichte f geschrieben werden als

$$f(s) = p \cdot f_D(s) + (1 - p) \cdot f_N(s) \quad (1.3)$$

Mit dem Satz von Bayes lässt sich die bedingte Ausfallwahrscheinlichkeit $P(X = 1 | S = s)$ dann errechnen zu

$$P(X = 1 | S = s) = \frac{p \cdot f_D(s)}{f(s)} \quad (1.4)$$

Abbildung 1.2 illustriert die bedingte Ausfallwahrscheinlichkeit - gegeben den Score s - gemäß (1.4).

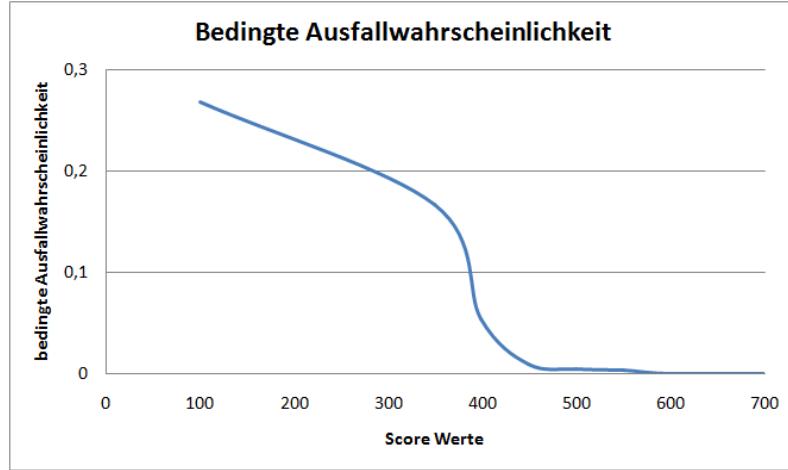


Abbildung 1.2: bedingte Ausfallwahrscheinlichkeit als Funktion der Score Werte

1.2 Einführung von Ratingklassen

Aufgrund der Stetigkeit der Scorevariablen S werden nur wenige bzw. keine Beobachtungen pro Realisation gezählt. Jedoch ist für das Backtesting eines Rating-systems eine gewisse Anzahl an Debitoren pro Realisation erforderlich. Zu diesem Zweck fordert das Baseler Komitee für Bankenaufsicht eine endliche Anzahl von Ratingklassen in einem Ratingsystem. Im Folgenden wird ein Übergang zu einer diskreten Dichte konstruiert, sodass den obigen Anforderungen genügt wird. Ein erstes Verfahren wird wie folgt konstruiert.

Es werden k Ratingklassen durch rekursiv definierte Scoregrenzen $s_{k-1} > s_{k-2} > \dots > s_1$ definiert, wobei k die Ratingklasse mit der höchsten Bonität angibt. Zu jeder Ratingklasse wird dann eine feste Ausfallwahrscheinlichkeit p_j , mit $j = 1, \dots, k$ bestimmt. Es gilt:

$$p_k < p_{k-1} < \dots < p_2 < p_1$$

Weiter wird angenommen, dass die bedingte Ausfallwahrscheinlichkeit $P(X = 1 | S = s)$ als Funktion in s fallend ist. Auf diesen Aspekt wird noch im Laufe der Arbeit (vgl. Abschnitt 2.3) näher eingegangen. Die Scoregrenzen werden festgelegt durch:

$$p_k = P(X = 1 | S \geq s_{k-1}) = \frac{p \int_{s_{k-1}}^{\infty} f_D(u) du}{\int_{s_{k-1}}^{\infty} f(u) du}$$

und

$$p_i = P(X = 1 \mid s_i > S \geq s_{i-1}) = \frac{p \int_{s_{i-1}}^{s_i} f_D(u) du}{\int_{s_{i-1}}^{s_i} f(u) du} \quad \text{für } i = 2, \dots, k-1. \quad (1.5)$$

Ein Debitor mit Score s wird anhand der Funktion

$$R(s) = \begin{cases} k & s \geq s_{k-1} \\ i & , \text{ falls } s_i > s \geq s_{i-1} \\ 1 & s_1 > s \end{cases} \quad (1.6)$$

einer Ratingklasse zugeordnet. Dabei wird z.B. ein Debitor mit Score größer oder gleich s_{k-1} der Ratingklasse k zugeordnet, sowie falls der Score echt kleiner s_1 ist der ersten Ratingklasse.

Die Ausfallwahrscheinlichkeit der „schlechtesten“ Ratingklasse 1 ist über die Ausfallwahrscheinlichkeiten der $k-1$ anderen Ratingklassen definiert. Diese wird festgelegt durch

$$p_1 = P(X = 1 \mid R(s) = 1) = P(X = 1 \mid s_1 > S) = \frac{p \int_{s_1}^{s_1} f_D(u) du}{\int_{-\infty}^{s_1} f(u) du}.$$

Des Weiteren lässt sich beobachten, dass die Abbildung $r \mapsto P(X = 1 \mid R(s) = r)$ monoton fällt in r , falls die Funktion $s \mapsto P[X = 1 \mid S = s]$ für fallende s fallend ist.

1.2.1 Bemerkung

Betrachtet man das System über mehrere Perioden, so müssen die Score-Grenzen mit der Zeit angepasst werden, damit die Ausfallwahrscheinlichkeiten der Ratingklassen konstant bleiben. Dann spricht man von einem *through – the – cycle*-Ratingsystem.

Ein weiteres Verfahren zur Konstruktion von Ratingklassen besteht darin, den Anteil der Non-Defaulter pro Ratingklasse konstant zu halten, d.h.:

Es werden k Ratingklassen gewählt, wobei die Ratingklasse k wieder die höch-

te Bonität angibt. Die relative Anzahl der Non-Defaulter pro Ratingklasse $0 < r_1, r_2, \dots, r_k$ mit $\sum_{i=1}^k r_i = 1$ bleibt konstant. Die Scoregrenzen $s_{k-1} > s_{k-2} > \dots > s_1$ werden dann rekursiv festgelegt durch

$$r_k = P(S \geq s_{k-1} \mid X = 0) = \int_{s_{k-1}}^{\infty} f_N(s) \, ds$$

und

$$r_i = P(s_i > S \geq s_{i-1} \mid X = 0) = \int_{s_{i-1}}^{s_i} f_N(s) \, ds, \quad \text{für } i = 2, \dots, k-1.$$

Die Abbildung der Scores auf die Ratingklasse ist analog zu (1.6) definiert. Somit folgt wegen $\sum_{i=1}^k r_i = 1$

$$P(R(s) = 1 \mid X = 0) = P(S < s_1 \mid X = 0) = r_1.$$

1.2.2 Bemerkung

Wird das System wieder über mehrere Perioden betrachtet, so müssen auch hier die Score-Grenzen mit der Zeit angepasst werden, damit der Anteil der Non-Defaulter konstant bleibt. Jedoch wird mit diesem Verfahren ein großer Shift der Score-Variablen vermieden. Es wird in diesem Fall von einem *point-in-time*-Ratingsystem gesprochen.

Im weiteren Verlauf der Arbeit wird allgemein von Ratingsystemen gesprochen, falls die Score Variable und ihre Einteilung in Ratingklassen gemeint ist.

1.3 Monotonie

Im vorherigen Kapitel wurde die Monotonie der bedingten Ausfallwahrscheinlichkeit $P(X = 1 \mid S = \cdot)$ vorausgesetzt. Dieses ist jedoch nicht immer gewährleistet. Ziel ist es, nun mit Hilfe eines statistischen Tests, die Bedingungen für die Monotonie näher zu analysieren. Zunächst wird ein optimaler Test bestimmt.

Dazu wird zufällig ein Debitor ausgewählt und sein Score notiert. Da noch keine

Informationen über seinen zukünftigen Status vorliegen, wird mit Hilfe des Scores auf die Zustandsvariable X geschlossen.

Anzunehmen ist, dass (s, x) eine Realisation von (S, X) ist. s ist beobachtet worden, x ist noch nicht bekannt. Es stellt sich die Frage, ob $x = 0$ oder $x = 1$ ist. Es wird eine Entscheidungsmenge E gewählt, sodass ein Ausfall angenommen wird, sofern der Score des Debtors in der Menge E liegt. Falls dieses nicht der Fall ist, wird von Solvenz des Debtors ausgegangen.

$$s \in E \implies \text{Annahme } x = 1$$

$$s \notin E \implies \text{Annahme } x = 0$$

Um die Menge E zu bestimmen, wird die bedingte Score Verteilung der Defaulter $P(S \in \cdot \mid X = 1)$ und die bedingte Score Verteilung der Non-Defaulter $P(S \in \cdot \mid X = 0)$ betrachtet. Ziel ist es, die Menge E so zu wählen, dass die bedingte Score Verteilung der Defaulter maximal und die bedingte Score Verteilung der Non-Defaulter minimal wird. Dazu wird die Hypothese H_0 (der Debitor ist zukünftiger Defaulter) bzw. die Alternative H_1 (der Debitor bleibt liquide) aufgestellt. Dann werden diese Hypothesen gegeneinander getestet.

$$H_0 : x = 1 \quad \text{gegen} \quad H_1 : x = 0$$

In einem solchen Test können zwei Fehler begangen werden. Zum Einen der Fehler 1. Art -die Hypothese H_0 wird ablehnt, obwohl diese wahr ist-, zum Anderen der Fehler 2. Art -die Hypothese H_0 wird beibehalten, obwohl die Alternative H_1 gilt.

- Fehler 1. Art: Debitor wird für solvent gehalten, obwohl er Defaulter ist
- Fehler 2. Art: Debitor wird als Defaulter angenommen, obwohl er solvent bleibt

Um nun optimal zu entscheiden, werden die Wahrscheinlichkeiten der beiden Fehlermöglichkeiten betrachtet.

Die Wahrscheinlichkeit des Fehlers 1. Art ergibt sich aus der bedingten Verteilung der Defaulter, unter der Annahme, dass der Score des Debtors kein Element der Menge E ist.

$$P(\text{Fehler 1. Art}) = P(S \notin E \mid X = 1)$$

Die Wahrscheinlichkeit des Fehlers 2. Art, ist die Wahrscheinlichkeit unter der Ver-

teilung der Non-Defaulter, dass der Debitor einen Score aus der Menge E hat.

$$P(\text{Fehler 2. Art}) = P(S \in E \mid X = 0)$$

Da der Fehler 1. Art kritischer ist, wird dieser nach oben hin durch eine Schranke α begrenzt, d.h. ein Test zum Niveau α durchgeführt. Dadurch ergibt sich als modifiziertes Ziel, einen gleichmäßig besten Test zum Niveau α zu finden, der den Fehler 2. Art minimiert.

$$\begin{aligned} P(\text{Fehler 1. Art}) &\leq \text{Schranke} \\ P(\text{Fehler 2. Art}) &\text{minimieren} \end{aligned}$$

Zunächst muss der sogenannte Likelihood Quotient und das zugehörige α -Fraktil eingeführt werden.

1.3.1 Definition

Seien f_D und f_N Dichtefunktionen von S , dann heißt

$$LR(S) \stackrel{\text{def}}{=} \frac{f_D}{f_N}(S)$$

Likelihood Quotient der Score Variablen S .

1.3.2 Definition

Das α -Fraktil von $LR(S)$ wird definiert durch

$$r_\alpha \stackrel{\text{def}}{=} \inf_{r \leq 0} \{P(LR(S) \leq r \mid X = 1) \geq 1 - \alpha\}.$$

Aufgrund der Stetigkeit kann das α -Fraktil durch Auflösung der folgenden Gleichung bestimmt werden.

$$1 - \alpha = P(LR(S) \leq r_\alpha \mid X = 1)$$

Mit Hilfe des Neymann-Pearson Lemmas folgt, dass

$$\varphi = \begin{cases} 1, & \text{für } LR(S) > r_\alpha \\ 0, & \text{sonst} \end{cases}$$

einen gleichmäßig besten Test zum Niveau α definiert. Dabei bedeutet $\varphi = 1$ die Ablehnung der Hypothese H_0 . Dieses Ergebnis wird auf das Problem übertragen.

Falls

- $S \notin E \iff \varphi = 1$
- $P(S \notin E \mid X = 1) \leq \alpha$

gilt, so folgt

$$P(LR(S) \leq r_\alpha \mid X = 0) \leq P(S \in E \mid X = 0). \quad (1.7)$$

Mit Hilfe von (1.7) kann nun die gesuchte Menge E definiert werden.

$$E \stackrel{\text{def}}{=} \{s : LR(s) \leq r_\alpha\} \quad (1.8)$$

Damit ist gezeigt, dass der gleichmäßig beste Test in der Betrachtung des Likelihood Quotienten liegt. Außerdem kann mit Hilfe der Menge E (1.8) zwischen Defaultern und Non-Defaultern unterschieden werden.

Um die Monotonie der Ausfallwahrscheinlichkeiten zu untersuchen, wird die Bayes Formel genutzt. Mit dieser kann gezeigt werden, dass die bedingte Ausfallwahrscheinlichkeit $P(X = 1 \mid S = s)$ genau dann monoton ist, wenn der Likelihood Quotient monoton ist.

Um diesen Aspekt weiter zu erläutern, wird die Eigenschaft der Linearität einer Score Variable eingeführt.

1.3.3 Definition

Die Score Variable S ist linear bzgl. der Verteilung $P(S \in \cdot \mid X = 1)$ und $P(S \in \cdot \mid X = 0)$, falls

$$\varphi = \begin{cases} 1 & \text{falls } S > r_\alpha \\ 0 & \text{sonst} \end{cases}$$

einen gleichmäßig besten Test zum Niveau α definiert.

Durch die Definition und (1.8) lässt sich folgende Bemerkung machen.

1.3.4 Bemerkung

Die Score Variable S ist genau dann linear bzgl. der Verteilung $P(S \in \cdot \mid X = 1)$ und $P(S \in \cdot \mid X = 0)$, wenn der Likelihood Quotient $LR(S)$ monoton ist und damit auch die bedingte Ausfallwahrscheinlichkeit.

In der Praxis ist es schwer, die Monotonie nachzuweisen. Jedoch ist es aus betriebswirtschaftlicher Sicht klar, dass die Monotonie der bedingten Ausfallwahrscheinlichkeit gegeben sein muss. Denn mit schlechter werdendem Score muss die bedingte Ausfallwahrscheinlichkeit des Debtors größer werden. Indirekt kann dieses schon als erstes Validierungskriterium des Ratings angesehen werden.

Kapitel 2

Validierungstechniken für die Trennschärfe von Ratingsystemen

Dieses Kapitel bezieht sich auf die Messung der Trennschärfe -die Fähigkeit zwischen Defaultern und Non-Defaultern zu unterscheiden- von Ratingsystemen. Wie bereits in der Einleitung erwähnt, ist die Trennschärfe ein Maß für die Güte eines Ratingsystems. Es gilt, je höher die Trennschärfe eines Ratingsystems, desto besser kann dieses zwischen Defaultern und Non-Defaultern differenzieren. Die Trennschärfe kann auf verschiedene Weise ermittelt werden. Einige in der Wirtschaft bedeutende Verfahren werden in diesem Kapitel vorgestellt. Zum Einen das *Cumulative Accuracy Profile*, zum Anderen die *Reciever Operating Characteristic*. Abschließend werden weitere Verfahren genannt.

2.1 Cumulative Accuracy Profile

Das Cumulative Accuracy Profile (*CAP*) ist ein graphisches Hilfsmittel zur Bestimmung der Trennschärfe von Ratingsystemen.

Mit F_D und F_N sei -wie im ersten Kapitel- die bedingte Verteilungsfunktion der Score Variablen S unter $X = 1$ bzw. $X = 0$ verstanden. p sei wiederum die totale Ausfallwahrscheinlichkeit der Debitoren. Somit kann mit Hilfe von (1.3) die unbedingte Verteilungsfunktion $F(s)$ der Score Variablen S bestimmt werden. Es ist:

$$F(s) = \int_{-\infty}^s f(u) du \stackrel{(1.3)}{=} (1-p)F_N(s) + pF_D(s) \quad (2.1)$$

Hiermit wird nun die *CAP*-Funktion definiert.

2.1.1 Definition

Die Funktion $CAP : [0, 1] \rightarrow [0, 1]$ definiert durch

$$CAP(u) \stackrel{\text{def}}{=} F_D(F^{-1}(u)), \quad u \in [0, 1] \quad (2.2)$$

heißt Cumulative Accuracy Profile (*CAP*)-Funktion, wobei $F^{-1}(u) \stackrel{\text{def}}{=} \sup_s \{P[S \leq s] < u\}$ die Pseudo-Inverse von F sei.

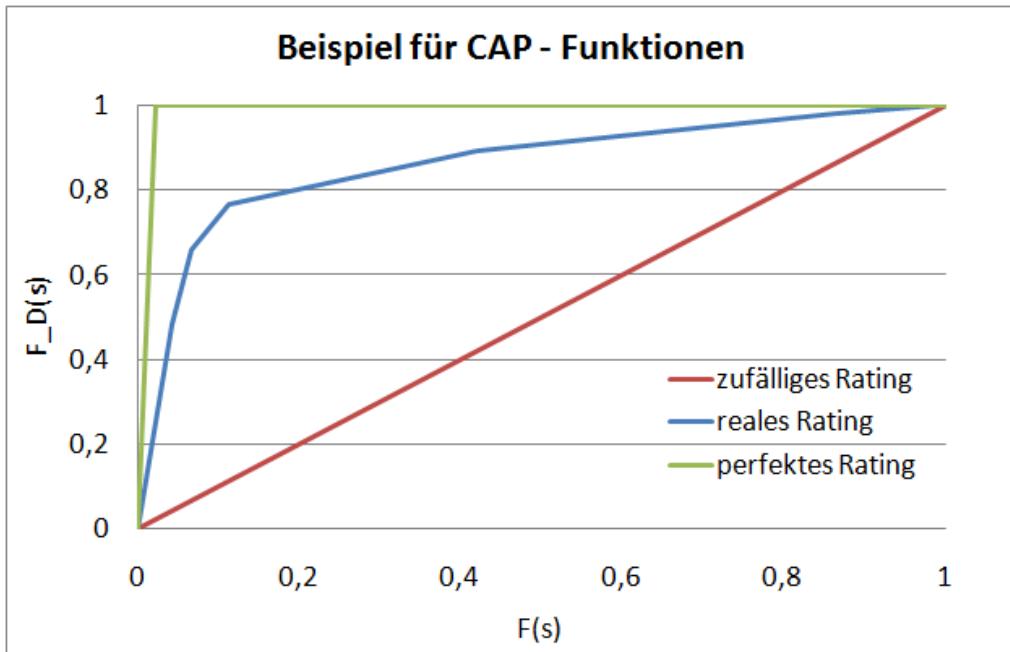


Abbildung 2.1: CAP - Kurve für Ratingsysteme

Alternativ erhält man die Kurve der *CAP*-Funktion, indem man die Punkte $(F(s), F_D(s))$ für $s \in \mathbb{R}$ plottet.

2.1.2 Bemerkung

Im *CAP*-Modell werden die Verteilungsfunktionen $F(s)$ und $F_D(s)$ auch als *Alarmrate* bzw. *Trefferquote* bezeichnet. Anschaulich gibt die *Alarmrate* $F(s)$ den relative Anteil aller Debitoren, kleiner oder gleich einem Score s , an. Mit der *Trefferquote* $F_D(s)$ ist der relative Anteil der Defaulten kleiner oder gleich einem Score s gemeint.

Die Abbildung 2.1 zeigt Beispiele für CAP-Funktionen. Diese Beispiele werden nun näher beschrieben.

- perfektes Ratingsystem

Der Graph eines perfekten Ratingsystems wächst linear vom Ursprung bis zum Punkt $(p, 1)$ und ist von $(p, 1)$ bis $(1, 1)$ konstant. Die Score Variable eines perfekten Ratingsystems ordnet allen Defaultern -besitzen einen Anteil von p an der Grundgesamtheit der Debitoren- im Vorhinein den niedrigsten Score zu. Somit unterscheidet das perfekte Ratingsystem nur zwischen Defaultern und Non-Defaultern. Der Definitionsbereich der bedingten Dichten f_D und f_N der Score Variablen ist disjunkt (Abbildung 2.2). Formal:

$$\{s : f_D(s) > 0\} \cap \{s : f_N(s) > 0\} = \emptyset \quad \text{für alle } s$$

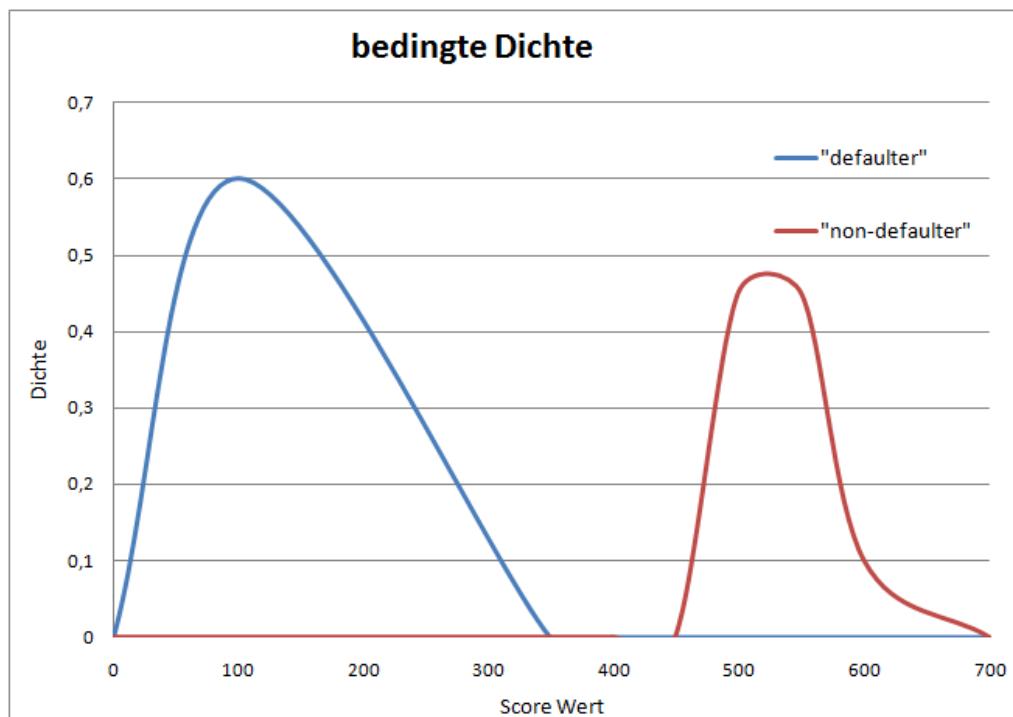


Abbildung 2.2: Bedingte Dichten bei perfektem Ratingsystem

- triviales Ratingsystem

Im Falle eines trivialen Ratingsystems besitzt das Ratingsystem keine Trennschärfe -keine Unterscheidung zwischen Defaultern und Non-Defaultern. Des-

halb sind die bedingten Dichten f_D und f_N der Score Variablen identisch. Die Kurve dieses Ratingsystems ist eine Gerade durch den Ursprung und $(1, 1)$. Anschaulich besitzt jeder Score den gleichen Anteil von Defaultern und Non-Defaultern. Die Score Variable ordnet die Debitoren zufällig den Score Werten zu - *Coin-Toss*.

- Ratingsystem in der Praxis

Ein reales Ratingsystem besitzt eine *CAP*-Kurve zwischen diesen beiden Extrema. Ein solches Ratingsystem besitzt Trennschärfe und liegt deshalb oberhalb der Kurve des trivialen Ratingmodells. Es kann aber nicht mit Sicherheit zwischen Defaulter und Non-Defaulter unterscheiden und ist somit unterhalb der Kurve des perfekten Ratingsystems. Auf die Konkavität der Kurve wird im weiteren Verlauf dieses Kapitels näher eingegangen.

2.1.3 Bemerkung

In der Praxis ist es unwahrscheinlich, dass ein Ratingsystem die *CAP*-Kurve eines perfekten Ratingsystems besitzt, da dies bedeutet, dass mit Sicherheit ein Defaulter vorhersehbar ist. Genauso unwahrscheinlich ist die Kurve eines trivialen Ratingsystems. Ein solches Ratingsystem ist nutzlos, da es nicht zwischen Defaultern und Non-Defaultern unterscheiden kann (Münze werfen ist günstiger).

Zur weiteren Untersuchung der *CAP*-Kurve -z.B. die Konkavität- wird die Ableitung der *CAP*-Funktion betrachtet. Zunächst wird die Ableitung der Pseudo-Inversen gebildet. Die Verteilungsfunktion F ist stetig und streng monoton auf dem Intervall $[0, 1]$. Weiter sei F in s differenzierbar mit $F'(s) \neq 0$. Dann ist F^{-1} in $F(s) = u$ differenzierbar und es gilt

$$F^{-1'}(u) = \frac{1}{F'(s)} = \frac{1}{F'(F^{-1}(u))} \quad (2.3)$$

Damit wird nun die Ableitung der *CAP*-Funktion bestimmt. Es gilt

$$CAP'(u) \stackrel{(2.3)}{=} \frac{f_D(F^{-1}(u))}{f(F^{-1}(u))}.$$

Mit der Bayes Formel (1.4) folgt

$$CAP'(u) = \frac{f_D(F^{-1}(u))}{f(F^{-1}(u))}$$

$$\begin{aligned}
 &\stackrel{(1.4)}{=} \frac{\frac{1}{p} P(X = 1 \mid S = F^{-1}(u)) \cdot f(F^{-1}(u))}{f(F^{-1}(u))} \\
 &= \frac{P(X = 1 \mid S = F^{-1}(u))}{p}
 \end{aligned} \tag{2.4}$$

Mit der Darstellung (2.4) folgt aus der Monotonie der bedingten Ausfallwahrscheinlichkeit die Konkavität der *CAP*-Funktion. Da die bedingte Ausfallwahrscheinlichkeit monoton fallend ist, fällt $CAP'(u)$ in u und somit ist die *CAP*-Funktion konkav.

Um nun eine Kennzahl zur Messung der Trennschärfe zu bekommen, wird die Fläche zwischen der Diagonalen und der *CAP*-Kurve betrachtet.

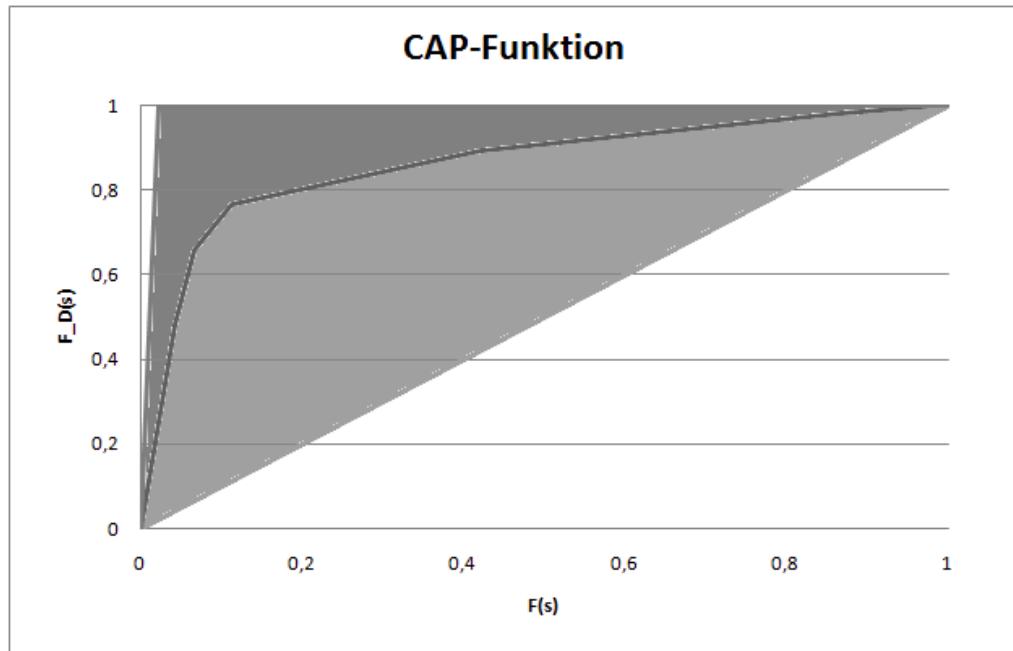


Abbildung 2.3: CAP - Kurve für Ratingsysteme

Die Fläche zwischen *CAP*-Kurve und der Diagonalen -hellgraue Fläche in Abbildung 2.3- ist gegeben durch

$$\int_0^1 CAP(u) du - \frac{1}{2}$$

Weiterhin berechnet sich die Fläche zwischen *CAP*-Kurve des perfekten Ratingsys-

tems und der Diagonalen -dunkelgraue Fläche in Abbildung 2.3- zu

$$\frac{1}{2} - \frac{p}{2}$$

Der Quotient aus diesen beiden Flächen bildet eine Kennzahl zur Bestimmung der Trennschärfe, die Accuracy Ratio.

2.1.4 Definition

Die Accuracy Ratio (AR) ist definiert als

$$\begin{aligned} AR &\stackrel{\text{def}}{=} \frac{\frac{1}{2} \int_0^1 CAP(u) du - \frac{1}{2}}{\frac{1}{2} - \frac{p}{2}} \\ &= \frac{2 \int_0^1 CAP(u) du - 1}{1 - p}. \end{aligned} \quad (2.5)$$

Die Kennzahl AR kann folgendermaßen anschaulich verstanden werden. Man wählt zwei zufällige Debitoren, einen aus der Grundgesamtheit der Defaulter, den anderen aus der Grundgesamtheit der Non-Defaulter. Die AR ergibt sich dann aus der Differenz der Wahrscheinlichkeit, dass der Score des Defaulters kleiner ist als der Score des Non-Defaulters und der Wahrscheinlichkeit, dass der Score des Defaulters größer ist als der Score des Non-Defaulters. Formal:

$$AR = P(S_D < S_N) - P(S_D > S_N), \quad (2.6)$$

wobei S_D und S_N unabhängig und F_D bzw. F_N verteilt sind. Der Beweis von (2.6) wird im nachfolgenden Kapitel geführt.

2.1.5 Bemerkung

Die Trennschärfe des Ratingsystems ist somit umso besser, je größer die Accuracy Ratio ist. Eine große AR bedeutet auch eine große Ausfallwahrscheinlichkeit für niedrige Score-Werte und umgekehrt.

2.2 Reciever Operating Characteristic

Ein weiteres graphisches Hilfsmittel zur Bestimmung der Trennschärfe ist die Reciever Operating Characteristic.

Zusätzlich zur Definition von Trefferquote und Alarmrate wird eine weitere Definition benötigt. Dabei handelt es sich um die Fehlalarmquote. Diese ist als die bedingte Wahrscheinlichkeit -der Score eines Non-Defaulters ist kleiner oder gleich einer bestimmten Score-Grenze s -

$$P(S \leq s \mid X = 0) = F_N(s)$$

definiert. Die Fehlalarmquote gibt den Anteil der Non-Defaulter unter einer Schranke s an. Damit ist es nun möglich, die *ROC*-Funktion zu definieren.

2.2.1 Definition

Die Funktion $ROC : [0, 1] \longrightarrow [0, 1]$ gegeben durch

$$ROC(u) \stackrel{\text{def}}{=} F_D(F_N^{-1}(u)), \quad u \in (0, 1) \quad (2.7)$$

heißt Reciever Operating Characteristic (*ROC*)-Funktion, wobei F_N^{-1} die Pseudo-Inverse von F_N sei.

2.2.2 Bemerkung

Die *ROC*-Funktion und *CAP*-Funktion unterscheiden sich darin, dass die *CAP*-Funktion durch $F(s)$ von der totalen Ausfallwahrscheinlichkeit p abhängt.

Nun wird auf die Charakterisierung der Kurve der *ROC*-Funktion eingegangen. Der Graph der *ROC*-Funktion kann durch Plotten von $(F_N(s), F_D(s))$, $s \in \mathbb{R}$ bestimmt werden.

Die in Abbildung 2.4 gezeigten Beispiele für *ROC*-Funktionen werden nun weiter beschrieben.

- perfektes Ratingsystem

Die *ROC*-Kurve ist eine Konstante ($ROC(u) \equiv 1$). Die Score Variable des perfekten Ratings ordnet allen Defaultern den „schlechtesten“ Score Wert zu,

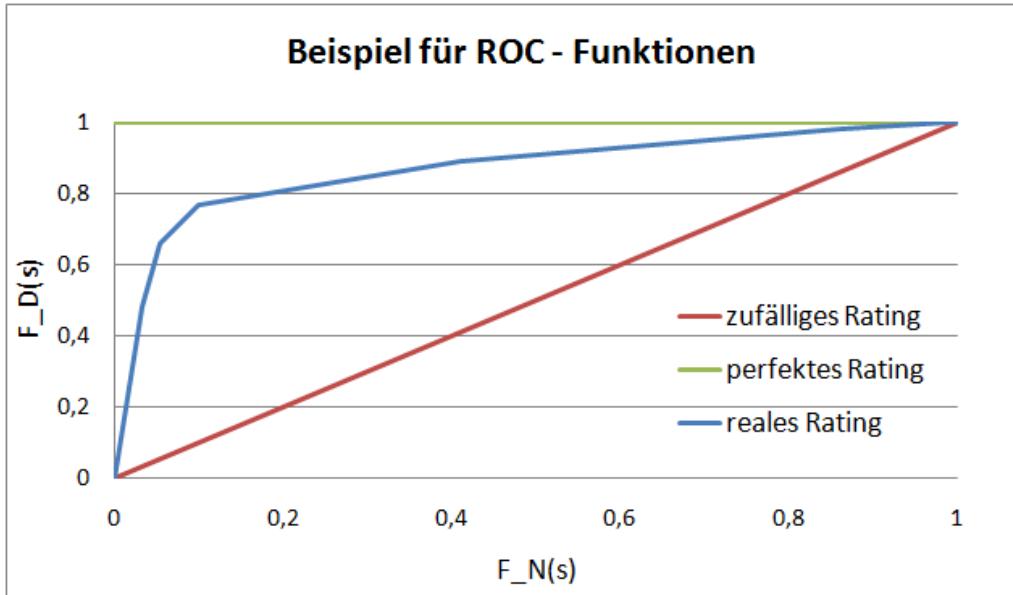


Abbildung 2.4: ROC - Kurve für Ratingsysteme

die Non-Defaulter werden auf alle weiteren Score Werte verteilt - keinem Non-Defaulter wird der „schlechteste“ Score Wert zugeordnet.

- triviales Ratingsystem

Der Graph eines trivialen Ratingsystems ist eine Gerade durch den Ursprung und den Punkt $(1, 1)$. Die Score Variable ordnet jedem Debitor einen zufälligen Score zu, die bedingten Dichten f_D und f_N sind deshalb identisch. ($\Rightarrow ROC(u) = u$)

- reales Ratingmodell

Der Graph eines realen Ratingsystems ist eine konkav Kurve zwischen den beiden Extrema - perfektes Rating und triviales Rating.

Zur weiteren Untersuchung der *ROC*-Kurve wird wieder die Ableitung der *ROC*-Funktion betrachtet. Die Ableitung ist gegeben durch

$$ROC'(u) = \frac{f_D(F_N^{-1}(u))}{f_N(F_N^{-1}(u))}, \quad u \in (0, 1). \quad (2.8)$$

Die Existenz der Ableitung der *ROC*-Funktion folgt mit derselben Argumentation wie bei der Ableitung der *CAP*-Funktion.

Der Quotient aus (2.8) entspricht dem Likelihood Quotienten, der bereits im Kapitel 1.3 -Monotonie- untersucht wurde. In Kapitel 1.3 wurde gezeigt, dass der Likelihood

Quotient genau dann monoton fallend ist, wenn die bedingte Ausfallwahrscheinlichkeit in s monoton fällt. Somit folgt aus der monoton fallenden Eigenschaft der bedingten Ausfallwahrscheinlichkeit die Konkavität der *ROC*-Kurve.

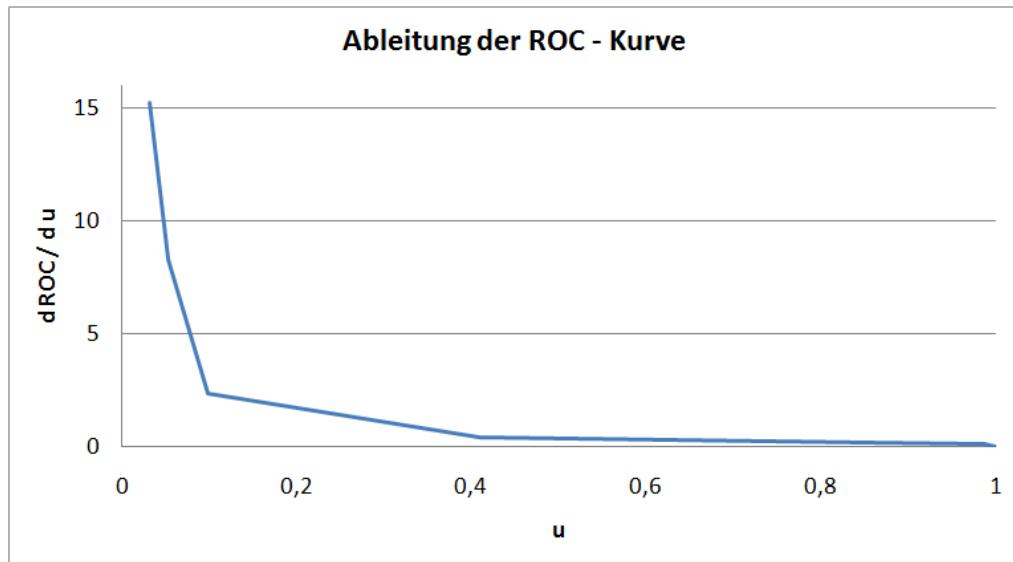


Abbildung 2.5: Ableitung der ROC - Kurve aus Abbildung 2.4

Für das *ROC*-Verfahren soll wie beim *CAP*-Verfahren eine einzelne Kennzahl zur Bestimmung der Trennschärfe definiert werden. Hierzu wird die Fläche unter der *ROC*-Kurve betrachtet.

2.2.3 Definition

Die Kennzahl *AUROC* (Area Under ROC) ist definiert als die Fläche unterhalb der *ROC*-Kurve

$$\begin{aligned} AUROC &\stackrel{\text{def}}{=} \int_0^1 ROC(u) \, du \\ &= \int_0^1 F_D(F_N^{-1}(s)) \, ds \end{aligned}$$

Die *AUROC* lässt sich auch als Wahrscheinlichkeit interpretieren und zwar so, dass der Score eines zufällig gewählten Non-Defaulters größer ist, als der Score eines zufällig gewählten ausgefallenen Debtors. Es gilt

$$AUROC = P(S_D < S_N), \quad (2.9)$$

wobei S_D und S_N unabhängig und F_D bzw. F_N verteilt sind.

Beweis:

Es sei \mathcal{S} die Menge aller möglichen Score Werte s . Dann gilt

$$\begin{aligned}
 AUROC &= \int_0^1 F_D(F_N^{-1}(s)) \, ds, \\
 &\stackrel{\text{Trafo}}{=} \int_{\mathcal{S}} F_D(s) \, dF_N(s) \\
 &= \int_{\mathcal{S}} F_D(S_N) \, dP \\
 &= E_P[P(S_D < S_N) \mid S_N] \\
 &\stackrel{\text{Eig. bed. EW}}{=} P(S_D < S_N) \tag{2.10}
 \end{aligned}$$

□

Die Maximierung der $AUROC$ dient somit der Maximierung der Trennschärfe des Ratingsystems, da mit größer werdender $AUROC$ die Wahrscheinlichkeit - $P(S_D < S_N)$ -, anhand des Score Wertes zwischen Defaultern und Non-Defaultern zu unterscheiden, steigt.

Weiterhin lassen sich die *Area Under ROC* und *Accuracy Ratio* über die Formel

$$AUROC = \frac{AR + 1}{2} \tag{2.11}$$

ineinander überführen.

Es gilt nämlich

$$\begin{aligned}
 AUROC &= \int_0^1 F_D(F_N^{-1}(s)) \, ds \\
 &\stackrel{\text{Trafo}}{=} \int_{\mathcal{S}} F_D(s) \, d(F_N(s)) \\
 &\stackrel{(2.1)}{=} \int_{\mathcal{S}} F_D(s) \, d\left(\frac{F(s) - pF_D(s)}{1 - p}\right) \\
 &= \int_{\mathcal{S}} \frac{F_D(s)}{1 - p} \, d(F(s)) - \int_{\mathcal{S}} \frac{pF_D(s)}{1 - p} \, d(F_D(s)) \\
 &= \frac{\int_{\mathcal{S}} F_D(s) \, d(F(s))}{1 - p} - \frac{p \int_{\mathcal{S}} F_D(s) \, d(F_D(s))}{1 - p} \\
 &\stackrel{\text{Trafo}}{=} \frac{\int_0^1 F_D(F^{-1}(u)) \, du}{1 - p} - \frac{p \int_0^1 F_D(F_D^{-1}(u)) \, du}{1 - p}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{\int_0^1 CAP(u) \, du}{1-p} - \frac{\frac{1}{2}p}{1-p} \\
&= \frac{1}{2} \left(\frac{2 \int_0^1 CAP(u) \, du - 1}{1-p} + \frac{1-p}{1-p} \right) \\
&\stackrel{(2.5)}{=} \frac{1}{2} (AR + 1)
\end{aligned}$$

Mit Hilfe von (2.11) ist es nun möglich, den ausstehenden Beweis von (2.6) zu führen.

Beweis:

$$\begin{aligned}
AR &\stackrel{(2.11)}{=} 2 \cdot AUROC - 1 \\
&\stackrel{(2.9)}{=} 2 \cdot P(S_D < S_N) - 1 \\
&= 2 \cdot P(S_D < S_N) - (P(S_D < S_N) + P(S_D > S_N)) \\
&= P(S_D < S_N) - P(S_D > S_N)
\end{aligned}$$

□

2.2.4 Bemerkung

(2.11) zeigt, dass die Accuracy Ratio -wie die Area under ROC- von den bedingten Dichten der Score Variablen S abhängt, aber nicht von der totalen Ausfallwahrscheinlichkeit p .

2.2.5 Bemerkung

Die ROC -Kurve kann auch als Diagramm des Fehlers 1. Art und Fehlers 2. Art interpretiert werden.

Testet man die Hypothesen

$$H_0 : X = 1 \quad \text{gegen} \quad H_1 : X = 0$$

entspricht $F_N(s)$ der Wahrscheinlichkeit des Fehlers 2. Art und $1 - F_D(s)$ der Wahr-

scheinlichkeit des Fehlers 1. Art, wobei die Testfunktion φ gegeben durch

$$\varphi = \begin{cases} 1, & \text{falls } S > s \\ 0, & \text{sonst} \end{cases}$$

zugrundeliegt. $\varphi = 0$ bedeutet das Annehmen der Hypothese H_0 .

2.3 Weitere Trennschärfemaße

In Kapitel 2.1 und 2.2 wurden die Gütemaße Cumulative Accuracy Profile und Receiver Operating Characteristic vorgestellt. Nun wird ein kurzer Überblick über weitere Trennschärfemaße gegeben. Zuerst wird der *Bayes Fehler* -wozu auch der Spezialfall *Klassifikationsfehler* zählt- vorgestellt. Daraufhin wird der *Pietra-Index* analysiert. Abschließend wird der -Entropie basierende- *bedingte Entropie Quotient* hergeleitet.

2.3.1 Bayes Fehler

Die *ROC*-Kurve kann als Diagramm der Fehlerarten bzgl. der verschiedenen Score Werte s gedeutet werden. Mit Hilfe der Fehleranalyse kann ein weiteres Trennschärfemaß entwickelt werden. Dazu wird der Gesamtfehler betrachtet, im Gegensatz zur separaten Analyse des Fehlers 1. Art bzw. 2. Art. Zunächst wird der Begriff des Klassifikators eingeführt.

2.3.1 Definition

Ein Klassifikator b ist eine Funktion, der Elemente aus einem Merkmalsraum -hier Score s - auf eine Menge von Mustern C abbildet.

$$b : \mathbb{R} \longrightarrow C$$

In diesem Fall wird $C = \{0, 1\}$ betrachtet, wobei $C = 0$ als Solvenz und $C = 1$ als Ausfall des Debitors gedeutet wird.

Der sogenannte Bayes Klassifikator ordnet die Merkmale anhand der a posteriori Wahrscheinlichkeit $-P(X = 1|S = s)$ - den Mustern zu. Das Merkmal wird dem Muster zugeordnet, für die die a posteriori Wahrscheinlichkeit am Größten ist (Bayes

Entscheidungsregel):

$$b(s) = 1 \iff P(X = 1|S = s) > P(X = 0|S = s) \quad (2.12)$$

$$b(s) = 0 \iff P(X = 0|S = s) > P(X = 1|S = s) \quad (2.13)$$

Aufgrund des Satzes von Bayes (1.4)

$$P(X = 1|S = s) = \frac{P(S = s|X = 1)P(X = 1)}{P(S = s|X = 1)P(X = 1) + P(S = s|X = 0)P(X = 0)}$$

kann die Bayes Entscheidungsregel -(2.12) und (2.13)- berechnet werden durch:

$$b(s) = 1 \iff P(S = s|X = 1)P(X = 1) > P(S = s|X = 0)P(X = 0)$$

$$b(s) = 0 \iff P(S = s|X = 0)P(X = 0) > P(S = s|X = 1)P(X = 1)$$

Weiter ist die Bayes Entscheidungsregel optimal bzgl. der Fehlerwahrscheinlichkeit. D.h. der Bayes Klassifikator minimiert die Fehlerwahrscheinlichkeit.

$$\begin{aligned} BF &\stackrel{\text{def}}{=} \min_{s \in S} \{P(\text{Fehler bei Score } s)\} \\ &= \min_{s \in S} \{P(X = 1)P(S > s | X = 1) + P(X = 0)P(S \leq s | X = 0)\} \\ &= \min_{s \in S} \{p(1 - F_D(s)) + (1 - p)F_N(s)\} \end{aligned} \quad (2.14)$$

Die Kennzahl BF wird Bayes Fehler genannt und ist ein Gütemaß für die Trennschärfe.

Der Bayes Fehler gibt die Fehlerwahrscheinlichkeit bei optimalem Einsatz des Ratingsystems wieder, also den Anteil der Debitoren, die bei optimaler Zuordnung falsch klassifiziert werden. Deshalb gilt, je kleiner der Bayes Fehler, desto höher ist die Trennschärfe des Ratingsystems.

2.3.2 Bemerkung

Bei einem perfekten Ratingsystem werden alle Debitoren richtig klassifiziert, der Bayes Fehler wäre somit gleich null und die Trennschärfe maximal.

Ein Spezialfall des Bayes Fehler ist der Klassifikationsfehler (KF). Bei diesem wird die totale Ausfallwahrscheinlichkeit p gleich $1/2$ gesetzt.

Für $p = \frac{1}{2}$ ist

$$\begin{aligned}
 KF = BF_{1/2} &= \min_{s \in S} \left\{ \frac{1}{2}(1 - F_D(s)) + \frac{1}{2}F_N(s) \right\} \\
 &= \frac{1}{2} \min_{s \in S} \{1 - F_D(s) + F_N(s)\} \\
 &= \frac{1}{2} (1 - \max_{s \in S} |F_D(s) - F_N(s)|) \in [0, 1]
 \end{aligned} \tag{2.15}$$

2.3.3 Bemerkung

Der Klassifikationsfehler ist nicht von der totalen Ausfallwahrscheinlichkeit p abhängig und eignet sich daher besser zum Vergleich von Ratingsystemen als die Bayes Fehler.

Mit der Schreibweise (2.15) des Klassifikationsfehlers lässt sich eine Beziehung zu einem weiteren Trennschärfemaß -Pietra-Index- aufweisen.

2.3.2 Pietra-Index

Mit Hilfe des Pietra-Index, der auch als Kolmogorov-Smirnoff-Teststatistik bekannt ist, wird getestet, ob die Verteilungsfunktionen F_D und F_N identisch sind.

2.3.4 Definition

Seien F_D und F_N die Verteilungsfunktionen der Defaulter bzw. Non-Defaulter, dann heißt

$$PI \stackrel{\text{def}}{=} \max_{s \in S} |F_D(s) - F_N(s)|$$

Pietra Index der Verteilungen F_D, F_N .

Getestet werden die Hypothesen

$$H_0 : F_D = F_N \quad , \quad H_1 : F_D \neq F_N.$$

Damit kann nun folgende Testfunktion definiert werden.

2.3.5 Bemerkung

Als Kolmogorov-Smirnoff Test wird folgender Test zum Niveau α verstanden.

$$\varphi = \begin{cases} 1, & \text{falls } PI > D_\alpha \\ 0, & \text{falls } PI \leq D_\alpha \end{cases}$$

Dabei ist D_α ein kritischer Wert, der mit Hilfe der Literatur¹ berechnet werden kann. Weiter bedeutet $\varphi = 1$ Ablehnung der Hypothese H_0 .

Wird die Hypothese angenommen ($\varphi = 0$), so sind die Verteilungsfunktionen identisch, d.h. $F_D \equiv F_N$. Die *ROC*-Kurve eines solchen Ratingsystems ist in Abbildung 2.6 gezeigt. Daraus folgt, dass die Score Variable und somit das Ratingsystem keinerlei Trennschärfe besitzt.

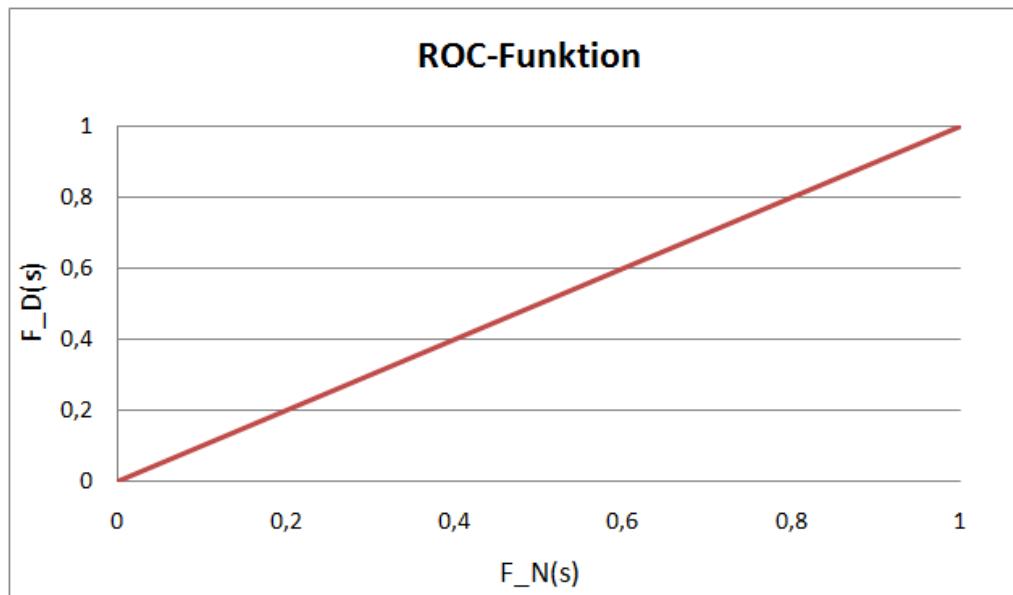


Abbildung 2.6: ROC - Kurve bei Identität der bedingten Vtlg.Fkt.

Dadurch besteht mit dem Pietra-Index ein weiteres Maß zur Messung der Trennschärfe.

¹ kritischer Wert ist der Tabelle der Kolmogorov-Smirnoff Verteilung entnehmbar, Anhang Tabelle 4.16

2.3.3 Bedingter Entropie Quotient

Im Folgenden werden die Entropie basierten Trennschärfemaße behandelt. Diese Trennschärfemaße betrachten den Informationsgewinn, der durch den Einsatz des Ratingsystems gewonnen wird. In diesem Zusammenhang wird Information als messbare Größe interpretiert, die dem Grad der Kenntnis über ein zukünftiges Ereignis entspricht.

Zunächst wird der Begriff der Entropie eingeführt.

2.3.6 Definition

Für ein dichotomes Ereignis -Ereignis tritt ein oder nicht- mit Eintrittswahrscheinlichkeit p wird durch

$$H(p) \stackrel{\text{def}}{=} -(p \log(p) + (1-p) \log(1-p)) \quad (2.16)$$

die sogenannte Entropie von p definiert.

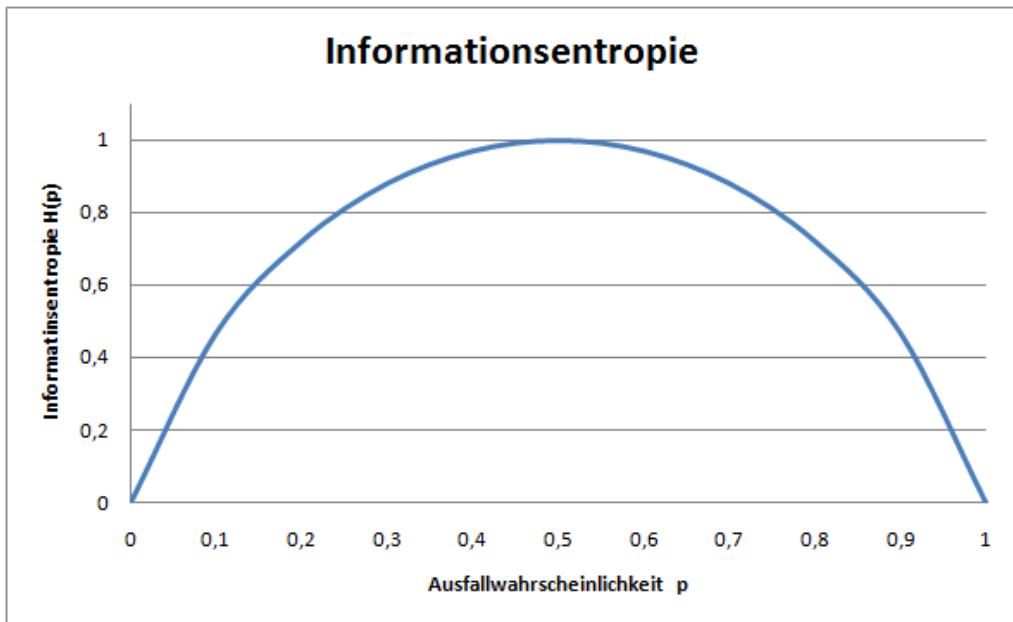


Abbildung 2.7: Informationsentropie als Funktion der Ausfallwahrscheinlichkeit

Anhand von Abbildung 2.7 kann deutlich gesehen werden, dass $H(p) \rightarrow 0$ strebt genau dann, wenn $p \rightarrow 0$ oder $p \rightarrow 1$. Somit ist die Entropie ein Maß der Unsicherheit

über den Ausgang eines Ereignisses.

Die Unsicherheit ist für eine Eintrittswahrscheinlichkeit von $p = 0,5$ am Größten (100 %).

Um nun die Trennschärfe eines Ratingsystems zu messen, wird die Entropie der bedingten Ausfallwahrscheinlichkeit betrachtet.

2.3.7 Definition

Sei $H(P(X = 1|S = s))$ die Entropie der bedingten Ausfallwahrscheinlichkeit, dann ist die über alle Scores gemittelte bedingte Entropie definiert als

$$H_S \stackrel{\text{def}}{=} E[H(P(X = 1|S = s))]. \quad (2.17)$$

Die gemittelte bedingte Entropie H_S gibt die verbliebene Unsicherheit nach Anwendung des Ratingsystems über den zukünftigen Ausfallstatus an. Falls der Wert der gemittelten bedingten Entropie H_S dicht bei null ist, deutet dieses auf eine hohe Trennschärfe des Ratingsystems hin.

Um eine vergleichbare Kennzahl des Entropiemaßes zu bekommen, wird die Entropie normiert.

$$\frac{H(p) - H_S}{H(p)} = 1 - \frac{H_S}{H(p)} \in [0, 1] \quad (2.18)$$

Die so gewonnene Kennzahl heißt bedingter Entropie Quotient (*CIER*). Der bedingte Entropie Quotient ist eine Kennzahl dafür, wie viele Informationen ein Ratingsystem trägt.

Falls bei der Anwendung des Ratingsystems keinerlei Informationen gewonnen werden, ist $CIER = 0$. Für ein ideales Ratingsystem ist $CIER = 1$. Somit lässt sich zusammenfassend sagen, je größer der Wert des *CIER* ist, desto größer ist die Trennschärfe des Ratingsystems.

Kapitel 3

Kalibrierung von Ratingsystemen

Von dem Begriff der Trennschärfe eines Ratingsystems ist der, der Kalibrierung eines Ratingsystems abzugrenzen. Während unter der Trennschärfe eines Ratingsystems die Fähigkeit *ex ante* einen Defaulter zu identifizieren verstanden wird, beschreibt die Kalibrierung die Zuweisung der Ausfallwahrscheinlichkeiten zu den einzelnen Ratingklassen. Somit kann es durchaus möglich sein, dass ein Ratingsystem trennscharf, aber nicht gut kalibriert ist. Der umgekehrte Fall -perfekt kalibriert, aber keinerlei Trennschärfe- ist ebenfalls möglich.

Ziel dieses Kapitels ist es nun, die Ausfallwahrscheinlichkeiten der verschiedenen Ratingklassen -Kapitel 1.2- zu überprüfen, d.h. ob diese richtig kalibriert sind. Zuerst wird ein Modell vorgestellt, mit dem eine solche Überprüfung ermöglicht wird. In einem zweiten Schritt wird auf die Kalibrierung eines trivialen Modells -der Unabhängigkeit der Ausfallereignisse- eingegangen. Anschließend wird die realitätsnahe Modellierung von abhängigen Ausfällen diskutiert.

3.1 Einfaktor-Modell

Sei n die Anzahl der Debitoren im Kunden-Portfeuille Ω . Diese unterteilen sich in k Ratingklassen mit jeweils n_r , $r = 1, \dots, k$ Debitoren. Die Zufallsvariable B_r beschreibe die Bonität des Debitors in der r -ten Ratingklasse und wird gegeben durch

$$B_r(i) = B_{ri} = \sqrt{v_r}Z + \sqrt{1-v_r}U_{ri} \quad r = 1, \dots, k \quad i = 1, \dots, n_r \quad (3.1)$$

Dabei sei die Zufallsvariable Z ein makroökonomischer Risikofaktor und die Zufallsvariable U_{ri} das firmenspezifische Risiko des i -ten Unternehmens. Z als auch U_{ri} seien $\mathcal{N}(0, 1)$ -verteilt. Der jeweilige Einfluss der Risikoparameter sei durch $v_r \in [0, 1]$ gesteuert. Des Weiteren seien für alle r, i Z und U_{ri} paarweise unabhängig und somit B_{ri} wiederum $\mathcal{N}(0, 1)$ -verteilt.

Zusätzlich sei eine Schranke c_r für die Bonität festgelegt. Falls ein Debitor unter diese Schranke fällt, fällt dieser aus. Somit werden die Ausfälle der r -ten Ratingklasse durch die Zufallsvariable

$$X_r(i) = X_{ri} = \begin{cases} 1 & \text{für } B_{ri} \leq c_r \\ 0 & \text{sonst} \end{cases}$$

beschrieben. Diese Schranke c_r wird eindeutig über die Ausfallwahrscheinlichkeit p_r der r -ten Ratingklasse festgelegt. Es gilt:

$$c_r = \Phi^{-1}(p_r), \quad (3.2)$$

denn für die Ausfallwahrscheinlichkeit der r -ten Ratingklasse gilt:

$$\begin{aligned} p_r &= P(X_{ri} = 1) = P(B_{ri} \leq c_r) = \Phi(c_r) \\ \iff c_r &= \Phi^{-1}(p_r) \end{aligned}$$

Somit folgt, dass $X_{ri} \sim \mathcal{B}(1, p_r)$. Die empirische Ausfallquote \bar{p}_r der r -ten Ratingklasse ist gegeben durch

$$\bar{p}_r \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n_r} X_{ri}.$$

Nun lässt sich folgendes beobachten.

1. Die Korrelation zweier Bonitätsvariablen B_{ri} und B_{sj} , $r \neq s \vee i \neq j$ ist gegeben durch

$$v_{rs} \stackrel{\text{def}}{=} \text{Cor}[B_{ri}, B_{sj}] = \sqrt{v_r v_s}.$$

Sowie die Korrelation innerhalb einer Ratingklasse durch

$$v_{rr} = \sqrt{v_r^2} = v_r,$$

denn mit (i) $\text{Var}[B_{ri}] = \text{Var}[B_{sj}] = 1$ und (ii) der Unabhängigkeit der U_{ri}, U_{sj}

und Z folgt:

$$\begin{aligned}
 v_{rs} &\stackrel{(i)}{=} \text{Cov}[B_{ri}, B_{sj}] \\
 &= E[B_{ri}B_{sj}] - \underbrace{E[B_{ri}]}_{=0} \cdot \underbrace{E[B_{sj}]}_{=0} \\
 &= E[(\sqrt{v_r}Z + \sqrt{1-v_r}U_{ri})(\sqrt{v_s}Z + \sqrt{1-v_s}U_{sj})] \\
 &\stackrel{(ii)}{=} \sqrt{v_r v_s} \underbrace{E[Z^2]}_{=1} + \sqrt{1-v_r} \sqrt{v_s} \underbrace{E[U_{ri}Z]}_{=0} + \\
 &\quad \sqrt{1-v_s} \sqrt{1-v_r} \underbrace{E[U_{ri}U_{sj}]}_{=0} + \sqrt{v_r} \sqrt{1-v_s} \underbrace{E[U_{sj}Z]}_{=0} \\
 &= \sqrt{v_r v_s}
 \end{aligned}$$

2. Die Korrelation zweier Ausfallvariablen X_{ri} und X_{sj} , $r \neq s \vee i \neq j$

$$v_{rs}^{Aus} \stackrel{\text{def}}{=} \text{Cor}(X_{ri}, X_{sj}) = \frac{\Phi_2(\Phi^{-1}(p_r), \Phi^{-1}(p_s); v_{rs}) - p_r p_s}{\sqrt{p_r(1-p_r)p_s(1-p_s)}},$$

wobei mit $\Phi_2(x, y, v) \stackrel{\text{def}}{=} P\{X \leq x, Y \leq y\} = P((X, Y) \in (-\infty, x] \times (-\infty, y])$ die bivariate Normalverteilung mit Korrelation v gemeint ist.

Denn mit (i) $B_{ri}, B_{sj} \sim \mathcal{N}(0, 1)$ und (ii) $X_{ri}, X_{sj} \sim \mathcal{B}(1, p_r)$ folgt:

$$\begin{aligned}
 v_{rs}^{Aus} &= \frac{E[(X_{ri} - \overbrace{E[X_{ri}]}^{=p_r})(X_{sj} - \overbrace{E[X_{sj}]}^{=p_s})]}{\sqrt{V[X_{ri}]V[X_{sj}]}} \\
 &\stackrel{(ii)}{=} \frac{E[X_{ri}X_{sj}] - p_r \overbrace{E[X_{sj}]}^{=p_s} - p_s \overbrace{E[X_{ri}]}^{=p_r} + p_r p_s}{\sqrt{p_r(1-p_r)p_s(1-p_s)}} \\
 &= \frac{E[X_{ri}X_{sj}] - p_r p_s}{\sqrt{p_r(1-p_r)p_s(1-p_s)}} \\
 &= \frac{P(B_{ri} \leq c_r, B_{sj} \leq c_s) - p_r p_s}{\sqrt{p_r(1-p_r)p_s(1-p_s)}} \\
 &\stackrel{(3.2)}{=} \frac{P((B_{ri}, B_{sj}) \in (-\infty, \Phi^{-1}(p_r)] \times (-\infty, \Phi^{-1}(p_s)])) - p_r p_s}{\sqrt{p_r(1-p_r)p_s(1-p_s)}} \\
 &\stackrel{(i)}{=} \frac{\Phi_2(\Phi^{-1}(p_r), \Phi^{-1}(p_s), v_{rs})}{\sqrt{p_r(1-p_r)p_s(1-p_s)}}
 \end{aligned}$$

3. Es gilt $v_{rs}^{Aus} = 0 \iff v_{rs} = 0$,

denn

$$\begin{aligned}
 v_{rs} &= 0 \\
 \iff B_{ri} \wedge B_{sj} \text{ unabhängig } \forall i \neq j \\
 \iff P(B_{ri} \leq \Phi^{-1}(p_r), B_{sj} \leq \Phi^{-1}(p_s)) &= \underbrace{P(B_{ri} \leq \Phi^{-1}(p_r))}_{= p_r} \underbrace{P(B_{sj} \leq \Phi^{-1}(p_s))}_{= p_s} \\
 \iff \underbrace{P(B_{ri} \leq \Phi^{-1}(p_r), B_{sj} \leq \Phi^{-1}(p_s)) - p_r p_s}_{= \Phi_2(\Phi^{-1}(p_r), \Phi^{-1}(p_s), v_{rs})} &= 0 \\
 \iff v_{rs}^{Aus} &= 0
 \end{aligned}$$

3.2 Verteilung einer Ausfallquote

Ziel ist es nun, eine Verteilung zu finden, gegen die die empirische Ausfallquote konvergiert.

Zunächst soll der triviale Fall von unabhängigen Ausfallereignissen untersucht werden.

A_r sei die Zufallsvariable, die die Anzahl der Ausfälle in der r -ten Ratingklasse beschreibt:

$$A_r = \sum_{i=1}^{n_r} X_{ri}, \quad (3.3)$$

wobei X_{ri} i.i.d. und $X_{ri} \sim \mathcal{B}(1, p_r)$ wie im vorherigen Abschnitt ist. Also ist A_r $\mathcal{B}(n_r, p_r)$ -verteilt. Weiterhin ist nach dem starken Gesetz der großen Zahlen,

$$\bar{p}_r = \frac{1}{n_r} A_r = \frac{1}{n_r} \sum_{i=1}^{n_r} X_{ri} \xrightarrow{n_r \rightarrow \infty} E[X_{ri}] = p_r \quad P-f.s.$$

Die Unabhängigkeit der Ausfallereignisse ist für die Praxis keine realistische Modellannahme. Daher wird nun der Fall der abhängigen Ereignisse analysiert. Da alle vorherigen Ergebnisse auf der Annahme der Unabhängigkeit beruhen, konvergiert die Ausfallquote im Fall der Abhängigkeit nicht gegen die Ausfallwahrscheinlichkeit. Viel mehr konvergiert die Ausfallquote in Verteilung gegen eine nicht deterministische Zufallsvariable, denn für die Varianz dieser Ausfallquote gilt

$$Var[\bar{p}_r] \xrightarrow{n_r \rightarrow \infty} v_{rr}^{Aus} p_r (1 - p_r).$$

Beweis:

$$\begin{aligned}
 \text{Var}[\bar{p}_r] &= \text{Var}\left[\frac{1}{n_r} \sum_{i=1}^{n_r} X_{ri}\right] = \frac{1}{n_r} \left(\sum_{i=1}^{n_r} \text{Var}[X_{ri}] + \sum_{j \neq k}^{n_r} \text{Cov}[X_{rj}, X_{rk}] \right) \\
 &= \frac{n_r p_r (1 - p_r)}{n_r^2} + \underbrace{\frac{1}{n_r^2} \sum_{j \neq k}^{n_r} v_{rr}^{\text{Aus}} p_r (1 - p_r)}_{(n_r^2 - n_r) v_{rr}^{\text{Aus}} p_r (1 - p_r)} \\
 &= \frac{n_r p_r (1 - p_r)}{n_r^2} + \frac{n_r - 1}{n_r} v_{rr}^{\text{Aus}} p_r (1 - p_r)
 \end{aligned}$$

Für abhängige Ausfallereignisse ist $v_r > 0$. Daraus folgt mit (1.), dass $v_{rr}^{\text{Aus}} > 0$ und somit auch $\lim_{n_r \rightarrow \infty} \text{Var}[\bar{p}_r] > 0$. Damit konvergiert die Varianz nicht gegen null und somit konvergiert die Ausfallquote nicht gegen einen konstanten Wert.

□

Die asymptotische Verteilung der Ausfallquote \bar{p}_r für $n_r \rightarrow \infty$ wird gegeben durch

$$\bar{p}_r \xrightarrow{V} g_r(Z) = \Phi\left(\frac{\Phi^{-1}(p_r) - \sqrt{v_r}Z}{\sqrt{1 - v_r}}\right) \quad (3.4)$$

Zum Nachweis von (3.4) wird auf [12] Seite 6 verwiesen.

Für $v_r > 0$ hat die Zufallsvariable $g_r(Z)$ die Verteilungsfunktion

$$F_r(x) \stackrel{\text{def}}{=} P(g_r(Z) \leq x) = \Phi\left(\frac{\Phi^{-1}(x)\sqrt{1 - v_r} - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right), \quad (3.5)$$

Diese wird auch als Probit-Verteilung bezeichnet.

Beweis:

$$\begin{aligned}
 P(g_r(Z) \leq x) &\stackrel{(3.4)}{=} P\left(\Phi\left(\frac{\Phi^{-1}(p_r) - \sqrt{v_r}Z}{\sqrt{1 - v_r}}\right) \leq x\right) \\
 &= P\left(-Z \leq \frac{\sqrt{1 - v_r}\Phi^{-1}(x) - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right) \\
 &= \Phi\left(\frac{\Phi^{-1}(x)\sqrt{1 - v_r} - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right)
 \end{aligned}$$

□

3.3 Test einer Ausfallwahrscheinlichkeit

Nun soll ein statistischer Test hergeleitet werden, der es ermöglicht, die Ausfallwahrscheinlichkeit einer Ratingklasse zu testen. Im ersten Teil wird kurz auf den trivialen Fall -unabhängige Ausfallereignisse- eingegangen. Daraufhin wird der realitätsnahe Fall von abhängigen Ereignissen untersucht. Zunächst werden einige Bezeichnungen eingeführt.

Mit p_r , $r = 1, \dots, k$, ist die tatsächliche Ausfallwahrscheinlichkeit der r -ten Ratingklasse gemeint. Unter π_r mit $r = 1, \dots, k$ wird die prognostizierte Ausfallwahrscheinlichkeit der r -ten Ratingklasse verstanden. \bar{p}_r mit $r = 1, \dots, k$ bezeichnet -wie im vorherigen Abschnitt- die empirische Ausfallquote.

Getestet wird, ob die tatsächliche Ausfallwahrscheinlichkeit p_r in $H_0 \subset \Theta$ ¹ (Hypothese) oder in H_1 (Alternative) liegt. Unterstützend ist hierbei die Teststatistik T mit dem Wertebereich \mathcal{X} . Der Wertebereich wird disjunkt in die Teilmenge A -Akzeptanzbereich- und die Teilmenge K -kritischer Bereich- aufgeteilt. Dann wird durch $\varphi : \mathcal{X} \rightarrow [0, 1]$ gegeben durch

$$\varphi = \begin{cases} 1, & \text{falls } T \in K \\ 0, & \text{Falls } T \in A \end{cases} \quad (3.6)$$

ein Test definiert, wobei $\varphi(x) = 1$ die Ablehnung der Hypothese bedeutet. Erfüllt die Gütfunktion² von φ

$$\beta_\varphi(\theta) \leq \alpha \quad \text{mit } \alpha \in [0, 1] \quad (3.7)$$

für alle $\theta \in H_0$, dann heißt φ Test zum Niveau α für H_0 gegen H_1 . Genügt die Gütfunktion zusätzlich

$$\beta_\varphi(\theta) \geq \alpha \quad \forall \theta \in H_1, \quad (3.8)$$

so wird von einem unverfälschten Test zum Niveau α gesprochen. Mit $P(T \in K | H_0)$ ist im Folgenden $\beta_\varphi(\theta)$ gemeint, falls $\theta \in H_0$ beliebig ist.

Unabhängige Ausfallereignisse

Im Fall von unabhängigen Ausfallereignissen wird die aus (3.3) bekannte Zufallsvariable für die Anzahl der Ausfälle in der r -ten Ratingklasse als geeignete Teststatistik T_r gewählt.

¹Der Parameterraum Θ ist die Menge der zu testenden Parameter, hier $\Theta \in [0, 1]$

²Die Gütfunktion von φ ist gegeben durch $\beta_\varphi : \theta \mapsto E_\theta \varphi$ mit $\theta \in \Theta$

Die Hypothese wird so gewählt, dass die tatsächliche Ausfallwahrscheinlichkeit p_r kleiner gleich der geschätzten Ausfallwahrscheinlichkeit π_r ist. Die Alternative lautet, dass die geschätzte Wahrscheinlichkeit zu niedrig geschätzt wurde, d.h.

$$H_0 = [0, \pi_r] \quad \text{und} \quad H_1 = (\pi_r, 1]$$

Sei α die vorgegebene Wahrscheinlichkeit des Fehlers 1. Art. Dann muss für den kritischen Bereich K_α gelten:

$$P(\bar{p}_r \in K_\alpha \mid H_0) = \alpha \quad (3.9)$$

Die Hypothese H_0 wird abgelehnt, falls die Ausfallquote \bar{p}_r in dem kritischen Bereich K_α liegt.

Der kritische Bereich für einen Test zum Niveau α ergibt sich mit Hilfe von (3.9) zu

$$K_\alpha \stackrel{\text{def}}{=} \left(\arg \max_{i=1, \dots, n_r} \left\{ \sum_{j=0}^i \mathcal{B}(n_r, p_r)(j) = 1 - \alpha \right\}, n_r \right].$$

Beweis:

Der kritische Bereich K_α ist nach oben beschränkt durch die Anzahl der Debitoren n_r der r -ten Ratingklasse. Für die untere Schranke von K_α gilt:

$$\begin{aligned} P(\bar{p}_r > \kappa_\alpha \mid H_0) = \alpha &\iff 1 - P(\bar{p}_r \leq \kappa_\alpha \mid H_0) = \alpha \\ &\iff F(\kappa_\alpha) = 1 - \alpha \\ &\iff \sum_{j=1}^{\kappa_\alpha} \mathcal{B}(n_r, p_r)(j) = 1 - \alpha \\ &\iff \kappa_\alpha = \arg \max_{i=1, \dots, n_r} \left\{ \sum_{j=0}^i \mathcal{B}(n_r, p_r)(j) = 1 - \alpha \right\} \end{aligned}$$

□

Der durch die Testfunktion φ mit

$$\varphi = \begin{cases} 1, & \text{falls } T \in K_\alpha \\ 0, & \text{Falls } T \in A \end{cases}$$

definierte Test zum Niveau α heißt Binomialtest für unabhängige Ausfallereignisse.

Abhängige Ausfallereignisse

Nun soll der Fall der abhängigen Ausfälle untersucht werden.

Es wird ein Test aufgestellt, der auf der Teststatistik \bar{p}_r basiert. Die Hypothesen aus dem unabhängigen Fall werden beibehalten.

$$H_0 = [0, \pi_r] \quad \text{und} \quad H_1 = (\pi_r, 1]$$

Aufgrund der asymptotischen Verteilung von \bar{p}_r im abhängigen Fall, wird die Anforderung an den kritischen Bereich \hat{K}_α umformuliert:

$$\lim_{n_r \rightarrow \infty} P(\bar{p}_r \in \hat{K}_\alpha \mid H_0) = \alpha.$$

Da $[0, \pi_r]$ gegen $(\pi_r, 1]$ getestet wird, führt eine geringe Ausfallquote \bar{p}_r zu einer Akzeptanz von H_0 . Deshalb genügt es, eine untere Schranke κ_α für den kritischen Bereich \hat{K}_α zu bestimmen. Es muss gelten:

$$\begin{aligned} \lim_{n_r \rightarrow \infty} P(\bar{p}_r > \kappa_\alpha \mid H_0) = \alpha &\iff 1 - \overbrace{\lim_{n_r \rightarrow \infty} P(\bar{p}_r \leq \kappa_\alpha \mid H_0)}^{= F_r(\kappa_\alpha)} = \alpha \\ &\iff \Phi\left(\frac{\sqrt{1 - v_r} \Phi^{-1}(\kappa_\alpha) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}}\right) = 1 - \alpha \\ &\iff \kappa_\alpha = \Phi\left(\frac{\sqrt{v_r} \Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r)}{\sqrt{1 - v_r}}\right) \end{aligned}$$

Also ist der kritische Bereich $\hat{K}_\alpha \stackrel{\text{def}}{=} (\kappa_\alpha, \infty)$.

Für die Berechnung von κ_α ist es nützlich, auf Quantiltabellen zurückzugreifen. Um dieses zu bekommen, wird eine passende Teststatistik gesucht. Hierzu wird

$$T_r \stackrel{\text{def}}{=} \frac{\sqrt{1 - v_r} \Phi^{-1}(\bar{p}_r) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}}. \quad (3.10)$$

als Teststatistik gewählt. Unter der Hypothese H_0 , also $p_r = \pi_r$, ist T_r asymptotisch $\mathcal{N}(0, 1)$ -verteilt.

Beweis:

Es gilt:

$$\begin{aligned}
 \lim_{n_r \rightarrow \infty} P(T_r \leq x) &= \lim_{n_r \rightarrow \infty} P\left(\frac{\sqrt{1-v_r}\Phi^{-1}(\bar{p}_r) - \Phi^{-1}(p_r)}{\sqrt{v_r}} \leq x\right) \\
 &= \lim_{n_r \rightarrow \infty} P\left(\bar{p}_r \leq \Phi\left(\frac{\sqrt{v_r}x + \Phi^{-1}(p_r)}{\sqrt{1-v_r}}\right)\right) \\
 &\stackrel{(3.4)}{=} P\left(\Phi\left(\frac{\Phi^{-1}(p_r) - \sqrt{v_r}Z}{\sqrt{1-v_r}}\right) \leq \Phi\left(\frac{\sqrt{v_r}x + \Phi^{-1}(p_r)}{\sqrt{1-v_r}}\right)\right) \\
 &= P(-Z \leq x) \stackrel{Z \sim \mathcal{N}(0,1)}{=} \Phi(x)
 \end{aligned}$$

□

Weiterhin gilt unter der Hypothese H_0 offensichtlich

$$\lim_{n_r \rightarrow \infty} P(T_r > \Phi^{-1}(1-\alpha)) = \alpha.$$

Somit definiert die Teststatistik T_r mit dem Ablehnbereich

$$\hat{K}_\alpha \stackrel{\text{def}}{=} (\Phi^{-1}(1-\alpha), \infty) \quad (3.11)$$

einen Test zum Niveau α .

Damit wurden nun zwei Tests vorgestellt. Der erste Test beruht auf der Statistik \bar{p}_r und ist verständlich, da die Hypothese direkt an der Ausfallquote getestet wird. Beim anderen Test wird die Teststatistik T_r verwendet, um die Berechnung des Quantils zu erleichtern. Diese beiden Tests sind jedoch äquivalent. Es gilt

$$\bar{p}_r \in \hat{K}_\alpha \iff T_r \in \tilde{K}_\alpha$$

Beweis:

$$\begin{aligned}
 T_r \in \tilde{K}_\alpha &\iff T_r > \Phi^{-1}(1 - \alpha) \\
 &\iff \frac{\sqrt{1 - v_r} \Phi^{-1}(\bar{p}_r) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}} > \Phi^{-1}(1 - \alpha) \\
 &\iff \bar{p}_r > \underbrace{\Phi\left(\frac{\Phi^{-1}(1 - \alpha)\sqrt{v_r} + \Phi^{-1}(\pi_r)}{\sqrt{1 - v_r}}\right)}_{= \kappa_\alpha} \\
 &\iff \bar{p}_r \in \hat{K}_\alpha
 \end{aligned}$$

□

Bisher galt das Interesse dem Fehler 1. Art. Im Folgenden wird der Fehler 2. Art intensiver analysiert. Eine erste Schranke für den Fehler 2. Art folgt aus der Unverfälschtheit des Tests.

Die Gütefunktion $\beta_\varphi(p_r)$ ist gegeben durch

$$\beta_\varphi(p_r) = 1 - \Phi\left(\frac{\sqrt{v_r}\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r) - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right) \quad (3.12)$$

Beweis:

$$\begin{aligned}
 \beta_\varphi(p_r) &= \lim_{n_r \rightarrow \infty} P(T_r \in \tilde{K}_\alpha \mid p_r) \\
 &= \lim_{n_r \rightarrow \infty} P\left(\frac{\sqrt{1 - v_r}\Phi^{-1}(\bar{p}_r) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}} > \Phi^{-1}(1 - \alpha) \mid p_r\right) \\
 &= \lim_{n_r \rightarrow \infty} P\left(\bar{p}_r > \Phi\left(\frac{\sqrt{v_r}\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r)}{\sqrt{1 - v_r}}\right) \mid p_r\right) \\
 &\stackrel{(3.4)}{=} 1 - P\left(g_r(Z) \leq \Phi\left(\frac{\sqrt{v_r}\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r)}{\sqrt{1 - v_r}}\right) \mid p_r\right) \\
 &\stackrel{(3.5)}{=} 1 - \Phi\left(\frac{\sqrt{1 - v_r}\Phi^{-1}\left(\Phi\left(\frac{\sqrt{v_r}\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r)}{\sqrt{1 - v_r}}\right)\right) - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right) \\
 &= 1 - \Phi\left(\frac{\sqrt{v_r}\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r) - \Phi^{-1}(p_r)}{\sqrt{v_r}}\right)
 \end{aligned}$$

□

Die Unverfälschtheit des Tests zum Niveau α folgt aus (3.8)

$$\beta_\varphi(p_r) \geq \alpha \quad \forall p_r \in H_1.$$

Beweis:

Sei die Alternative H_1 gültig.

$$\beta_\varphi(p_r) = 1 - \Phi \left(\frac{\sqrt{v_r} \Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r) - \Phi^{-1}(p_r)}{\sqrt{v_r}} \right)$$

Da $p_r \leq \pi_r$ und Φ^{-1} eine streng monotone Funktion ist, folgt:

$$\begin{aligned} \beta_\varphi(p_r) &\leq 1 - \Phi \left(\frac{\sqrt{v_r} \Phi^{-1}(1 - \alpha) + \Phi^{-1}(\pi_r) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}} \right) \\ &= 1 - (1 - \alpha) = \alpha \end{aligned}$$

□

Der Fehler 2. Art (3.13) ist durch $1 - \alpha$ beschränkt. Diese obere Schranke ist jedoch sehr groß, sodass mit großer Wahrscheinlichkeit die Hypothese H_0 angenommen wird, obwohl H_1 gilt. Deshalb wird ein adäquateres Verfahren gesucht, das den Fehler 2. Art weiter einschränkt.

Zunächst wird der Modellfehler $p_r - \pi_r$ untersucht. Bei größer werdendem Modellfehler, wird die Differenz $\Phi^{-1}(\pi_r) - \Phi^{-1}(p_r)$ aus (3.12) kleiner, da die Funktion Φ^{-1} streng monoton steigend ist. Deshalb wächst die Gütfunktion (3.12) bei wachsendem Modellfehler. Der Fehler 2. Art, der durch die Gütfunktion definiert ist, wird kleiner.

$$P(\text{Fehler 2. Art}) = 1 - \beta_\varphi(p_r) \quad (3.13)$$

Dadurch erscheint es sinnvoll, den Test so zu konstruieren, dass die Wahrscheinlichkeit für den Fehler 1. Art $\leq \alpha$ ist und der Modellfehler $p_r - \pi_r \geq$ einer Schranke c mit der Wahrscheinlichkeit $1 - \beta$ erkannt wird.

Es wird der Test basierend auf der Teststatistik T_r gewählt. Der kritische Bereich $K_{\alpha,\beta}$ muss den neuen Anforderungen angepasst werden. Für $p_r - \pi_r \geq c$, soll T_r mit einer Wahrscheinlichkeit von $1 - \beta$ in $K_{\alpha,\beta}$ liegen, wobei für den Grenzfall $p_r - \pi_r = c$

Gleichheit gelten soll. Somit gilt:

$$\begin{aligned} \lim_{n_r \rightarrow \infty} P(T_r \in K_{\alpha,\beta} \mid p_r - \pi_r = c) &= 1 - \beta \\ \iff \lim_{n_r \rightarrow \infty} P(T_r > \iota_\beta \mid p_r - \pi_r = c) &= 1 - \beta \end{aligned} \quad (3.14)$$

Daraus folgt:

$$\lim_{n_r \rightarrow \infty} P(T_r > \iota_\beta \mid p_r - \pi_r \geq c) \geq 1 - \beta$$

Beweis:

Es gilt $p_r \geq \pi_r + c$.

$$\begin{aligned} \lim_{n_r \rightarrow \infty} P(T_r > \iota_\beta \mid p_r) &= 1 - \Phi \left(\frac{\sqrt{v_r} \Phi^{-1}(\iota_\beta) + \Phi^{-1}(\pi_r) - \Phi^{-1}(p_r)}{\sqrt{v_r}} \right) \\ &= 1 - \Phi \left(\frac{\sqrt{v_r} \Phi^{-1}(\iota_\beta) + \Phi^{-1}(\pi_r) - \Phi^{-1}(\pi_r + c)}{\sqrt{v_r}} \right) \\ &\stackrel{(3.14)}{=} 1 - \beta \end{aligned}$$

□

Die untere Grenze ι_β des kritischen Bereiches $K_{\alpha,\beta}$ wird mit Hilfe von (3.14) berechnet zu

$$\iota_\beta = \frac{\sqrt{v_r} \Phi^{-1}(\pi_r + c) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}}$$

Damit erfüllt der kritische Bereich

$$K_{\alpha,\beta} \stackrel{\text{def}}{=} (\iota_\beta, \infty)$$

die Anforderung an den Fehler 2. Art. Jedoch ist mit dem Ablehngebereich $K_{\alpha,\beta}$ die Begrenzung des Fehlers 1. Art durch α nicht mehr gewährleistet, da im Allgemeinen der Ablehngebereich \tilde{K}_α aus (3.11) nicht mit $K_{\alpha,\beta}$ übereinstimmt. Damit den Bedingungen beider Fehlerarten gerecht wird, wird der Entscheidungsbereich - Wertebereich der Teststatistik- in drei Zonen unterteilt. Die erste Zone -grüne Zone- bedeutet eine sofortige Annahme der Hypothese. Liegt die Teststatistik in der zweiten -gelben Zone- so kann keine genaue Entscheidung getroffen werden, da nur einer der beiden Fehlerarten die Anforderung erfüllt. Bei der letzten Zone -rote Zone- wird die Hypothese eindeutig abgewiesen. Um die Zonen näher zu bestimmen, werden im Folgenden drei Fälle unterschieden.

1. Es gilt $\Phi^{-1}(1 - \alpha) > \iota_\beta$.

Dann wird der Akzeptanzbereich definiert als

$$A_{\alpha,\beta}^1 \stackrel{\text{def}}{=} [0, \iota_\beta].$$

Gilt $T_r \in A_{\alpha,\beta}^1$, wird die Hypothese H_0 ohne Weiteres angenommen. $A_{\alpha,\beta}^1$ entspricht der grünen Zone. Befindet sich die Teststatistik T_r im Intervall $(\iota_\beta, \Phi^{-1}(1 - \alpha)]$, so kann keine genaue Entscheidung getroffen werden. Die Anforderung an den Fehler 1. Art wird eingehalten, jedoch wird die Beschränkung der Wahrscheinlichkeit des Fehlers 2. Art nicht erreicht. Das Intervall $(\iota_\beta, \Phi^{-1}(1 - \alpha)]$ ist die gelbe Zone. Liegt die Teststatistik T_r allerdings im kritischen Bereich $K_{\alpha,\beta}^1 \stackrel{\text{def}}{=} (\Phi^{-1}(1 - \alpha), \infty)$, so wird die Hypothese H_0 ohne Weiteres verworfen, rote Zone.

2. Es gilt $\Phi^{-1}(1 - \alpha) < \iota_\beta$.

In diesem Fall ergibt sich der Akzeptanzbereich -grüne Zone- $A_{\alpha,\beta}^2$ zu

$$A_{\alpha,\beta}^2 \stackrel{\text{def}}{=} [0, \Phi^{-1}(1 - \alpha)].$$

Fällt die Teststatistik T_r in das Intervall $(\Phi^{-1}(1 - \alpha), \iota_\beta]$, so lässt sich erneut keine genaue Entscheidung treffen. Deshalb wird dieses Intervall der gelben Zone zugeordnet. Der Fehler 2. Art ist beschränkt, allerdings besitzen die Anforderungen an den Fehler 1. Art keine Gültigkeit mehr. Die rote Zone entspricht dem kritischen Bereich $K_{\alpha,\beta}^2 \stackrel{\text{def}}{=} (\iota_\beta, \infty)$, der eine eindeutige Ablehnung der Hypothese zulässt.

3. Es gilt $\Phi^{-1}(1 - \alpha) = \iota_\beta$.

Dieser Fall stellt den Idealfall dar. Der Wertebereich \mathcal{X} der Teststatistik T_r wird disjunkt in den Akzeptanzbereich (grüne Zone)

$$A_{\alpha,\beta}^3 \stackrel{\text{def}}{=} [0, \iota_\beta]$$

sowie den kritischen Bereich (rote Zone)

$$K_{\alpha,\beta}^3 \stackrel{\text{def}}{=} (\iota_\beta, \infty)$$

aufgeteilt. Es existiert keine gelbe Zone. Somit lässt sich eindeutig eine Entscheidung für das Akzeptieren bzw. Verwerfen der Hypothese treffen.

Da sich bei gegebenem Fehler 1. Art bzw. Fehler 2. Art die Bereichsgrenzen durch die Wahl der Modellfehlerschranke c verändern lassen, scheint es sinnvoll, die Schranke c so zu wählen, dass die Bereiche denen im Idealfall (3.Fall) entsprechen. Dadurch erhält man einen optimalen Test sowie eindeutige Entscheidungsregeln. Die Schranke wird berechnet zu

$$c = \Phi(\sqrt{v_r}(\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\beta)) + \Phi^{-1}(\pi_r)) - \pi_r \quad (3.15)$$

Beweis:

Es gilt:

$$\begin{aligned} \Phi^{-1}(1 - \alpha) &= \iota_\beta \\ \iff \Phi^{-1}(1 - \alpha) &= \frac{\sqrt{v_r}\Phi^{-1}(\beta) + \Phi^{-1}(\pi_r + c) - \Phi^{-1}(\pi_r)}{\sqrt{v_r}} \\ \iff \Phi^{-1}(\pi_r + c) &= \sqrt{v_r}\Phi^{-1}(1 - \alpha) - \sqrt{v_r}\Phi^{-1}(\beta) + \Phi^{-1}(\pi_r) \\ \iff c &= \Phi(\sqrt{v_r}\Phi^{-1}(1 - \alpha) - \sqrt{v_r}\Phi^{-1}(\beta) + \Phi^{-1}(\pi_r)) - \pi_r \end{aligned}$$

□

Ein weiteres Verfahren, den Test zu optimieren, besteht darin, den bisherigen Test in zwei Tests aufzuteilen. Zuerst wird ein Test der Hypothesen

$$H_0 = [0, \pi_r] \quad \text{vs.} \quad H_1 = (\pi_r, \infty)$$

basierend auf der Teststatistik T_r zum Niveau α durchgeführt.

Dann wird ein Test der Hypothesen

$$H'_0 = [\pi_r + c, \infty) \quad \text{vs.} \quad H'_1 = [0, \pi_r + c)$$

zum Niveau β durchgeführt. Der Test wird analog zum vorherigen Test kontruiert.

Die Teststatistik wird definiert als

$$\tilde{T}_r \stackrel{\text{def}}{=} \frac{\sqrt{1 - v_r}\Phi^{-1}(\bar{p}_r) - \Phi^{-1}(\pi_r + c)}{\sqrt{v_r}}.$$

Unter der Hypothese H'_0 ist \tilde{T}_r standardnormalverteilt. Damit ergibt sich die untere Schranke des kritischen Bereiches zu $\kappa_\beta = \Phi^{-1}(1 - \beta)$ und somit

$$K_\beta \stackrel{\text{def}}{=} (\Phi^{-1}(1 - \beta), \infty).$$

Die geschätzte Ausfallwahrscheinlichkeit π_r wird angenommen, falls die Hypothese H_0 des ersten Tests angenommen wird und die Hypothese H'_0 des zweiten Tests abgelehnt wird. Ist dieses nicht der Fall, wird die geschätzte Ausfallwahrscheinlichkeit verworfen.

Mit Hilfe des ersten Tests wird die Wahrscheinlichkeit für den Fehler 1. Art durch α beschränkt. Durch den zweiten Test erhält man aufgrund des Hypothesentausches eine Schranke für den Fehler 2. Art in Höhe von β . Somit werden beide Fehler adäquat eingeschränkt.

Kapitel 4

Datenanalyse der Fahrzeugwerk Bernard KRONE GmbH

Die Fahrzeugwerk Bernard KRONE GmbH ist einer der führenden Hersteller von LKW-Anhängern und Aufliegern in Europa. Das Fahrzeugwerk gehört zur Bernard KRONE Holding GmbH & Co.KG. Im Geschäftsjahr 2006/2007 belief sich der Umsatz auf 832 Mio. Euro.

Die Produktion der Fahrzeugwerke ist auftragsbezogen. Bevor ein Fahrzeug für die technische Überprüfung bzw. Produktion freigegeben wird, wird zunächst die Bonität des Auftraggebers überprüft. Bei der Überprüfung wird auf die Bonitätsbewertung bzw. Risikoeinschätzung der CreditReform e.V. Nordhorn zurückgegriffen. Dabei wird dem Debitor ein Score s bzw. Ratingklasse r zugeordnet.

Die möglichen Score Werte bzw. Ratingklassen können der Tabelle 4.1 entnommen werden. Weiter zeigt Tabelle 4.1 die von der CreditReform prognostizierten Ausfallwahrscheinlichkeiten π_r . Dabei gibt π_r die Wahrscheinlichkeit an, mit der ein Debitor in der r -ten Ratingklasse vorraussichtlich innerhalb eines Jahres ausfällt.

r	1	2	3	4	5	6	7
s	199 - 200	201 - 350	351 - 399	400 - 449	450 - 499	500 - 550	551 - 600
π_r	26,87%	15,46%	6,04%	1,46%	0,73%	0,32%	0,07%

Tabelle 4.1: Score-Werte s und ihre Einteilung in Ratingklassen

Zur Validierung des Ratingsystems wurde eine Datenerhebung durchgeführt. Die Datenerhebung fand mit Hilfe des Unternehmens-Informationssystems mySAP ERP statt. Die hierfür geschriebene Transaktionsmaske ist in Abbildung 4.1 zu sehen. Der

Beobachtungszeitraum bezog sich vom 01.Juli 2007 bis zum 30.Juni 2008. Erfasst wurden alle Debitoren, denen ein Rating zugeordnet wurde. Danach wurde weiter gefiltert, ob mit diesen Debitoren ein Geschäftsvorgang im Beobachtungszeitraum abgeschlossen wurde bzw. inwieweit offene Forderungen gegenüber dem Debitor bestehen. Tabelle 4.2 zeigt die Ergebnisse der Erhebung.

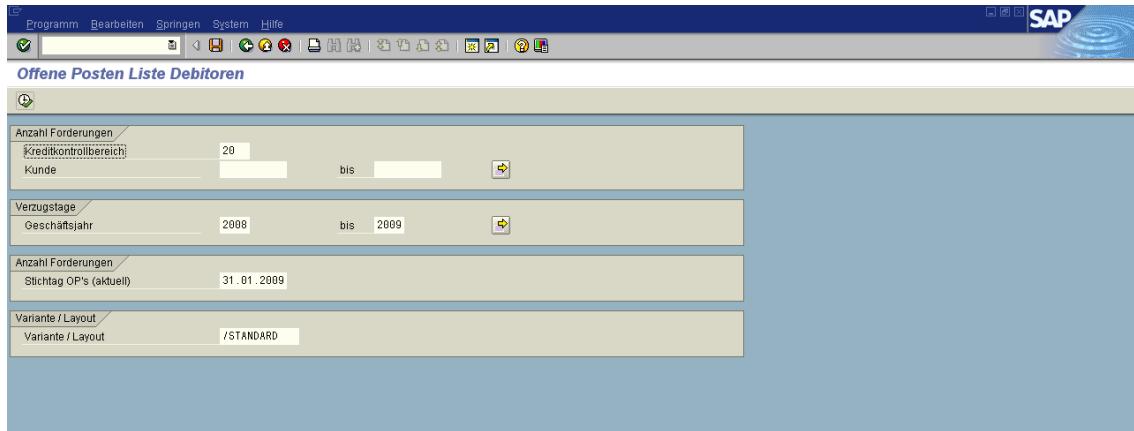


Abbildung 4.1: Transaktionsmaske mySAP ERP

r	1	2	3	4	5	6	7	gesamt
Debitoren	201	120	222	1460	2102	588	58	4751

Tabelle 4.2: Datenerhebung

Weiterhin ist für die Validierung die Anzahl der Ausfälle relevant. Dabei wird nach Basel II des Basler Ausschusses für Bankenaufsicht ein Ausfall verzeichnet, falls ein Debitor mehr als 90 Tage in Zahlungsverzug ist. Die Ergebnisse sind in Tabelle 4.3 aufgelistet.

r	1	2	3	4	5	6	7	gesamt
Defaulter	54	20	12	14	10	2	0	112
Non-defaulter	147	100	210	1446	2092	586	58	4639

Tabelle 4.3: Trennung in Defaulter und Non-Defaulter

Danach werden die kumulierten Wahrscheinlichkeiten \bar{F}_D und \bar{F}_N berechnet. Dazu werden die Daten aus 4.2 und 4.3 herangezogen. Die Verteilungsfunktion \bar{F} lässt sich ebenfalls über diese Daten bestimmen. Es besteht auch die Möglichkeit, diese mit Hilfe der Formel (2.1) zu ermitteln.

$$\bar{F}(r) = (1 - \bar{p})\bar{F}_N(r) + \bar{p}\bar{F}_D(r)$$

r	1	2	3	4	5	6	7
$\bar{F}_D(r)$	48,21	66,07	76,79	89,29	98,21	100	100
$\bar{F}_N(r)$	3,17	5,32	9,85	41,02	86,12	98,75	100
$\bar{F}(r)$	4,23	6,76	11,43	42,16	86,4	98,78	100

Tabelle 4.4: kum. Wahrscheinlichkeiten, Angaben in %

4.1 Trennschärfeberechnung des Ratingsystems

Damit Aussagen über die Trennschärfe des Ratings der Fa. KRONE getroffen werden können, werden die Ergebnisse aus dem dritten Kapitel auf die Daten der Datenerhebung angewandt.

Cumulative Accuracy Profile

Zuerst wird dazu das Cumulative Accuracy Profile untersucht (Kapitel 2.1). Die Werte werden mit Hilfe der Gleichung (2.2)

$$CAP(u) \stackrel{\text{def}}{=} \bar{F}_D(\bar{F}^{-1}(u)), \quad u \in [0, 1]$$

und den Daten aus Tabelle 4.4 berechnet.

u	0,0423	0,0676	0,1143	0,4216	0,864	0,9878	1
$CAP(u)$	0,4821	0,6607	0,7679	0,8929	0,9821	1	1

Tabelle 4.5: Funktionswerte $CAP(u)$

Die Daten aus Tabelle 4.5 zeigen, dass ein hoher Prozentanteil der Defaulter in die „schlechten“ Ratingklassen eingeordnet werden. Beispielsweise befinden sich unter den 4,23% „schlechtesten“ Debitoren bereits 48,23% der Defaulter.

Weiterhin ist es nun möglich, den Graph der CAP -Funktion zu plotten und das Trennschärfmaß Accuracy Ratio zu berechnen. In Abbildung 4.2 ist der Graph der Funktionswerte aus Tabelle 4.5 zu sehen.

Die Accuracy Ratio (2.5) ist gegeben durch

$$AR = \frac{2 \int_0^1 CAP(u) du - 1}{1 - \bar{p}},$$

wobei $\bar{p} = \frac{112}{4751} = 2,36\%$ die totale Ausfallwahrscheinlichkeit der Datenerhebung beschreibt. Weiter werden für die Berechnung die Wahrscheinlichkeiten $\bar{f}(r)$ bzw.

$\bar{f}_D(r)$ für $r = 1, \dots, 7$ benötigt. Diese geben die Wahrscheinlichkeit wieder, mit der ein Debitor bzw. Defaulter in der r -ten Ratingklasse ist. Mit Hilfe der Tabelle 4.3 ergibt sich:

r	1	2	3	4	5	6	7	Σ
$f_D(r)$	48,21	17,86	10,71	12,5	8,93	1,79	0	1
$f_N(r)$	3,17	2,16	4,53	31,17	45,1	12,63	1,25	1
$f(r)$	4,23	2,53	4,67	30,73	44,24	12,38	1,22	1

Tabelle 4.6: Anteil in den Ratingklasse, Angaben in %

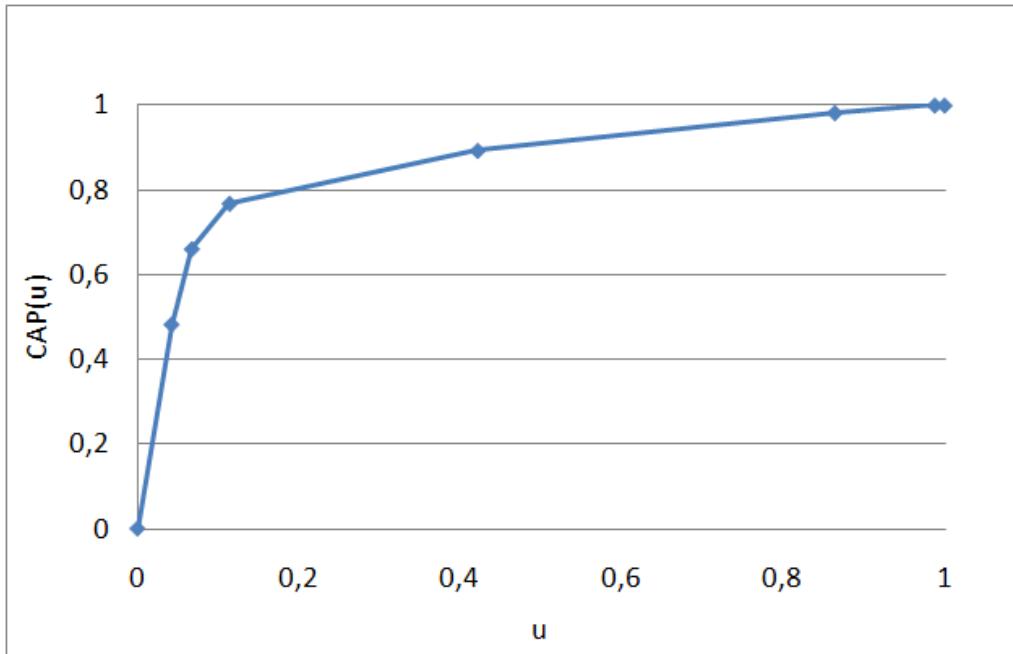


Abbildung 4.2: CAP - Kurve des Ratingsystems der Fa. KRONE

Mit den Werten aus Tabelle 4.6 berechnet sich die Fläche unter der CAP -Kurve zu

$$\begin{aligned} \int_0^1 CAP(u) \, du &= \sum_{r=1}^7 \left(\frac{1}{2} \bar{f}(r) \bar{f}_D(r) + \bar{f}(r) \bar{F}_D(1-r) \right) \\ &= 0,8625. \end{aligned}$$

Damit besteht nun die Möglichkeit, die Accuracy Ratio zu berechnen. Es ist

$$AR = \frac{2 \cdot 0,8625 - 1}{1 - 0,0236} = 0,7432.$$

Der Wert der AR des CreditReform Ratings¹ liegt in diesem Beispiel bei $AR_{CR} = 0,6849$.

Reciever Operating Characteristic

Als nächstes wird das ROC Verfahren untersucht. Mit den Daten aus Tabelle 4.3 und der ROC -Funktion (2.7)

$$ROC(u) \stackrel{\text{def}}{=} F_D(F_N^{-1}(u))$$

werden die Werte berechnet (Tabelle 4.7).

u	0,0317	0,0532	0,0985	0,4102	0,8612	0,9875	1
$ROC(u)$	0,4821	0,6607	0,7679	0,8929	0,9821	1	1

Tabelle 4.7: Funktionswerte $ROC(u)$

Geplottet siehe Abbildung 4.3.

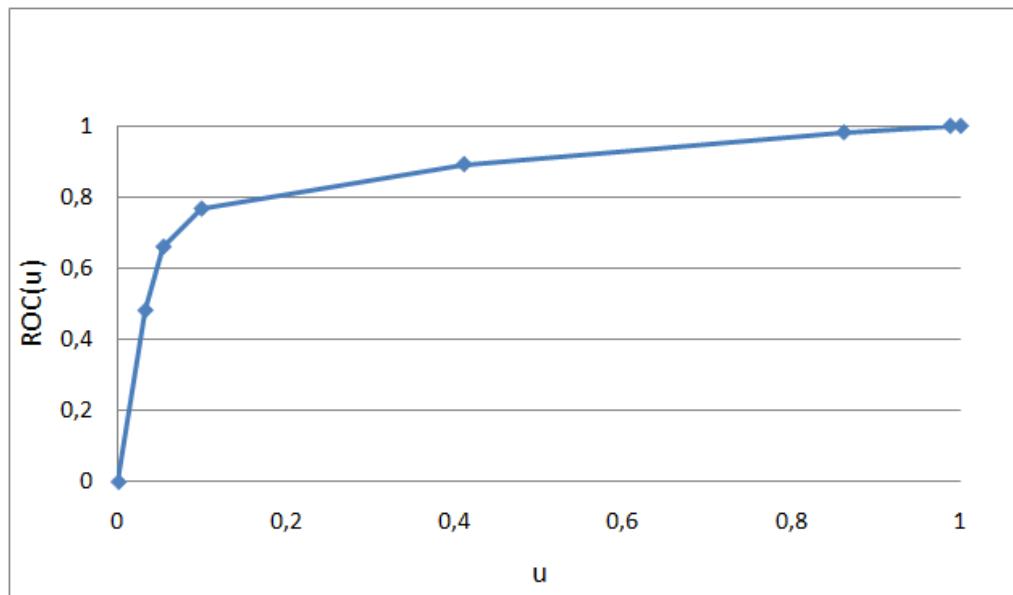


Abbildung 4.3: ROC - Kurve des Ratingsystems der Fa. KRONE

Für eine optimale Einteilung der Ratingklassen, muss die ROC-Kurve konkav sein. Da dieses anhand des Graphens nicht genau ersichtlich ist, wird die Ableitung $ROC'(u)$ betrachtet. Die Ableitung der ROC -Funktion wird gemäß (2.8) herge-

¹siehe Anhang, Berechnung der Gütemaße des CreditReform Ratings

leitet.

$$ROC'(u) = \frac{\bar{f}_D(\bar{F}_N^{-1}(u))}{\bar{f}_N(\bar{F}_N^{-1}(u))}$$

u	0,0317	0,0532	0,0985	0,4102	0,8612	0,9875	1
$ROC'(u)$	0,1521	0,0827	0,0236	0,004	0,002	0,0014	0

Tabelle 4.8: Ableitung von $ROC(u)$

Anhand der Tabelle 4.8 ist offensichtlich, dass die Ableitung, die dem Likelihood Quotienten entspricht, streng monoton fallend ist. Somit ist der Graph der ROC -Funktion konkav und die Einteilung der Ratingklassen testtheoretisch optimal (siehe dazu Kapitel 1.3).

Bei diesem Validierungsverfahren ist das Maß der Trennschärfe die $AUROC$. Die $AUROC$ ist die Fläche unter der ROC -Kurve und wird mit Hilfe der Daten aus Tabelle 4.6 berechnet.

$$\begin{aligned} AUROC &= \int_0^1 ROC(u) \, du \\ &= \sum_{r=1}^7 \left(\frac{1}{2} \bar{f}_N(r) + \bar{f}_N(r) + \bar{F}_D(r-1) \right) \\ &= 0,8726 \end{aligned}$$

Interpretiert man die $AUROC$ als Wahrscheinlichkeit, gemäß (2.9), so bedeutet dieses Resultat, dass das Rating der Fa. KRONE einen Defaulter mit 87,26% von einem Non-Defaulter unterscheiden kann.

Zum Vergleich ist die $AUROC_{CR}$ des CreditReform Ratings ² gleich 0,8425.

Bayes Fehler

In Kapitel 2.3.1 wurde der Bayes Fehler hergeleitet. Dieses ist ein Maß für die Fehlerwahrscheinlichkeit bei optimalem Einsatz des Ratingsystems. Zunächst wird der Geamtfehler gemäß (2.14) berechnet.

$$P(Fehler(r)) = \bar{p}(1 - \bar{F}_D(r)) + (1 - \bar{p})\bar{F}_N(r) \quad \text{für } r = 1, \dots, 7$$

Die Fehlerwahrscheinlichkeit der Ratingklassen sind in Tabelle 4.9 zusammengefasst.

Für die Berechnung wurden die Daten der Tabelle 4.3 entnommen.

Der Bayes Fehler (2.14) minimiert die Fehlerwahrscheinlichkeit. Somit werden auch

²siehe Anhang, Berechnung der Gütemaße des CreditReform Ratings

r	1	2	3	4	5	6	7
$P(\text{Fehler}(r))$	4,31	6	10,17	40,31	94,13	96,42	97,64

Tabelle 4.9: Fehlerwahrscheinlichkeit der Ratinklassen, in %

bei optimalem Einsatz des Ratings der Fa. KRONE noch immer 4,31% der Debitoren falsch klassifiziert bzw. eingeteilt.

Pietra-Index

Für den Fall identischer Verteilungsfunktionen besitzt das Ratingsystem keine Trennschärfe. Um dieses zu widerlegen, wird der Pietra-Index zum Testen der Identität der Verteilungsfunktion herangezogen. Zunächst wird $|\bar{F}_D - \bar{F}_N|$ bestimmt.

r	1	2	3	4	5	6	7
$ \bar{F}_D(r) - \bar{F}_N(r) $	0,5405	0,6075	0,6693	0,4826	0,1210	0,0125	0

Tabelle 4.10: Abstand der Verteilungsfunktionen

Das Maximum der Werte aus Tabelle 4.10 ist der Pietra-Index. Dieser wird nun als Teststatistik zum Testen der Hypothesen

$$H_0 : \bar{F}_D = \bar{F}_N \quad \text{gegen} \quad H_1 : \bar{F}_D \neq \bar{F}_N$$

verwendet. Der Test wird zum Niveau 1% durchgeführt (Bemerkung 2.3.4). Der kritische Wert berechnet sich zu

$$D = \frac{1,63}{\sqrt{4751}} = 0,0236$$

Da die Teststatistik (Pietra-Index = 0,6693) demnach im kritischen Bereich liegt, kann die Hypothese mit einer Sicherheit von 99% verworfen werden. D.h. die Verteilungsfunktionen sind nicht identisch und somit besitzt das Ratingsystem der Fa. KRONE Trennschärfe.

Informationsentropie

Der Informationsgewinn, den man mit dem Einsatz eines Ratingsystems erhält, berechnet man mit Hilfe von Entropie Quotienten $CIER$ (2.18). Dazu wird nun der Entropie Quotient der Fa. KRONE berechnet. Zuerst wird die Entropie der totalen Ausfallwahrscheinlichkeit $p = 0,0236$ mit Hilfe von (2.16) bestimmt.

$$\begin{aligned} H(0,0236) &= -(0,0236 \cdot \log_2(0,0236) + (1 - 0,0236) \cdot \log_2(1 - 0,0236)) \\ &= 0,1612 \end{aligned} \tag{4.1}$$

Dieses bedeutet, dass die Ungewissheit über den Zustand des Debtors 16,12% beträgt. Da durch das Ratingsystem mehr Informationen des Debtors bekannt sind als durch die durchschnittliche Ausfallwahrscheinlichkeit, wird die Entropie der bedingten Ausfallwahrscheinlichkeit p_r für alle $r = 1, \dots, 7$ bestimmt (Tabelle 4.11).

r	1	2	3	4	5	6	7
$H(\bar{p}_r)$	0	0,0328	0,0439	0,0782	0,3036	0,6501	0,8396

Tabelle 4.11: Entropie der bedingten empirischen Ausfallwahrscheinlichkeit

Die bedingte Entropie H_R ergibt sich gemäß (2.17) als Erwartungswert über alle Entropiewerte der bedingten Ausfallwahrscheinlichkeiten. Somit folgt mit den Werten der Tabelle 4.11:

$$\begin{aligned} E(H(p_r)) &= \sum_{r=1}^7 \bar{f}(r) H(\bar{p}_r) \\ &= 0,1137 \end{aligned} \tag{4.2}$$

Der bedingte Entropie Quotient (2.18) ergibt sich mit den Werten aus (4.1) und (4.2) zu:

$$CIER = 1 - \frac{0,1137}{0,1612} = 0,2946$$

Damit lässt sich abschließend zusammenfassen, dass das Ratingmodell der Fa.KRONE einen Informationsgewinn von 29,46% liefert. Zum Vergleich besitzt das Ratingsystem der CreditReform³ einen Informationsgewinn in Höhe von 24,65%.

4.2 Test einer Ausfallwahrscheinlichkeit

Nachdem im vorangegangenen Abschnitt lediglich die Trennschärfe des Ratingsystems behandelt wurde, wird nun getestet, inwieweit die tatsächliche Ausfallwahrscheinlichkeit mit der prognostizierten Ausfallwahrscheinlichkeit übereinstimmt. Da der asymptotische Test eine Ratingklassengröße von mehr als 500 Debitoren voraussetzt, wird die Ratingklasse 5 zum Testen der Ausfallwahrscheinlichkeit als Beispiel herangezogen. Als Ausfallkorrelation der Transport- bzw. Logistikbranche wird $v_5 = 0,0184$ angenommen⁴.

³siehe Anhang, Berechnung der Gütemaße des CreditReform Ratings

⁴Ausfallkorrelation Branchenklasse Verkehr des VDA

Die Datenerhebung beinhaltete $n = 4751$ Debitoren, die in $k = 7$ Ratingklassen unterteilt waren. Davon wurden $n_5 = 2102$ Debitoren der Ratingklasse 5 zugeordnet (Tabelle 4.2). Während des Beobachtungszeitraumes wurde eine Ausfallquote von

$$\bar{p}_5 = \frac{10}{2102} = 0,00476$$

registriert (Tabelle 4.3). Die prognostizierte Ausfallwahrscheinlichkeit kann den Unterlagen der CreditReform entnommen werden (Tabelle 4.1) und beträgt für die Ratingklasse 5 $\pi_5 = 0,0073$. Mit diesen Daten ist es nun möglich, einen Test zum Niveau $\alpha = 5\%$ festzulegen.

Der Parameterraum Θ des zu testenden Parameters entspricht

$$\Theta = [0, 1]$$

Getestet wird nun, ob die tatsächliche Ausfallwahrscheinlichkeit p_5 in der Hypothese H_0 oder in der Alternative H_1 liegt. Da es im Interesse des Kreditors liegt, das Ausfallrisiko nicht zu unterschätzen, wird die Hypothese H_0 so gewählt, dass die prognostizierte Ausfallwahrscheinlichkeit mit hoher vorgegebener Wahrscheinlichkeit $(1 - \alpha)$ angenommen werden kann. Damit ergeben sich die Hypothesen

$$H_0 : p_5 \leq 0,0073 \quad \text{und} \quad H_1 : p_5 > 0,0073.$$

Die Teststatistik für diesen Test (3.10) wurde in Kapitel 3.3 hergeleitet und lautet für die Ratingklasse 5:

$$\begin{aligned} T_5 &= \frac{\sqrt{1 - 0,0184} \Phi^{-1}(0,00476) - \Phi^{-1}(0,0073)}{\sqrt{0,0184}} \\ &= -0,9353 \end{aligned} \tag{4.3}$$

Um zu überprüfen, ob die Hypothese abgelehnt wird oder nicht, wird der Akzeptanzbereich sowie der kritische Bereich $K_{5\%}$ bestimmt. Mit Hilfe von (3.11) folgt

$$K_{5\%} = \underbrace{(\Phi^{-1}(0,95), \infty)}_{= 1,6449}. \tag{4.4}$$

Mit (4.3) und (4.4) folgt,

$$T_5 \notin K_{5\%}.$$

Deshalb gilt für die Testfunktion (3.6)

$$\varphi(T_5) = 0.$$

Dieses Resultat bedeutet, dass die Hypothese H_0 mit 95%-iger Sicherheit angenommen wird und somit die tatsächliche Ausfallwahrscheinlichkeit p_5 der prognostizierten Ausfallwahrscheinlichkeit $\pi_5 = 0,0073$ entspricht.

Um noch Informationen über den Fehler 2. Art zu bekommen, wird der Modellfehler $p_5 - \pi_5$ betrachtet. Nach Kapitel 3.3 liefert das Gleichsetzen der kritischen Grenzen bei gegebenem Fehler 1. Art bzw. Fehler 2. Art optimale Entscheidungsregeln. Mit Hilfe von (3.15) wird eine obere Schranke für den Modellfehler bestimmt. Bei einem Fehler 1. Art von 5% und einem Fehler 2. Art von 50% berechnet sich diese zu

$$c = \Phi \left(\sqrt{0,0184} (\Phi^{-1}(0,95) - \Phi^{-1}(0,5)) + \Phi^{-1}(0,0073) \right) - 0,0073 = 0,0059$$

Das bedeutet, dass ein Modellfehler $p_5 - \pi_5 \geq 0,0059$ mit einer Wahrscheinlichkeit von 50% erkannt wird.

Zum Abschluss sei noch gesagt, dass die prognostizierten Ausfallwahrscheinlichkeiten π_4 und π_6 der Ratingklasse 4 und 6 durch analoge Rechnungen angenommen werden. Ebenfalls wird die Ausfallwahrscheinlichkeit π_7 der siebten Ratingklasse angenommen, da eine Ausfallwahrscheinlichkeit von 0% in jedem Fall akzeptiert wird. Über die übrigen Ausfallwahrscheinlichkeiten lässt sich aufgrund der geringen Datenanzahl keine adäquate Aussage treffen.

Fazit

Ziel dieser Arbeit bestand darin, das Ratingsystem eines mittelständigen Unternehmens zu validieren. Besonders wurde dazu auf quantitative Validierung eingegangen.

Nachdem ein Modell entwickelt wurde, wurde die Gütemaße der Trennschärfe hergeleitet. Besonders bewährt haben sich dabei das Cumulative Accuracy Profile und die Receiver Operating Characteristic. Diese sind sehr gut erforscht und liefern gleichzeitig auch sehr anschauliche Ergebnisse über die Trennschärfe eines Ratingsystems. In einer solchen Arbeit dürfen weitere Gütemaße -wie die Bayessche Fehler oder Pietra Index- jedoch nicht fehlen, weswegen diese ebenfalls im zweiten Kapitel vorgestellt wurden.

Ein weiterer Schwerpunkt der Arbeit lag auf der Kalibrierung der Ausfallwahrscheinlichkeiten. Dabei ist es imminent, die Abhängigkeit der Ausfallereignisse zu modellieren. Dazu wurde auf den asymptotischen Test im Ein-Faktor Modell zurückgegriffen. Die Abhängigkeit wurde durch die Korrelation der Ausfallvariablen modelliert. Des Weiteren ist die Grundvoraussetzung für einen solchen Test eine gewisse Anzahl an Daten.

Grundlage für die Anwendung der gewonnenen Erkenntnisse war es, eine gewisse Datenqualität zu gewährleisten. Dabei war das Unternehmens-Informationssystem -SAP- eine hervorragende Unterstützung. Durch eine geschriebene Transaktionsmaske konnte auf die Datenbank der Buchhaltung zugegriffen und nach den gewünschten Kennzahlen gefiltert werden. Diese ließen sich dann ohne Weiteres zur Weiterverarbeitung in ein Tabellenkalkulationsprogramm exportieren.

Die damit berechneten Gütemaße wiesen stets einen „besseren“ -jedoch keinen signifikant besseren- Wert auf⁵. Damit wurden die vorgegebenen Ratingeinstufungen der CreditReform optimal umgesetzt.

Beim Testen der Ausfallwahrscheinlichkeit konnten aufgrund der Datenmenge ledig-

⁵z.B. $AR_{KRONE} = 0,8625$; $AR_{CR} = 0,6849$ & $AUROC_{KRONE} = 0,8726$; $AUROC_{CR} = 0,8425$

lich die Ratingklassen 4,5 und 6 getestet werden, da ein solcher Test eine mindest Anzahl von 500 Debitoren fordert, um zu gerechten Ergebnissen zu gelangen. Die prognostizierten Ausfallwahrscheinlichkeiten dieser Ratingklassen wurden alle durch den Test angenommen ⁶.

Damit lässt sich als Resultat dieser Arbeit der Fa. KRONE eine sehr gute Umsetzung der Ratings bescheinigen.

⁶Test zum Niveau 5% bei einem „Fehler 2. Art“ in Höhe von 50%

Anhang

Berechnung der Gütemaße der Trennschärfe des CreditReform Ratings

Auf Grundlage der im vierten Kapitel betrachteten Datenerhebung werden nun die Trennschärfemaße des CreditReform Ratings berechnet. Die Datenerhebung ergab $n = 4751$ Debitoren, die folgendermaßen den Ratingklassen ($r = 1, \dots, 7$) zugeteilt wurden.

r	1	2	3	4	5	6	7	gesamt
Debitoren	201	120	222	1460	2102	588	58	4751

r	1	2	3	4	5	6	7
π_r	26,87%	15,46%	6,04%	1,46%	0,73%	0,32%	0,07%

Tabelle 4.12: prognostizierte Ausfallwahrscheinlichkeit der CreditReform

Mit den prognostizierten Ausfallwahrscheinlichkeiten π_r (Tabelle 4.12) und Tabelle 4.2 ergeben sich die theoretisch prognostizierten Ausfälle zu

r	1	2	3	4	5	6	7
Defaulter	54,0087	18,552	13,4088	21,316	15,3446	1,8816	0,00406

Tabelle 4.13: theoretische Ausfälle des CreditReform Ratings

Die kumulierten Wahrscheinlichkeiten \bar{F}_D , \bar{F}_N sowie \bar{F} berechnen sich zu Tabelle 4.14.

Weiter werden für die Berechnung die Wahrscheinlichkeiten $\bar{f}(r)$ bzw. $\bar{f}_D(r)$ für $r = 1, \dots, 7$ benötigt. Diese geben die Wahrscheinlichkeit wieder, mit der ein Debitor bzw. Defaulter in der r -ten Ratingklasse ist. Mit Hilfe der Tabelle 4.13 werden die Werte aus 4.15 berechnet.

Die totale Ausfallwahrscheinlichkeit berechnet sich zu $p = \frac{124,51576}{4751} = 2,62\%$.

r	1	2	3	4	5	6	7
$F_D(r)$	43,36	58,26	69,02	86,14	98,46	99,97	100
$F_N(r)$	4,23	6,76	11,43	42,16	86,40	98,78	100
$F(r)$	4,23	6,76	11,43	42,16	86,4	98,78	100

Tabelle 4.14: kum. Wahrscheinlichkeiten, Angaben in %

r	1	2	3	4	5	6	7	Σ
$f_D(r)$	43,36	14,89	10,77	17,11	12,32	1,51	0,03	1
$f_N(r)$	3,18	2,19	4,51	31,10	45,10	12,67	1,25	1
$f(r)$	4,23	2,53	4,67	30,73	44,24	12,38	1,22	1

Tabelle 4.15: Anteil in den Ratingklasse, Angaben in %

Damit können nun -analog zu den Berechnungen aus dem vierten Kapitel- die Gütemaße Accuracy Ratio, Area under the ROC, usw. berechnet werden. Es folgt:

$$AR_{CR} = 0,6849$$

$$AUROC_{CR} = 0,8425$$

$$CIER_{CR} = 0,2465$$

Kritische Werte des Kolmogorov-Smirnoff Tests

Datenanzahl n	Signifikanzniveau	0,10	0,05	0,01
1		0,950	0,975	0,995
2		0,776	0,842	0,929
3		0,642	0,708	0,828
4		0,564	0,624	0,733
5		0,510	0,565	0,669
6		0,470	0,521	0,618
7		0,438	0,486	0,577
8		0,411	0,457	0,543
9		0,388	0,432	0,514
10		0,368	0,410	0,490
11		0,352	0,391	0,468
12		0,338	0,375	0,450
13		0,325	0,361	0,433
14		0,314	0,349	0,418
15		0,304	0,338	0,404
16		0,295	0,328	0,392
17		0,286	0,318	0,381
18		0,278	0,309	0,371
19		0,272	0,301	0,363
20		0,264	0,294	0,356
25		0,240	0,270	0,320
30		0,220	0,240	0,290
35		0,210	0,230	0,270
> 35		$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

Tabelle 4.16: Kritische Werte des Kolmogorov-Smirnoff Tests

Literaturverzeichnis

- [1] Alsmeyer, G. (2005) *Wahrscheinlichkeitstheorie, Skripten zur Mathematischen Statistik Nr. 30*. 4. Auflage ,Münster
- [2] Alsmeyer, G. (2006) *Mathematische Statistik, Skripten zur Mathematischen Statistik Nr. 36*. 3. Auflage ,Münster
- [3] Tasche, D. (2009) *Estimating discriminatory power and PD curves when the number of defaults is small*. Preprint, Technische Universität München
- [4] Tasche, D. (2006) *Validation of internal rating systems and PD estimates*. Preprint, Technische Universität München
- [5] Pluta, K. & Tasche, D. (2006) *Thinking positively*. Preprint, Technische Universität München
- [6] Engelmann, B. & Rauhmeier, R. (2006) *The Basel II Risk Parameters, Estimation, Validation and Stress Testing*. Springer 1. Auflage, Berlin
- [7] Engelmann, B., Hayden, E. & Tasche, D. (2006) *Measuring the discriminative power of rating systems*. Working Paper, Discussion paper No. 01/2003, Banking and Financial Supervision, Deutsche Bundesbank
- [8] Henking, A., Bluhm, C. & Fahrmeier, L. (2006) *Kreditrisikomessung, Statistische Grundlagen, Methoden und Modellierung*. Springer 1. Auflage , Berlin
- [9] Tasche, D. (2002) *Remarks on the monotonicity of default probabilities*. Working Paper, Technische Universität München
- [10] Thonabauer, G. & Nösslinger, B. (2004) *Leitfadenreihe zum Kreditrisiko, Ratingmodelle und -validierung*. Österreichische Nationalbank, Wien
- [11] Cormann, U. (2005) *Backtesting von Kreditrisikomodellen*. Bachelor Thesis, Siegen

- [12] Huschens, S. (2004) *Backtesting von Ausfallwahrscheinlichkeiten*. Dresdner Beiträge zu Quantitativen Verfahren Technische Universität Dresden
- [13] Höse, S. & Huschens, S. (2003) *Sind interne Ratingsysteme im Rahmen von Basel II evaluierbar?, Zur Schätzung von Ausfallwahrscheinlichkeiten durch Ausfallquoten*. Zeitschrift für Betriebswirtschaft 73 Gabler, Münschen
- [14] Oehler, A. (2002) *Kreditrisikomanagement, Kernbereiche, Aufsicht und Entwicklungstendenzen*. Schäffer-Poeschel 2. überarb. Auflage, Stuttgart