

Dissertation

**Qualitätsvergleiche kalibrierter
Wahrscheinlichkeitsprognosen mit Anwendung
auf die internationale Ratingindustrie**

Simon Neumärker

4. September 2018

Betreuer: Prof. Dr. Walter Krämer

Technische Universität Dortmund

Fakultät Statistik

Inhaltsverzeichnis

Abbildungsverzeichnis	vi
Tabellenverzeichnis	vii
Symbolverzeichnis	viii
1 Überblick und Motivation	1
2 Grundbegriffe von Wahrscheinlichkeitsprognosen	7
3 Skalarwertige Qualitätskriterien	17
3.1 Gini-Koeffizient und AUC	17
3.2 Score-Regeln	22
3.3 Kalibrierte Prognosen	33
3.4 Anreizerhaltende Skill-Scores	38
4 Halbordnungen bei identischen Schuldnerkollektiven	47
4.1 Motivation	47
4.2 Refinement-Ordnung	50
4.3 Vardeman-Meeden-Ordnung	59
4.4 Lorenz- und ROC-Ordnung	63

5	Halbordnungen bei unterschiedlichen Schuldnerkollektiven	67
5.1	Motivation	67
5.2	Vardeman-Meeden-Ordnung	70
5.3	Verallgemeinerte Lorenz-Ordnung	74
6	Gewichtete monotone Regression	83
6.1	Motivation	83
6.2	Schätzung isotoner Ausfallwahrscheinlichkeiten	85
7	Anwendung im Bereich der Ratingindustrie	93
7.1	Gewichtete monotone Regression	94
7.2	Vardeman-Meeden-Ordnung	98
7.3	Verallgemeinerte Lorenz-Ordnung	100
8	Fazit und Zusammenfassung	103
A	Grafiken	106
B	Tabelle	108
C	Verträgliche Schätzungen	110
D	Alternative verträgliche Schätzung	114
	Eidesstattliche Versicherung	123

Abbildungsverzeichnis

3.1	Beispiel: Lorenzkurve und ROC-Kurve	22
3.2	Erwartete Score-Funktion des Logarithmus-Scores mit Tangente . . .	30
3.3	Beispiel: Skill-Score-optimale Prognose	44
4.1	Dominanz nach der Refinement-Ordnung zwischen zwei kalibrierten Wahrscheinlichkeitsprognostikern für $k = 4$	58
5.1	Beispiel: Lorenzkurve und ROC-Kurve bei verschiedenen Schuldner- populationen	69
5.2	Score-Funktionen des hyperbolischen und des sphärischen Scores . . .	79
5.3	Beispiel: Verallgemeinerte Lorenz-Ordnung und Skill-Score	80
6.1	Beispiel: Lorenzkurve bei nicht isotonen Ausfallanteilen und zusam- mengefassten Ratingklassen	91
7.1	RMBS: Monotone Regression der empirischen Ausfallanteile	97
7.2	RMBS: Integrierte empirische Verteilungsfunktionen	100
7.3	RMBS: Empirische Lorenzkurven (Verallgemeinerte Lorenz-Ordnung)	101
A.1	Beispiel: Lorenzkurven von drei (kalibrierten) Wahrscheinlichkeits- prognostikern auf gleichen Schuldnerpopulationen	106

A.2	Beispiel: Lorenzkurven der Nichtausfälle bei verschiedenen Schuldnerpopulationen	107
A.3	RMBS: Empirische Verteilungsfunktionen der auf den Nichtausfall und der auf den Ausfall bedingten Verteilungen (VM-Ordnung) . . .	107

Tabellenverzeichnis

2.1	Bivariate Verteilung von Ausfallvariable und Ausfallprognose	10
3.1	Score-Funktionen und erwartete Score-Funktionen ausgewählter Score-Regeln	28
3.2	Beispiel: Erwartete Skill-Scores bei Skill-Score-optimalen Prognosen .	45
4.1	Beispiel: Prognoseverteilungen von drei (kalibrierten) Prognostikern auf gleichen Schuldnerpopulationen	48
4.2	Beispiel: Erwartungswerte ausgewählter Score-Regeln von drei (kalibrierten) Prognostikern auf gleichen Schuldnerpopulationen	48
4.3	Beispiel: Trennschärfe von kalibrierten Prognosen	52
7.1	RMBS: Ratingverteilung und geschätzte Ausfallwahrscheinlichkeiten .	96
7.2	RMBS: Dominanzbeziehungen nach der VM-Ausfallordnung	98
7.3	RMBS: Dominanzbeziehungen nach der VM-Nichtausfallordnung . . .	99
7.4	RMBS: Werte ausgewählter Score-Regeln und Skill-Scores	102
B.1	Ratingskala von Moody's für langfristige Anleihen	108
B.2	Marktanteile der in der EU registrierten Ratingagenturen	109

Symbolverzeichnis

Symbole für Wahrscheinlichkeitsprognosen

a_1, \dots, a_k	Zulässige Wahrscheinlichkeitsprognosen
$\hat{a}_1, \dots, \hat{a}_k$	Geschätzte Wahrscheinlichkeitsprognosen
$\tilde{a}_1, \dots, \tilde{a}_k$	Empirische Ausfallanteile in den Ratingklassen R_1, \dots, R_k
$E(X)$	Erwartungswert der Zufallsvariablen X
$g(\pi) = E(s(\Theta, \pi))$	Erwartete Score-Funktion von S mit $\Theta \sim \text{Bin}(1, \pi)$
$GB(\vec{\pi})$	Verallgemeinerte Binomialverteilung mit Parametervektor $\vec{\pi}$
G, AUC, AR	Gini-Koeffizient, Area under the curve, Accuracy ratio
L, ROC, CAP	Lorenzkurve, ROC-Kurve, CAP-Kurve
\hat{L}	Empirische Lorenzkurve
$L_{\hat{a}_1, \dots, \hat{a}_k}$	Lorenzkurve der (empirisch kalibrierten) Prognosen $\hat{a}_1, \dots, \hat{a}_k$
\mathbb{L}_{rnn}	Menge aller regulären nulloptimalen nichtnegativen Score-Regeln
$n \in \mathbb{N}$	Anzahl der bewerteten Schuldner/Kredite
P^X	Verteilung der Zufallsvariablen X
$p_1 = \sum_{j=1}^k r(1, a_j)$	Gesamtausfallwahrscheinlichkeit in der Schuldnerpopulation
\hat{p}_1	Empirischer Gesamtausfallanteil in der Schuldnerpopulation

$p(\theta a_j)$	Zähldichte von $P^{\Theta X}$
$\vec{\pi} = (\pi_1, \dots, \pi_n)'$	Vektor der Ausfallwahrscheinlichkeiten
$\vec{\pi}_{(i)}$	$\vec{\pi}$ ohne den i -ten Eintrag
$q(a_j)$	Zähldichte von P^X
$q(a_j \theta)$	Zähldichte von $P^{X \Theta}$
$\hat{q}(\hat{a}_j), \hat{r}(\theta, \hat{a}_j)$	Empirische relative Häufigkeitsfunktionen von P^X und $P^{\Theta, X}$
$\hat{q}(R_j)$	Empirischer Anteil der Beobachtungen mit Ratingklasse R_j
$\{R_1, \dots, R_k\}$	Menge der zulässigen Ratingkategorien
$\vec{r} = (r_1, \dots, r_n)'$	Vektor der Ratings, $r_i \in \{R_1, \dots, R_k\}$
$r(\theta, a_j)$	Bivariate Wahrscheinlichkeitsfunktion der Ausfallvariablen Θ und der Prognose X
$S(\vec{\theta}, \vec{x})$	Score-Regel für $\vec{\theta}$ und \vec{x}
$s(\theta_i, x_i)$	Score-Funktion von S für θ_i und x_i
$S(\Theta, X) = E(s(\Theta, X))$	Erwartungswert der Score-Funktion von S unter Θ und X
$SK(\vec{\theta}, \vec{x})$	Skill-Score von S für $\vec{\theta}$ und \vec{x}
$\Theta \sim \text{Bin}(1, p_1)$	Zufallsvariable des Ausfalls mit P^Θ einer Bernoulliverteilung mit Parameter $p_1 \in [0, 1]$ und Realisation θ
$\vec{\theta} = (\theta_1, \dots, \theta_n)'$	Vektor der realisierten Ausfallvariablen, $\theta_i \in \{0, 1\}$
$\mathbb{W} = \{a_1, \dots, a_k\}$	Menge der zulässigen Wahrscheinlichkeitsprognosen
$\widehat{\mathbb{W}} = \{\hat{a}_1, \dots, \hat{a}_k\}$	Menge der geschätzten Wahrscheinlichkeitsprognosen
$w_{\vec{\pi}}$	Zähldichte der $GB(\vec{\pi})$
X	Zufallsvariable der Wahrscheinlichkeitsprognose mit Zähldichte q und Träger \mathbb{W}
X_T, X_P	Trivialprognose, perfekte Prognose
$\vec{x} = (x_1, \dots, x_n)'$	Vektor der geschätzten Wahrscheinlichkeitsprognosen, $x_i \in \widehat{\mathbb{W}}$
$x_{(i)}$	i -te geordnete Beobachtung aus \vec{x}
$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	Arithmetisches Mittel von x_1, \dots, x_n

Symbole für den Vergleich von Prognostikern

$q^A(a_j)$	Zähldichte von P^X des Wahrscheinlichkeitsprognostikers A ; ebenso werden ggf. andere Symbole hochgestellt indiziert, um die Zugehörigkeit zum Prognostiker A herauszustellen
\geq_L	Relation der Lorenz-Ordnung
\geq_{pROC}	Relation der partiellen ROC-Ordnung
\geq_R	Relation der Refinement-Ordnung
\geq_{ROC}	Relation der ROC-Ordnung
$\geq_{SD(n)}$	Relation der stochastischen Dominanz n -ter Ordnung
\geq_{VL}	Relation der verallgemeinerten Lorenz-Ordnung
$\geq_{VM(d)}$	Relation der Vardeman-Meeden-Ausfallordnung
$\geq_{VM(nd)}$	Relation der Vardeman-Meeden-Nichtausfallordnung

Mengen und mathematische Symbole

$ \mathbb{A} $	Mächtigkeit der Menge \mathbb{A}
$\mathbb{A} \subset \mathbb{B}$	Menge \mathbb{A} ist gleich Menge \mathbb{B} oder echte Teilmenge von \mathbb{B}
f', f'', f''' und $f^{(n)}$	Erste, zweite, dritte und n -te Ableitung von $f(x)$ nach x
$\mathbb{N} = \{1, 2, \dots\}$	Menge der natürlichen Zahlen
$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	Menge der natürlichen Zahlen mit Null
$P(\cdot)$	Wahrscheinlichkeitsfunktion
\mathbb{R}	Menge der reellen Zahlen
$\mathbb{R}_{>0}$	Menge der echt positiven reellen Zahlen
$\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$	Abschluss der reellen Zahlen
\vec{x}'	Transponierter Vektor von \vec{x}
\mathbb{Z}	Menge der ganzen Zahlen
$\mathbf{1}(\cdot)$	Indikatorfunktion

Kapitel 1

Überblick und Motivation

Auch wenn es nur der Wetterbericht ist, der die morgige Regenwahrscheinlichkeit in Dortmund mit 20 % beziffert, heutzutage sind Wahrscheinlichkeitsprognosen in vielen Bereichen des täglichen Lebens zu finden. Nicht nur in der Meteorologie, sondern auch in der Medizin sowie in der Versicherungs- und Bankenbranche sind Wahrscheinlichkeitsprognosen essentiell. So prognostiziert der Mediziner die Wahrscheinlichkeit für einen Schlaganfall, die Versicherung die Wahrscheinlichkeit für den Eintritt eines Schadens und die Bank die Wahrscheinlichkeit für die Zahlungsunfähigkeit eines Schuldners. Und der Buchmacher hält eine Wahrscheinlichkeit für den Sieg des BVBs beim nächsten Heimspiel bereit.

Diese Arbeit befasst sich mit dem Qualitätsvergleich von Wahrscheinlichkeitsprognosen im Bereich Wirtschaft. Dabei stehen insbesondere Kreditausfallprognosen im Mittelpunkt der Untersuchung. Eng verbunden mit solchen Wahrscheinlichkeitsprognosen sind sogenannte Ratings. Diese werden nicht nur von Banken zur Bonitätsprüfung potentieller Schuldner individuell und bankintern erstellt, sie sind auch der zentrale Output von internationalen Ratingagenturen, siehe Tabelle B.1 im Anhang. Solche Agenturen bewerten Anleihen, Firmen oder sogar ganze Staaten

hinsichtlich des Ausfallrisikos. Ihre Bewertung ist für ein Unternehmen oder einen Staat entscheidend am Geldmarkt, etwa für die Höhe der aufzubringenden Zinsen für Schuldverschreibungen.

Die größten internationalen Ratingagenturen sind Standard & Poor's, Moody's und Fitch. Zusammen besitzen diese drei Agenturen einen EU-Marktanteil von über 90 %, vgl. Tabelle B.2 im Anhang. Dabei kommt es nicht selten vor, dass das Ausfallrisiko eines bestimmten Unternehmens von den Agenturen unterschiedlich eingeschätzt wird. Die Bewertung von Ausfallprognosen widmet sich also Fragen der Art: Welche dieser Ratingagenturen macht den besten Job?

Neben der methodischen Herausforderung ist die Bewertung von Ratingagenturen von enormem wirtschaftlichen und gesellschaftlichen Interesse. Die Bedeutung dieser Agenturen hat bereits Friedman (1995) in einem Artikel der New York Times herausgestellt:

„In fact, you could almost say that we live again in a two-superpower world.

There is the U.S. and there is Moody's. The U.S. can destroy a country by leveling it with bombs; Moody's can destroy a country by downgrading its bonds.“

Die Macht und damit auch die Verantwortung der Ratingagenturen, wurden in der Finanzkrise ab 2007 offensichtlich. Den Ratingagenturen wird eine große Mitschuld an der Entwicklung der Krise gegeben, da sie unter anderem unrealistisch gute Ratings im Bereich der hypothekenbesicherten Wertpapiere vergeben haben sollen. Dadurch seien falsche Anreize auf dem Markt gesetzt worden (Financial Crisis Inquiry Commission, 2011, S. xxv).

Bei der Erstellung von Wahrscheinlichkeitsprognosen und Ratings kommen statistische Verfahren, wie zum Beispiel die logistische Regression, die Diskriminanzanalyse oder der Random-Forest zum Einsatz (Hosmer et al., 2013; Fraley & Raftery, 2002; Breiman, 2001).

Ohne jedoch die Entstehung der Prognosen genauer zu betrachten, wird in dieser Arbeit untersucht, wie die Qualität von prognostizierten Ausfallwahrscheinlichkeiten adäquat beurteilt werden kann. Hierbei sind zwei Kriterien entscheidend: Kalibrierung und Trennschärfe. Eine Wahrscheinlichkeitsprognose ist kalibriert, falls sie mit der tatsächlichen Ausfallwahrscheinlichkeit in der so bewerteten Schuldnermenge übereinstimmt. Diese Eigenschaft ist jedoch nicht hinreichend für eine gute Prognose: Angenommen in einer Population fallen 10 % der Schuldner aus. Die Prognose, die jedem Schuldner eine Ausfallwahrscheinlichkeit von 10 % zuordnet, ist dann zwar kalibriert, jedoch wertlos. Diese Prognose besitzt nämlich keine Trennschärfe, d. h. Ausfälle und Nichtausfälle sind anhand der Prognose nicht unterscheidbar. Kalibrierung ist daher in vielen Arbeiten als notwendige Bedingung für Wahrscheinlichkeitsprognosen vorausgesetzt, sodass sich konkurrierende Prognosen nur hinsichtlich ihrer Trennschärfe vergleichen müssen (Gneiting et al., 2007). Hier untersucht die vorliegende Arbeit, wie sich Wahrscheinlichkeitsprognosen bei vorliegender Kalibrierung vergleichen lassen.

Anreizerhaltende Score-Regeln sind in der Literatur das favorisierte Mittel zum Vergleich von Verteilungsprognosen (Gneiting et al., 2007; Gneiting & Raftery, 2007). Diese skalarwertigen Qualitätskriterien messen den Abstand zwischen Prognose und realisiertem Ereignis und belohnen dabei die ehrliche Prognose eines Prognostikers (eine genaue Definition gibt Kapitel 3). Aus Score-Regeln können

sogenannte Skill-Scores konstruiert werden. Diese kommen zum Einsatz, wenn zwei konkurrierende Prognostiker unterschiedliche Populationen von Schuldnern bewerten (Murphy, 1973; Wilks, 2006, S. 305). Bereits Gneiting & Raftery (2007) fordern in ihrem Ausblick, die Eigenschaften von Skill-Scores genauer zu untersuchen. Dazu wird in dieser Arbeit der Frage nachgegangen, inwiefern Skill-Scores von anreizerhaltenden Score-Regeln selbst auch anreizerhaltend sind.

Halbordnungen zum Vergleich von Wahrscheinlichkeitsprognosen sind von besonderem Nutzen, da sie Dominanz in einer ganzen Klasse von Score-Regeln garantieren können (DeGroot & Fienberg, 1983; Vardeman & Meeden, 1983). Allerdings sind solche Halbordnungen bislang nur für den Fall vorhanden, dass die zu vergleichenden Prognostiker die gleichen Schuldnerkollektive bewerten. Hier zeigen Krämer & Neumärker (2016) neue Wege auf, indem sie eine Halbordnung für Wahrscheinlichkeitsprognosen bei unterschiedlichen Schuldnerkollektiven definieren. Die weitergehende Untersuchung dieser Halbordnung ist auch Gegenstand dieser Arbeit.

Der weitere Aufbau der Arbeit ist wie folgt: Kapitel 2 definiert die grundlegenden Begriffe für den Umgang mit Wahrscheinlichkeitsprognosen, wie die Kalibrierung sowie die entsprechende Notation. In Kapitel 3 werden skalarwertige Qualitätskriterien für Wahrscheinlichkeitsprognosen definiert und deren Eigenschaften beleuchtet. Die Klasse der anreizerhaltenden Score-Regeln wird vorgestellt, und abschließend wird eine neue anreizerhaltende Score-Regel definiert. Das Kapitel 4 untersucht Halbordnungen bei identischen Schuldnerkollektiven. In Kapitel 5 wird die Restriktion identischer Schuldnerkollektive aufgegeben. Dazu wird eine neue Halbordnung, die verallgemeinerte Lorenz-Ordnung definiert und untersucht. Das Kapitel 6 befasst sich mit der Schätzung von monoton steigenden Ausfallwahrscheinlichkeiten zu

gegeben Ratingklassen. Dazu wird ein Bezug zwischen der monotonen Regression und der Lorenz-Ordnung hergestellt. Ein Anwendungsbeispiel in Kapitel 7 veranschaulicht die in dieser Arbeit herausgearbeiteten Ergebnisse durch einen Vergleich der Ratingagenturen Standard & Poor's, Moody's und Fitch. Ein Fazit in Kapitel 8 fasst die wichtigsten Ergebnisse der Arbeit zusammen.

Kapitel 2

Grundbegriffe von Wahrscheinlichkeitsprognosen

Formal gesehen, ist eine Wahrscheinlichkeitsprognose die Vorhersage einer Bernoulli-Verteilung und damit ein Spezialfall allgemeiner Verteilungsprognosen (Gneiting et al., 2007). Verteilungsprognosen liefern in der Regel mehr Informationen als einfache Punkt- oder Intervallprognosen. Statt ein binäres Ereignis direkt vorherzusagen (Punktprognose), kann es daher sinnvoll sein, eine Wahrscheinlichkeit anzugeben, mit der dieses Ereignis vermutlich eintritt.

Im Folgenden werden Wahrscheinlichkeitsprognosen für Kreditausfälle betrachtet. Eine Wahrscheinlichkeitsprognose ist hier also eine prognostizierte Wahrscheinlichkeit für das Ereignis, dass ein Kreditnehmer bis zu einem gewissen Zeitpunkt bzw. in einer festgelegten Periode, seinen finanziellen Verpflichtungen nicht nachkommen kann und daher ausfällt. Die Situation ließe sich aber auch auf Vorhersagen anderer binärer Ereignisse übertragen, wie beispielsweise dem Eintritt eines Versicherungsschadens oder der Insolvenz eines Unternehmens.

Die Grundannahmen sind wie folgt: Eine Ausfallwahrscheinlichkeit ist als Eigenschaft einer Schuldnermenge zu verstehen. Die Schuldnerpopulation, also die Menge aller betrachteten Schuldner, legt die Gesamtausfallwahrscheinlichkeit $p_1 \in (0, 1)$ fest. Liegen von einem einzelnen Kreditnehmer (kurz: Kredit) aus dieser Population keine weiteren Informationen vor, so entspricht die Ausfallwahrscheinlichkeit dieses Kredits gerade der Gesamtausfallwahrscheinlichkeit (a priori). Ein Wahrscheinlichkeitsprognostiker hat jedoch nicht das Ziel, die Gesamtausfallwahrscheinlichkeit zu schätzen und diese jedem Kredit zuzuordnen. Diese Prognose hätte nämlich keinen Wert, da sich Ausfälle und Nichtausfälle anhand dieser Prognose, der sogenannten Trivialprognose, nicht unterscheiden lassen. Die Aufgabe eines Prognostikers liegt vielmehr darin, anhand von weiteren Informationen des Individuums, wie dem Job, dem Vermögen, dem Wohnort etc. ein Rating bzw. eine Wahrscheinlichkeitsprognose zu erstellen und so Klassen von Krediten zu bilden. Die Wahrscheinlichkeit für den Ausfall des Kredits sollte dann von der so erstellten Prognose abhängig sein, d. h. „gut“ bewertete Kredite sollten seltener ausfallen als „schlecht“ bewertete. Das Beste ist es, lediglich zwei Schuldnerklassen zu bilden und a priori alle Nichtausfälle mit einer 0 %- und alle Ausfälle mit einer 100 %-Ausfallprognose zu bewerten. Dies ist in der Praxis zwar kaum möglich, jedoch möchte ein Prognostiker diesem Ideal möglichst gut entsprechen. Insbesondere interessiert hier also die bedingte Ausfallwahrscheinlichkeit gegeben einer Prognose bzw. Ratingklasse.

Damit ist auch klar, dass zwei Prognostiker, die den gleichen Kredit mit verschiedenen Ausfallprognosen versehen, eventuell beide richtig (kalibriert) prognostizieren. Je nachdem, auf welchem Informationsstand die Ausfallprognose beruht, können sich verschiedene bedingte Ausfallwahrscheinlichkeiten für ein und denselben Kredit ergeben.

Definition 2.1.

- (i) Die **Ausfallvariable** $\Theta \sim P^\Theta = \text{Bin}(1, p_1)$ sei eine mit $p_1 \in (0, 1)$ bernoulliverteilte Zufallsvariable, die den Ausfall ($\Theta = 1$) oder den Nichtausfall ($\Theta = 0$) eines Kredits anzeigt.
- (ii) Die diskrete **Prognoseverteilung** P^X habe die Zähldichte q und den Träger $\mathbb{W} := \{a_1, \dots, a_k\}$ mit $0 \leq a_1 < a_2 < \dots < a_k \leq 1$ und $k \in \mathbb{N}$. Hierbei ist \mathbb{W} die Menge der möglichen Wahrscheinlichkeitsprognosen, aus der ein Prognostiker seine Prognosen wählen kann.
- (iii) Ein **Wahrscheinlichkeitsprognostiker** ist über die gemeinsame Verteilung $P^{\Theta, X}$ von Ausfallvariable und Prognose definiert. Diese bivariate Verteilung habe die Wahrscheinlichkeitsfunktion $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ mit

$$r(\theta, a_j) := P(\Theta = \theta, X = a_j), \quad \theta \in \{0, 1\}, \quad j = 1, \dots, k,$$

der gemeinsamen Wahrscheinlichkeit, dass ein Kredit ausfällt bzw. nicht ausfällt und die Prognose $a_j \in \mathbb{W}$ erhält.

Beispielsweise macht der US Wetterdienst Prognosen für die Regenwahrscheinlichkeit, die einem Vielfachen von 10 % entsprechen, d. h. es ist $k = 11$. Nach Tabelle B.1 im Anhang besitzt Moody's 21 Ratingklassen ($k = 21$).

Durch Betrachtung der Randverteilungen von $P^{\Theta, X}$ werden sofort zwei grundlegende Eigenschaften offensichtlich: Die Gesamtausfallwahrscheinlichkeit ergibt sich durch $p_1 = \sum_{j=1}^k r(1, a_j)$ und für die Zähldichte der Prognosen folgt entsprechend $q(a_j) = r(0, a_j) + r(1, a_j)$, $j = 1, \dots, k$. Tabelle 2.1 fasst diese Situation zusammen.

Tabelle 2.1: Bivariate Verteilung der Ausfallvariablen Θ und der Ausfallprognose X mit den zugehörigen Randverteilungen.

		X				Σ
		a_1	a_2	\dots	a_k	
Θ	0	$r(0, a_1)$	$r(0, a_2)$	\dots	$r(0, a_k)$	$1 - p_1$
	1	$r(1, a_1)$	$r(1, a_2)$	\dots	$r(1, a_k)$	p_1
Σ		$q(a_1)$	$q(a_2)$	\dots	$q(a_k)$	1

Für einen Wahrscheinlichkeitsprognostiker ist es unter anderem erstrebenswert, wenn P^X und P^Θ eine hohe (positive) Abhängigkeit aufweisen, denn Ausfälle sollten selbstredend abhängig von der Prognose sein und genauso umgekehrt. Von besonderem Interesse sind daher die bedingten Verteilungen $P^{\Theta|X}$ und $P^{X|\Theta}$. Dabei ist $P^{\Theta|X}$ die bedingte Verteilung der Ausfallvariablen gegeben der Prognose, die für $q(a_j) > 0$ definiert ist, mit Zähldichte

$$p(\theta | a_j) := \frac{r(\theta, a_j)}{q(a_j)}, \quad \theta \in \{0, 1\}, \quad j = 1, \dots, k.$$

Und ebenso ist $P^{X|\Theta}$ die bedingte Verteilung der Prognose gegeben der Ausfallvariablen für $p_1 \in (0, 1)$ definiert durch die Zähldichte

$$q(a_j | \theta) := \frac{r(\theta, a_j)}{\sum_{i=1}^k r(\theta, a_i)}, \quad \theta \in \{0, 1\}, \quad j = 1, \dots, k.$$

Neben der Abhängigkeit zwischen P^X und P^Θ , welche später unter dem Begriff der Trennschärfe eingehend untersucht wird, sollten die Prognosen eines Wahrscheinlichkeitsprognostikers kalibriert sein. Dies ist der Fall, wenn in einer mit der Ausfallwahrscheinlichkeit x % bewerteten Menge von Krediten, die wahre Ausfallwahrscheinlichkeit bei x % liegt. Folgend wird diese zentrale Eigenschaft über die bedingte Verteilung der Ausfallvariable gegeben der Prognose definiert, also über $P^{\Theta|X}$.

Definition 2.2. Die Prognosen eines Wahrscheinlichkeitsprognostikers mit bivariater Wahrscheinlichkeitsfunktion $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ heißen

(i) **kalibriert**, wenn für die Zähldichte von $P^{\Theta|X}$ gilt:

$$p(1 | a_j) = \frac{r(1, a_j)}{q(a_j)} = a_j \quad \forall j = 1, \dots, k \text{ mit } q(a_j) > 0.$$

(ii) **semikalibriert**, falls $p(1 | \cdot)$ monoton steigend ist (Vardeman & Meeden, 1983).

Sind die Prognosen eines Wahrscheinlichkeitsprognostikers (semi-)kalibriert, so wird auch der Prognostiker als (semi-)kalibriert bezeichnet.

Die Semikalibrierung ist eine schwächere Forderung als die Kalibrierung selbst. Siehe Krämer (2017) für eine ausführliche Untersuchung von ROC- und Lorenzkurven bei semikalibrierten Wahrscheinlichkeitsprognosen.

Die Kalibrierung eines Prognostikers hat weitreichende Folgen für dessen bivariate Verteilung von Ausfällen und Prognosen, wie das folgende Resultat zeigt:

Korollar 2.3. Gegeben sei ein kalibrierter Wahrscheinlichkeitsprognostiker mit bivariater Wahrscheinlichkeitsfunktion $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ und einer Gesamtausfallwahrscheinlichkeit von $0 < p_1 < 1$.

(i) Jede der drei folgenden Verteilungen legt die bivariate Verteilung von Ausfallvariable und Prognose $P^{\Theta, X}$ eindeutig fest: P^X , $P^{X|\Theta=0}$ und $P^{X|\Theta=1}$.

(ii) Es gilt $a_j r(0, a_j) = (1 - a_j) r(1, a_j)$ für alle $j = 1, \dots, k$ und damit insbesondere $r(1, 0) = r(0, 1) = 0$, d. h. die Prognosen 0 und 1 sind mit Wahrscheinlichkeit Eins korrekt.

- (iii) Der Erwartungswert der Wahrscheinlichkeitsprognosen entspricht der Gesamtausfallwahrscheinlichkeit, d. h. $\sum_{j=1}^k a_j q(a_j) = p_1$. (Die Randverteilungen der bivariaten Verteilung $P^{\Theta, X}$ haben gleiche Erwartungswerte.)
- (iv) Für $q(1) \neq p_1$ lässt sich die Gesamtausfallwahrscheinlichkeit über die Erwartungswerte der bedingten Prognoseverteilungen bestimmen:

$$p_1 = \frac{\sum_{j=1}^k a_j q(a_j | 0)}{1 + \sum_{j=1}^k a_j q(a_j | 0) - \sum_{j=1}^k a_j q(a_j | 1)}.$$

Beweis: Aus der Definition der Verteilungen und der Definition 2.2 folgen direkt (i) und (ii).

- (iii) Für $a_j > 0$ gilt wegen der Kalibrierung $q(a_j) = \frac{r(1, a_j)}{a_j}$, damit folgt

$$\sum_{j=1}^k a_j q(a_j) = \sum_{\{a_j > 0\}} a_j q(a_j) = \sum_{\{a_j > 0\}} a_j \frac{r(1, a_j)}{a_j} \stackrel{(ii)}{=} \sum_{j=1}^k r(1, a_j) = p_1.$$

- (iv) Die Behauptung folgt mit (ii):

$$\begin{aligned} p_1 &\stackrel{(ii)}{=} \sum_{j=1}^k a_j q(a_j) = \sum_{j=1}^k a_j \frac{r(0, a_j)}{1 - p_1} (1 - p_1) + \sum_{j=1}^k a_j \frac{r(1, a_j)}{p_1} p_1 \\ &= \sum_{j=1}^k a_j q(a_j | 0) (1 - p_1) + \sum_{j=1}^k a_j q(a_j | 1) p_1 \\ &\stackrel{q(1) \neq p_1}{\iff} p_1 = \frac{\sum_{j=1}^k a_j q(a_j | 0)}{1 + \sum_{j=1}^k a_j q(a_j | 0) - \sum_{j=1}^k a_j q(a_j | 1)} \end{aligned}$$

□

Gneiting & Ranjan (2013) definieren marginale Kalibrierung für stetige Verteilungsvorhersagen. Die Eigenschaft $\sum_{j=1}^k a_j q(a_j) = p_1$ aus Korollar 2.3 (iii) ist dem ähnlich und könnte als marginale Kalibrierung für Vorhersagen bei binären Ereignissen definiert werden. Dass diese Eigenschaft jedoch nicht hinreichend für einen „guten“ Prognostiker ist, wird bei Betrachtung der Trivialprognose deutlich:

Definition 2.4. *Gegeben sei ein Wahrscheinlichkeitsprognostiker mit bivariater Wahrscheinlichkeitsfunktion $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ und einer Gesamtausfallwahrscheinlichkeit von $0 < p_1 < 1$. Die Prognose des Wahrscheinlichkeitsprognostikers heißt **Trivialprognose**, falls $q(p_1) = 1$.*

Offensichtlich ist die Trivialprognose kalibriert, da

$$p(1 | p_1) = \frac{r(1, p_1)}{q(p_1)} = r(1, p_1) = p_1.$$

Jedoch eignet sich die Trivialprognose nicht zur Unterscheidung zwischen Ausfällen und Nichtausfällen, weshalb sie die schlechteste aller kalibrierten Vorhersagen darstellt (Gegenstand des Kapitels 3.3). Im Gegensatz dazu steht der Prognostiker, der den Ausfall bzw. Nichtausfall stets sicher vorhersagt:

Definition 2.5. *Gegeben sei ein Wahrscheinlichkeitsprognostiker mit bivariater Wahrscheinlichkeitsfunktion $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ und einer Gesamtausfallwahrscheinlichkeit von $0 < p_1 < 1$. Die Prognose des Wahrscheinlichkeitsprognostikers heißt **perfekte Prognose**, falls $r(1, 1) = p_1$ und $r(0, 0) = 1 - p_1$ gilt.*

Bemerkung 2.6. ***Trennschärfe** ist die Fähigkeit zwischen Ausfällen und Nichtausfällen in einer Schuldnerpopulation anhand der Wahrscheinlichkeitsprognosen*

unterscheiden zu können. Offensichtlich besitzt die Trivialprognose keinerlei Trennschärfe, da alle Kredite dieselbe Ausfallwahrscheinlichkeit (die Gesamtausfallwahrscheinlichkeit) zugeordnet bekommen. Die perfekte Prognose hat hingegen maximale Trennschärfe. Auch jeder andere Wahrscheinlichkeitsprognostiker, der exakt zwischen Ausfällen und Nichtausfällen diskriminiert, hat maximale Trennschärfe. Jedoch ist von denen nur der mit der perfekten Prognose kalibriert.

Ist nicht die theoretische bivariate Verteilung eines Prognostikers relevant, sondern eine Population von $n \in \mathbb{N}$ bewerteten Krediten, so können die Begriffe Kalibrierung, Trivialprognose und perfekte Prognose auch für das Vorliegen einer bewerteten Schuldnerpopulation definiert werden. Diese beiden Situationen sind inhaltlich zwar eng miteinander verbunden, methodisch jedoch strikt voneinander zu trennen. Die unterschiedlichen Sichtweisen und Methoden werden folgend motiviert.

Seien für jeden Kredit in der Population ein Rating $r_i \in \{R_1, \dots, R_k\}$ und die Ausfallvariable $\theta_i \in \{0, 1\}$ gegeben, $i = 1, \dots, n$. Dabei ist n die Größe der Population und θ_i ist für $i = 1, \dots, n$ definiert als

$$\theta_i := \begin{cases} 1, & \text{Kredit } i \text{ fällt aus} \\ 0, & \text{Kredit } i \text{ fällt nicht aus.} \end{cases}$$

Weiterhin sei jedes Rating R_j mit einer Ausfallwahrscheinlichkeit a_j verknüpft. Diese Wahrscheinlichkeit ist entweder bekannt oder sie wird aus $\vec{r} := (r_1, \dots, r_n)'$ und $\vec{\theta} := (\theta_1, \dots, \theta_n)'$ geschätzt. Folgend wird daher mit einer geschätzten Wahrscheinlichkeitsprognose $x_i \in \widehat{W} := \{\hat{a}_1, \dots, \hat{a}_k\}$ gearbeitet. Beispielsweise könnte \hat{a}_j der empirische Ausfallanteil in Ratingklasse R_j oder die entsprechende Schätzung mittels der monotonen Regression (Kapitel 6) sein.

Definition 2.7. Gegeben sei eine Population von Krediten mit Ausfällen $\vec{\theta} \in \{0, 1\}^n$ und einem Gesamtausfallanteil von $\hat{p}_1 := \frac{1}{n} \sum_{i=1}^n \theta_i \in [0, 1]$.

(i) Die Wahrscheinlichkeitsprognosen $\vec{x} := (x_1, \dots, x_n)' \in \widehat{\mathbb{W}}^n$ heißen **empirisch kalibriert** für die Ausfälle $\vec{\theta}$, wenn für alle $j = 1, \dots, k$ mit $\sum_{i=1}^n \mathbb{1}_{\{\hat{a}_j\}}(x_i) > 0$ gilt:

$$\frac{\sum_{i=1}^n \theta_i \mathbb{1}_{\{\hat{a}_j\}}(x_i)}{\sum_{i=1}^n \mathbb{1}_{\{\hat{a}_j\}}(x_i)} = \hat{a}_j.$$

(ii) Der Vektor der Wahrscheinlichkeitsprognosen $\vec{x}_T := \vec{x} \in \widehat{\mathbb{W}}^n$ heißt **Trivialprognose**, wenn $x_i = \hat{p}_1$ für alle $i = 1, \dots, n$.

(iii) Der Vektor der Wahrscheinlichkeitsprognosen $\vec{x}_P := \vec{x} \in \widehat{\mathbb{W}}^n$ heißt **perfekte Prognose**, wenn $x_i = \theta_i$ für alle $i = 1, \dots, n$.

Ein kalibrierter Wahrscheinlichkeitsprognostiker (Definition 2.2) liefert im Allgemeinen keine empirisch kalibrierten Wahrscheinlichkeitsprognosen (Definition 2.7). Trotzdem sind Resultate für empirisch kalibrierte Wahrscheinlichkeitsprognosen praxisrelevant und hilfreich in der Anwendung von Kapitel 7. Wird die Ausfallwahrscheinlichkeit einer Ratingklasse beispielsweise durch den zugehörigen empirischen Ausfallanteil oder das Verfahren der sogenannten monotonen Regression geschätzt (Kapitel 6), dann sind die so generierten Wahrscheinlichkeitsprognosen (ex post) empirisch kalibriert.

Kapitel 3

Skalarwertige Qualitätskriterien

In diesem Kapitel werden skalarwertige Qualitätskriterien für Wahrscheinlichkeitsprognosen vorgestellt und untersucht. Dabei wird zwischen Kriterien unterschieden, welche auf grafischen Methoden beruhen und den sogenannten Score-Regeln.

3.1 Gini-Koeffizient und AUC

Folgende Qualitätskriterien basieren auf grafischen Methoden, welche lediglich die Verteilung der Prognosen und nicht die Prognosen selbst benutzen. Daher können die Prognosen hier auch als Ratings auf der Ordinalskala verstanden werden.

Die Lorenzkurve von Max Otto Lorenz (1905) ist auch heute noch ein Standardinstrument zur Visualisierung der Ungleichheit von Einkommens- und Vermögensverteilungen einer Gesellschaft. So ist beispielsweise, der auf der Lorenzkurve basierende Gini-Koeffizient eine wichtige Maßzahl im Armuts- und Reichtumsbericht der Bundesregierung (Bundesministerium für Arbeit und Soziales, 2017). Darüber hinaus ist die Lorenzkurve in den jährlich erscheinenden Berichten großer Ratingagenturen, wie zum Beispiel in dem von Standard & Poor's zu finden (Vazza & Kraemer, 2017).

Hier wird die Lorenzkurve im Kontext der Wahrscheinlichkeitsprognosen definiert.

Definition 3.1. Sei $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ die bivariate Wahrscheinlichkeitsfunktion eines Wahrscheinlichkeitsprognostikers und $p_1 > 0$ die Gesamtausfallwahrscheinlichkeit. Die **Lorenzkurve** $L : [0, 1] \rightarrow [0, 1]$ ist der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j q(a_i), \sum_{i=1}^j q(a_i | 1) \right)', \quad j = 0, \dots, k,$$

wobei $\sum_{i=1}^0 q(a_i) := \sum_{i=1}^0 q(a_i | 1) := 0$ ist.

Für ein $u \in [0, 1]$ besagt die Lorenzkurve, dass die $u \cdot 100$ % bestbewerteten Kredite einen Anteil von $L(u) \cdot 100$ % an den Ausfällen besitzen. Damit eignet sich die Lorenzkurve zur Visualisierung der Trennschärfe eines Prognostikers.

Die Lorenzkurve verbindet die Punkte $(0, 0)'$ und $(1, 1)'$. Sie ist auf ihrem Definitionsbereich konvex, falls der Prognostiker semikalibriert ist (Krämer & Posch, 2015).

Beispiel 3.2. Abbildung 3.1 (a) zeigt die Lorenzkurve eines kalibrierten Prognostikers für den gilt: Ein Fünftel aller Kredite erhalten die beste Ausfallprognose $a_1 = 0$ und fallen nicht aus, d. h. $(0.2, 0)'$ ist ein Punkt auf der Lorenzkurve. Weiter besitzen drei Fünftel der Kredite die Prognosen $a_1 = 0$ oder $a_2 = 0.25$, und auf diese Kredite entfällt ein Viertel aller Ausfälle, d. h. $(0.6, 0.25)'$ ist ein Punkt auf der Lorenzkurve. Zusammen mit den (offensichtlichen) Punkten $(0, 0)'$ und $(1, 1)'$ ergibt sich so der gesamte Polygonzug der Lorenzkurve.

Bemerkung 3.3.

- (i) Während bei der Ungleichheitsmessung von Einkommen eine Lorenzkurve nahe der Winkelhalbierenden erstrebenswert ist (geringe Ungleichheit der Einkommen), ist bei der Bewertung von Ausfallprognosen eine möglichst durchgeboogene Lorenzkurve wünschenswert (hohe Trennschärfe des Prognostikers).
- (ii) Die hier definierte Lorenzkurve ist ein P-P-Plot (probability-probability-plot), da sie zwei Verteilungsfunktionen gegeneinander abbildet (Gnanadesikan, 1997, S. 228). Die Lorenzkurve zeigt somit an, wie die Prognosen der ausgefallenen Kredite in der Menge der Prognosen liegen.
- (iii) Ebenso wäre ein Polygonzug denkbar, der anstatt, der auf den Ausfall bedingten Prognoseverteilung $q(\cdot | 1)$, die auf den Nichtausfall bedingte Prognoseverteilung $q(\cdot | 0)$ verwendet. Dieser Polygonzug verläuft also durch die Punkte

$$\left(\sum_{i=1}^j q(a_i), \sum_{i=1}^j q(a_i | 0) \right)', \quad j = 0, \dots, k,$$

und wird zumeist als **Lorenzkurve der Nichtausfälle** bezeichnet (Krämer & Posch, 2015). Sie zeigt an, wie die Prognosen der nicht ausgefallenen Kredite in der Menge der Prognosen liegen.

In der Bankenwelt wird häufig statt der Lorenzkurve das sogenannte Cumulative Accuracy Profile (CAP) verwendet (Engelmann et al., 2003). Das **Cumulative Accuracy Profile** $CAP : [0, 1] \rightarrow [0, 1]$ ist der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j q(a_{k-i+1}), \sum_{i=1}^j q(a_{k-i+1} | 1) \right)', \quad j = 0, \dots, k,$$

sodass sich dieses von der Lorenzkurve lediglich durch eine veränderte Summationsreihenfolge unterscheidet. Alle für die Lorenzkurve hergeleiteten Resultate aus Kapitel 5 und 6 lassen sich ohne viel Arbeit auch für das Cumulative Accuracy Profile formulieren.

Auf Basis des CAP wird das **Accuracy Ratio** als $AR := A_R/A_P$ definiert, wobei A_R die Fläche zwischen CAP und Winkelhalbierender und A_P die Fläche zwischen Winkelhalbierender und der CAP des perfekten Prognostikers ist.

Das bekannteste skalarwertige Qualitätskriterium, welches auf der Lorenzkurve beruht, ist sicherlich der Gini-Koeffizient.

Definition 3.4. Der **Gini-Koeffizient** $G \in [0, 1]$ zur Lorenzkurve L ist das Doppelte der Fläche zwischen Winkelhalbierender und Lorenzkurve

$$G := 1 - 2 \int_0^1 L(u) du.$$

Ein hoher Gini-Koeffizient ist bei der Bewertung von Kreditausfallprognosen wünschenswert, vgl. Bemerkung 3.3 (i) und Abbildung 3.1 (a).

Die ROC-Kurve veranschaulicht, ähnlich wie die Lorenzkurve, die Trennschärfe eines Prognostikers.

Definition 3.5. Sei $r : \{0, 1\} \times \mathbb{W} \rightarrow [0, 1]$ die bivariate Wahrscheinlichkeitsfunktion eines Wahrscheinlichkeitsprognostikers und $0 < p_1 < 1$.

(i) Die **ROC-Kurve** (receiver operating characteristic curve) $ROC : [0, 1] \rightarrow [0, 1]$ ist der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j q(a_{k-i+1} | 0), \sum_{i=1}^j q(a_{k-i+1} | 1) \right)', \quad j = 0, \dots, k.$$

(ii) Die **AUC** (area under the curve) ist die Fläche unterhalb der ROC-Kurve.

Die ROC-Kurve vergleicht die auf den Ausfall bedingte Prognoseverteilung $P^{X|\Theta=1}$ und die auf den Nichtausfall bedingte Prognoseverteilung $P^{X|\Theta=0}$ bezüglich ihrer Lage, siehe Abbildung 3.1 (b). Für einen Prognostiker mit maximaler Trennschärfe sind diese beiden Verteilungen in ihrer Lage vollkommen separiert. Für die Trivialprognose sind die Verteilungen identisch. Allgemein gilt: Für ein $u \in [0, 1]$ existiert ein Rating, sodass $u \cdot 100$ % der Nichtausfälle und $ROC(u) \cdot 100$ % der Ausfälle dieses oder ein „schlechteres“ Rating erhalten. Demnach ist eine stark (nach oben) durchgebogene ROC-Kurve für einen Prognostiker wünschenswert.

Über die Beziehung $AR = 2AUC - 1$ hängen die Fläche unter der ROC-Kurve und das Accuracy Ratio direkt voneinander ab (Engelmann et al., 2003).

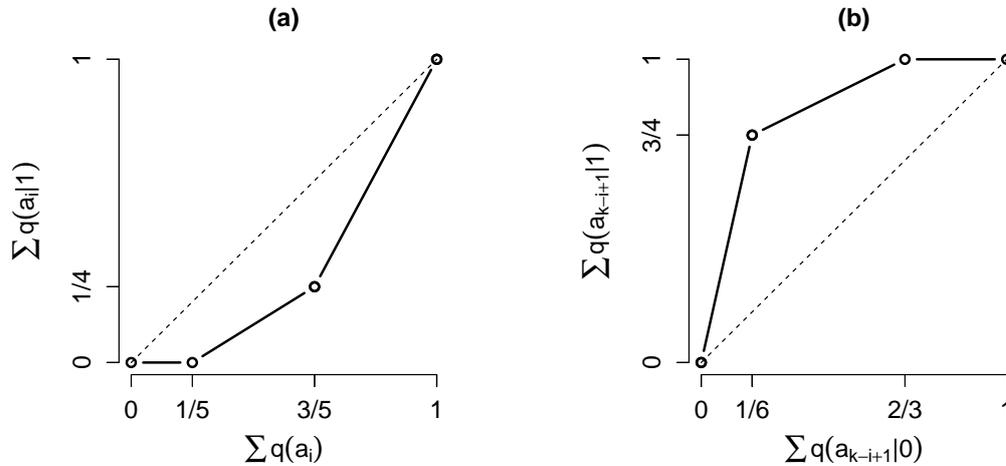


Abbildung 3.1: Lorenzkurve (a) und ROC-Kurve (b) eines kalibrierten Wahrscheinlichkeitsprognostikers mit $W = \{0, 0.25, 0.75\}$ und $q(0) = \frac{1}{5}$, $q(0.25) = q(0.75) = \frac{2}{5}$ auf einer Schuldnerpopulation mit $p_1 = 0.4$. Damit ist $q(0|0) = \frac{1}{3}$, $q(0.25|0) = \frac{1}{2}$, $q(0.75|0) = \frac{1}{6}$ und $q(0.25|1) = \frac{1}{4}$, $q(0.75|1) = \frac{3}{4}$. Die Lorenz- bzw. ROC-Kurve der Trivialprognose entspricht der Winkelhalbierenden (gestrichelte Linie).

3.2 Score-Regeln

Im Gegensatz zu dem Gini-Koeffizienten und der AUC ist bei einer Score-Regel vor allem nicht die theoretische bivariate Verteilung des Prognostikers relevant, sondern eine Population von bewerteten Krediten (vorhergesagte Wahrscheinlichkeiten). Für jeden Kredit in der Population sind eine prognostizierte Ausfallwahrscheinlichkeit $x_i \in [0, 1]$ und die Ausfallvariable $\theta_i \in \{0, 1\}$ gegeben, $i = 1, \dots, n$. Dabei ist n die Größe der Population und θ_i ist für $i = 1, \dots, n$ definiert als

$$\theta_i := \begin{cases} 1, & \text{Kredit } i \text{ fällt aus} \\ 0, & \text{Kredit } i \text{ fällt nicht aus.} \end{cases}$$

Sei $\vec{x} := (x_1, \dots, x_n)'$ der Vektor der Prognosen und $\vec{\theta} := (\theta_1, \dots, \theta_n)'$ der Vektor der Ausfälle. Eine Score-Regel beurteilt die Qualität der Prognosen, indem sie den Abstand zwischen den Prognosen in \vec{x} und den Ausfällen in $\vec{\theta}$ misst.

Definition 3.6. Sei $\vec{x} \in [0, 1]^n$ der Vektor der Wahrscheinlichkeitsprognosen und $\vec{\theta} \in \{0, 1\}^n$ der Vektor der Ausfälle.

(i) Eine **Score-Funktion** ist eine Abbildung $s : \{0, 1\} \times [0, 1] \rightarrow \overline{\mathbb{R}}$.

(ii) Eine Score-Funktion heißt **regulär**, falls für $\theta \in \{0, 1\}$ und $x \in [0, 1]$ gilt:

$$s(\theta, x) \in \begin{cases} \mathbb{R} \cup \{\infty\}, & x = 1 - \theta \\ \mathbb{R}, & \text{sonst.} \end{cases}$$

(iii) Eine **Score-Regel** (auch **mittlerer Score** genannt) ist eine Abbildung

$S : \{0, 1\}^n \times [0, 1]^n \rightarrow \overline{\mathbb{R}}$ zu einer regulären Score-Funktion s mit:

$$S(\vec{\theta}, \vec{x}) := \frac{1}{n} \sum_{i=1}^n s(\theta_i, x_i).$$

Eine Score-Funktion heißt **nulloptimal**, wenn $s(\theta, x) = 0 \iff x = \theta$ und sie wird **nichtnegativ** genannt, wenn $s(\theta, x) \geq 0 \forall x \in [0, 1], \theta \in \{0, 1\}$.

Annahme 3.7. In dieser Arbeit werden ausschließlich Score-Regeln mit regulärer und nichtnegativer Score-Funktion betrachtet. Außerdem seien Score-Regeln negativ orientiert, d. h. ein Prognostiker möchte den Wert einer Score-Regel minimieren.

Die Menge aller regulären nichtnegativen nulloptimalen Score-Regeln sei mit \mathbb{L}_{rnn} bezeichnet.

Die Score-Funktion misst die Abweichung zwischen der Prognose und der Ausfallvariable. Hierbei ist $s(\theta, x)$ auch als Verlust interpretierbar, der entsteht, wenn x prognostiziert und θ beobachtet wird.

Die Auswahl an Score-Regeln in der Literatur ist groß. Zu den populärsten Score-Regeln gehören der *Brier-Score*, der *Logarithmus-Score* sowie der *sphärische Score*. Definition 3.8 gibt einen Auszug von Score-Funktionen bekannter Score-Regeln.

Definition 3.8. *Eine Score-Regel S , wie in Definition 3.6, heißt ...*

(i) **Brier-Score**, wenn $s(\theta, x) = (x - \theta)^2$ (Brier, 1950).

(ii) **Logarithmus-Score**, wenn

$$s(\theta, x) = \begin{cases} -\ln(|x + \theta - 1|), & x \neq 1 - \theta \\ \infty, & x = 1 - \theta \end{cases} \quad (\text{Good, 1952}).$$

(iii) **sphärischer Score**, wenn $s(\theta, x) = 1 - |x + \theta - 1| \left(\sqrt{x^2 + (1 - x)^2} \right)^{-1}$ (Savage, 1971).

(iv) **absoluter Score**, wenn $s(\theta, x) = |x - \theta|$.

Die Score-Regeln aus Definition 3.8 sind nulloptimal, nichtnegativ und regulär.

Dawid & Musio (2014) geben einen Überblick über Score-Regeln bei allgemeinen Verteilungsprognosen. So ist beispielsweise der Logarithmus-Score auch definiert über die Score-Funktion

$$s(\theta, Q) = -\ln p(\theta), \quad (3.1)$$

wobei p die Dichte der prognostizierten Verteilung Q und θ eine Realisation der zu prognostizierenden Verteilung ist. Mit $Q = \text{Bin}(1, x)$ folgt für $x \neq 1 - \theta$ sofort die Formulierung aus Definition 3.8: $-\ln(x\theta + (1 - x)(1 - \theta)) = -\ln(|x + \theta - 1|)$.

Definition 3.9. Sei $\Theta \sim \text{Bin}(1, \pi)$ eine bernoulliverteilte Zufallsvariable mit Parameter $\pi \in [0, 1]$. Eine Score-Funktion s heißt **anreizerhaltend** (proper), falls

$$E(s(\Theta, \pi)) \leq E(s(\Theta, x)) \quad \text{für alle } x \in [0, 1]. \quad (3.2)$$

Die Score-Funktion s heißt **streng anreizerhaltend** (strictly proper), falls in Gleichung (3.2) für alle $x \in [0, 1] \setminus \{\pi\}$ sogar echte Ungleichheit besteht.

Falls ein Wahrscheinlichkeitsprognostiker $\Theta \sim \text{Bin}(1, \pi)$ vermutet, dann sollte er auch den Anreiz haben diese Prognose zu äußern. Daher belohnt eine anreizerhaltende Score-Regel in dieser Situation die Prognose $x = \pi$. Diese Prognose minimiert dann den erwarteten Verlust (Bröcker & Smith, 2007).

Bemerkung 3.10. Für $\Theta \sim \text{Bin}(1, \pi)$, $\pi \in [0, 1]$ und $x \in [0, 1]$ ist die numerische Zufallsvariable $s(\Theta, x)$ quasiintegrierbar, da $s(\cdot, \cdot) \geq 0$ ist. Daher existiert der Erwartungswert von $s(\Theta, x)$, auch wenn dieser eventuell nicht endlich ist (Bauer, 2001, S. 14-16). Allerdings gilt stets $E(s(\Theta, \pi)) < \infty$, da s regulär ist (Annahme 3.7).

Wird eine Score-Regel als *anreizerhaltend*, *regulär*, *nulloptimal* oder *nichtnegativ* bezeichnet, so ist damit gemeint, dass die zugehörige Score-Funktion die entsprechenden Eigenschaften aufweist.

Das folgende Beispiel zeigt, dass der Brier-Score im Gegensatz zum absoluten Score streng anreizerhaltend ist.

Beispiel 3.11.

(i) *Der Brier-Score ist streng anreizerhaltend, da für eine feste aber beliebige Prognose $x \in [0, 1]$ und ein bernoulliverteiltes Θ gilt*

$$E((x - \Theta)^2) = x^2 + E(\Theta)(1 - 2x). \quad (3.3)$$

Betrachtung der ersten beiden Ableitungen von Gleichung (3.3) nach x liefert, dass die eindeutige Minimalstelle bei $x = E(\Theta)$ liegt. Dies zeigt die Gültigkeit von Gleichung (3.2) (mit echter Ungleichheit) für den Brier-Score.

(ii) *Für den absoluten Score gilt $E(|x - \Theta|) = x + E(\Theta)(1 - 2x)$. Dieser Ausdruck verhält sich monoton in x , sodass das Minimum im Allgemeinen nicht durch $x = E(\Theta)$ erreicht wird. Genauer: Falls $E(\Theta) < 0.5$, so ist $x = 0$ Minimalstelle und falls $E(\Theta) > 0.5$, so ist diese $x = 1$. Für $E(\Theta) = 0.5$ ist jedes $x \in [0, 1]$ eine Minimalstelle. Der absolute Score ist also nicht anreizerhaltend, da Gleichung (3.2) hier nicht erfüllt ist.*

Parry et al. (2012) betrachten Score-Regeln, die nicht nur von der Dichte der prognostizierten Verteilung an der Stelle θ abhängen (vgl. Gleichung (3.1)), sondern auch von den ersten Ableitungen der prognostizierten Dichte (lokal anreizerhaltende Score-Regeln höherer Ordnung). Dabei betrachten sie insbesondere auch die erwartete Score-Funktion, welche im Folgenden definiert wird.

Definition 3.12. Sei $\Theta \sim \text{Bin}(1, \pi)$ mit $\pi \in [0, 1]$ sowie S eine Score-Regel mit Score-Funktion s . Die Funktion $g : [0, 1] \rightarrow \mathbb{R}$ mit

$$g(\pi) := E(s(\Theta, \pi)) = \pi s(1, \pi) + (1 - \pi)s(0, \pi)$$

heißt **erwartete Score-Funktion** von S .

Bemerkung 3.13. In Definition 3.9, Beispiel 3.11 und Definition 3.12 wird der Erwartungswert der Score-Funktion in Abhängigkeit einer festen Wahrscheinlichkeitsprognose bestimmt. Mit einer Prognoseverteilung gemäß Definition 2.1 ergibt sich der **Erwartungswert der Score-Funktion** zu:

$$S(\Theta, X) := E(s(\Theta, X)) = \sum_{j=1}^k \sum_{l=0}^1 s(l, a_j)r(l, a_j).$$

Ist der Prognostiker kalibriert, so lässt sich der Ausdruck vereinfachen zu

$$S(\Theta, X) = E(g(X)) = \sum_{j=1}^k g(a_j)q(a_j),$$

was sofort mit Definition 2.2 und 3.12 folgt. Ist der Wahrscheinlichkeitsprognostiker kalibriert, so hängt der Erwartungswert einer Score-Regel also allein von der Verteilung der Prognosen ab (DeGroot & Eriksson, 1985). Diese Beobachtung ist mit Korollar 2.3 (i) wenig überraschend.

Tabelle 3.1 gibt einen Überblick über die Score-Funktionen und erwarteten Score-Funktionen verschiedener Score-Regeln mit ersten und zweiten Ableitungen.

Tabelle 3.1: Score-Funktionen und erwartete Score-Funktionen verschiedener Score-Regeln mit ersten und zweiten Ableitungen für $x \in [0, 1]$, bzw. für den Logarithmus-Score für $x \in (0, 1)$.

Score-Regel	$s(\theta, x)$	$g(x)$	$g'(x)$	$g''(x)$
Brier	$(\theta - x)^2$	$x(1 - x)$	$1 - 2x$	-2
Logarithmus	$-\ln(1 - \theta - x)$	$(x - 1) \ln(1 - x) - x \ln(x)$	$\ln\left(\frac{1-x}{x}\right)$	$\frac{-1}{x(1-x)}$
Absolut	$ x - \theta $	$2x(1 - x)$	$2 - 4x$	-4
Hyberbolisch	$\sinh(x) \sinh(1 - x) + (x - \theta) \sinh(2x - 1)$	$\sinh(x) \sinh(1 - x)$	$-\sinh(2x - 1)$	$-2 \cosh(2x - 1)$
Sphärisch	$1 - x + \theta - 1 \left(\sqrt{x^2 + (1 - x)^2}\right)^{-1}$	$1 - \sqrt{x^2 + (1 - x)^2}$	$\frac{1 - 2x}{\sqrt{x^2 + (1 - x)^2}}$	$\frac{-1}{(x^2 + (1 - x)^2)^{\frac{3}{2}}}$

Nach Savage (1971) hängt eine anreizerhaltende Score-Regel eindeutig von ihrer erwarteten Score-Funktion ab:

Satz 3.14 (Savage, 1971; McCarthy, 1956). *Eine Score-Regel S mit Score-Funktion $s : \{0, 1\} \times [0, 1] \rightarrow \overline{\mathbb{R}}$ ist genau dann anreizerhaltend, wenn*

$$s(\theta, x) = g(x) + (\theta - x)g'(x), \quad (3.4)$$

wobei $g : [0, 1] \rightarrow \mathbb{R}$ eine konkave Funktion und g' ein Subgradient von g ist, d. h. für alle $x, y \in [0, 1]$ gilt $g(y) \leq g(x) + g'(x)(y - x)$. Dabei sei g' reellwertig mit Ausnahme, dass $g'(0) = \infty$ und $g'(1) = -\infty$ zulässig sei.

Ist g differenzierbar auf $(0, 1)$, so ist g' eindeutig und die erste Ableitung von g . Die konkave Funktion g ist dann außerdem die erwartete Score-Funktion von S . Die Score-Regel ist streng anreizerhaltend genau dann, wenn g aus Gleichung (3.4) streng konkav ist.

Bemerkung 3.15. *Score-Regeln können alternativ so definiert werden, dass große Werte der Score-Regel für einen Prognostiker erstrebenswert sind. Satz 3.14 ließe sich dann so formulieren, dass die zu einer (streng) anreizerhaltenden Score-Regel gehörende erwartete Score-Funktion (streng) konvex ist (Gneiting & Raftery, 2007).*

Mit der Savage-Repräsentation wird auch deutlich, dass $s(1, x)$ bzw. $s(0, x)$ der Tangente von g im Punkt $(x, g(x))'$ an der Stelle 1 bzw. 0 entspricht, siehe Abbildung 3.2.

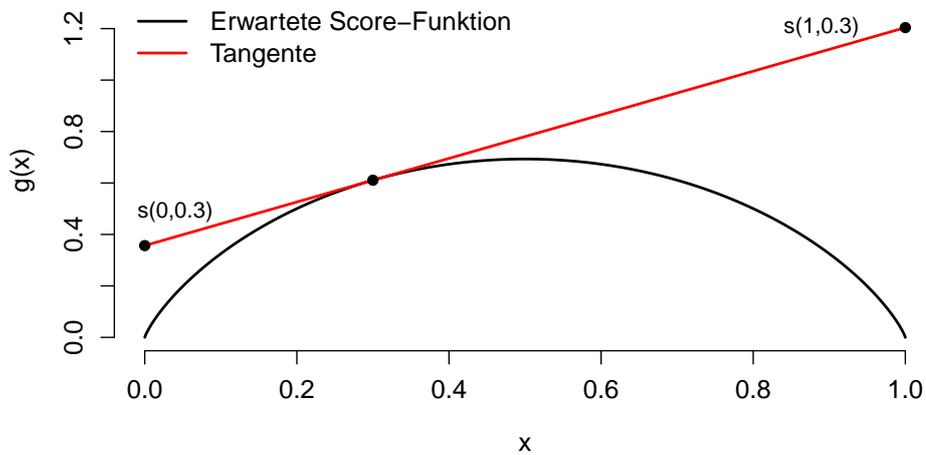


Abbildung 3.2: Erwartete Score-Funktion des Logarithmus-Scores g mit eingezeichneter Tangente an der Stelle 0.3. Der Wert der Score-Funktion $s(0, 0.3) = 0.357$ bzw. $s(1, 0.3) = 1.204$ entspricht dem Wert der Tangente $g(0.3) + (x - 0.3)g'(0.3)$ an der Stelle $x = 0$ bzw. $x = 1$ (Gneiting & Raftery, 2007).

Mit Satz 3.14 können auch (streng) anreizerhaltende Score-Regeln aufgestellt werden, da eine reellwertige (streng) konkave Funktion g eine (streng) anreizerhaltende Score-Regel über Gleichung (3.4) definiert. So wird folgend der hyperbolische Score eingeführt.

Definition 3.16. Die Score-Regel S , wie in Definition 3.6, mit Score-Funktion $s : \{0, 1\} \times [0, 1] \rightarrow \mathbb{R}$ in der Form

$$s(\theta, x) = \sinh(x) \sinh(1 - x) + (x - \theta) \sinh(2x - 1)$$

heißt **hyperbolischer Score**.

Korollar 3.17. Der hyperbolische Score ist streng anreizerhaltend mit erwarteter Score-Funktion $g(x) = \sinh(x) \sinh(1 - x)$.

Beweis: Definiere die erwartete Score-Funktion als

$$g(x) := \sinh(x) \sinh(1-x) \stackrel{(*)}{=} \frac{1}{2}(\cosh(1) - \cosh(2x-1)). \quad (3.5)$$

Dabei gilt $(*)$ wegen der allgemeinen Summenformel für den Kosinus-Hyperbolicus $\forall x, y \in \mathbb{R}$: $\cosh(x) - \cosh(y) = 2 \sinh\left(\frac{x+y}{2}\right) \sinh\left(\frac{x-y}{2}\right)$. Die auf $[0, 1]$ beschränkte erwartete Score-Funktion hat die Ableitungen

$$\begin{aligned} g'(x) &= -\sinh(2x-1), \\ g''(x) &= -2 \cosh(2x-1) < 0, \end{aligned}$$

sodass g streng konkav auf dem Intervall $[0, 1]$ ist. Unter Ausnutzung der Savage-Repräsentation (Gleichungen (3.4)) folgt

$$s(\theta, x) = \sinh(x) \sinh(1-x) + (x-\theta) \sinh(2x-1).$$

Also ist der hyperbolische Score nach Konstruktion über Satz 3.14 streng anreizerhaltend. □

Bemerkung 3.18. *Der hyperbolische Score ist nulloptimal, nichtnegativ und regulär.*

Mit Annahme 3.7 gilt für die erwartete Score-Funktion einer Score-Regel $S \in \mathbb{L}_{rnn}$:

- (i) $g(0) = g(1) = 0$ (folgt mit $s(1, 1) = s(0, 0) = 0$),
- (ii) $g(x) \geq 0 \forall x \in [0, 1]$ (folgt mit $s(\theta, x) \geq 0 \forall x \in [0, 1], \theta \in \{0, 1\}$).

Diese einheitliche Form erleichtert den Vergleich von Score-Regeln und ist dabei wenig restriktiv, da sich Score-Regeln zumeist passend (linear) transformieren lassen. Gneiting & Raftery (2007) betonen, dass anreizerhaltende Score-Regeln bei einer linearen Transformation anreizerhaltend bleiben. Dawid (1998) spricht daher bei solch transformierten Score-Regeln von äquivalenten Score-Regeln.

Die Score-Regeln aus Definition 3.8 werden in der Literatur häufig etwas anders definiert. Beispielsweise ist auch

$$s(\theta, x) = \frac{|x + \theta - 1|}{\sqrt{x^2 + (1 - x)^2}}$$

eine gebräuchliche Definition der Score-Funktion des sphärischen Scores (Krämer & Güttler, 2008). Da es sich dabei jedoch lediglich um eine lineare Transformation der Score-Regel aus Definition 3.8 handelt, werden die verschiedenen Formen (äquivalente Score-Regeln) hier nicht gesondert betrachtet.

Murphy (1973) definiert sogenannte Skill-Scores, die im Gegensatz zu einfachen Score-Regeln, den Gesamtausfallanteil einer Schuldnerpopulation berücksichtigen:

Definition 3.19. Sei S eine Score-Regel und \vec{x}_T bzw. \vec{x}_P der Vektor der *Trivialprognose* bzw. der *perfekten Prognose* für die Ausfälle $\vec{\theta} \in \{0, 1\}^n$ bei einem Gesamtausfallanteil von $\hat{p}_1 \in (0, 1)$. Dann ist der zu S gehörende **Skill-Score** $SK : \{0, 1\}^n \times [0, 1]^n \rightarrow \overline{\mathbb{R}}$ für $\vec{\theta}$ und die Prognosen $\vec{x} \in [0, 1]^n$ definiert durch

$$SK(\vec{\theta}, \vec{x}) := \frac{S(\vec{\theta}, \vec{x}) - S(\vec{\theta}, \vec{x}_T)}{S(\vec{\theta}, \vec{x}_P) - S(\vec{\theta}, \vec{x}_T)}.$$

Ein Skill-Score misst also die Qualität einer Prognose im direkten Vergleich zur Trivialprognose der Schuldnerpopulation. Er nimmt den optimalen Wert Eins an, wenn \vec{x} die perfekte Prognose ist und den Wert Null, falls \vec{x} der Trivialprognose entspricht.

Statt der Trivialprognose kann auch eine beliebige Referenzprognose in der Definition des Skill-Scores verwendet werden (Gneiting & Raftery, 2007). Da jedoch in dieser Arbeit der Vergleich verschiedener Schuldnerpopulationen im Fokus steht, wird an der Trivialprognose als Referenz festgehalten.

3.3 Kalibrierte Prognosen

Dieser Abschnitt untersucht Qualitätskriterien im Falle, dass die Ausfallprognosen kalibriert sind. Folgender Satz zeigt, dass die Trivialprognose im Sinne anreizerhaltender Score-Regeln, die schlechteste aller kalibrierten Prognosen ist.

Satz 3.20. *Sei $P^{\Theta, X}$ die bivariate Verteilung eines kalibrierten Wahrscheinlichkeitsprognostikers gemäß Definition 2.1. Für eine anreizerhaltende Score-Regel S (mit konkaver erwarteter Score-Funktion) gilt dann:*

$$(i) \quad S(\Theta, X) \leq S(\Theta, X_T),$$

$$(ii) \quad S(\Theta, X) \geq S(\Theta, X_P),$$

wobei X_T die Trivialprognose und X_P die perfekte Prognose für Θ ist.

Beweis: Teil (i) folgt sofort mit Bemerkung 3.13 und der Jensen-Ungleichung, da g eine konkave Funktion ist:

$$S(\Theta, X) = E(s(\Theta, X)) = E(g(X)) \leq g(E(X)) = g(p_1) = S(\Theta, X_T).$$

Mit q der Zähldichte von P^X gilt aufgrund der Kalibrierung $p_1 = \sum_{j=1}^k a_j q(a_j)$. Somit ergibt sich (ii):

$$\begin{aligned} S(\Theta, X_P) &= \sum_{j=1}^k a_j q(a_j) g(1) + \sum_{j=1}^k (1 - a_j) q(a_j) g(0) \\ &= \sum_{j=1}^k q(a_j) (a_j g(1) + (1 - a_j) g(0)) \\ &\stackrel{g \text{ konkav}}{\leq} \sum_{j=1}^k q(a_j) g(a_j) = S(\Theta, X). \end{aligned}$$

□

Bemerkung 3.21. Satz 3.20 lässt sich auch für realisierte Werte von Score-Regeln formulieren: Der Wert einer anreizerhaltenden Score-Regel (mit konkaver erwarteter Score-Funktion), also $S(\vec{\theta}, \vec{x})$, ist bei empirisch kalibrierten Wahrscheinlichkeitsprognosen maximal für die Trivialprognose und minimal für die perfekte Prognose. Der Nachweis ist analog zu dem von Satz 3.20.

Mit Satz 3.20 ist es wenig überraschend, dass für eine anreizerhaltende Score-Regel und kalibrierte Prognosen $SK \in [0, 1]$ und für nicht kalibrierte Prognosen $SK \in [-\infty, 1]$ gilt.

Aus Satz 3.20 folgt auch, dass der Erwartungswert einer anreizerhaltenden Score-Regel mit erwarteter Score-Funktion g für kalibrierte Prognosen im Intervall

$$[p_1 g(1) + (1 - p_1) g(0), g(p_1)]$$

liegt. Für nulloptimale Score-Regeln ergibt sich entsprechend $[0, g(p_1)]$. Dieses Intervall hängt offensichtlich von der Gesamtausfallwahrscheinlichkeit p_1 ab. Also strebt

der Erwartungswert einer nulloptimalen Score-Regel für $p_1 \downarrow 0$ zwangsläufig gegen den (optimalen) Wert Null. Dies gilt insbesondere auch für die Trivialprognose. Damit ist klar, dass sich bei unterschiedlichen Schuldnerpopulationen (mit unterschiedlichen Gesamtausfallwahrscheinlichkeiten) einfache Score-Regeln nicht zum Vergleich der Vorhersagequalität eignen. Diese Überlegung motiviert im Nachhinein die Definition und Verwendung von Skill-Scores, wenn Prognostiker unterschiedliche Schuldnerpopulationen bewerten.

Bei Vorliegen einer bewerteten Schuldnerpopulation können die theoretischen Verteilungen und die darauf basierenden Konstrukte dieses Kapitels (z. B. die Lorenzkurve) durch die zugehörigen empirischen Verteilungen geschätzt werden. Beispielsweise wird folgend die empirische Lorenzkurve definiert, dazu vorab:

Zu gegebenen Ausfällen $\vec{\theta} \in \mathbb{R}^n$ und Prognosen $\vec{x} \in \widehat{\mathbb{W}}^n = \{\hat{a}_1, \dots, \hat{a}_k\}^n$ wird die bivariate Wahrscheinlichkeitsfunktion eines Prognostikers geschätzt durch $\hat{r} : \{0, 1\} \times [0, 1] \rightarrow [0, 1]$, mit

$$\begin{aligned}\hat{r}(1, \hat{a}_j) &:= \frac{1}{n} \sum_{i=1}^n \theta_i \mathbb{1}_{\{\hat{a}_j\}}(x_i), \quad j = 1, \dots, k, \\ \hat{r}(0, \hat{a}_j) &:= \frac{1}{n} \sum_{i=1}^n (1 - \theta_i) \mathbb{1}_{\{\hat{a}_j\}}(x_i), \quad j = 1, \dots, k.\end{aligned}$$

Der Gesamtausfallanteil in der Population ist dann $\hat{p}_1 := \frac{1}{n} \sum_{i=1}^n \theta_i = \sum_{j=1}^k \hat{r}(1, \hat{a}_j)$. Damit sind die Größen $\hat{q}(\hat{a}_j)$, $\hat{p}(\theta | \hat{a}_j)$ und $\hat{q}(\hat{a}_j | \theta)$ entsprechend wie in Kapitel 2 definiert.

Definition und Korollar 3.22. Gegeben seien empirisch kalibrierte Wahrscheinlichkeitsprognosen $\vec{x} \in \widehat{W}^n$ mit $0 \leq \hat{a}_1 < \dots < \hat{a}_k \leq 1$ für die Ausfälle $\vec{\theta} \in \{0, 1\}^n$ und ein Gesamtausfallanteil von $\hat{p}_1 > 0$. Dann ist die **empirische Lorenzkurve**, also der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j \hat{q}(\hat{a}_i), \sum_{i=1}^j \hat{q}(\hat{a}_i | 1) \right)', \quad j = 0, \dots, k,$$

gleich dem Polygonzug durch die Punkte

$$\left(\frac{i}{n}, \frac{\sum_{j=1}^i x_{(j)}}{\sum_{j=1}^n x_j} \right)', \quad i = 0, \dots, n,$$

wobei $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ die geordneten Wahrscheinlichkeitsprognosen sind.

Beweis: Es ist sofort ersichtlich, dass beide Polygonzüge ihre Steigung nur an maximal $k - 1$ Stellen ändern, da $x_i \in \{\hat{a}_1, \dots, \hat{a}_k\}, \forall i = 1, \dots, n$. Der y -Koordinatenwert der Lorenzkurve für ein beliebiges $j \in \{1, \dots, k - 1\}$ lautet nach Definition

$$\sum_{i=1}^j \hat{q}(\hat{a}_i | 1) = \frac{\sum_{i=1}^j \hat{r}(1, \hat{a}_i)}{\hat{p}_1} = \frac{\frac{1}{n} \sum_{i=1}^j \sum_{j=1}^n \theta_j \mathbb{1}_{\{\hat{a}_i\}}(x_j)}{\hat{p}_1}.$$

Aufgrund der Kalibrierung ist $\sum_{j=1}^n \theta_j \mathbb{1}_{\{\hat{a}_i\}}(x_j) = \sum_{j=1}^n \hat{a}_i \mathbb{1}_{\{\hat{a}_i\}}(x_j)$, vgl. Definition 2.7.

Damit folgt

$$\sum_{i=1}^j \hat{q}(\hat{a}_i | 1) = \frac{\frac{1}{n} \sum_{i=1}^j \sum_{j=1}^n x_i \mathbb{1}_{\{\hat{a}_i\}}(x_j)}{\hat{p}_1} = \frac{\frac{1}{n} \sum_{\{i: x_i \leq \hat{a}_j\}} x_i}{\hat{p}_1}.$$

Da Korollar 2.3 (iii) auch bei empirischer Kalibrierung gilt, folgt die Behauptung. \square

Murphy (1988) hat bereits den Zusammenhang zwischen Skill-Scores und dem Korrelationskoeffizienten untersucht. Unter der Annahme der Kalibrierung lässt sich eine weitergehende Beziehung herausstellen:

Satz 3.23. *Gegeben seien empirisch kalibrierte Wahrscheinlichkeitsprognosen $\vec{x} \in \widehat{\mathbb{W}}^n$ für die Ausfälle $\vec{\theta} \in \{0, 1\}^n$ mit einem Gesamtausfallanteil von $\hat{p}_1 \in (0, 1)$. Dann entspricht der Brier-Skill-Score (BSK) dem quadrierten Korrelationskoeffizienten (von Bravais und Pearson) zwischen \vec{x} und $\vec{\theta}$:*

$$BSK(\vec{\theta}, \vec{x}) = \left(\frac{\sum_{i=1}^n (x_i - \hat{p}_1)(\theta_i - \hat{p}_1)}{\sqrt{\sum_{i=1}^n (x_i - \hat{p}_1)^2 \sum_{i=1}^n (\theta_i - \hat{p}_1)^2}} \right)^2.$$

Beweis: Aus der Kalibrierung der Prognosen folgt:

$$\frac{1}{n} \sum_{i=1}^n \theta_i^2 = \frac{1}{n} \sum_{i=1}^n \theta_i = \frac{1}{n} \sum_{i=1}^n x_i, \quad (*)$$

$$\frac{1}{n} \sum_{i=1}^n x_i \theta_i = \frac{1}{n} \sum_{i=1}^n x_i^2. \quad (**)$$

Somit ist

$$\begin{aligned} BSK(\vec{\theta}, \vec{x}) &= \frac{\frac{1}{n} \sum_{i=1}^n x_i(1 - x_i) - \frac{1}{n} \sum_{i=1}^n x_i \left(1 - \frac{1}{n} \sum_{i=1}^n x_i\right)}{\frac{1}{n} \sum_{i=1}^n \theta_i(1 - \theta_i) - \frac{1}{n} \sum_{i=1}^n x_i \left(1 - \frac{1}{n} \sum_{i=1}^n x_i\right)} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2}{\frac{1}{n} \sum_{i=1}^n x_i - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2} \\ &\stackrel{(*)}{=} \frac{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2}{\frac{1}{n} \sum_{i=1}^n \theta_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \theta_i\right)^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right)^2}{\left(\frac{1}{n} \sum_{i=1}^n \theta_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \theta_i\right)^2\right) \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right)} \\
&\stackrel{(\star\star)}{=} \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i \theta_i - \frac{1}{n} \sum_{i=1}^n x_i \frac{1}{n} \sum_{i=1}^n \theta_i\right)^2}{\left(\frac{1}{n} \sum_{i=1}^n \theta_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \theta_i\right)^2\right) \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right)}.
\end{aligned}$$

□

3.4 Anreizerhaltende Skill-Scores

Der Skill-Score einer anreizerhaltenden Score-Regel ist *im Allgemeinen* nicht anreizerhaltend (Gneiting & Raftery, 2007; Jolliffe & Stephenson, 2008; Mason, 2008). Als bekanntes Beispiel wird in der Literatur zumeist der nicht anreizerhaltende Brier-Skill-Score genannt (Murphy, 1973). Dieses Kapitel geht der Frage nach, ob der Skill-Score einer anreizerhaltenden Score-Regel, wie in Definition 3.19 und 3.9 eingeführt, überhaupt anreizerhaltend sein kann.

Mit dieser Problematik beschäftigt sich bereits Murphy (1973), indem er den erwarteten Brier-Skill-Score betrachtet. Für die Trivialprognose unterstellt Murphy eine feste Wahrscheinlichkeit (Referenz), gegen die der Wahrscheinlichkeitsprognostiker antritt. Die Annahme einer fixen Trivialprognose wird im Folgenden aufgeweicht, indem die Trivialprognose als zufälliger Gesamtausfallanteil in der Schuldnerpopulation angesehen wird, vgl. Definition 2.7.

Der Skill-Score einer nulloptimalen Score-Regel S mit Score-Funktion s und erwarteter Score-Funktion g hat zu gegebenen Prognosen $\vec{x} \in \mathbb{W}^n$ und Ausfällen $\vec{\theta} \in \{0, 1\}^n$ die Form

$$\begin{aligned} SK(\vec{\theta}, \vec{x}) &:= \frac{S(\vec{\theta}, \vec{x}) - S(\vec{\theta}, \vec{x}_T)}{S(\vec{\theta}, \vec{x}_P) - S(\vec{\theta}, \vec{x}_T)} \\ &= 1 - \frac{S(\vec{\theta}, \vec{x})}{S(\vec{\theta}, \vec{x}_T)} \\ &= 1 - \frac{1}{n} \sum_{i=1}^n \frac{s(\theta_i, x_i)}{g(\hat{p}_1)}, \end{aligned}$$

wobei $\hat{p}_1 \in (0, 1)$ dem Gesamtausfallanteil in der Schuldnerpopulation entspricht. Wichtig hierbei ist: Für $\hat{p}_1 = 0$ oder $\hat{p}_1 = 1$ ist der Skill-Score nicht definiert, da $g(0) = g(1) = 0$ ist. Daher ist ein erwarteter Skill-Score mit zufälligem Gesamtausfallanteil nur unter der Bedingung, dass mindestens einer und nicht alle Kredite ausfallen, sinnvoll definiert.

Seien $\Theta_1, \dots, \Theta_n$ unabhängig bernoulliverteilte Zufallsvariablen mit $\Theta_i \sim \text{Bin}(1, \pi_i)$, $\pi_i \in (0, 1)$, $i = 1, \dots, n$. Definiere $\vec{\pi} := (\pi_1, \dots, \pi_n)'$ und $\bar{\Theta} := \frac{1}{n} \sum_{i=1}^n \Theta_i$. Der erwartete Skill-Score einer Score-Regel $S \in \mathbb{L}_{rnn}$ ist dann:

$$E(SK \mid \bar{\Theta} \in (0, 1)) = 1 - \frac{1}{n} \sum_{i=1}^n E \left(\frac{s(\Theta_i, x_i)}{g(\bar{\Theta})} \mid \bar{\Theta} \in (0, 1) \right).$$

Falls auch $x_i \in (0, 1)$, $i = 1, \dots, n$, so ist der erwartete Skill-Scores garantiert endlich.

Da die Wahrscheinlichkeitsprognosen x_1, \dots, x_n nicht zufällig sind, spielt die bivariate Verteilung von Ausfallvariable und Prognose hier keine Rolle. Diese Überlegung führt zur folgenden Definition anreizerhaltender Skill-Scores.

Definition 3.24. Seien $\Theta_1, \dots, \Theta_n$ unabhängig bernoulliverteilte Zufallsvariablen mit $\Theta_i \sim \text{Bin}(1, \pi_i)$, $\pi_i \in (0, 1)$, $i = 1, \dots, n$. Der Skill-Score der Score-Regel $S \in \mathbb{L}_{rnn}$ heißt **anreizerhaltend**, falls für alle $(x_1, \dots, x_n)' \in (0, 1)^n$ gilt:

$$\sum_{i=1}^n E \left(\frac{s(\Theta_i, \pi_i)}{g(\bar{\Theta})} \mid \bar{\Theta} \in (0, 1) \right) \leq \sum_{i=1}^n E \left(\frac{s(\Theta_i, x_i)}{g(\bar{\Theta})} \mid \bar{\Theta} \in (0, 1) \right). \quad (3.6)$$

Dabei ist s die Score-Funktion und g die erwartete Score-Funktion von S . Der Skill-Score von S heißt **streng anreizerhaltend**, falls in Gleichung (3.6) für alle $x \in (0, 1)^n \setminus \{(\pi_1, \dots, \pi_n)'\}$ sogar echte Ungleichheit besteht.

Definition 3.24 könnte als irrelevant bezeichnet werden, denn wie die folgenden Ergebnisse zeigen, sind Skill-Scores von anreizerhaltenden Score-Regeln $S \in \mathbb{L}_{rnn}$ mit zweifach differenzierbarer erwarteter Score-Funktion selbst nicht anreizerhaltend.

Nach Definition 3.24 hängt der erwartete Skill-Score also von der gemeinsamen Verteilung von Θ_i und $\bar{\Theta}$ ab, $i = 1, \dots, n$. Weil die Ausfallwahrscheinlichkeiten (möglicherweise) unterschiedlich sind, folgt $\sum_{i=1}^n \Theta_i$ einer verallgemeinerten Binomialverteilung mit Wahrscheinlichkeitsvektor $\vec{\pi}$, geschrieben $\sum_{i=1}^n \Theta_i \sim \text{GB}(\vec{\pi})$. Diese Verteilung hängt vom Wahrscheinlichkeitsvektor $\vec{\pi} \in (0, 1)^n$ ab und hat die Zähldichte

$$w_{\vec{\pi}}(m) := P \left(\sum_{i=1}^n \Theta_i = m \right) = \sum_{A \in B_m} \prod_{i \in A} \pi_i \prod_{j \in A^c} (1 - \pi_j), \quad m = 0, \dots, n,$$

dabei ist B_m das Mengensystem, das aus allen m -elementigen Teilmengen besteht, die aus dem Träger $\{1, 2, \dots, n\}$ gebildet werden können.

Satz 3.25. *In der Situation von Definition 3.24 gilt für $i = 1, \dots, n$*

$$\begin{aligned} & E \left(\frac{s(\Theta_i, x_i)}{g(\bar{\Theta})} \mid \bar{\Theta} \in (0, 1) \right) \\ &= \sum_{m=1}^{n-1} \frac{s(0, x_i) \cdot w_{\vec{\pi}(i)}(m) \cdot (1 - \pi_i) + s(1, x_i) \cdot w_{\vec{\pi}(i)}(m-1) \cdot \pi_i}{g\left(\frac{m}{n}\right) \cdot (1 - w_{\vec{\pi}}(0) - w_{\vec{\pi}}(n))}. \end{aligned} \quad (3.7)$$

Dabei entspricht $\vec{\pi}(i)$ dem Vektor $\vec{\pi}$ ohne den i -ten Eintrag und $w_{\vec{\pi}}$ der Zähl-dichte der verallgemeinerten Binomialverteilung mit Wahrscheinlichkeitsvektor $\vec{\pi}$.

Beweis: Zunächst einmal gilt für $i = 1, \dots, n$

$$\begin{aligned} & E \left(\frac{s(\Theta_i, x_i)}{g(\bar{\Theta})} \mid \bar{\Theta} \in (0, 1) \right) \\ &= \sum_{m=1}^{n-1} \frac{s(0, x_i)}{g\left(\frac{m}{n}\right)} P \left(\Theta_i = 0, \bar{\Theta} = \frac{m}{n} \mid \bar{\Theta} \in (0, 1) \right) \\ &+ \sum_{m=1}^{n-1} \frac{s(1, x_i)}{g\left(\frac{m}{n}\right)} P \left(\Theta_i = 1, \bar{\Theta} = \frac{m}{n} \mid \bar{\Theta} \in (0, 1) \right). \end{aligned}$$

Hierbei ist für $\Theta_i = 0$ und ein beliebiges $m \in \{1, \dots, n-1\}$:

$$\begin{aligned} P \left(\Theta_i = 0, \bar{\Theta} = \frac{m}{n} \mid \bar{\Theta} \in (0, 1) \right) &= P \left(\Theta_i = 0, \bar{\Theta} = \frac{m}{n} \mid \sum_{j=1}^n \Theta_j \neq 0, \sum_{j=1}^n \Theta_j \neq n \right) \\ &= \frac{P \left(\Theta_i = 0, \bar{\Theta} = \frac{m}{n}, \sum_{j=1}^n \Theta_j \neq 0, \sum_{j=1}^n \Theta_j \neq n \right)}{P \left(\sum_{j=1}^n \Theta_j \neq 0, \sum_{j=1}^n \Theta_j \neq n \right)} \\ &= \frac{P \left(\Theta_i = 0, \sum_{j=1}^n \Theta_j = m \right)}{1 - P \left(\sum_{j=1}^n \Theta_j = 0 \right) - P \left(\sum_{j=1}^n \Theta_j = n \right)} \\ &= \frac{P \left(\sum_{j \in \{1, \dots, n\} \setminus \{i\}} \Theta_j = m \right) (1 - \pi_i)}{1 - P \left(\sum_{j=1}^n \Theta_j = 0 \right) - P \left(\sum_{j=1}^n \Theta_j = n \right)}. \end{aligned}$$

Da jedoch $\sum_{j=1}^n \Theta_j \sim \text{GB}(\vec{\pi})$ und $\sum_{j \in \{1, \dots, n\} \setminus \{i\}} \Theta_j \sim \text{GB}(\vec{\pi}_{(i)})$ folgt:

$$P\left(\Theta_i = 0, \bar{\Theta} = \frac{m}{n} \mid \bar{\Theta} \in (0, 1)\right) = \frac{w_{\vec{\pi}_{(i)}}(m) \cdot (1 - \pi_i)}{1 - w_{\vec{\pi}}(0) - w_{\vec{\pi}}(n)}.$$

Analog kann gezeigt werden $P\left(\Theta_i = 1, \bar{\Theta} = \frac{m}{n} \mid \bar{\Theta} \in (0, 1)\right) = \frac{w_{\vec{\pi}_{(i)}}(m-1) \cdot \pi_i}{1 - w_{\vec{\pi}}(0) - w_{\vec{\pi}}(n)}$. Damit folgt die Behauptung. \square

Damit ist der Weg frei zur Bestimmung der im Sinne des erwarteten Skill-Scores optimalen Prognosen.

Korollar 3.26. *In der Situation von Definition 3.24 maximieren die Wahrscheinlichkeitsprognosen*

$$x_i^* = \frac{\sum_{m=1}^{n-1} \frac{w_{\vec{\pi}_{(i)}}(m-1) \cdot \pi_i}{g\left(\frac{m}{n}\right)}}{\sum_{m=1}^{n-1} \frac{w_{\vec{\pi}_{(i)}}(m-1) \cdot \pi_i + w_{\vec{\pi}_{(i)}}(m) \cdot (1 - \pi_i)}{g\left(\frac{m}{n}\right)}}, \quad i = 1, \dots, n,$$

den Erwartungswert des Brier-Skill-Scores bzw. des Logarithmus-Skill-Scores, wobei g die erwartete Score-Funktion der entsprechenden Score-Regel ist.

Beweis: Differenzieren von Gleichung (3.7) nach x_i führt zur notwendigen Bedingung für das Vorliegen der optimalen Prognose x_i^*

$$\sum_{m=1}^{n-1} \frac{s'(0, x_i^*) \cdot w_{\vec{\pi}_{(i)}}(m) \cdot (1 - \pi_i) + s'(1, x_i^*) \cdot w_{\vec{\pi}_{(i)}}(m-1) \cdot \pi_i}{g\left(\frac{m}{n}\right) \cdot (1 - w_{\vec{\pi}}(0) - w_{\vec{\pi}}(n))} = 0. \quad (3.8)$$

Mit der Score-Funktion des Brier- bzw. Logarithmus-Scores, vgl. Tabelle 3.1 folgt die Behauptung. Da die zweite Ableitung von Gleichung (3.7) nach x_i größer als Null ist, minimiert x_i^* den Ausdruck aus Gleichung (3.7) bzw. maximiert den entsprechenden erwarteten Skill-Score. \square

Für den Brier- und Logarithmus-Skill-Score gilt im Allgemeinen $x_i^* \neq \pi_i$, was leicht durch Gegenbeispiele gezeigt werden kann. Dies ist für den Brier-Skill-Score nicht überraschend, da dieser bekanntermaßen nicht anreizerhaltend ist. Es gilt jedoch auch folgendes Resultat:

Korollar 3.27. *Sei $S \in \mathbb{L}_{rnn}$ eine anreizerhaltende Score-Regel mit zweifach differenzierbarer erwarteter Score-Funktion. Dann ist der zu S gehörende Skill-Score nach Definition 3.24 nicht anreizerhaltend.*

Beweis: Gleichung (3.8) muss für anreizerhaltende Skill-Scores insbesondere für $x_i^* = \pi_i$ erfüllt sein. Für $n = 2$ und $\pi_1 = \pi_2 \in (0, 1)$ ist dies dann äquivalent zu

$$\begin{aligned} & s'(0, \pi_1) \cdot w_{\bar{\pi}(1)}(1) \cdot (1 - \pi_1) + s'(1, \pi_1) \cdot w_{\bar{\pi}(1)}(0) \cdot \pi_1 = 0 \\ \iff & s'(0, \pi_1) \cdot \pi_1(1 - \pi_1) + s'(1, \pi_1) \cdot (1 - \pi_1)\pi_1 = 0 \\ \iff & s'(0, \pi_1) = -s'(1, \pi_1). \end{aligned}$$

Mit Gleichung (3.4) ist dies wiederum äquivalent zu

$$(1 - \pi_1)g''(\pi_1) = \underbrace{\pi_1 g''(\pi_1)}_{\neq 0} \iff \pi_1 = 0.5.$$

Also ist die notwendige Bedingung aus Gleichung (3.8) hier nicht für jedes $\pi_1 \in (0, 1)$ erfüllt. \square

Beispiel 3.28. *Seien $\Theta_1, \dots, \Theta_{30}$ unabhängig bernoulliverteilte Zufallsvariablen mit $\Theta_i \sim \text{Bin}(1, \pi_i = \frac{i}{100})$ für $i = 1, \dots, 30$. Dieses Beispiel untersucht die folgenden Fragen, jeweils für den Brier- und den Logarithmus-Skill-Score.*

- 1) *Wie sehen die Skill-Score optimalen Prognosen x_i^* aus?*

2) Wie verhält sich der Skill-Score, wenn als Prognose nicht $x_i = \pi_i$, sondern $x_i = x_i^*$ verwendet wird?

Zu 1): Die Skill-Score optimalen Prognosen werden mit Korollar 3.26 bestimmt. Abbildung 3.3 zeigt, dass die Abweichung zwischen x_i^* und π_i mit wachsendem $\pi_i \in [0.01, 0.3]$ betragsmäßig größer wird und dass $x_i^* < \pi_i$ gilt. Außerdem ist die Abweichung für den Logarithmus-Skill-Score hier betragsmäßig kleiner als für den Brier-Skill-Score. Vereinfacht ausgedrückt, verletzt der Logarithmus-Skill-Score die Anreizerhaltung in diesem Beispiel also weniger stark als der Brier-Skill-Score. Es sei darauf hingewiesen, dass der Brier-Skill-Score (bei fester Referenzprognose) asymptotisch anreizerhaltend ist (Murphy, 1973; Jolliffe & Stephenson, 2008).

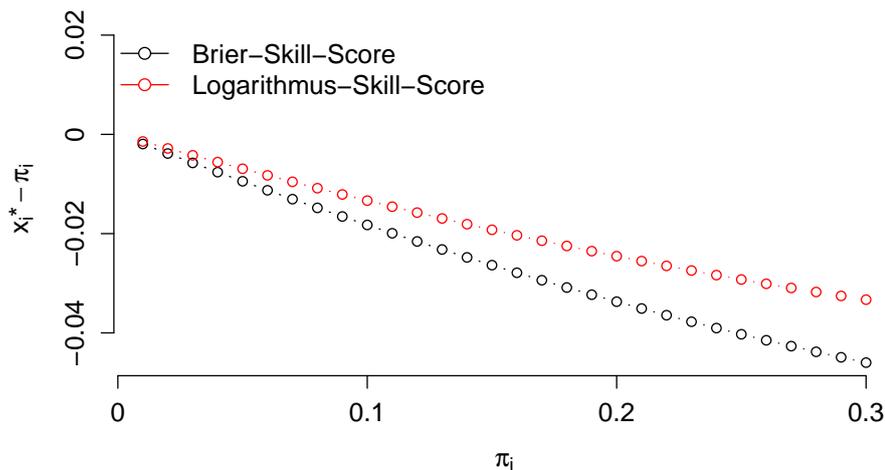


Abbildung 3.3: Gegeben die Situation aus Definition 3.24 mit $n = 30$ und $\pi_i = \frac{i}{100}$: Abweichung zwischen der Ausfallwahrscheinlichkeit π_i und der Skill-Score-optimalen Prognose x_i^* .

Zu 2): Tabelle 3.2 zeigt die erwarteten Skill-Scores (Logarithmus und Brier), jeweils für $\vec{x} = \vec{\pi}$ und $\vec{x} = \vec{x}^*$, vgl. Satz 3.25. Zum Vergleich zeigt \overline{SK}_{sim} die simulierten mittleren Skill-Scores. Mit Korollar 3.26 ist es nicht überraschend, dass die erwarteten und realisierten Skill-Scores mit $\vec{x} = \vec{x}^*$ größer sind als die mit $\vec{x} = \vec{\pi}$.

Tabelle 3.2: Seien $\Theta_i \sim \text{Bin}(1, \pi_i = \frac{i}{100})$ unabhängig für $i = 1, \dots, 30$. Die erwarteten Skill-Scores für $\vec{x} = \vec{\pi}$ und $\vec{x} = \vec{x}^*$ sind mithilfe von Satz 3.25 bestimmt. Dabei ist \vec{x}^* der Vektor der Skill-Score-optimalen Prognosen gemäß Korollar 3.26. Zum Vergleich zeigt $\overline{SK}_{\text{sim}}$ die mittleren realisierten Skill-Scores von 10^6 Simulationen.

	Brier-Skill-Score		Logarithmus-Skill-Score	
	$E(SK \bar{\Theta} \in (0, 1))$	$\overline{SK}_{\text{sim}}$	$E(SK \bar{\Theta} \in (0, 1))$	$\overline{SK}_{\text{sim}}$
$\vec{x} = \vec{\pi}$	0.0040	0.0041	0.0228	0.0229
$\vec{x} = \vec{x}^*$	0.0118	0.0119	0.0268	0.0268

Kapitel 4

Halbordnungen bei identischen Schuldnerkollektiven

Dieses Kapitel untersucht verschiedene Halbordnungen zum Vergleich von Wahrscheinlichkeitsprognostikern bei identischen Schuldnerkollektiven.

4.1 Motivation

Ein Kreditausfall ist für eine Bank vor allem dann ungünstig, wenn dieser Kredit mit einer geringen Ausfallwahrscheinlichkeit bewertet wurde. Behr & Güttler (2004, S. 149/150) verdeutlichen, dass der Logarithmus-Score einen solchen Fehler stärker bestraft als der Brier-Score, weshalb der Logarithmus-Score dem Brier-Score im Kreditrisikobereich vorzuziehen sei. Auch wenn solche Überlegungen bei der Wahl einer geeigneten Score-Regel berücksichtigt werden sollten, sind sie nur bedingt hilfreich, da beliebig viele skalare Maße, wie z. B. Score-Regeln existieren. Die Wahl für ein bestimmtes skalares Maß ist daher letztendlich immer ein Kompromiss. Dabei ist es problemlos möglich, dass sich zwei Maße bezüglich der gemessenen Vorhersagequalität widersprechen, wie folgendes Beispiel zeigt:

Beispiel 4.1. Seien die Wahrscheinlichkeitsprognostiker A , B und C gemäß Tabelle 4.1 definiert. Tabelle 4.2 zeigt die zugehörigen erwarteten Brier-, Logarithmus- und sphärischen Scores.

Tabelle 4.1: Prognoseverteilungen von drei (kalibrierten) Prognostikern mit einer Gesamtausfallwahrscheinlichkeit von 10 % (gleiche Schuldnerpopulationen).

	a_i			
	0	0.1	0.2	1
$q^A(a_i)$	0.532	0.240	0.190	0.038
$q^B(a_i)$	0.626	0.020	0.320	0.034
$q^C(a_i)$	0.563	0.330	0.050	0.057

Tabelle 4.2: Erwartungswerte vom Brier-, Logarithmus- und sphärischen Score für drei kalibrierte Wahrscheinlichkeitsprognostiker, deren Prognosen gemäß Tabelle 4.1 verteilt sind. Die Erwartungswerte sind auf drei Nachkommastellen gerundet.

	Erwarteter Score		
	Logarithmus	Brier	Sphärisch
Prognostiker A	0.173	0.052	0.059
Prognostiker B	0.167	0.053	0.059
Prognostiker C	0.132	0.038	0.043

Offensichtlich widersprechen sich der Logarithmus- und der Brier-Score bei der Bewertung der Prognostiker A und B . Nach dem Brier-Score ist A besser als B , hingegen ist nach dem Logarithmus-Score B besser als A . Prognostiker C ist bezüglich aller drei Score-Regeln besser als A und B .

Von besonderem Interesse ist daher die Frage: Ist C im Sinne aller anreizerhaltenden Score-Regeln besser als A und B ? Um Fragen solcher Art zu beantworten, werden in diesem Kapitel verschiedene Halbordnungen vorgestellt. Es zeigt sich, dass die hier gestellte Frage zu verneinen ist, jedoch gilt: C ist besser als A bezüglich aller anreizerhaltenden Score-Regeln (Fortsetzung in Beispiel 4.7).

Für zwei Wahrscheinlichkeitsprognostiker A und B , die mittels einer Ordnungsrelation verglichen werden, sei die folgende Notation definiert:

- (i) $A \geq B$, falls A so gut wie B oder besser im Sinne der Ordnungsrelation ist.
- (ii) $A > B$, falls A besser als B im Sinne der Ordnungsrelation ist.

Weiter sei $A = B$, falls A und B die gleiche bivariate Verteilung von Ausfallvariable und Prognose besitzen.

Definition 4.2. Seien A , B und C drei beliebige Wahrscheinlichkeitsprognostiker. Eine Ordnungsrelation „ \geq “ auf der Menge der Wahrscheinlichkeitsprognostiker bildet eine **Präordnung**, wenn die Eigenschaften

- (i) **Reflexivität:** $A \geq A$,
- (ii) **Transitivität:** Falls $A \geq B$ und $B \geq C$, so folgt $A \geq C$

erfüllt sind. Die Ordnungsrelation bildet eine **Halbordnung**, wenn neben (i) und (ii) auch die Eigenschaft

- (iii) **Antisymmetrie:** Falls $A \geq B$ und $B \geq A$, so folgt $A = B$

erfüllt ist. Die Ordnungsrelation bildet eine **Totalordnung**, wenn neben (i)-(iii) auch die Eigenschaft

- (iv) **Totalität:** Es gilt $A \geq B$ oder $B \geq A$

erfüllt ist.

Im Folgenden ist die in Kapitel 3 eingeführte Notation mit dem Prognostiker spezifischen Buchstaben indiziert. So ist beispielsweise $q^A(a_1)$ die Wahrscheinlichkeit,

dass Prognostiker A die Prognose a_1 abgibt. Für zwei Wahrscheinlichkeitsprognostiker A und B auf demselben Schuldnerkollektiv folgt dann insbesondere für die Gesamtausfallwahrscheinlichkeit: $p_1 = \sum_{i=1}^k r^A(1, a_i) = \sum_{i=1}^k r^B(1, a_i)$.

4.2 Refinement-Ordnung

Die Refinement-Ordnung ist eine Halbordnung für kalibrierte Wahrscheinlichkeitsprognostiker auf derselben Schuldnerpopulation. Mathematisch liefert diese Halbordnung ein Konzept zum Vergleich zweier diskreter Verteilungen mit gleichem Erwartungswert. Die Refinement-Ordnung wurde von DeGroot & Fienberg (1981) vorgestellt und basiert auf einer Transformation der Prognoseverteilung mittels einer spaltenstochastischen Matrix.

Definition 4.3. *Eine Matrix $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$, heißt **spaltenstochastisch** genau dann, wenn alle Einträge nichtnegativ sind und sich die Einträge in jeder Spalte zu Eins aufsummieren.*

Definition 4.4 (DeGroot & Fienberg, 1981). *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf derselben Schuldnerpopulation. A dominiert B bezüglich der **Refinement-Ordnung** ($A \geq_R B$) genau dann, wenn eine spaltenstochastische Matrix $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$ existiert, sodass die Gleichungen*

$$\sum_{j=1}^k M_{ij} q^A(a_j) = q^B(a_i), \quad i = 1, \dots, k, \quad (4.1)$$

$$\sum_{j=1}^k M_{ij} a_j q^A(a_j) = a_i q^B(a_i), \quad i = 1, \dots, k, \quad (4.2)$$

erfüllt sind.

Die Refinement-Ordnung ist ein Ordnungskonzept zum Vergleich von kalibrierten Prognoseverteilungen mit gleichem Erwartungswert hinsichtlich der Spreizung der Prognosen (gleichbedeutend mit Trennschärfe).

Zur besseren Interpretation lassen sich Gleichung (4.1) und (4.2) umschreiben, so dass für zwei kalibrierte Vorhersager folgt

$$\begin{aligned} r^B(0, a_i) + r^B(1, a_i) &= \sum_{j=1}^k M_{ij}(r^A(0, a_j) + r^A(1, a_j)), & i = 1, \dots, k, \\ r^B(1, a_i) &= \sum_{j=1}^k M_{ij}r^A(1, a_j), & i = 1, \dots, k. \end{aligned}$$

Subtraktion der beiden Gleichungen voneinander liefert eine alternative Definition der Refinement-Ordnung: $A \geq_R B$ ist äquivalent zu, für zwei kalibrierte Wahrscheinlichkeitsprognostiker A und B auf derselben Schuldnerpopulation existiert eine spaltenstochastische Matrix $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$, sodass gilt:

$$\underbrace{\begin{pmatrix} r^B(0, a_1) & r^B(1, a_1) \\ \vdots & \vdots \\ r^B(0, a_k) & r^B(1, a_k) \end{pmatrix}}_{=:R^B} = \begin{pmatrix} M_{11} & \dots & M_{1k} \\ \vdots & \ddots & \vdots \\ M_{k1} & \dots & M_{kk} \end{pmatrix} \underbrace{\begin{pmatrix} r^A(0, a_1) & r^A(1, a_1) \\ \vdots & \vdots \\ r^A(0, a_k) & r^A(1, a_k) \end{pmatrix}}_{=:R^A}.$$

Die bivariate Verteilung von B ergibt sich also aus der bivariaten Verteilung von A über die Transformation mit einer spaltenstochastischen Matrix M . Dabei wirkt M auf beide Spalten in R^A gleich. Wegen der Kalibrierung kann die so entstehende Prognoseverteilung von B daher nur weniger trennscharf sein als die von A .

Beispiel 4.5. *Tabelle 4.3 zeigt die Prognoseverteilungen von fünf kalibrierten Wahrscheinlichkeitsprognostikern. Offensichtlich ist Prognostiker B trennschärfer als die Trivialprognose von Prognostiker A, d. h. $B \geq_R A$. Die Prognostiker C und D teilen die Kredite in die Ausfallklassen 5 %, 15 % und 45 % bzw. 5 %, 10 % und 45 % auf und sind wiederum trennschärfer als Prognostiker B, d. h. $C \geq_R B$ und $D \geq_R B$. Die perfekte Prognose von Prognostiker E besitzt maximale Trennschärfe und dominiert daher alle anderen Prognostiker bezüglich der Refinement-Ordnung.*

Tabelle 4.3: Verteilung der Ausfallwahrscheinlichkeiten von fünf kalibrierten Wahrscheinlichkeitsprognostikern. Die Gesamtausfallwahrscheinlichkeit in der Schuldnerpopulation beträgt 20 % (entspricht dem Erwartungswert der Prognoseverteilungen).

Ausfallwahrscheinlichkeit	Verteilung der Kredite auf die Ausfallklassen				
	A	B	C	D	E
0 %	0	0	0	0	0.8
5 %	0	0	0.25	0.1	0
10 %	0	0.5	0	0.6	0
15 %	0	0	0.50	0	0
20 %	1	0	0	0	0
30 %	0	0.5	0	0	0
45 %	0	0	0.25	0.3	0
100 %	0	0	0	0	0.2

Die Prognostiker C und D sind nicht bezüglich der Trennschärfe (Refinement-Ordnung) vergleichbar: Weder gilt $C \geq_R D$ noch $D \geq_R C$. Die Refinement-Ordnung ist also auf der Menge der kalibrierten Wahrscheinlichkeitsprognosen keine Totalordnung (sondern lediglich eine Halbordnung).

Die Suffizienz-Ordnung liefert ein Konzept zum Vergleich zweier Wahrscheinlichkeitsprognostiker ohne Kalibrierung (DeGroot & Fienberg, 1983). Dabei dominiert Prognostiker A den Prognostiker B nach der Suffizienz-Ordnung, wenn für eine

spaltenstochastische Matrix $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$ gilt:

$$q^B(a_i | \theta) = \sum_{j=1}^k M_{ij} q^A(a_j | \theta), \quad \forall i = 1, \dots, k, \quad \theta \in \{0, 1\}.$$

Bei vorliegender Kalibrierung der beiden Prognostiker reduziert sich die Suffizienz-Ordnung zur Refinement-Ordnung.

Die Refinement-Ordnung ist eng mit anderen Vergleichskriterien verbunden, welche im Folgenden dargelegt werden. Satz 4.6 beschreibt unter anderem den Zusammenhang zur Lorenzkurve.

Satz 4.6. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf derselben zu bewertenden Schuldnerpopulation. Sei L^A und L^B die Lorenzkurve von A und B . Weiter sei g eine auf $[0, 1]$ stetige und konkave Funktion, dann gilt:*

$$\begin{aligned} (i) \quad A \geq_R B &\iff L^A(x) \leq L^B(x) \quad \forall x \in [0, 1], \\ (ii) \quad A \geq_R B &\iff \sum_{i=1}^{j-1} (a_j - a_i)(q^A(a_i) - q^B(a_i)) \geq 0 \quad \forall j = 2, \dots, k-1, \\ (iii) \quad A \geq_R B &\iff \sum_{i=1}^k g(a_i)q^A(a_i) \leq \sum_{i=1}^k g(a_i)q^B(a_i). \end{aligned}$$

Beweis: (i) DeGroot & Eriksson (1985). (ii) DeGroot & Fienberg (1981). (iii) Hardy et al. (1929). □

Mit Satz 4.6 (iii) und Bemerkung 3.13 folgt für zwei kalibrierte Wahrscheinlichkeitsprognostiker A und B auf derselben zu bewertenden Population von Krediten:

$$A \geq_R B \iff S(\Theta, X^A) \leq S(\Theta, X^B) \tag{4.3}$$

für alle anreizerhaltenden Score-Regeln S (Krämer, 2006).

Beispiel 4.7 (Fortsetzung von Beispiel 4.1).

Seien die Prognostiker A , B und C wie in Beispiel 4.1 gegeben. Abbildung A.1 im Anhang zeigt die zugehörigen Lorenzkurven. Hier ist $C \geq_L A$, sodass mit Satz 4.6 der Prognostiker C im Erwartungswert besser ist als Prognostiker A bezüglich aller anreizerhaltenden Score-Regeln, vgl. Gleichung (4.3).

Im Folgenden wird die Refinement-Ordnung für den Spezialfall $\mathbb{W} = \{a_1, a_2, a_3, a_4\}$ ausführlich diskutiert. Dabei seien sowohl die perfekte Prognose als auch die Trivialprognose möglich und $a_3 \in (p_1, 1)$ beliebig aber fest, sodass $\mathbb{W} = \{0, p_1, a_3, 1\}$ folgt.

Wie kann ein kalibrierter Wahrscheinlichkeitsprognostiker nun seine Prognosen wählen und welchen Einfluss hat diese Wahl auf den Vergleich mit einem anderen kalibrierten Prognostiker bezüglich der Refinement-Ordnung? Diese Fragen beantwortet der folgende Satz.

Satz 4.8. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf derselben Schuldnerpopulation mit Gesamtausfallwahrscheinlichkeit $0 < p_1 < 1$. Sei $\mathbb{W} = \{0, p_1, a_3, 1\}$ die zulässige Menge der Wahrscheinlichkeitsprognosen mit $a_3 \in (p_1, 1)$. Mit den Grenzen*

$$b_1 := \frac{p_1}{a_3} (q^B(p_1) - q^A(p_1)) + q^B(a_3), \quad (4.4)$$

$$b_2 := \frac{1 - p_1}{1 - a_3} (q^B(p_1) - q^A(p_1)) + q^B(a_3) \quad (4.5)$$

gilt:

$$(i) \quad A \geq_R B \iff q^A(a_3) \leq \min \{b_1, b_2\},$$

$$(ii) \quad B \geq_R A \iff q^A(a_3) \geq \max \{b_1, b_2\}.$$

(iii) Es besteht nach der Refinement-Ordnung keine Dominanz zwischen A und B genau dann, wenn

$$q^A(a_3) \in \begin{cases} (b_1, b_2), & q^A(p_1) < q^B(p_1) \\ (b_2, b_1), & q^A(p_1) > q^B(p_1). \end{cases}$$

Beweis: Für die Prognoseverteilungen von A und B gilt $\sum_{j=1}^4 q^A(a_j) = \sum_{j=1}^4 q^B(a_j) = 1$ und aufgrund der Kalibrierung zusätzlich $\sum_{j=1}^4 a_j q^A(a_j) = \sum_{j=1}^4 a_j q^B(a_j) = p_1$. Damit sind pro Prognostiker zwei Gleichungen gegeben. Diese führen für A zu einer Prognoseverteilung der Form $\vec{q}^A := (q^A(a_1), \dots, q^A(a_4))' \in \mathbb{R}^4$ mit:

$$\vec{q}^A = \begin{pmatrix} (1 - p_1)(1 - \lambda_1^A) - \lambda_2^A(1 - a_3) \\ \lambda_1^A \\ \lambda_2^A \\ p_1(1 - \lambda_1^A) - \lambda_2^A a_3 \end{pmatrix}, \quad (4.6)$$

wobei

$$0 \leq \lambda_1^A \leq 1, \quad (4.7)$$

$$0 \leq \lambda_2^A \leq \frac{p_1(1 - \lambda_1^A)}{a_3} \quad (4.8)$$

erfüllt sein müssen (genauso für B indiziert). Sei nun $A \geq_R B$, dann gelten wegen Satz 4.6 (ii) die Ungleichungen:

$$\begin{aligned} (a_2 - a_1)(q^A(a_1) - q^B(a_1)) &\geq 0, \\ (a_3 - a_1)(q^A(a_1) - q^B(a_1)) + (a_3 - a_2)(q^A(a_2) - q^B(a_2)) &\geq 0. \end{aligned}$$

Einsetzen der Lösung aus Gleichung (4.6) und Umstellen der Gleichungen nach λ_2^A

liefert Grenzen für gegebene λ_1^B , λ_2^B -Werte in Abhängigkeit von λ_1^A

$$\lambda_2^A \leq \underbrace{\frac{p_1}{a_3}(\lambda_1^B - \lambda_1^A) + \lambda_2^B}_{=:b_1(\lambda_1^A)},$$

$$\lambda_2^A \leq \underbrace{\frac{1-p_1}{1-a_3}(\lambda_1^B - \lambda_1^A) + \lambda_2^B}_{=:b_2(\lambda_1^A)},$$

sodass gilt:

$$A \geq_R B \iff \lambda_2^A \leq \min \{b_1(\lambda_1^A), b_2(\lambda_1^A)\},$$

$$B \geq_R A \iff \lambda_2^A \geq \max \{b_1(\lambda_1^A), b_2(\lambda_1^A)\},$$

wobei die Gleichungen (4.7) und (4.8) erfüllt seien, damit die Verteilungen wohldefiniert sind. Damit lässt sich auch leicht erkennen, dass genau dann keine Dominanz zwischen A und B vorliegt, wenn

$$\lambda_2^A \in \begin{cases} (b_1(\lambda_1^A), b_2(\lambda_1^A)), & \lambda_1^A < \lambda_1^B \\ (b_2(\lambda_1^A), b_1(\lambda_1^A)), & \lambda_1^A > \lambda_1^B. \end{cases}$$

□

Bemerkung 4.9.

- (i) Satz 4.8 verdeutlicht auch, dass für den Fall $\mathbb{W} = \{a_1, a_2, a_3, a_4\}$ die komplette bivariate Verteilung $r(\cdot, \cdot)$ eines kalibrierten Wahrscheinlichkeitsprognostikers durch zwei Einträge des Vektors $(q(a_1), q(a_2), q(a_3), q(a_4))'$ festgelegt ist.

(ii) Hier gilt nach Voraussetzung $a_3 \neq p_1$. Für $a_3 = p_1$, also $W = \{0, p_1, 1\}$, ist die Refinement-Ordnung eine Totalordnung. Von zwei kalibrierten Wahrscheinlichkeitsprognostikern auf demselben Schuldnerkollektiv ist in diesem Fall der mit weniger Wahrscheinlichkeitsmasse auf p_1 dominant.

Abbildung 4.1 veranschaulicht die Aussage von Satz 4.8. Hier ist

- $p_1 = 0.1$ die Gesamtausfallwahrscheinlichkeit in der Schuldnerpopulation,
- $W = \{0, 0.1, 0.2, 1\}$ die zulässige Menge der Wahrscheinlichkeitsprognosen,
- $\vec{q}^B = (0.553, 0.190, 0.220, 0.307)'$ die Prognoseverteilung des kalibrierten Wahrscheinlichkeitsprognostikers B .

Wegen der Gleichungen (4.6), (4.7) und (4.8) gehört jeder Punkt auf den markierten Flächen zu genau einer kalibrierten Prognoseverteilung. Die Prognoseverteilung von B wird eindeutig durch den Punkt $(q^B(0.1), q^B(0.2))' = (0.190, 0.220)'$ festgelegt, dieser ist blau eingezeichnet. Genauso legt $(q^A(0.1), q^A(0.2))'$ die Prognoseverteilung von A eindeutig fest und je nach Ausprägung, gilt entweder $A \geq_R B$ oder $B \geq_R A$ oder A und B lassen sich nicht bezüglich der Refinement-Ordnung vergleichen. Die zugehörigen Flächen in der Abbildung sind durch die Grenzen b_1 und b_2 aus Satz 4.8 voneinander getrennt.

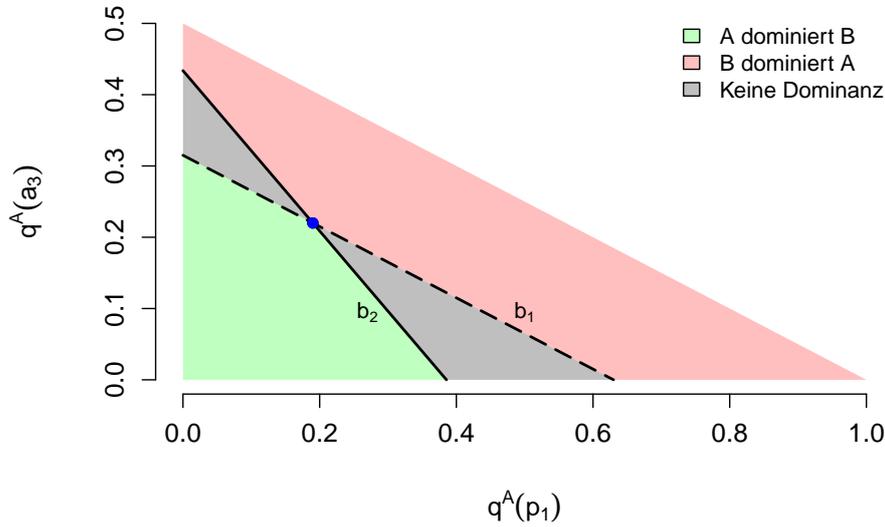


Abbildung 4.1: Dominanz nach der Refinement-Ordnung zwischen zwei kalibrierten Wahrscheinlichkeitsprognostikern A und B . Hier gilt $p_1 = 0.1$, $W = \{0, 0.1, 0.2, 1\}$ und $\vec{q}^B = (0.553, 0.190, 0.220, 0.307)'$. Die Dominanzbeziehung ist in Abhängigkeit von $(q^A(0.1), q^A(0.2))'$ zu verstehen. Die Grenzen b_1 und b_2 aus Satz 4.8 sind eingezeichnet. Offensichtlich: Die Prognose von A entspricht der perfekten Prognose im Punkt $(0, 0)'$ und der Trivialprognose im Punkt $(1, 0)'$.

4.3 Vardeman-Meeden-Ordnung

Die Vardeman-Meeden-Ordnung ist eine Halbordnung für Wahrscheinlichkeitsprognostiker, wenn diese identische Schuldnerpopulationen bewerten. Unabhängig von Kalibrierung ist die Frage sinnvoll, welcher von zwei Prognostikern den Ausfällen bzw. Nichtausfällen a priori die größeren bzw. kleineren Ausfallprognosen zuordnet. Im Mittelpunkt stehen daher die bedingten Prognoseverteilungen gegeben dem Ausfall bzw. dem Nichtausfall eines Kredits. Die Vardeman-Meeden-Ordnung verwendet für den Vergleich dieser Verteilungen das Konzept der stochastischen Dominanz.

Definition 4.10. Seien X^A und X^B Zufallsvariablen mit Verteilungsfunktionen F^A und F^B . Dann gilt: X^A dominiert X^B im Sinne der **stochastischen Dominanz n -ter Ordnung** ($n \in \mathbb{N}$), $(X^A \geq_{SD(n)} X^B)$, wenn

$$F_{(n)}^A(x) \leq F_{(n)}^B(x) \quad \forall x \in \mathbb{R}$$

mit $F_{(n)}^A(x) := \int_{-\infty}^x F_{(n-1)}^A(u) du$ und $F_{(1)}^A(x) := F^A(x)$ und für X^B entsprechend definiert.

Aus stochastischer Dominanz n -ter Ordnung folgt stochastische Dominanz $(n+1)$ -ter Ordnung. Eine umgekehrte Folgerung ist im Allgemeinen nicht möglich.

Vardeman & Meeden (1983) wenden das Prinzip der stochastischen Dominanz bei der Bewertung von Wahrscheinlichkeitsprognosen an. Die Idee dabei ist, dass Ausfälle hohe Ausfallprognosen und Nichtausfälle kleine Ausfallprognosen zugeordnet bekommen sollten.

Definition 4.11. Seien A und B Wahrscheinlichkeitsprognostiker.

(i) A dominiert B im Sinne der **Vardeman-Meeden-Ausfallordnung** (VM-Ausfallordnung) genau dann, wenn

$$A \geq_{VM(d)} B \quad :\Leftrightarrow \quad \sum_{i=1}^j q^A(a_i | 1) \leq \sum_{i=1}^j q^B(a_i | 1) \quad \forall j = 1, \dots, k.$$

(ii) A dominiert B im Sinne der **Vardeman-Meeden-Nichtausfallordnung** (VM-Nichtausfallordnung) genau dann, wenn

$$A \geq_{VM(nd)} B \quad :\Leftrightarrow \quad \sum_{i=1}^j q^A(a_i | 0) \geq \sum_{i=1}^j q^B(a_i | 0) \quad \forall j = 1, \dots, k.$$

(iii) A dominiert B im Sinne der **Vardeman-Meeden-Ordnung** (VM-Ordnung) genau dann, wenn

$$A \geq_{VM(d)} B \quad \text{und} \quad A \geq_{VM(nd)} B.$$

Die VM-Ausfallordnung entspricht stochastischer Dominanz erster Ordnung in der auf den Ausfall bedingten Prognoseverteilung. Die VM-Nichtausfallordnung entspricht stochastischer Dominanz erster Ordnung in der auf den Nichtausfall bedingten Prognoseverteilung.

Bemerkung 4.12. Bei gleichen Schuldnerkollektiven ist die VM-Ordnung eine Halbordnung und die VM-Ausfallordnung und die VM-Nichtausfallordnung sind Präordnungen. Wird zusätzlich die Kalibrierung der Prognostiker unterstellt, so erfüllen auch die VM-Ausfallordnung und die VM-Nichtausfallordnung die Antisymmetrie, sodass diese dann Halbordnungen sind.

In der Diskussion um die VM-Ordnung stellen Vardeman & Meeden (1983) folgende Beziehung zur Refinement-Ordnung heraus:

Satz 4.13. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf derselben zu bewertenden Schuldnerpopulation, dann gilt:*

$$(i) \quad A \geq_{VM(d)} B \implies A \geq_R B,$$

$$(ii) \quad A \geq_{VM(nd)} B \implies A \geq_R B.$$

Beweis: Siehe Satz 2.1 in Vardeman & Meeden (1983). □

Folgender Satz liefert eine zur Refinement-Ordnung (Definition 4.4) ähnliche Betrachtungsweise für die VM-Ausfallordnung.

Satz 4.14. *Seien A und B Wahrscheinlichkeitsprognostiker, dann gilt:*

$$A \geq_{VM(d)} B \iff q^B(a_i | 1) = \sum_{j=1}^k M_{ij} q^A(a_j | 1) \quad \forall i = 1, \dots, k,$$

wobei $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$ eine spaltenstochastische obere Dreiecksmatrix ist.

Beweis: „ \Leftarrow “: Definiere $\vec{e}_i \in \mathbb{R}^{1 \times k}$ für $i = 1, \dots, k$ als Zeilenvektor, dessen erste i Einträge gleich 1 und die übrigen gleich 0 sind. Da M eine spaltenstochastische obere Dreiecksmatrix ist, gilt:

$$\vec{e}_i M = (\underbrace{1, \dots, 1}_{i \text{ Einträge}}, c_{i+1}^{(i)}, \dots, c_k^{(i)}) \in \mathbb{R}^{1 \times k},$$

wobei $0 \leq c_j^{(i)} \leq 1$, $j = i + 1, \dots, k$. Sei $\vec{q}_1^A := (q^A(a_1 | 1), \dots, q^A(a_k | 1))' \in \mathbb{R}^k$, womit für $i = 1, \dots, k$ beliebig folgt:

$$\sum_{j=1}^i q^B(a_j | 1) \stackrel{\text{n. V.}}{=} \vec{e}_i M \vec{q}_1^A \geq \sum_{j=1}^i q^A(a_j | 1).$$

„ \implies “: Sei nun $A \geq_{VM(d)} B$ und seien $X^{A|1}$ und $X^{B|1}$ die Zufallsvariablen der bedingten Prognoseverteilungen mit Zähldichte $q^A(\cdot | 1)$ und $q^B(\cdot | 1)$. Sei die gemeinsame Verteilung von $X^{B|1}$ und $X^{A|1}$ für $i, j = 1, \dots, k$ über die Komonotoniecopula definiert, also über

$$F_{ij} := P(X^{B|1} \leq a_i, X^{A|1} \leq a_j) := \min \left\{ \sum_{m=1}^j q^A(a_m | 1), \sum_{m=1}^i q^B(a_m | 1) \right\}.$$

Dann gilt wegen der stochastischen Dominanz $F_{ij} = F_{jj}$ für alle $i \geq j$. Daher ist die Matrix $Q \in \mathbb{R}^{k \times k}$ mit Einträgen $Q_{ij} := P(X^{B|1} = a_i, X^{A|1} = a_j)$ für $i, j = 1, \dots, k$ eine obere Dreiecksmatrix. Und die Matrix $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$ mit Einträgen

$$M_{ij} := \begin{cases} P(X^{B|1} = a_i | X^{A|1} = a_j), & q^A(a_j | 1) > 0 \\ 0, & q^A(a_j | 1) = 0 \end{cases}, \quad i, j = 1, \dots, k,$$

ist eine spaltenstochastische obere Dreiecksmatrix. Damit folgt

$$\sum_{j=1}^k M_{ij} q^A(a_j | 1) = \sum_{j=1}^k Q_{ij} = q^B(a_i | 1), \quad \forall i = 1, \dots, k,$$

und damit die Behauptung. □

Offensichtlich gilt $A \geq_{VM(nd)} B \iff q^A(a_i | 0) = \sum_{j=1}^k M_{ij} q^B(a_j | 0) \forall i = 1, \dots, k$ mit $M = (M_{ij})_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k}$ einer spaltenstochastischen oberen Dreiecksmatrix (folgt wegen umgekehrter stochastischer Dominanz).

Krämer (2005) hält die VM-Ordnung bei kalibrierten Prognosen auf gleichen Schuldnerkollektiven für praktisch irrelevant und zeigt, dass die VM-Ordnung nur in bestimmten Fällen möglich ist:

Satz 4.15 (Krämer, 2005). *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf derselben zu bewertenden Schuldnerpopulation, dann gilt:*

- (i) *Falls $q^A(0) = q^B(0) = 0$, so können A und B nicht nach der VM-Ausfallordnung sortiert werden.*
- (ii) *Falls $q^A(1) = q^B(1) = 0$, so können A und B nicht nach der VM-Nichtausfallordnung sortiert werden.*

Beweis: Krämer (2005). Ähnlich zum Beweis von Satz 5.4 in dieser Arbeit. □

4.4 Lorenz- und ROC-Ordnung

Die in diesem Unterkapitel vorgestellten Ordnungen basieren auf der Lorenzkurve und der ROC-Kurve und benötigen daher grundsätzlich keine expliziten Ausfallwahrscheinlichkeiten, vgl. Kapitel 3.1. Da jedoch in dieser Arbeit der Fokus auf der Bewertung von Wahrscheinlichkeitsprognosen liegt, sind die Definitionen entsprechend formuliert.

Definition 4.16. *Seien A und B Wahrscheinlichkeitsprognostiker mit Lorenzkurven L^A und L^B sowie ROC-Kurven ROC^A und ROC^B .*

- (i) *A dominiert B im Sinne der **Lorenz-Ordnung** genau dann, wenn*

$$A \geq_L B \quad :\iff \quad L^A(x) \leq L^B(x) \quad \forall x \in [0, 1].$$

- (ii) *A dominiert B im Sinne der **ROC-Ordnung** genau dann, wenn*

$$A \geq_{ROC} B \quad :\iff \quad ROC^A(x) \geq ROC^B(x) \quad \forall x \in [0, 1].$$

Bemerkung 4.17. *Wird in Definition 4.16 Kalibrierung vorausgesetzt, so sind die Lorenz-Ordnung und die ROC-Ordnung bei identischen Schuldnerkollektiven Halbordnungen. Ohne Kalibrierung ist jedoch die Antisymmetrie verletzt: Werden beispielsweise in der Situation von Abbildung 3.1 alle Prognosen halbiert, so ist der Prognostiker zwar nicht mehr kalibriert, aber die Lorenzkurve und die ROC-Kurve bleiben unverändert. Aus der Gleichheit der Kurven kann dann nicht auf die Gleichheit der Prognosen geschlossen werden. Ohne Kalibrierung sind die Lorenz- und die ROC-Ordnung also nur Präordnungen (Reflexivität, Transitivität sind gegeben).*

Da sich zwei Lorenzkurven bzw. ROC-Kurven schneiden können, ist die Totalität hier offensichtlich nicht erfüllt.

Falls sich die ROC-Kurven der beiden Prognostiker schneiden, kann die partielle Fläche unter der ROC-Kurve betrachtet werden (partial AUC), um einen Vergleich der Trennschärfe zu erhalten (McClish, 1989).

Definition 4.18. *Seien A und B Wahrscheinlichkeitsprognostiker mit ROC-Kurven ROC^A und ROC^B . A dominiert B im Sinne der **partiellen ROC-Ordnung** genau dann, wenn*

$$A \geq_{pROC} B \quad :\iff \quad \int_0^x ROC^A(t)dt \geq \int_0^x ROC^B(t)dt \quad \forall x \in [0, 1].$$

Offensichtlich gelten die Folgerungen $A \geq_{ROC} B \implies A \geq_{pROC} B$ und $A \geq_{pROC} B \implies AUC^A \geq AUC^B$, sodass sich diese Ordnung (Halbordnung, falls die Schuldnerkollektive identisch und die Prognosen kalibriert sind) in den Kontext bestehender Bewertungskriterien einordnen lässt. Wie man leicht durch Gegenbeispiele zeigt, gibt es hier keine eindeutige Beziehung zur Refinement-Ordnung sowie zur VM-Ordnung.

Krämer & Posch (2015) zeigen, dass die Lorenz- und ROC-Ordnung bei identischen Schuldnerpopulationen und kalibrierten Prognosen äquivalent zueinander sind. Außerdem sind diese Halbordnungen dann äquivalent zur Refinement-Ordnung, vgl. Satz 4.6.

Falls die empirische Lorenzkurve von A überall kleiner oder gleich der von B ist, gilt die **empirische Lorenz-Ordnung** $A \geq_{\widehat{L}} B$. Genauso sind alle in Kapitel 4 vorgestellten Halbordnungen für empirische Größen definiert, d. h. $A \geq_{\widehat{R}} B$, $A \geq_{\widehat{ROC}} B$, $A \geq_{\widehat{VM(d)}} B$, $A \geq_{\widehat{VM(nd)}} B$ usw.

Kapitel 5

Halbordnungen bei unterschiedlichen Schuldnerkollektiven

In diesem Kapitel werden kalibrierte Wahrscheinlichkeitsprognostiker verglichen, die (möglicherweise) unterschiedliche Schuldnerkollektive bewerten. Dazu wird zum einen die VM-Ordnung betrachtet und zum anderen die verallgemeinerte Lorenz-Ordnung definiert und untersucht.

5.1 Motivation

Liegen beim Vergleich von Wahrscheinlichkeitsprognostikern keine gemeinsamen Populationen von Kreditnehmern vor, so hat dies weitreichende Folgen für alle bislang vorgestellten Bewertungskriterien: Die Lorenz- und die ROC-Ordnung sowie die VM-Ordnung erfüllen, selbst bei kalibrierten Prognosen, nicht länger die Antisymmetrie und sind daher lediglich Präordnungen. Die Refinement-Ordnung ist, unter diesen allgemeineren Annahmen, nicht mehr anwendbar.

Insbesondere sind die Ordnungsprinzipien nun auf einer größeren Menge von kalibrierten Wahrscheinlichkeitsprognostikern definiert. Wurden bislang nur bivariate Verteilungen $P^{\Theta, X}$ mit identischer Randverteilung P^{Θ} verglichen, sind jetzt auch unterschiedliche Randverteilungen für P^{Θ} möglich.

Beispiel 5.1 zeigt, dass bei unterschiedlichen Schuldnerkollektiven und kalibrierten Prognosen die Lorenz-Ordnung und die ROC-Ordnung (Definition 4.16) nur Präordnungen sind (Reflexivität und Transitivität sind gegeben).

Beispiel 5.1. *Bei verschiedenen Schuldnerpopulationen verletzen die Lorenz- und ROC-Ordnung die Antisymmetrie: Betrachte einen kalibrierten Wahrscheinlichkeitsprognostiker auf einer Schuldnerpopulation mit einer Gesamtausfallwahrscheinlichkeit von $p_1 = 0.2$ und $\mathbb{W} = \{a_1, a_2, a_3\}$.*

- (1) *Mit den Prognosen $\mathbb{W} = \{0, 0.125, 0.375\}$ und $q(a_i)$ wie in Abbildung 3.1 (halbierte Prognosen) ergibt sich ebenfalls die Lorenzkurve aus Abbildung 3.1 (a).*
- (2) *Mit den Prognosen $\mathbb{W} = \{0, \frac{1}{9}, \frac{9}{17}\}$ und $q(a_i | 1)$ wie in Abbildung 3.1 erhält man die ROC-Kurve aus Abbildung 3.1 (b).*

Bei unterschiedlichen Schuldnerpopulationen gibt es also in der Menge der kalibrierten Wahrscheinlichkeitsprognosen verschiedene Verteilungen $P^{\Theta, X}$, welche die gleiche Lorenz- bzw. ROC-Kurve besitzen. Daher kann hier aus der Gleichheit der Kurven nicht auf die Gleichheit der Prognostiker geschlossen werden (Antisymmetrie ist verletzt).

Beispiel 5.2 veranschaulicht, dass bei unterschiedlichen Schuldnerpopulationen die Lorenz-Ordnung nicht mehr sinnvoll ist und dass die Lorenz- und ROC-Ordnung nicht mehr äquivalent zueinander sind.

Beispiel 5.2. *Abbildung 5.1 (a) zeigt, dass die optimale Lorenzkurve von Prognostiker B bis $1 - p_1^B$ bei Null verläuft und dann linear zum Punkt $(1, 1)'$ ansteigt. Infolgedessen ist es bei verschiedenen Populationen von Krediten problemlos möglich, dass die Lorenzkurve eines nicht optimalen Prognostikers unterhalb der eines optimalen Prognostikers verläuft. Die ROC-Ordnung bleibt in dieser Situation sinnvoll, es gilt $A \geq_{ROC} B$, vgl. Abbildung 5.1 (b). Doch auch die ROC-Ordnung ist hier lediglich eine Präordnung, da beispielsweise zwei perfekte Prognostiker auf verschiedenen Populationen dieselbe (optimale) ROC-Kurve besitzen (Antisymmetrie ist verletzt).*

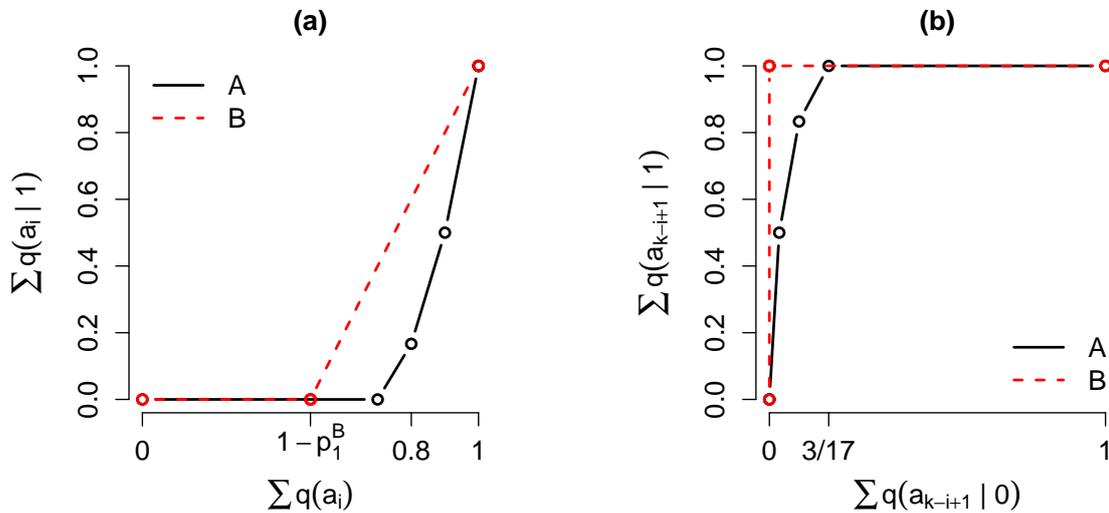


Abbildung 5.1: (a) Lorenzkurve des perfekten Prognostikers B bei einer Gesamtausfallwahrscheinlichkeit von $p_1^B = 0.5$ im Vergleich zur Lorenzkurve des Prognostikers A (kalibriert) mit $q^A(0) = 0.7$, $q^A(0.25) = q^A(0.5) = q^A(0.75) = 0.1$ und $p_1^A = 0.15$. (b) ROC-Kurven von A und B .

In den folgenden Kapiteln 5.2 und 5.3 sind Bewertungskriterien von Interesse, welche die Eigenschaften einer Schuldnerpopulation (hier die Gesamtausfallwahrscheinlichkeit) geeignet berücksichtigen.

5.2 Vardeman-Meeden-Ordnung

Die VM-Ordnung berücksichtigt die Gesamtausfallwahrscheinlichkeit in einer Population von Kreditnehmern, da die Verteilungen $q(a_i | 1) = r(a_i, 1)/p_1$ und $q(a_i | 0) = r(a_i, 1)/(1 - p_1)$ von der Gesamtausfallwahrscheinlichkeit p_1 abhängen. Damit ist die VM-Ordnung grundsätzlich geeignet, um Wahrscheinlichkeitsprognosen bei (möglicherweise) unterschiedlichen Schuldnerkollektiven zu bewerten. Jedoch verletzen z. B. zwei perfekte Prognostiker auf verschiedenen Populationen die Antisymmetrie der Ordnung, sodass die VM-Ordnung hier nur eine Präordnung ist.

Weiter folgt aus der VM-Ausfallordnung bzw. der VM-Nichtausfallordnung bei verschiedenen Schuldnerkollektiven nicht länger die Lorenz-Ordnung, vgl. Satz 4.13.

Auch die VM-Ausfallordnung und die VM-Nichtausfallordnung sind nun, selbst bei Kalibrierung, lediglich Präordnungen. Wird jedoch die Menge der möglichen Wahrscheinlichkeitsprognosen $W = \{a_1, \dots, a_k\}$ geeignet restringiert, so stellen die VM-Ausfallordnung und die VM-Nichtausfallordnung Halbordnungen dar:

Korollar 5.3. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf Schuldnerpopulationen mit Gesamtausfallwahrscheinlichkeiten $0 < p_1^A \leq p_1^B < 1$, dann gilt:*

(i) *Für $0 < a_1 < \dots < a_k \leq 1$ ist die VM-Ausfallordnung eine Halbordnung.*

(ii) *Für $0 \leq a_1 < \dots < a_k < 1$ ist die VM-Nichtausfallordnung eine Halbordnung.*

Beweis:

(i) Dass die VM-Ausfallordnung bei unterschiedlichen Populationen die Reflexivität und die Transitivität erfüllt, ist leicht zu sehen. Wie folgend gezeigt wird, ist durch $a_1 > 0$ garantiert, dass aus $A \geq_{VM(d)} B$ und $B \geq_{VM(d)} A$ auch $A = B$ folgt (Antisymmetrie):

Sei $A \geq_{VM(d)} B$ und $B \geq_{VM(d)} A$, so folgt $q^A(a_j | 1) = q^B(a_j | 1) \forall j = 1, \dots, k$. Also folgt mit der Kalibrierung $a_j q^A(a_j) = \frac{p_1^A}{p_1^B} a_j q^B(a_j) \forall j = 1, \dots, k$. Damit und mit $a_1 > 0$ folgt

$$\sum_{j=1}^k q^A(a_j) = \frac{p_1^A}{p_1^B} \sum_{j=1}^k q^B(a_j) \iff p_1^A = p_1^B.$$

(ii) Analog zu (i) folgt $q^A(a_j | 0) = q^B(a_j | 0) \forall j = 1, \dots, k$, sodass mit der Kalibrierung und mit $a_k < 1$ gilt: $\sum_{j=1}^k q^A(a_j) = \frac{1-p_1^A}{1-p_1^B} \sum_{j=1}^k q^B(a_j) \iff p_1^A = p_1^B. \quad \square$

Mit Korollar 5.3 folgt: Falls $a_1 > 0$ oder $a_k < 1$ gilt, so ist die VM-Ordnung auf der Menge der kalibrierten Prognosen eine Halbordnung bei unterschiedlichen Schuldnerkollektiven. Es reicht jedoch auch aus, wenn aus der Menge der kalibrierten Wahrscheinlichkeitsprognosen die perfekte Prognose ausgeschlossen wird. Auf dieser eingeschränkten Menge ist die VM-Ordnung eine Halbordnung bei unterschiedlichen Schuldnerkollektiven. Diese Aussage lässt sich ähnlich wie Korollar 5.3 nachweisen, unter Ausnutzung, dass $\sum_{i=1}^k a_i q(a_i) = \sum_{i=1}^k a_i^2 q(a_i)$ hier nur gilt, falls die perfekte Prognose vorliegt.

Wie bei identischen Schuldnerkollektiven, ist eine Dominanzaussage mittels der VM-Ordnung auch hier nur selten möglich: Satz 4.15 lässt sich unter Berücksichtigung der Gesamtausfallwahrscheinlichkeit wie folgt verallgemeinern.

Satz 5.4. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker auf Schuldnerpopulationen mit Gesamtausfallwahrscheinlichkeiten $0 < p_1^A \leq p_1^B < 1$, dann gilt:*

(i) *Falls $q^A(0) = q^B(0) = 0$, so ist $A >_{VM(d)} B$ nicht möglich.*

(ii) *Falls $q^A(1) = q^B(1) = 0$, so ist $B >_{VM(nd)} A$ nicht möglich.*

Beweis:

(i) Der Nachweis erfolgt durch Widerspruch: Sei $q^A(0) = q^B(0) = 0$, $p_1^A \leq p_1^B$ und $A >_{VM(d)} B$. Weiter sei o. B. d. A. $0 = a_1 < a_2 < \dots < a_k = 1$ und $q^A(a_2 | 1) < q^B(a_2 | 1)$.

$$\begin{aligned} \implies r^A(a_2, 1) &< r^B(a_2, 1) \underbrace{\frac{p_1^A}{p_1^B}}_{\leq 1} \leq r^B(a_2, 1) \\ \implies q^A(a_2) &< q^B(a_2). \quad (\star) \end{aligned}$$

Aus $A >_{VM(d)} B$ folgt weiter $q^A(a_2 | 1) + q^A(a_3 | 1) \leq q^B(a_2 | 1) + q^B(a_3 | 1)$.

$$\begin{aligned} \implies r^A(a_2, 1) + r^A(a_3, 1) &\leq \left(r^B(a_2, 1) + r^B(a_3, 1) \right) \underbrace{\frac{p_1^A}{p_1^B}}_{\leq 1} \\ \implies a_2 q^A(a_2) + a_3 q^A(a_3) &\leq a_2 q^B(a_2) + a_3 q^B(a_3) \\ \implies a_3 q^A(a_3) &\leq \underbrace{a_2}_{< a_3} \underbrace{(q^B(a_2) - q^A(a_2))}_{\substack{(\star) \\ > 0}} + a_3 q^B(a_3) \\ &< a_3 (q^B(a_2) - q^A(a_2)) + a_3 q^B(a_3) \\ \implies q^A(a_2) + q^A(a_3) &< q^B(a_2) + q^B(a_3). \end{aligned}$$

Analog wie in Krämer (2005) kann sukzessiv $\sum_{i=2}^k q^A(a_i) < \sum_{i=2}^k q^B(a_i)$ gefolgert werden, was im Widerspruch zu $q^A(0) = q^B(0) = 0$ steht.

(ii) Sei $q^A(1) = q^B(1) = 0$, $B >_{VM(nd)} A$ und o. B. d. A. $\sum_{i=1}^{k-2} q^A(a_i | 0) < \sum_{i=1}^{k-2} q^B(a_i | 0)$

$$\begin{aligned} \implies & 1 - q^A(a_{k-1} | 0) < 1 - q^B(a_{k-1} | 0) \\ \implies & (1 - a_{k-1})q^A(a_{k-1}) > (1 - a_{k-1})q^B(a_{k-1}) \underbrace{\frac{1 - p_1^A}{1 - p_1^B}}_{\geq 1} \\ \implies & q^A(a_{k-1}) > q^B(a_{k-1}). \quad (***) \end{aligned}$$

Aus $B >_{VM(nd)} A$ folgt weiter

$$\begin{aligned} & \sum_{i=1}^{k-3} q^A(a_i | 0) \leq \sum_{i=1}^{k-3} q^B(a_i | 0) \\ \implies & 1 - q^A(a_{k-1} | 0) - q^A(a_{k-2} | 0) \leq 1 - q^B(a_{k-1} | 0) - q^B(a_{k-2} | 0) \\ \implies & (1 - a_{k-2})q^A(a_{k-2}) \geq \underbrace{(1 - a_{k-1})}_{> 1 - a_{k-2}} (q^B(a_{k-1}) - q^A(a_{k-1})) \\ & \quad + (1 - a_{k-2})q^B(a_{k-2}) \\ \stackrel{(***)}{\implies} & q^A(a_{k-2}) + q^A(a_{k-1}) > q^B(a_{k-2}) + q^B(a_{k-1}). \end{aligned}$$

Wieder kann sukzessiv $\sum_{i=1}^{k-1} q^A(a_i) > \sum_{i=1}^{k-1} q^B(a_i)$ gefolgert werden, was im Widerspruch zu $q^A(1) = q^B(1) = 0$ steht. \square

Korollar 5.5. *Seien A und B kalibrierte Wahrscheinlichkeitsprognostiker mit $0 = a_1 < a_2 < \dots < a_k = 1$ sowie $0 < p_1^A < 1$ und $0 < p_1^B < 1$, dann gilt:*

(i) *Falls $q^A(0) = q^B(0) = 0$ und $q^A(a_2) > 0 = q^B(a_2)$, so sind $A \geq_{VM(d)} B$ und $B \geq_{VM(nd)} A$ nicht möglich.*

(ii) *Falls $q^A(1) = q^B(1) = 0$ und $q^A(a_{k-1}) > 0 = q^B(a_{k-1})$, so sind $B \geq_{VM(d)} A$ und $A \geq_{VM(nd)} B$ nicht möglich.*

Beweis: (i) Seien A, B wie gefordert mit $q^A(0) = q^B(0) = 0$ und $q^A(a_2) > 0 = q^B(a_2)$. Damit ist vor allem $\sum_{i=1}^2 q^A(a_i | 1) > 0 = \sum_{i=1}^2 q^B(a_i | 1)$, so dass $A \geq_{VM(d)} B$ nicht möglich ist. Genauso gilt $\sum_{i=1}^2 q^A(a_i | 0) > 0 = \sum_{i=1}^2 q^B(a_i | 0)$, weshalb $B \geq_{VM(nd)} A$ nicht möglich ist. Teil (ii) folgt analog. \square

Bemerkung 5.6. *In der Situation von Korollar 5.5 ist auch keine stochastische Dominanz (VM-Ordnung) höherer Ordnung für die bedingten Prognoseverteilungen möglich.*

5.3 Verallgemeinerte Lorenz-Ordnung

Wie bereits gesehen, ist keines der genannten Ordnungsprinzipien (VM-Ordnung, Lorenz-Ordnung, ROC-Ordnung) bei unterschiedlichen Schuldnerkollektiven und kalibrierten Prognosen im Allgemeinen eine Halbordnung. Diese Lücke schließen Krämer & Neumärker (2016), indem sie die verallgemeinerte Lorenz-Ordnung definieren.

Abbildung 5.1 (a) macht deutlich: Die Lorenzkurve der perfekten Prognose strebt bei steigender Gesamtausfallwahrscheinlichkeit gegen die Lorenzkurve der Trivialprognose. Daher beweist insbesondere *der* Prognostiker eine hohe Trennschärfe, dessen Lorenzkurve bei größerer Gesamtausfallwahrscheinlichkeit unterhalb der des Konkurrenten verläuft. Diese Überlegung führt zur

Definition 5.7. *Seien A und B Wahrscheinlichkeitsprognostiker auf (möglicherweise) verschiedenen Schuldnerkollektiven mit $p_1^A > 0$ und $p_1^B > 0$. A dominiert B im Sinne der **verallgemeinerten Lorenz-Ordnung** genau dann, wenn*

$$A \geq_{VL} B \quad :\iff \quad A \geq_L B \quad \underline{\text{und}} \quad p_1^A \geq p_1^B.$$

Für $p_1^A = p_1^B$ reduziert sich die verallgemeinerte Lorenz-Ordnung auf die Lorenz-Ordnung aus Definition 4.16. Dass die verallgemeinerte Lorenz-Ordnung bei kalibrierten Prognosen die Eigenschaften Transitivität, Reflexivität und Antisymmetrie erfüllt und somit eine Halbordnung ist, ist offensichtlich.

Bemerkung 5.8. *Krämer & Posch (2015) betrachten bei verschiedenen Schuldnerkollektiven die Lorenzkurve der Ausfälle (Definition 3.1) und die Lorenzkurve der Nichtausfälle (Bemerkung 3.3 (iii)). Für einen Prognostiker ist es dann erstrebenswert, wenn für ihn Dominanz bezüglich beider Lorenzkurven besteht. In Beispiel 5.2 gilt dies für keinen der beiden Prognostiker (Abbildung A.2 im Anhang zeigt die Lorenzkurven der Nichtausfälle).*

Folgender Satz aus Krämer & Neumärker (2016) stellt einen Zusammenhang zwischen der verallgemeinerten Lorenz-Ordnung und Skill-Scores her.

Satz 5.9. *Seien A und B empirisch kalibrierte Wahrscheinlichkeitsprognostiker mit $A \geq_{\widehat{VL}} B$. Dann ist A mindestens genauso gut wie B , im Sinne des Brier-Skill-Scores und des Logarithmus-Skill-Scores.*

Dieses Resultat wird im Folgenden auf eine größere Klasse von Score-Regeln erweitert, dazu vorab:

Korollar 5.10. Seien $\vec{x} \in \widehat{W}^n$ empirisch kalibrierte Wahrscheinlichkeitsprognosen für die Ausfälle $\vec{\theta} \in \{0, 1\}^n$. Weiter sei $S \in \mathbb{L}_{rnn}$ eine anreizerhaltende Score-Regel mit zweifach differenzierbarer erwarteter Score-Funktion $g(x)$ und $xg''(x)$ monoton fallend auf $[0, 1]$. Dann gilt für alle $c \in [0, 1]$ und $\hat{p}_1 = \frac{1}{n} \sum_{i=1}^n \theta_i > 0$:

$$\frac{1}{n} \sum_{i=1}^n g(\hat{p}_1)g(cx_i) \geq \frac{1}{n} \sum_{i=1}^n g(c\hat{p}_1)g(x_i). \quad (5.1)$$

Beweis: Für $c = 1$ ist die Gleichheit in Gleichung (5.1) offensichtlich, für $c = 0$ folgt sie sofort mit $g(0) = 0$. Sei nun $c \in (0, 1)$. Zunächst gilt aufgrund der empirischen Kalibrierung $\hat{p}_1 = \frac{1}{n} \sum_{i=1}^n x_i$. Falls $f(x) := g(\hat{p}_1)g(cx) - g(c\hat{p}_1)g(x)$ eine konvexe Funktion auf $[0, 1]$ ist, so folgt die Gültigkeit von Gleichung (5.1) direkt mit der Jensen-Ungleichung. Zeige hier also $f''(x) = c^2g(\hat{p}_1)g''(cx) - g(c\hat{p}_1)g''(x) \geq 0$, $\forall x \in [0, 1]$. Da g konkav ist, gilt für alle $c \in (0, 1)$:

$$g(c\hat{p}_1) \geq cg(\hat{p}_1) + \underbrace{(1-c)g(0)}_{=0} = cg(\hat{p}_1). \quad (\star)$$

Damit ist

$$\begin{aligned} f''(x) &= c^2g(\hat{p}_1)g''(cx) + \underbrace{g(c\hat{p}_1)}_{\substack{(\star) \\ \geq cg(\hat{p}_1)}} \underbrace{(-g''(x))}_{\geq 0} \\ &\geq c^2g(\hat{p}_1)g''(cx) + cg(\hat{p}_1)(-g''(x)) \\ &= \underbrace{cg(\hat{p}_1)}_{\geq 0} (cg''(cx) - g''(x)). \end{aligned}$$

Also ist eine hinreichende Bedingung für $f''(x) \geq 0$ gegeben durch

$$cg''(cx) - g''(x) \geq 0 \quad \iff \quad xg''(x) \leq cxg''(cx). \quad (5.2)$$

Für $c \in (0, 1)$ und $x \in [0, 1]$ ist Gleichung (5.2) äquivalent dazu, dass $xg''(x)$ eine monoton fallende Funktion auf $[0, 1]$ ist, was nach Voraussetzung gilt. \square

Satz 5.11. *Sei $S \in \mathbb{L}_{rnn}$ eine anreizerhaltende Score-Regel mit zweifach differenzierbarer erwarteter Score-Funktion $g(x)$ und $xg''(x)$ monoton fallend auf $[0, 1]$. Weiter seien A und B empirisch kalibrierte Prognostiker mit $A \geq_{\widehat{V}_L} B$. Dann ist A mindestens genauso gut wie B , im Sinne des zu S gehörenden Skill-Scores.*

Beweis: Sei \hat{p}_1^A bzw. \hat{p}_1^B der zu Prognostiker A bzw. B gehörende Gesamtausfallanteil in der Schuldnerpopulation. Mit Satz 4.6 ist leicht ersichtlich, dass hier im Falle $\hat{p}_1^A = \hat{p}_1^B$ der Prognostiker A mindestens genauso gut ist wie B für alle Skill-Scores, die auf einer anreizerhaltenden Score-Regel beruhen. Sei nun aber o. B. d. A. $0 < \hat{p}_1^B < \hat{p}_1^A$ und $c := \hat{p}_1^B / \hat{p}_1^A < 1$. Betrachte einen weiteren empirisch kalibrierten Vorhersager A^* mit möglichen Vorhersagen $\hat{a}_i^* := c\hat{a}_i$ und Häufigkeiten $\hat{q}^{A^*}(a_i^*) = \hat{q}^A(\hat{a}_i)$. Wegen der Invarianz der Lorenzkurve unter positiver Skalierung, haben A und A^* die gleichen empirischen Lorenzkurven und die empirische Lorenzkurve von B liegt nach Voraussetzung unter der von A^* . Weiter gilt nach Konstruktion $\hat{p}_1^{A^*} = \hat{p}_1^B$, sodass A^* nicht schlechter ist als B für jede anreizerhaltende Score-Regel, was hier insbesondere auch für die zugehörigen Skill-Scores (SK) gilt. Diese sind zwar nicht anreizerhaltend, jedoch für gleiche Gesamtausfallanteile äquivalent zu den zugrundeliegenden Score-Regeln. Damit bleibt die folgende erste Ungleichheit zu zeigen

$$SK^A \stackrel{!}{\geq} SK^{A^*} \stackrel{\text{Satz 4.6}}{\geq} SK^B.$$

Die Behauptung $SK^{A^*} \leq SK^A$ lässt sich unter Ausnutzen der empirischen Kalibrierung für $S \in \mathbb{L}_{rnn}$ umschreiben zu

$$\frac{\frac{1}{n} \sum_{i=1}^n g(cx_i^A) - g(c\hat{p}_1^A)}{-g(c\hat{p}_1^A)} \leq \frac{\frac{1}{n} \sum_{i=1}^n g(x_i^A) - g(\hat{p}_1^A)}{-g(\hat{p}_1^A)}, \quad (5.3)$$

wobei x_i^A die Wahrscheinlichkeitsprognosen von Prognostiker A sind, $i = 1, \dots, n$. Gleichung (5.3) ist äquivalent zu $\frac{1}{n} \sum_{i=1}^n (g(\hat{p}_1^A)g(cx_i^A) - g(c\hat{p}_1^A)g(x_i^A)) \geq 0$. Diese Ungleichung ist wegen Korollar 5.10 hinreichend erfüllt, da $xg''(x)$ eine monoton fallende Funktion auf $[0, 1]$ ist. \square

Satz 5.9 folgt mit Satz 5.11, denn es gilt: Für den Brier-Score ist $xg''(x) = -2x$ monoton fallend auf $[0, 1]$. Für den Logarithmus-Score ist $xg''(x) = \frac{1}{x-1}$ monoton fallend auf $(0, 1)$ und f aus dem Beweis von Korollar 5.10 ist stetig und somit auch konvex auf $[0, 1]$.

Die Funktion $xg''(x)$ zu einer anreizerhaltenden Score-Regel $S \in \mathbb{L}_{rnn}$ mit zweifach differenzierbarer erwarteter Score-Funktion g , ist genau dann monoton fallend auf $[0, 1]$, wenn

$$-\int xg''(x)dx = g(x) - g'(x)x = s(0, x)$$

konvex auf $[0, 1]$ ist, vgl. Satz 3.14 (Savage-Darstellung).

Bemerkung 5.12.

- (i) Neben dem Logarithmus-Score und dem Brier-Score ist die Aussage von Satz 5.9 beispielsweise auch für den hyperbolischen Score aus Kapitel 3.2 gültig, denn $s(0, x)$ ist konvex auf $[0, 1]$, siehe Abbildung 5.2 (a).

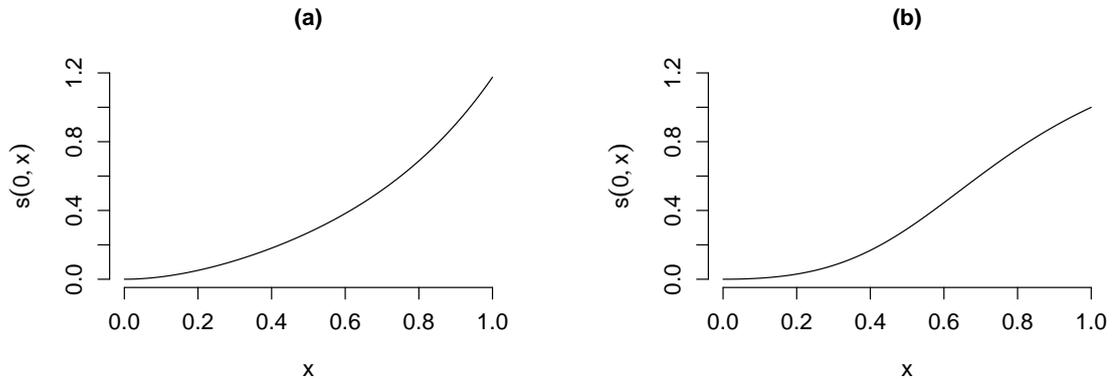


Abbildung 5.2: (a) Die Score-Funktion $s(0, x)$ des hyperbolischen Scores ist konvex auf $[0, 1]$, sodass die Voraussetzung aus Satz 5.11 erfüllt ist. (b) Hingegen ist $s(0, x)$ des sphärischen Scores nicht konvex, sodass Satz 5.11 für diesen nicht anwendbar ist.

- (ii) Eine Score-Regel mit konvexer Funktion $s(0, x)$ bestraft Abweichungen zwischen $\theta = 0$ und der Prognose mit wachsender Prognose „besonders stark“, d. h.

$$\Delta_d s(0, x) \leq \Delta_d s(0, y) \quad \forall x \leq y$$

mit $\Delta_d s(0, x) := s(0, x + d) - s(0, x) \geq 0$ für ein $d \in [0, 1 - x]$. Die Konvexität von $s(0, x)$ ist in diesem Sinne also durchaus erstrebenswert.

Satz 5.11 ist ohne zusätzliche Annahmen nicht auf alle (streng) anreizerhaltenden Score-Regeln erweiterbar, wie folgendes Beispiel zeigt.

Beispiel 5.13. Sei A ein Prognostiker mit Prognosen $x_1 = \dots = x_{500} = 0.1$ und $x_{501} = \dots = x_{1000} = 0.4$ und B ein Prognostiker mit Prognosen $y_i = 0.5x_i + 0.01$ für $i = 1, \dots, 1000$ (beide empirisch kalibriert). Weiter sei eine anreizerhaltende Score-Regel gegeben, mit (konkaver) erwarteter Score-Funktion

$$g(x) := 0.5^8 - (x - 0.5)^8.$$

Abbildung 5.3 zeigt die zu A und B gehörenden empirischen Lorenzkurven. Damit ist $A \geq_{\widehat{VL}} B$, jedoch gilt für die Skill-Scores konstruiert aus g :

$$SK^A = 0.08 < 0.11 = SK^B.$$

Somit ist hier die Folgerung $A \geq_{\widehat{VL}} B \implies SK^A \geq SK^B$ nicht für jede anreizerhaltende Score-Regel $S \in \mathbb{L}_{rnn}$ richtig, aber für solche mit $xg''(x)$ monoton fallend (hier ist $xg''(x) = -56x(x - 0.5)^6$ nicht monoton fallend auf $[0, 1]$). Für den Brier-Skill-Score gilt hingegen: $BSK^A = 0.12 > 0.05 = BSK^B$.

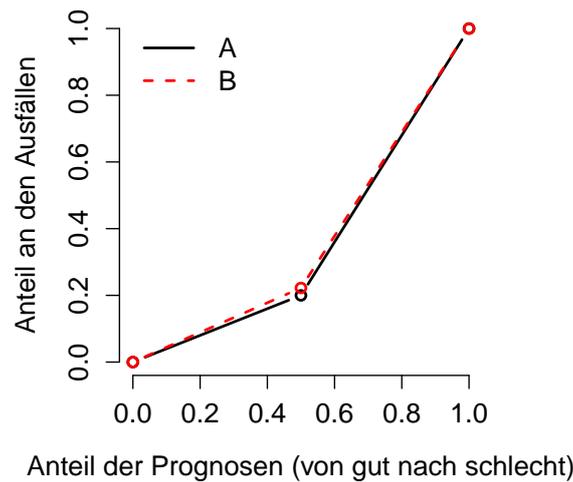


Abbildung 5.3: Empirische Lorenzkurven von Prognostiker A mit Wahrscheinlichkeitsprognosen $x_1 = \dots = x_{500} = 0.1$ und $x_{501} = \dots = x_{1000} = 0.4$ sowie Prognostiker B mit Prognosen $y_i = 0.5x_i + 0.01$ für $i = 1, \dots, 1000$ (beide empirisch kalibriert). Damit ist $\hat{p}_1^A = 0.25 > \hat{p}_1^B = 0.135$, sodass $A \geq_{\widehat{VL}} B$ folgt.

Korollar 5.14. Sei $S \in \mathbb{L}_{rnn}$ eine anreizerhaltende Score-Regel mit dreifach differenzierbarer erwarteter Score-Funktion g , welche $g(x) = g(1 - x) \forall x \in [0, 1]$ erfülle, dann ist:

$$s''(0, x) \geq 0 \quad \forall x \in [0, 1] \quad \iff \quad s''(1, x) \geq 0 \quad \forall x \in [0, 1].$$

Beweis: Wegen der Symmetrieeigenschaft $g(x) = g(1 - x) \forall x \in [0, 1]$, ist

$$g''(x) = g''(1 - x) \quad \text{und} \quad g'''(x) = -g'''(1 - x) \quad (\star)$$

und für ein beliebiges $x \in [0, 1]$ folgt mit Gleichung (3.4):

$$\begin{aligned} s''(0, 1 - x) &= -g''(1 - x) - (1 - x)g'''(1 - x) \geq 0 \\ &\iff g''(1 - x) \leq -(1 - x)g'''(1 - x) \\ &\stackrel{(\star)}{\iff} g''(x) \leq (1 - x)g'''(x) \\ &\iff -g''(x) + (1 - x)g'''(x) \geq 0 \\ &\stackrel{(3.4)}{\iff} s''(1, x) \geq 0. \end{aligned}$$

Da $x \in [0, 1]$ beliebig ist, folgt die Behauptung. □

Wegen Korollar 5.14 ist für eine anreizerhaltende Score-Regel $S \in \mathbb{L}_{rnn}$ mit um 0.5 symmetrischer erwarteter Score-Funktion die Forderung, dass $s(0, x)$ konvex ist (nötig für Satz 5.11) äquivalent dazu, dass $s(1, x)$ konvex ist.

Kapitel 6

Gewichtete monotone Regression

Die meisten Methoden aus den Kapiteln 3 bis 5 benutzen explizite Ausfallwahrscheinlichkeiten. Daher ist es von praktischem Interesse, einer durch Ratings bewerteten Schuldnerpopulation explizite Ausfallwahrscheinlichkeiten zuzuordnen bzw. diese *ex post* zu schätzen. Dazu wird in diesem Kapitel die gewichtete monotone Regression vorgestellt. Es wird gezeigt, dass dieses Verfahren im vorliegenden Kontext Lorenz-optimal zur Schätzung monoton steigender Ausfallwahrscheinlichkeiten in empirisch kalibrierten Klassen ist.

6.1 Motivation

Bei der Bewertung von Kreditratings ist die vom Prognostiker unterstellte Menge der Ausfallwahrscheinlichkeiten $W = \{a_1, \dots, a_k\}$ in der Regel unbekannt. Für den i -ten Kredit in der Schuldnerpopulation sei hier also nur die Ratingklasse $r_i \in \{R_1, \dots, R_k\}$ und der Wert der Ausfallvariablen $\theta_i \in \{0, 1\}$ gegeben, $i = 1, \dots, n$.

Weiter sei jedes Rating mit einer Ausfallwahrscheinlichkeit über eine unbekannte Funktion f verknüpft. Dabei sei f monoton steigend (**isoton**) in den Ratingklassen, sodass $f(R_j) = a_j$ für $j = 1, \dots, k$ und $0 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq 1$ ist.

Es ist in diesem Kapitel also nicht von Interesse, ein Rating bzw. eine Ausfallwahrscheinlichkeit anhand von individuellen Schuldnermerkmalen zu erstellen. Dies ist Aufgabe der Ratingagentur. Stattdessen ist das Ziel, den gegebenen Ratings r_i ex post eine Ausfallwahrscheinlichkeit x_i für $i = 1, \dots, n$ zuzuordnen bzw. gleichbedeutend den Ratingkategorien R_j eine Ausfallwahrscheinlichkeit \hat{a}_j für $j = 1, \dots, k$. Also werden die Ausfallwahrscheinlichkeiten $a_j = f(R_j)$ für $j = 1, \dots, k$ geschätzt und zwar so, dass diese erstens in den Ratingklassen monoton steigend sind (da f isoton) und dass diese zweitens zu empirischer Kalibrierung (Definition 2.7) führen.

Dazu ist der empirische Ausfallanteil in einer Ratingkategorie R_j für $j = 1, \dots, k$ definiert als

$$\tilde{a}_j := \begin{cases} \frac{\sum_{i=1}^n \theta_i \mathbb{1}_{\{R_j\}}(r_i)}{\sum_{i=1}^n \mathbb{1}_{\{R_j\}}(r_i)}, & \text{falls } \sum_{i=1}^n \mathbb{1}_{\{R_j\}}(r_i) > 0 \\ 0, & \text{sonst.} \end{cases}$$

Ein erster naiver Ansatz zur Lösung des obigen Problems ist, die Ausfallwahrscheinlichkeiten der Ratingkategorien durch \tilde{a}_j zu schätzen, falls $\sum_{i=1}^n \mathbb{1}_{\{R_j\}}(r_i) > 0$. Diese Prognosen sind dann zwar nach Konstruktion empirisch kalibriert, jedoch nicht zwangsläufig monoton steigend in den Ratingklassen. Konsequenterweise werden Ratingklassen oftmals zusammengefasst, sodass deren Ausfallanteile monoton steigend sind. Beispielsweise arbeitet Krämer (2017) mit der reduzierten Ratingskala AAA, AA, A, BBB, BB, B, CCC-C von Standard & Poor's. Diese Skala ist inhaltlich sinnvoll, da nur Ratings mit gleichen Buchstaben zusammengefasst werden, vgl. Tabelle 7.1. Die Art und Weise, wie die Ratingkategorien zusammengefasst werden,

hat jedoch offensichtlich Einfluss auf die Spreizung der empirischen Ausfallanteile. Würde man beispielsweise alle Ratings in einen Topf werfen, würde man in diesem Topf nur noch den empirischen Gesamtausfallanteil (die Trivialprognose) beobachten. Daher ist es aus mathematischer Sicht interessant, welche Zusammenlegung der Ratingklassen die Spreizung der geschätzten Ausfallwahrscheinlichkeiten (die isotonen empirischen Ausfallanteile) maximiert. Um diese Frage zu beantworten, wird folgend die gewichtete monotone Regression vorgestellt.

6.2 Schätzung isotoner

Ausfallwahrscheinlichkeiten

Die monotone Regression wird zum Beispiel zur Klassifikation beim maschinellen Lernen (Zadrozny & Elkan, 2002) oder zur Erstellung von Dosis-Wirkungs-Kurven (Stylianou & Flournoy, 2002) angewandt. In dieser Arbeit, aber auch beispielsweise in Härdle et al. (2007), wird die monotone Regression als nichtparametrisches Verfahren zur Schätzung monotoner Ausfallwahrscheinlichkeiten genutzt. Die gewichtete monotone Regression ist eine gewichtete Kleinste-Quadrate-Schätzung unter Restriktion (Robertson et al., 1988, S. 4).

Definition 6.1. Der Vektor $(a_1^*, \dots, a_k^*)' \in \mathbb{R}^k$ ist die **Schätzung der monotonen Regression** zum Vektor $(\tilde{a}_1, \dots, \tilde{a}_k)' \in \mathbb{R}^k$ mit Gewichten $\hat{q}(R_j) \in \mathbb{R}$, $j = 1, \dots, k$, falls

$$(a_1^*, \dots, a_k^*)' = \arg \min_{(z_1, \dots, z_k)' \in \mathbb{R}^k} \sum_{j=1}^k \hat{q}(R_j) (z_j - \tilde{a}_j)^2, \quad (6.1)$$

unter den Nebenbedingungen $z_j \leq z_{j+1}$ für $j = 1, \dots, k - 1$.

Zur Schätzung isotoner Ausfallwahrscheinlichkeiten sind die **Gewichte der Regression** als Anteile der Beobachtungen in den Ratingklassen definiert, also

$$\hat{q}(R_j) := n^{-1} \sum_{i=1}^n \mathbb{1}_{\{R_j\}}(r_i), \quad j = 1, \dots, k.$$

Bemerkung 6.2. *Die Schätzungen der monotonen Regression für die Ausfallwahrscheinlichkeiten sind zwar verzerrt, jedoch konsistent, d. h. $a_i^* \xrightarrow{n \rightarrow \infty} a_i$. Dies ist intuitiv, da für $q(a_j) > 0$ auch $\tilde{a}_i \xrightarrow{n \rightarrow \infty} a_i$ gilt und $a_1 \leq \dots \leq a_k$ ist (Robertson et al., 1988, S. 42 & S. 391/392).*

Zur Lösung des Minimierungsproblems (6.1) existieren eine Vielzahl von Lösungsalgorithmen. Nachfolgend wird der intuitive Pool-Adjacent-Violators-Algorithmus vorgestellt (Barlow et al., 1972, S. 13-15). Die Grundidee des Algorithmus ist es, benachbarte Ratingklassen geeignet zusammenzufassen, falls deren Ausfallanteile nicht isoton sind. Den so zusammengefassten Ratingklassen wird dann der gemeinsame mittlere Ausfallanteil zugeordnet. Sukzessives Zusammenfassen der Ratingklassen und Mitteln der entsprechenden Ausfallanteile (diese Schritte werden auch als das Poolen von Ausfallanteilen bezeichnet) führen letztlich zu einer isotonen Folge von geschätzten Ausfallwahrscheinlichkeiten.

Definition 6.3. *In der Situation von Definition 6.1 sei a_1^*, \dots, a_k^* die Lösung des Minimierungsproblems (6.1). Der **Pool-Adjacent-Violators-Algorithmus** (PAV-Algorithmus) erhält diese Lösung in folgenden Schritten:*

- (1) Setze $a_i^* := \tilde{a}_i \geq 0$ und $q^*(R_i) := \hat{q}(R_i)$ für $i = 1, \dots, k$ und $a_{k+1}^* := -1$.

(2) Sei $l := \min\{i : a_i^* > a_{i+1}^*\}$. Falls $l = k$, so ist a_1^*, \dots, a_k^* die Lösung des Minimierungsproblems (6.1). Falls $l < k$, so setze

$$a_l^* := a_{l+1}^* := \frac{q^*(R_l)a_l^* + q^*(R_{l+1})a_{l+1}^*}{q^*(R_l) + q^*(R_{l+1})},$$

$$q^*(R_l) := q^*(R_{l+1}) := q^*(R_l) + q^*(R_{l+1})$$

und gehe zu (2).

Die Lösung des Minimierungsproblems (6.1) kann auch grafisch mithilfe der kumulierten Summenkurven konstruiert werden (Silvapulle & Sen, 2005, S. 48).

Definition 6.4.

(i) Die **kumulierte Summenkurve** (cumulative sum diagram (CSD)) trägt den kumulierten Anteil der Beobachtungen in den Ratingklassen (beste Ratingklassen zuerst) gegen die kumulierte Summe der gewichteten Ausfallanteile zu den Ratingklassen (beste Ratingklassen zuerst) gegeneinander ab. Sie ist der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j \hat{q}(R_i), \sum_{i=1}^j \hat{q}(R_i) \tilde{a}_i \right)', \quad j = 0, \dots, k. \quad (6.2)$$

(ii) Die **größte konvexe Minorante** (greatest convex minorant (GCM)) der kumulierten Summenkurve ist der Graph des Maximums über alle konvexen Funktionen, deren Graphen unterhalb der kumulierten Summenkurve liegen:

$$GCM(x) := \max\{u(x) : u \text{ ist konvex und } u(x) \leq CSD(x) \forall x\}.$$

Robertson et al. (1988, S. 7/8) liefern den Zusammenhang zwischen der GCM und der Lösung der monotonen Regression. Sie zeigen, dass die Lösungen a_1^*, \dots, a_k^* des Minimierungsproblems (6.1) der monotonen Regression durch die Steigungen der GCM festgelegt sind. Konkret ist a_j^* die Steigung der GCM auf dem Intervall $\left(\sum_{i=1}^{j-1} \hat{q}(R_i), \sum_{i=1}^j \hat{q}(R_i)\right)$ für $j = 1, \dots, k$ (falls $\hat{q}(R_j) > 0$).

Andererseits ist damit aber auch die GCM der kumulierten Summenkurve über die Lösung des Minimierungsproblems (6.1) a_1^*, \dots, a_k^* definiert: Die GCM ist der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j \hat{q}(R_i), \sum_{i=1}^j \hat{q}(R_i) a_i^*\right)', \quad j = 0, \dots, k. \quad (6.3)$$

Im Folgenden wird die GCM aus Gleichung (6.3) mit der Lorenzkurve der geschätzten Ausfallwahrscheinlichkeiten a_1^*, \dots, a_k^* in Verbindung gebracht, vgl. Definition (3.1). So wird gezeigt, dass die mittels der monotonen Regression erhaltene Lösung, die also dem Rating R_j die geschätzte Ausfallwahrscheinlichkeit a_j^* aus Gleichung (6.1) zuordnet, Lorenz-optimal in empirisch kalibrierten Klassen ist.

In dieser Situation sei eine feste Population von bewerteten Krediten gegeben, mit Ausfällen $\theta_i \in \{0, 1\}$ und Ratings $r_i \in \{R_1, \dots, R_k\}$, $i = 1, \dots, n$. Diese gegebene Population sei mit (P) bezeichnet. Somit legt (P) vor allem die empirischen Ausfallanteile in den Ratingklassen $\tilde{a}_1, \dots, \tilde{a}_k$ und die empirischen relativen Häufigkeiten der Ratingklassen $\hat{q}(R_1), \dots, \hat{q}(R_k)$ fest.

Definition 6.5. Sei (P) eine Population von bewerteten Krediten mit Ausfällen $\theta_i \in \{0, 1\}$ und Ratings $r_i \in \{R_1, \dots, R_k\}$, $i = 1, \dots, n$. Die geschätzten Ausfallwahrscheinlichkeiten $\hat{a}_1, \dots, \hat{a}_k$ heißen **verträglich** mit (P) , wenn:

$$(a) \quad 0 \leq \hat{a}_1 \leq \dots \leq \hat{a}_k \leq 1,$$

(b) Mit der Zuordnung $r_i = R_j \implies x_i = \hat{a}_j$ für $i = 1, \dots, n$ und $j = 1, \dots, k$ ist $(x_1, \dots, x_n)'$ empirisch kalibriert für $(\theta_1, \dots, \theta_n)'$.

In Definition 6.5 sichert die Bedingung (b) zum einen die empirische Kalibrierung einer verträglichen Schätzung und zum anderen, dass die durch die Ratings vorgegebenen Klassen erhalten bleiben oder zusammengefasst werden.

Bemerkung 6.6.

(i) Die empirischen Ausfallanteile $\tilde{a}_1, \dots, \tilde{a}_k$ erfüllen (b) aus Definition 6.5. Sie sind genau dann verträglich mit (P) , wenn $\tilde{a}_1 \leq \dots \leq \tilde{a}_k$.

(ii) Die Trivialprognose, also $\hat{a}_i := \hat{p}_1 \forall i = 1, \dots, k$, ist eine mit (P) verträgliche Schätzung.

Anhang C untersucht die Menge der mit (P) verträglichen Schätzungen und zeigt, dass die Schätzung der monotonen Regression a_1^*, \dots, a_k^* verträglich ist.

Die Lorenzkurve $L_{\hat{a}_1, \dots, \hat{a}_k}$ einer mit (P) verträglichen Schätzung $\hat{a}_1, \dots, \hat{a}_k$ ist, wegen der empirischen Kalibrierung von $(x_1, \dots, x_n)'$ mit Zuordnung wie in Definition 6.5 (b), gerade der Polygonzug durch die Punkte

$$\left(\sum_{i=1}^j \hat{q}(R_i), \frac{1}{\hat{p}_1} \sum_{i=1}^j \hat{q}(R_i) \hat{a}_i \right)', \quad j = 0, \dots, k, \quad (6.4)$$

wobei $\hat{p}_1 = \sum_{i=1}^k \tilde{a}_i \hat{q}(R_i) > 0$ der Gesamtausfallanteil in der betrachteten Population ist, vgl. Definition 3.1.

Mit dieser Notation wird nun die Lorenz-Optimalität der monotonen Regression in der Menge der verträglichen Schätzungen herausgestellt:

Satz 6.7. *Seien $\hat{a}_1, \dots, \hat{a}_k$ mit (P) verträglich und a_1^*, \dots, a_k^* die Lösungen des Minimierungsproblems (6.1) der monotonen Regression sowie $\hat{p}_1 > 0$. Dann gilt für die zugehörigen Lorenzkurven: $L_{a_1^*, \dots, a_k^*}(x) \leq L_{\hat{a}_1, \dots, \hat{a}_k}(x)$ für alle $x \in [0, 1]$.*

Beweis: Offensichtlich entspricht die empirische Lorenzkurve $L_{\tilde{a}_1, \dots, \tilde{a}_k}$ (Notation wie in Gleichung (6.4)) der kumulierten Summenkurve aus Gleichung (6.2) mit auf $[0, 1]$ skaliertes y -Achse. Daher gilt auch hier mit a_1^*, \dots, a_k^* der Lösung des Minimierungsproblems (6.1), dass $L_{a_1^*, \dots, a_k^*}$ die größte konvexe Minorante von $L_{\tilde{a}_1, \dots, \tilde{a}_k}$ ist, vgl. Gleichung (6.3). Das Zusammenfassen von Ratingkategorien und Mitteln der zugehörigen Ausfallanteile kann jedoch nur eine Lorenzkurve liefern, die durch eine Teilmenge der von $L_{\tilde{a}_1, \dots, \tilde{a}_k}$ vorgegebenen Punkte verläuft. Damit ist $L_{a_1^*, \dots, a_k^*}$ offensichtlich optimal in dieser Menge von Lorenzkurven. \square

Beispiel 6.8. *Angenommen von $n = 80$ Krediten besitzen 30 das beste Rating R_1 , 20 das mittlere Rating R_2 und nochmal 30 das schlechte Rating R_3 . Dazugehörend seien die (nicht isotonen) empirischen Ausfallanteile $\tilde{a}_1 = 0.2$, $\tilde{a}_2 = 0$ und $\tilde{a}_3 = 0.8$ beobachtet. Abbildung 6.1 (a) zeigt die Lorenzkurve der empirischen Ausfallanteile $L_{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3}$. Diese ist nicht konvex, da die Ausfallanteile nicht isoton sind, vgl. Definition 3.1. Das Zusammenfassen von Ratingkategorien kann zu Isotonie in den empirischen Ausfallanteilen (der zusammengefassten Kategorien) führen. Hier existieren vier Möglichkeiten, die drei Ratingkategorien zusammenzufassen und zwar*

$\{\{R_1\}, \{R_2\}, \{R_3\}\}$, $\{\{R_1, R_2\}, \{R_3\}\}$, $\{\{R_1\}, \{R_2, R_3\}\}$ und $\{\{R_1, R_2, R_3\}\}$. Diese Zusammenlegungen liefern jeweils verschiedene (gepoolte) Ausfallanteile. So ergeben sich die vier in Abbildung 6.1 (b) dargestellten Lorenzkurven. Abbildung 6.1 (c) zeigt die konvexen Lorenzkurven aus Abbildung 6.1 (b). Die größte konvexe Minorante von $L_{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3}$ ist optimal (am stärksten durchgebogen) in der Klasse aller konvexen Lorenzkurven, die sich aus dem Zusammenfassen der Ratingkategorien ergeben. Diese entspricht hier der empirischen Lorenzkurve nach dem Zusammenfassen der Ratingkategorien zu $\{\{R_1, R_2\}, \{R_3\}\}$, d. h. die Lorenz-optimalen geschätzten Ausfallwahrscheinlichkeiten lauten $a_1^* = a_2^* = \frac{0.2 \cdot 30 + 0.20}{50} = 0.12$ und $a_3^* = 0.8$. Dies sind auch die Schätzungen nach der monotonen Regression.

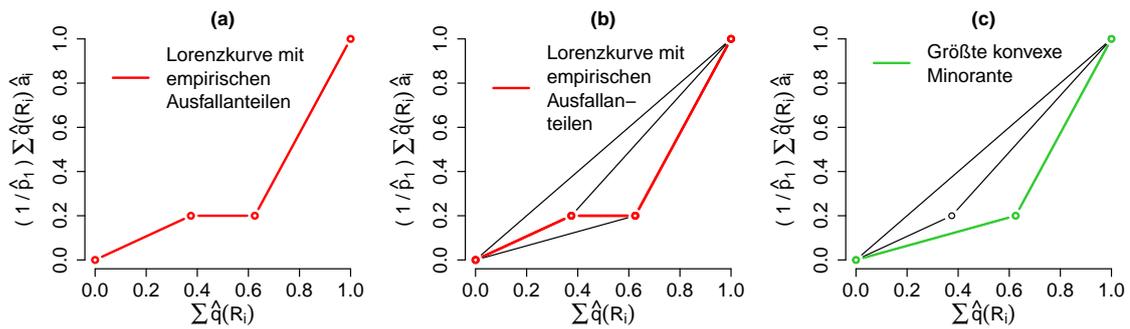


Abbildung 6.1: (a) Lorenzkurve der empirischen Ausfallanteile. (b) Lorenzkurven der zusammengefassten Ratingkategorien. (c) Einschränkung auf konvexe Lorenzkurven. In dieser Menge ist die größte konvexe Minorante von $L_{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3}$ optimal.

Aus Satz 6.7 und Satz 4.6 folgt auch, dass die mittels monotoner Regression geschätzten Ausfallwahrscheinlichkeiten optimal im Sinne aller streng anreizerhaltenden Score-Regeln sind (unter allen verträglichen Schätzungen). Dies stellen auch Brümmer & Du Preez (2013) heraus. Ihre Argumentation ist allerdings wesentlich technischer und verwendet keine Halbordnungen.

Fawcett & Niculescu-Mizil (2007) bringen die monotone Regression mit der ROC-Kurve in Zusammenhang. Sie zeigen, dass die geschätzten Ausfallwahrscheinlichkeiten aus der monotonen Regression ROC-Dominanz über alle anderen verträglichen Schätzungen besitzen.

Ebenso können auch Brücken zur Refinement-Ordnung (vgl. Kapitel 4) und zum Mean-Preserving-Spread (Rothschild & Stiglitz, 1970) geschlagen werden. Auch im Sinne dieser Kriterien sind die aus der monotonen Regression resultierenden geschätzten Ausfallwahrscheinlichkeiten optimal in der Menge aller verträglichen Schätzungen.

Kapitel 7

Anwendung im Bereich der Ratingindustrie

Standard & Poor's (S&P), Moody's und Fitch sind die größten drei internationalen Ratingagenturen. Sie gehören zu den zehn Ratingagenturen, die als Nationally Recognized Statistical Rating Organization (NRSRO) eingetragen sind (U.S. Securities and Exchange Commission, 2017, S. 7). Diese förmliche Anerkennung wird von der United States Securities and Exchange Commission (SEC), einer US-Börsenaufsichtsbehörde vergeben und soll die Qualität der Ratings garantieren. Dazu verlangt die SEC von den eingetragenen Agenturen unter anderem eine regelmäßige Offenlegung ihrer Ratinghistorie. Aus dieser Historie werden hier Ratings für *durch private Wohnimmobilien hypothekenbesicherte Wertpapiere* (residential mortgage-backed securities (RMBS)) mit einer Laufzeit von zehn Jahren betrachtet (Fitch Ratings, 2017, S. 7; Moody's, 2017, S. 8; Standard & Poor's, 2017, S. 14).

Diese Informationen werden in diesem Kapitel für einen exemplarischen Vergleich von S&P, Moody's und Fitch genutzt. Dabei hat die Anwendung vorrangig das Ziel, die vorgestellten Methoden zu veranschaulichen. Auf Signifikanztests, z. B. für die Gleichheit von Lorenzkurven, wird daher verzichtet, siehe etwa Dardanoni & Forcina (1999) oder Blöchlinger & Leippold (2018).

Tabelle 7.1 zeigt die Verteilung der Ratings für RMBS getrennt für S&P, Moody's und Fitch. Die Gesamtausfallraten von 41.55 % (S&P), 40.09 % (Moody's) und 34.83 % (Fitch) sind auffällig hoch und auch damit begründet, dass die Finanzkrise ab 2007 im Untersuchungszeitraum liegt, vgl. Kapitel 1.

7.1 Gewichtete monotone Regression

Tabelle 7.1 zeigt auch, dass die empirischen Ausfallanteile der RMBS nicht monoton steigend in den 21 Ratingklassen sind. Vor allem in den Ratingkategorien mit erhöhtem Ausfallrisiko (B bis C) ist die Monotonieeigenschaft deutlich verletzt. Diese Beobachtung wird hier nicht weiter untersucht. Ein Grund dafür könnte jedoch die verhältnismäßig geringe Besetzung dieser Klassen sein.

Zur Schätzung monoton steigender Ausfallwahrscheinlichkeiten wird hier die gewichtete monotone Regression genutzt. Dabei werden die relativen Häufigkeiten der Ratingklassen als Gewichte der Regression verwendet. So ergeben sich die geschätzten Ausfallwahrscheinlichkeiten (empirisch kalibriert) aus Tabelle 7.1, welche außerdem in Abbildung 7.1 dargestellt sind.

Seien die gegebenen Ratingklassen von gut nach schlecht mit den Werten $1, 2, \dots, 21$ assoziiert. Die optimalen Zusammenlegungen der Menge $\{1, 2, \dots, 21\}$ sind nach dem PAV-Algorithmus mit den Daten aus Tabelle 7.1 gegeben durch

$$\{\{1\}, \{2\}, \{3\}, \{4, 5, 6\}, \{7\}, \{8, 9\}, \{10\}, \{11, 12, 13\}, \{14\}, \{15\}, \{16, \dots, 21\}\} \quad (\text{S\&P})$$

$$\{\{1\}, \{2, 3\}, \{4\}, \{5, 6\}, \{7\}, \{8, 9\}, \{10, \dots, 21\}\} \quad (\text{Moody's})$$

$$\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9, 10\}, \{11, \dots, 14\}, \{15, \dots, 21\}\} \quad (\text{Fitch}).$$

Die Zusammenlegungen zeigen an, welche Ratingklassen durch die monotone Regression dieselbe geschätzte Ausfallwahrscheinlichkeit zugeordnet bekommen: Dies sind die Ratings, die gemeinsam in einer Menge des Mengensystems stehen, und die geschätzte Ausfallwahrscheinlichkeit entspricht dem gewichteten arithmetischen Mittel der empirischen Ausfallanteile in den entsprechenden Klassen. Beispielsweise wird den 8151 von S&P bewerteten Wertpapieren aus den Ratingklassen AA-, A+ und A die Ausfallwahrscheinlichkeit

$$a_4^{*\text{S\&P}} = a_5^{*\text{S\&P}} = a_6^{*\text{S\&P}} = \frac{1588}{8151}0.6247 + \frac{1941}{8151}0.6394 + \frac{4622}{8151}0.5887 = 0.6078$$

zugeordnet, vgl. Tabelle 7.1.

Die in Kapitel 6 herausgestellte Lorenz-Optimalität der monotonen Regression wird anhand der Schätzung für S&P im Beispiel D.1 im Anhang verdeutlicht.

Tabelle 7.1: Verteilung der Ratings von S&P, Moody's und Fitch im Bereich RMBS bei einer Laufzeit von 10 Jahren. Empirische Ausfallanteile in den Ratingklassen \tilde{a}_j und die mittels der monotonen Regression geschätzten Ausfallwahrscheinlichkeiten a_j^* .

j	Standard & Poor's				Moody's				Fitch					
	Rating	Anzahl der Ratings am 31.12.06	Ausfallanteile im Zeitraum 31.12.06 bis 31.12.16 in %	\tilde{a}_j	a_j^*	Rating	Anzahl der Ratings am 31.12.06	Ausfallanteil im Zeitraum 31.12.06 bis 31.12.16 in %	\tilde{a}_j	a_j^*	Rating	Anzahl der Ratings am 31.12.06	Ausfallanteil im Zeitraum 31.12.06 bis 31.12.16 in %	\tilde{a}_j
1	AAA	39073	21.45	21.45	Aaa	27008	22.00	22.00	AAA	23407	AAA	23407	21.60	21.60
2	AA+	2605	42.65	42.65	Aa1	2531	52.00	43.76	AA+	1192	AA+	1192	27.80	27.80
3	AA	5410	43.99	43.99	Aa2	3618	38.00	43.76	AA	2739	AA	2739	34.00	34.00
4	AA-	1588	62.47	60.78	Aa3	1816	50.00	50.00	AA-	779	AA-	779	39.70	39.70
5	A+	1941	63.94	60.78	A1	1473	58.00	53.75	A+	970	A+	970	41.60	41.60
6	A	4622	58.87	60.78	A2	3566	52.00	53.75	A	2556	A	2556	47.10	47.10
7	A-	2235	68.10	68.10	A3	2035	64.00	64.00	A-	1073	A-	1073	49.30	49.30
8	BBB+	2273	74.66	70.34	Baa1	2023	68.00	68.00	BBB+	1218	BBB+	1218	57.10	57.10
9	BBB	4554	68.18	70.34	Baa2	3359	68.00	68.00	BBB	2550	BBB	2550	60.60	60.23
10	BBB-	2522	79.78	79.78	Baa3	2164	78.00	76.33	BBB-	1119	BBB-	1119	59.40	60.23
11	BB+	967	84.80	82.47	Ba1	882	84.00	76.33	BB+	558	BB+	558	74.60	74.43
12	BB	2669	81.87	82.47	Ba2	1088	79.00	76.33	BB	1379	BB	1379	75.20	74.43
13	BB-	174	78.74	82.47	Ba3	186	58.00	76.33	BB-	93	BB-	93	64.50	74.43
14	B+	85	83.53	83.53	B1	85	49.00	76.33	B+	96	B+	96	71.90	74.43
15	B	1881	91.12	91.12	B2	304	77.00	76.33	B	909	B	909	87.70	75.68
16	B-	53	92.45	91.94	B3	70	50.00	76.33	B-	45	B-	45	53.30	75.68
17	CCC+	0	0	91.94	Caa1	49	37.00	76.33	CCC+	0	CCC+	0	0	75.68
18	CCC	87	93.10	91.94	Caa2	33	48.00	76.33	CCC	149	CCC	149	38.30	75.68
19	CCC-	6	83.33	91.94	Caa3	19	63.00	76.33	CCC-	3	CCC-	3	0	75.68
20	CC	3	66.67	91.94	Ca	74	41.00	76.33	CC	97	CC	97	55.70	75.68
21	C	0	0	91.94	C	26	69.00	76.33	C	231	C	231	66.20	75.68
	Gesamt	72748	41.55	41.55		52409	40.09	40.09		41163		41163	34.83	34.83

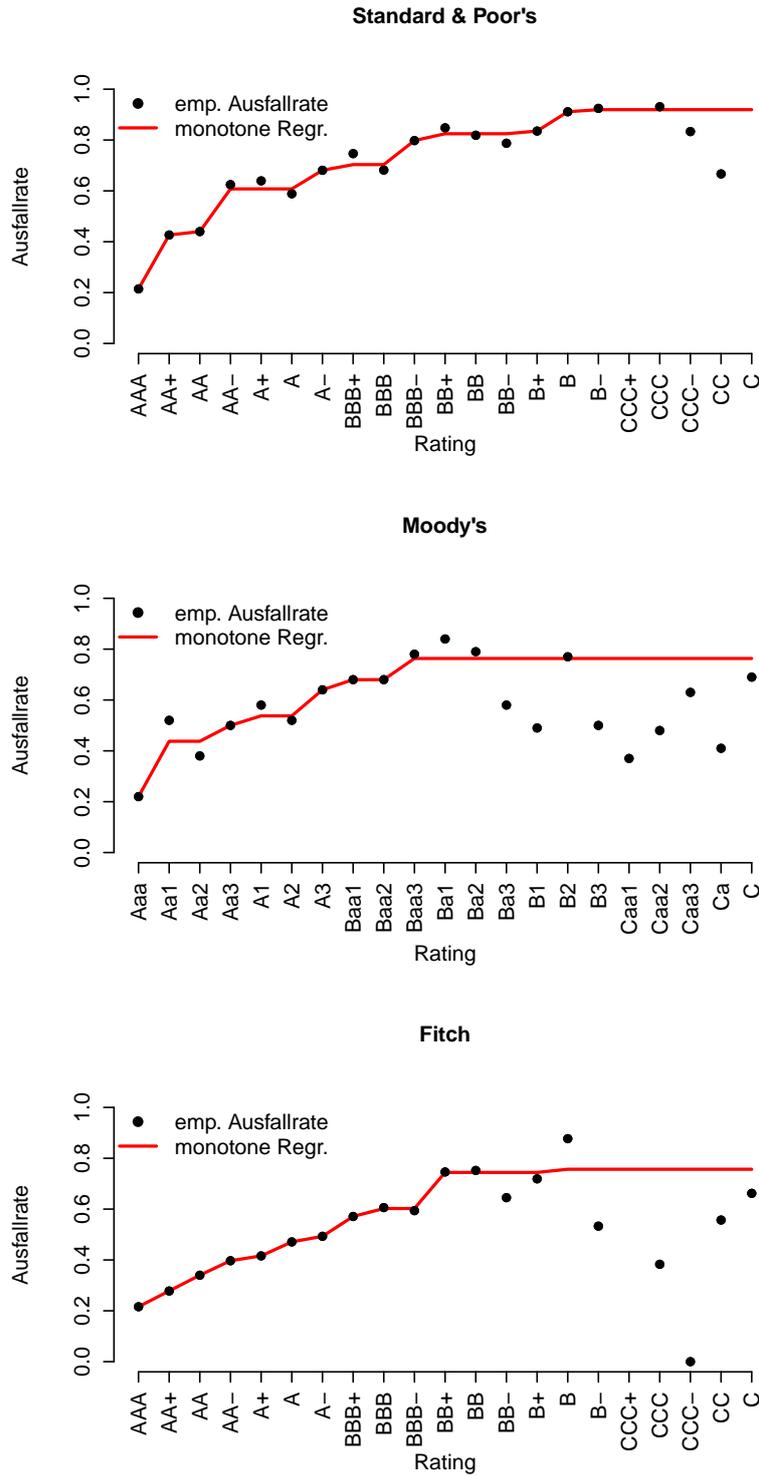


Abbildung 7.1: Empirische Ausfallanteile und monotone Regression im Bereich RMBS (residential mortgage-backed securities) für eine Laufzeit von 10 Jahren. *Bemerkung: Die gefitteten Werte der monotonen Regression sind lediglich zur besseren Visualisierung linear miteinander verbunden, zwischen den Ratings sind die geschätzten Ausfallwahrscheinlichkeiten nicht definiert.*

7.2 Vardeman-Meeden-Ordnung

In diesem Unterkapitel werden die mittels der monotonen Regression geschätzten Ausfallwahrscheinlichkeiten von Fitch, Moody's und S&P aus Tabelle 7.1 mithilfe der Vardeman-Meeden-Ordnung verglichen.

Wie Tabelle 7.1 zu entnehmen, lassen sich die Gesamtausfallanteile im Bereich RMBS (Laufzeit 10 Jahre) aufsteigend ordnen: Fitch (34.83 %), Moody's (40.09 %) und S&P (41.55 %). Außerdem hat keine der Agenturen ein Rating abgegeben, welches mit der geschätzten Ausfallwahrscheinlichkeit 0 oder 1 verknüpft ist. Damit sind die Voraussetzungen von Satz 5.4 und Korollar 5.5 gegeben und es können die meisten Beziehungen nach der VM-Ausfall- und Nichtausfallordnung ohne weitere Berechnungen direkt ausgeschlossen werden. Tabelle 7.2 und 7.3 veranschaulichen dies und zeigen, dass bezüglich der VM-Ausfallordnung keinerlei Ordnung zwischen S&P, Moody's und Fitch besteht und dass bezüglich der VM-Nichtausfallordnung lediglich Moody's durch Fitch dominiert wird. Dies resultiert aus der Betrachtung der zugehörigen bedingten Prognoseverteilungen (Abbildung A.3 im Anhang).

Tabelle 7.2: Dominanzbeziehungen nach der VM-Ausfallordnung für den Bereich RMBS (Laufzeit 10 Jahre). Es besteht nach der VM-Ausfallordnung keinerlei Ordnung zwischen S&P, Moody's und Fitch. ($\times \triangleq$ Keine Dominanz, $\checkmark \triangleq$ Dominanz)

	wird dominiert			
dominiert		S&P	Moody's	Fitch
S&P	—	\times (Korollar 5.5)	\times (Korollar 5.5)	
Moody's	\times (Satz 5.4)	—		\times
Fitch	\times (Satz 5.4)	\times (Satz 5.4)		—

Tabelle 7.3: Dominanzbeziehungen nach der VM-Nichtausfallordnung für den Bereich RMBS (Laufzeit 10 Jahre). Fitch dominiert Moody's, ansonsten besteht keine Ordnung zwischen S&P, Moody's und Fitch. ($\times \triangleq$ Keine Dominanz, $\checkmark \triangleq$ Dominanz)

wird dominiert	S&P	Moody's	Fitch
dominiert			
S&P	——	\times (Satz 5.4)	\times (Satz 5.4)
Moody's	\times (Korollar 5.5)	——	\times (Satz 5.4)
Fitch	\times (Korollar 5.5)	\checkmark	——

Abbildung 7.2 zeigt die integrierten empirischen Verteilungsfunktionen, der auf den Ausfall bedingten Prognoseverteilungen und einen Ausschnitt dieser. Diese sind für $x \in [0, 1]$ durch

$$F_{(2)}(x) := \int_{-\infty}^x \sum_{a_j^* \leq t} \hat{q}(a_j^*|1) dt \tag{7.1}$$

definiert, wobei $\hat{q}(a_j^*|1)$ die empirische relative Häufigkeit von a_j^* in der auf den Ausfall bedingten Prognoseverteilung ist ($j = 1, \dots, 21$), vgl. Tabelle 7.1. In Abbildung 7.2 ist $F_{(2)}^M(x) \leq F_{(2)}^F(x)$ für alle $x \in [0, 1]$, wobei $F_{(2)}^M, F_{(2)}^F$ für Moody's und Fitch nach Gleichung (7.1) definiert sind. Moody's dominiert daher Fitch bezüglich der Ausfälle im Sinne der stochastischen Dominanz zweiter Ordnung.

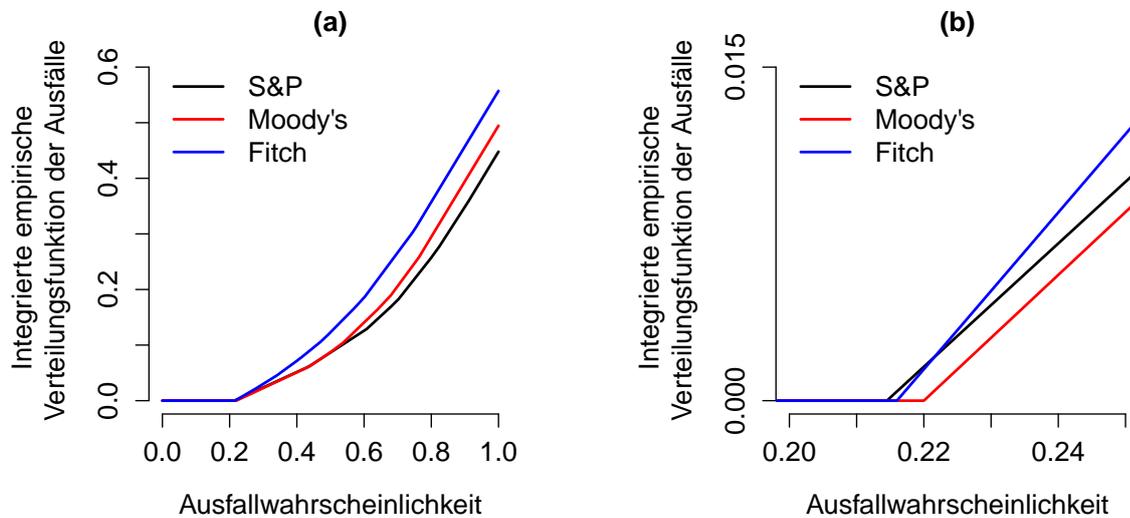


Abbildung 7.2: (a) Integrierte empirische Verteilungsfunktionen, der auf den Ausfall bedingten Prognoseverteilungen. (b) Der Ausschnitt zeigt, dass die integrierte Verteilungsfunktion von Moody's auf dem Intervall $[0.20, 0.25]$ kleiner oder gleich denjenigen von S&P und Fitch ist.

7.3 Verallgemeinerte Lorenz-Ordnung

Die Refinement-Ordnung (Definition 4.4) ist für die Ratinghistorie im Bereich RMBS (Tabelle 7.1) nicht anwendbar, da die Agenturen unterschiedliche Populationen bewertet haben. Die verallgemeinerte Lorenz-Ordnung (Definition 5.7) stellt eine Erweiterung der Refinement-Ordnung für unterschiedliche Schuldnerpopulationen dar. Daher werden folgend die mittels der monotonen Regression geschätzten Ausfallwahrscheinlichkeiten aus Tabelle 7.1 mithilfe der verallgemeinerten Lorenz-Ordnung verglichen. Die zugehörigen empirischen Lorenzkurven zeigt Abbildung 7.3.

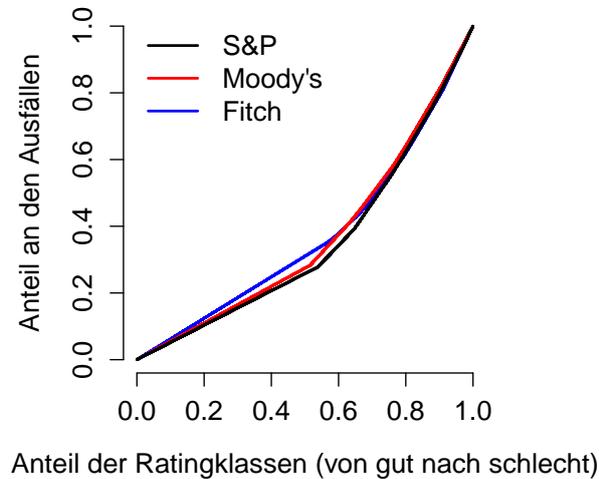


Abbildung 7.3: Empirische Lorenzkurven von S&P, Moody's und Fitch. Die zugehörigen Gesamtausfallraten sind $\hat{p}_1^{S\&P} = 41.55\%$, $\hat{p}_1^M = 40.09\%$ und $\hat{p}_1^F = 34.83\%$ (gerundet auf zwei Nachkommastellen).

Die nur schwach durchgebogenen empirischen Lorenzkurven in Abbildung 7.3 zeigen die geringe Trennschärfe in den gemachten Ratings. Dies ist nicht verwunderlich, da die Sparte der RMBS zur Zeit der Finanzkrise 2007 ungenügend präzise bewertet war (Financial Crisis Inquiry Commission, 2011, S. xxv). Dennoch ist ein systematischer Unterschied zwischen den Agenturen vorhanden: Die Lorenzkurve von S&P liegt auf dem Intervall $(0, 1)$ unterhalb der Lorenzkurve von Moody's. Da außerdem der Gesamtausfallanteil in der von S&P bewerteten Schuldnerpopulation mit $\hat{p}_1^{S\&P} = 41.55\%$ größer ist als der von Moody's mit $\hat{p}_1^M = 40.09\%$, wird Moody's von S&P im Sinne der (empirischen) verallgemeinerten Lorenz-Ordnung dominiert. Wegen Satz 5.11 ist S&P hier daher auch besser als Moody's im Sinne aller Skill-Scores, die auf einer anreizerhaltenden Score-Regel mit konvexer Score-Funktion (für $\theta = 0$) basieren.

Tabelle 7.4 veranschaulicht dies und zeigt Werte ausgewählter Score-Regeln, die die Voraussetzung von Satz 5.11 erfüllen. Insbesondere hat S&P hier bessere Skill-Score-Werte als Moody's, was auch direkt wegen der Dominanz nach der (empirischen) verallgemeinerten Lorenz-Ordnung folgt.

Tabelle 7.4: Werte ausgewählter Score-Regeln und Skill-Scores für den Bereich RMBS (Laufzeit 10 Jahre).

	Score-Regel			Skill-Score		
	Brier	Logarith- mus	Hyper- bolisch	Brier	Logarith- mus	Hyper- bolisch
S&P	0.1861	0.5566	0.2058	0.2337	0.1800	0.2215
Moody's	0.1983	0.5841	0.2186	0.1743	0.1326	0.1646
Fitch	0.1941	0.5750	0.2143	0.1448	0.1104	0.1373

Kapitel 8

Fazit und Zusammenfassung

Wahrscheinlichkeitsprognosen finden Anwendung in verschiedensten Bereichen, wie der Meteorologie, der Medizin sowie der Versicherungs- und Bankenbranche. In dieser Arbeit liegt der Fokus auf der Bewertung von Wahrscheinlichkeitsprognosen in der Ratingindustrie. Dazu wurden verschiedene Konzepte zum Vergleich von Kreditausfallprognosen vorgestellt, untersucht und erweitert.

Für die Klasse der anreizerhaltenden Score-Regeln wurden Resultate für kalibrierte Wahrscheinlichkeitsprognosen formuliert. Beispielsweise entspricht der Brier-Skill-Score bei empirischer Kalibrierung gerade dem quadrierten Korrelationskoeffizienten von Bravais-Pearson. Die Savage-Darstellung ermöglicht auch die Konstruktion von anreizerhaltenden Score-Regeln. Auf diese Weise wurde der (anreizerhaltende) hyperbolische Score definiert. Außerdem wurde nachgewiesen, dass Skill-Scores anreizerhaltender Score-Regeln, wie der Brier- oder Logarithmus-Skill-Score, im Allgemeinen nicht anreizerhaltend sind.

Unterschiedliche Score-Regeln können sich hinsichtlich der gemessenen Vorhersagequalität widersprechen. Hier helfen Halbordnungen, die bezüglich einer gesamten Klasse von Score-Regeln Dominanz garantieren können. Halbordnungen zum Vergleich von Wahrscheinlichkeitsprognosen bei identischen Schuldnerkollektiven, wie die Refinement-Ordnung, die Vardeman-Meeden-Ordnung und die Lorenz-Ordnung, sind in der Literatur gut erforscht. Darauf aufbauend wurde insbesondere die Refinement-Ordnung bei Vorliegen von vier Ratingkategorien untersucht. Ein hierfür formulierter Dominanzsatz ermöglicht auch die Visualisierung der Dominanzbeziehung von zwei Prognostikern.

Wird die Restriktion auf identische Schuldnerkollektive aufgegeben, so sind die genannten Ordnungen entweder nicht mehr anwendbar oder reduzieren sich zu Präordnungen. Daher wurde unter der Annahme kalibrierter Prognosen bei unterschiedlichen Schuldnerkollektiven eine neue Halbordnung, die verallgemeinerte Lorenz-Ordnung definiert und untersucht. Ein zentrales Ergebnis für kalibrierte Prognosen ist: Besteht Dominanz nach der verallgemeinerten Lorenz-Ordnung, so ist der dominierende Prognostiker auch im Sinne wichtiger Skill-Scores, wie dem Brier-Skill-Score besser als sein Konkurrent.

Die monotone Regression ist ein nichtparametrisches Verfahren, das zur Schätzung von Ausfallwahrscheinlichkeiten zu gegebenen Ratingklassen genutzt werden kann. Hier wurde gezeigt, dass dieses Verfahren Lorenz-optimal in einer Klasse von verträglichen Schätzungen ist.

Eine Anwendung auf die internationale Ratingindustrie veranschaulicht die in dieser Arbeit herausgearbeiteten Ergebnisse. Der Vergleich umfasst die

Ratings der führenden Ratingagenturen Standard & Poor's, Moody's und Fitch. Die verwendeten Daten wurden von den Agenturen im Rahmen der Anerkennung als NRSRO (nationally recognized statistical rating organization) veröffentlicht. Da die Agenturen unterschiedliche Schuldnerkollektive in der betrachteten Sparte bewertet haben (residential mortgage backed securities), wurde der Vergleich insbesondere mittels der verallgemeinerten Lorenz-Ordnung durchgeführt. Hierbei dominiert Standard & Poor's den Konkurrenten Moody's im Sinne der verallgemeinerten Lorenz-Ordnung.

Ausblick: Die Relevanz, Ratingagenturen zu überwachen und zu bewerten, motiviert eine Vielzahl von weiteren zukünftigen Forschungsfragen. Neben aussagekräftigen empirischen Untersuchungen sind viele methodische Fragen offen. So sollten beispielsweise anreizerhaltende Qualitätskriterien für den Vergleich von Wahrscheinlichkeitsprognosen bei unterschiedlichen Schuldnerpopulationen entwickelt werden. Weiter ist es interessant, die Restriktion auf kalibrierte Wahrscheinlichkeitsprognosen fallen zu lassen. Einen Ansatz dazu bieten Krämer & Posch (2015), die Halbordnungen für semikalibrierte Prognosen betrachten.

Anhang A

Grafiken

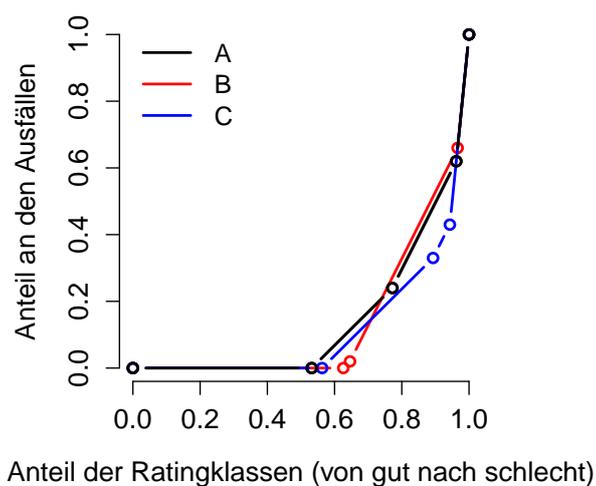


Abbildung A.1: Lorenzkurven von drei kalibrierten Wahrscheinlichkeitsprognostikern auf gleichen Schuldnerpopulationen, vgl. Tabelle 4.1. Hier ist $C \geq_L A$, sodass mit Satz 4.6 der Prognostiker C besser ist als Prognostiker A bezüglich aller anreizerhaltender Score-Regeln (im Erwartungswert).

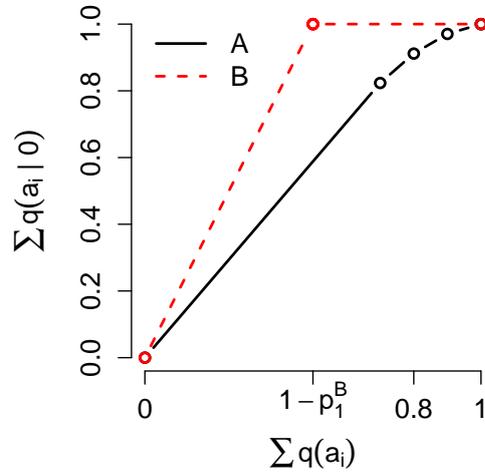


Abbildung A.2: Lorenzkurve der Nichtausfälle des perfekten Prognostikers B bei einer Gesamtausfallwahrscheinlichkeit von $p_1^B = 0.4$ im Vergleich zur Lorenzkurve der Nichtausfälle des Prognostikers A (kalibriert) mit $q^A(0) = 0.7$, $q^A(0.25) = q^A(0.5) = q^A(0.75) = 0.1$ und $p_1^A = 0.15$.

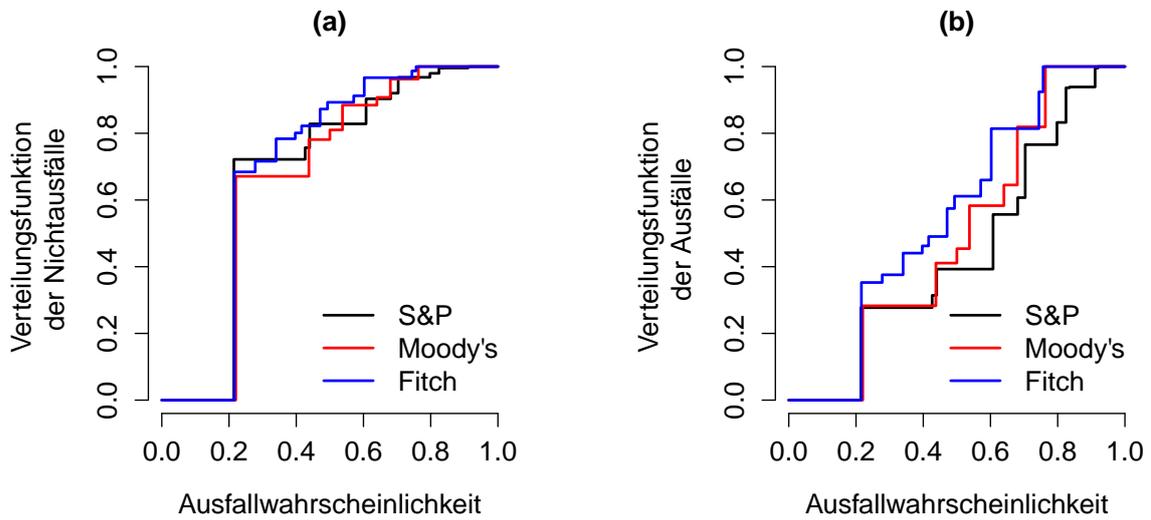


Abbildung A.3: (a) Empirische Verteilungsfunktionen, der auf den Nichtausfall bedingten Prognoseverteilungen. (b) Empirische Verteilungsfunktionen, der auf den Ausfall bedingten Prognoseverteilungen.

Anhang B

Tabelle

Tabelle B.1: Ratingskala von Moody's für langfristige Anleihen mit einer Laufzeit von mindestens einem Jahr (Emery, 2017). Zur weiteren Differenzierung innerhalb der Ratingkategorien fügt Moody's den Ratings Aa bis Caa eine Zahl aus $\{1, 2, 3\}$ hinzu, beispielsweise ist Baa1 besser als Baa2 und Baa2 ist besser als Baa3, so ergeben sich 21 mögliche Ratings.

- Aaa Anleihen bewertet mit Aaa sind von höchster Qualität und haben das geringste Kreditausfallrisiko.
- Aa Anleihen bewertet mit Aa sind von guter Qualität und haben ein sehr geringes Kreditausfallrisiko.
- A Anleihen bewertet mit A sind von mittel guter Qualität und haben ein geringes Kreditausfallrisiko.
- Baa Anleihen bewertet mit Baa sind von mittlerer Qualität und haben ein mäßiges Kreditausfallrisiko, sie sind leicht spekulativ.
- Ba Anleihen bewertet mit Ba sind spekulativ und besitzen ein substantielles Kreditausfallrisiko.
- B Anleihen bewertet mit B sind spekulativ und besitzen ein hohes Kreditausfallrisiko.
- Caa Anleihen bewertet mit Caa sind spekulativ und besitzen ein sehr hohes Kreditausfallrisiko.
- Ca Anleihen bewertet mit Ca sind sehr spekulativ. Der Zahlungsausfall ist wahrscheinlich, mit etwas Aussicht auf Erholung.
- C Anleihen bewertet mit C sind, die am schlechtesten bewerteten. Der Zahlungsausfall steht meist unmittelbar bevor, mit geringer Aussicht auf Erholung.

Tabelle B.2: Marktanteile der in der EU registrierten Ratingagenturen (European Securities and Markets Authority, 2017).

Kreditratingagentur	Marktanteil in %
S&P Global Ratings	46.26
Moody's Investors Service	31.27
Fitch Ratings	15.65
DBRS Ratings	1.87
CERVED Rating Agency	0.97
AM Best Europe Rating Services	0.90
The Economist Intelligence Unit	0.69
CreditReform Rating	0.53
Scope Ratings	0.46
GBB-Rating	0.35
Assekurata	0.23
Euler Hermes Rating	0.22
Capital Intelligence Ratings	0.13
ICAP	0.12
ModeFinance	0.08
Spread Research	0.07
Dagong Europe Credit Rating	0.07
ARC Ratings	0.05
Axesor Rating	0.03
CRIF Ratings	0.03
BCRA Credit Rating Agency	0.02
EuroRating	0.01
INC Rating	< 0.01
European Rating Agency	< 0.01
Rating-Agentur Expert RA GmbH	< 0.01
Total	100

Anhang C

Verträgliche Schätzungen

Bezeichne mit $V_{(P)}$ die Menge aller mit (P) verträglichen Schätzungen. Im Folgenden werden verträgliche Schätzungen und ihre Verteilungen näher untersucht. Dazu wird das Hilfskonstrukt der monotonen Partitionierung definiert.

Definition C.1.

(i) Eine **monotone Partitionierung** einer Menge $\{1, \dots, k\}$, $k \in \mathbb{N}$, ist ein Mengensystem $\mathcal{M}_l = \{\mathbb{M}_1, \dots, \mathbb{M}_l\}$ mit Elementen

$$\mathbb{M}_j = \{b_{j-1} + 1, b_{j-1} + 2, \dots, b_j\} = \mathbb{N} \cap [b_{j-1} + 1, b_j]$$

mit $b_j \in \mathbb{N}$ für $j = 1, \dots, l$ und $0 = b_0 < b_1 < \dots < b_{l-1} < b_l = k$.

(ii) Seien $0 \leq \hat{a}_1 \leq \dots \leq \hat{a}_k \leq 1$ und für $i = 1, \dots, k$ sei $u_i^{(1)} := \min\{j : \hat{a}_i = \hat{a}_j\}$ und $u_i^{(2)} := \max\{j : \hat{a}_i = \hat{a}_j\}$. Damit ist

$$\mathcal{M}_l^{\{\hat{a}_1, \dots, \hat{a}_k\}} := \bigcup_{i=1}^k \{\mathbb{N} \cap [u_i^{(1)}, u_i^{(2)}]\}$$

die zu $\hat{a}_1, \dots, \hat{a}_k$ **gehörende monotone Partitionierung**.

Seien $\hat{a}_1, \dots, \hat{a}_k$ mit (P) verträglich und sei $\{\hat{a}_1^{(<)}, \dots, \hat{a}_l^{(<)}\}$ die Menge, der sich voneinander unterscheidenden Wahrscheinlichkeitsprognosen in $\{\hat{a}_1, \dots, \hat{a}_k\}$ mit $\hat{a}_1^{(<)} < \dots < \hat{a}_l^{(<)}$, $l \leq k$. Dann gilt mit $\mathcal{M}_l^{\{\hat{a}_1, \dots, \hat{a}_k\}} = \{\mathbb{M}_1, \dots, \mathbb{M}_l\}$ und (b) aus Definition 6.5 sowie der Monotonie der geschätzten Prognosen, dass die relative Häufigkeit von $\hat{a}_j^{(<)}$ gegeben ist durch:

$$\hat{q}(\hat{a}_j^{(<)}) = \sum_{s \in \mathbb{M}_j} \hat{q}(R_s), \quad j = 1, \dots, l.$$

Wegen der empirischen Kalibrierung gilt weiter

$$\begin{aligned} \hat{q}(\hat{a}_j^{(<)}) \hat{a}_j^{(<)} &= \hat{r}(1, \hat{a}_j^{(<)}) = \sum_{s \in \mathbb{M}_j} \hat{r}(1, R_s) = \sum_{s \in \mathbb{M}_j} \hat{q}(R_s) \tilde{a}_s \\ \iff \hat{a}_j^{(<)} &= \frac{\sum_{s \in \mathbb{M}_j} \hat{q}(R_s) \tilde{a}_s}{\sum_{s \in \mathbb{M}_j} \hat{q}(R_s)}. \end{aligned}$$

Also sind mit (P) verträgliche Schätzungen und ihre Verteilung eindeutig über die zugehörige monotone Partitionierung festgelegt. Diese Überlegung führt direkt zu Korollar C.2, das die Menge der verträglichen Schätzungen über geeignete monotone Partitionierungen definiert.

Korollar C.2. *Sei (P) eine Population von bewerteten Krediten mit Ausfällen $\theta_i \in \{0, 1\}$ und Ratings $r_i \in \{R_1, \dots, R_k\}$, $i = 1, \dots, n$. Die geschätzten Ausfallwahrscheinlichkeiten $\hat{a}_1 \leq \dots \leq \hat{a}_k$ sind genau dann verträglich mit (P), wenn eine monotone Partitionierung \mathcal{M}_l existiert, sodass für $j = 1, \dots, l$ gilt*

$$\hat{q}(\hat{a}_j^{(<)}) = \sum_{s \in \mathbb{M}_j} \hat{q}(R_s) \quad \text{und} \quad \hat{a}_j^{(<)} = \frac{\sum_{s \in \mathbb{M}_j} \hat{q}(R_s) \tilde{a}_s}{\sum_{s \in \mathbb{M}_j} \hat{q}(R_s)},$$

wobei $\{\hat{a}_1^{(<)}, \dots, \hat{a}_l^{(<)}\}$ die Menge, der sich voneinander unterscheidenden Wahrscheinlichkeitsprognosen in $\{\hat{a}_1, \dots, \hat{a}_k\}$ mit $\hat{a}_1^{(<)} < \dots < \hat{a}_l^{(<)}$ ist.

Beweis: „ \implies “: Klar mit Vorüberlegung.

„ \impliedby “: Bedingung (a) aus Definition 6.5 ist nach Voraussetzung erfüllt. Bedingung (b) aus Definition 6.5, also die empirische Kalibrierung, ist erfüllt, da $\hat{a}_j^{(<)}$ ein empirischer Ausfallanteil in (möglicherweise) zusammengefassten Ratingklassen ist, $j = 1, \dots, k$. \square

Korollar C.2 zeigt auch, dass sich verträgliche Schätzungen als gewichtete Mittel der empirischen Ausfallanteile darstellen lassen.

Korollar C.3. *Sei (P) eine Population von bewerteten Krediten mit Ausfällen $\theta_i \in \{0, 1\}$ und Ratings $r_i \in \{R_1, \dots, R_k\}$, $i = 1, \dots, n$, dann gilt:*

$$(i) \quad |\mathbb{V}_{(P)}| \leq 2^{k-1}.$$

(ii) *Die Lösung a_1^*, \dots, a_k^* des Minimierungsproblems (6.1) der monotonen Regression ist verträglich mit (P) .*

Beweis:

(i) Sei zunächst $\tilde{a}_1 < \dots < \tilde{a}_k$, dann gehört jede monotone Partitionierung der Menge $\{1, \dots, k\}$ gemäß Korollar C.2 zu genau einer mit (P) verträglichen Schätzung. Weiter sei $\mathcal{M}_l = \{\mathbb{M}_1, \dots, \mathbb{M}_l\}$ die monotone Partitionierung der Menge $\{1, \dots, k\}$, $k \in \mathbb{N}$, definiert über

$$h_i := \begin{cases} 1, & \text{falls } \exists j \in \{1, \dots, l\} \text{ mit } i \in \mathbb{M}_j \text{ und } i+1 \in \mathbb{M}_j \\ 0, & \text{sonst} \end{cases} \quad (\text{C.1})$$

für $i = 1, \dots, k-1$. Der Vektor $\vec{h} := (h_1, \dots, h_{k-1})'$ legt die monotone Partitionierung also eindeutig fest. Offensichtlich gibt es jedoch 2^{k-1} verschiedene

Möglichkeiten den Vektor h zu bilden. Daher gibt es ebenso viele monotone Partitionierungen bzw. verschiedene verträgliche Schätzungen.

Sei nun beispielsweise $\tilde{a}_1 > \tilde{a}_2$, so liefert der Vektor $h = (0, \dots, 0)' \in \mathbb{R}^{k-1}$ gemäß Gleichung (C.1) keine verträgliche Schätzung (Monotonie der Prognosen ist verletzt), sodass insgesamt $|\mathbb{V}_{(P)}| \leq 2^{k-1}$ gilt.

(ii) Die Aussage folgt aus Schritt (2) des PAV-Algorithmus und Korollar C.2. \square

Anhang D

Alternative verträgliche Schätzung

Beispiel D.1. *Tabelle 7.4 umfasst die Werte ausgewählter Score-Regeln, jeweils für S&P, Moody's und Fitch. Beispielsweise ist der Logarithmus-Score von S&P gerade $S_{\ln}^{S\&P} = 0.5566$. Den Berechnungen dieser Score-Werte liegen die mittels der monotonen Regression geschätzten Ausfallwahrscheinlichkeiten \tilde{a}_j^* aus Tabelle 7.1 zugrunde. Würden die Ausfallwahrscheinlichkeiten stattdessen als Ausfallanteile gemäß der Zusammenlegung*

$$\{\{1\}, \{2\}, \{3, 4\}, \{5, 6\}, \{7\}, \{8, 9\}, \{10\}, \{11, 12, 13\}, \{14\}, \{15\}, \{16, \dots, 21\}\}$$

geschätzt (unterscheidet sich zur optimalen Zusammenlegung nur durch die Aufteilung der Klassen 3, 4, 5 und 6), so wären die geschätzten Ausfallwahrscheinlichkeiten

$$\begin{aligned}\tilde{a}_3^{S\&P} &= \tilde{a}_4^{S\&P} = \frac{5410}{6998}0.4399 + \frac{1588}{6998}0.6247 = 0.4818, \\ \tilde{a}_5^{S\&P} &= \tilde{a}_6^{S\&P} = \frac{1941}{6563}0.6394 + \frac{4622}{6563}0.5887 = 0.6037.\end{aligned}$$

Die geschätzten Ausfallwahrscheinlichkeiten der Klassen AAA, AA+ sowie A bis C blieben unverändert (Zusammenlegung ist dort gleich), siehe Tabelle 7.1. Mit dieser leicht veränderten Zusammenlegung ergäbe sich also auch eine verträgliche Schätzung der Ausfallwahrscheinlichkeiten. Mit der würde S&P einen Logarithmus-Score von $\tilde{S}_{\ln}^{S\&P} = 0.5577 > 0.5566$ erzielen. Dass dieser Score-Wert schlechter ist, ist nicht überraschend, da die Schätzung mittels der monotonen Regression nach Satz 6.7 Lorenz-optimal in der Menge der verträglichen Schätzungen ist. Und Lorenz-Optimalität ist hier äquivalent zur Optimalität für alle anreizerhaltenden Score-Regeln, also auch für den Logarithmus-Score, vgl. Satz 4.6.

Literaturverzeichnis

- Barlow, R. E., Bartholomew, J. M., Bremner, J. M. & Brunk, H. D. (1972). *Statistical inference under order restrictions*. New York: John Wiley & Sons.
- Bauer, H. (2001). *Wahrscheinlichkeitstheorie* (5. Aufl.). Berlin: Walter de Gruyter.
- Behr, P. & Güttler, A. (2004). *Interne und externe Ratings*. Frankfurt am Main: Bankakademie.
- Blöchlinger, A. & Leippold, M. (2018). Are ratings the worst form of credit assessment except for all the others? *Journal of Financial and Quantitative Analysis*, 53 (1), 299-334.
- Breiman, L. (2001). Random forest. *Machine Learning*, 45, 5-32.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78 (2), 1-3.
- Bröcker, J. & Smith, L. A. (2007). Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22, 382-388.
- Brümmer, N. & Du Preez, J. (2013). The PAV algorithm optimizes binary proper scoring rules. *arXiv preprint arXiv:1304.2331*. Zugriff am 03.09.2018 auf <https://arxiv.org/pdf/1304.2331.pdf>.

- Bundesministerium für Arbeit und Soziales. (2017). *Lebenslagen in Deutschland* (Bericht Nr. 5). Zugriff am 13.08.2018 auf <https://www.bmas.de/DE/Service/Medien/Publicationen/a306-5-armuts-und-reichtumsbericht.html>.
- Dardanoni, V. & Forcina, A. (1999). Inference for Lorenz curve orderings. *Econometrics Journal*, 12, 44-75.
- Dawid, A. P. (1998). *Coherent measures of discrepancy, uncertainty and dependence, with applications to bayesian predictive experimental design* (Bericht Nr. 139). Department of Statistical Science University College London.
- Dawid, A. P. & Musio, M. (2014). Theory and applications of proper scoring rules. *Metron*, 72 (2), 169-183.
- DeGroot, M. H. & Eriksson, E. A. (1985). Probability forecasting, stochastic dominance and the Lorenz curve. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith (Hrsg.), *Bayesian statistics 2* (S. 99-118). Amsterdam: Elsevier.
- DeGroot, M. H. & Fienberg, S. E. (1981). *Assessing probability assessors: Calibration and refinement* (Bericht Nr. 205). Department of Statistics Carnegie-Mellon University Pittsburgh. Zugriff am 03.09.2018 auf <http://www.dtic.mil/dtic/tr/fulltext/u2/a104174.pdf>.
- DeGroot, M. H. & Fienberg, S. E. (1983). The comparison and evaluation of forecasters. *The Statistician*, 32, 12-22.
- Emery, K. (2017). *Rating symbols and definitions*. Moody's Investor Service. Zugriff am 06.11.2017 auf <https://www.moodys.com/Pages/amr002002.aspx>.
- Engelmann, B., Hayden, E. & Tasche, D. (2003). Testing rating accuracy. *Risk*, 16 (1), 82-86.

- European Securities and Markets Authority. (2017). *Report on CRA market share calculation*. Zugriff am 13.08.2018 auf https://www.esma.europa.eu/sites/default/files/library/esma33-5-209_-_cra_market_share_calculation_2017.pdf.
- Fawcett, T. & Niculescu-Mizil, A. (2007). PAV and the ROC convex hull. *Machine Learning*, 68, 97-106.
- Financial Crisis Inquiry Commission. (2011). *The financial crisis inquiry report, authorized edition: Final report of the National Commission on the causes of the financial and economic crisis in the United States*. Washington D.C.: Public Affairs.
- Fitch Ratings. (2017). *Form NRSRO exhibit 1: Credit ratings performance statistics*. Zugriff am 13.08.2018 auf https://www.sec.gov/Archives/edgar/data/1652282/000114420417017197/v462764_nrsro-ex99e1.pdf.
- Fraley, C. & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97 (458), 611-631.
- Friedman, T. L. (1995). Foreign affairs; don't mess with Moody's. *The New York Times*. Zugriff am 03.09.2018 auf <https://www.nytimes.com/1995/02/22/opinion/foreign-affairs-don-t-mess-with-moody-s.html>.
- Gnanadesikan, R. (1997). *Methods for statistical data analysis of multivariate observations* (2. Aufl.). New York: John Wiley & Sons.
- Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society*, 69, 243-268.

- Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102 (477), 359-378.
- Gneiting, T. & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7, 1747-1782.
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society*, 14 (1), 107-114.
- Härdle, W., Moro, R. & Schäfer, D. (2007). *Estimating probabilities of default with support vector machines* (Bericht Nr. 35). SFB 649 Discussion Paper. Zugriff am 03.09.2018 auf <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2007-035.pdf>.
- Hardy, G. H., Littlewood, J. E. & Pólya, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger of Mathematics*, 58, 145-152.
- Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied logistic regression* (3. Aufl.). Hoboken: John Wiley & Sons.
- Jolliffe, I. T. & Stephenson, D. B. (2008). Proper scores for probability forecasts can never be equitable. *Monthly Weather Review*, 136 (4), 1505-1510.
- Krämer, W. (2005). On the ordering of probability forecasts. *Sankhyā: The Indian Journal of Statistics*, 67 (4), 662-669.
- Krämer, W. (2006). Evaluating probability forecasts in terms of refinement and strictly proper scoring rules. *Journal of Forecasting*, 25, 223-226.
- Krämer, W. (2017). On assessing the relative performance of default predictions. *Journal of Forecasting*, 36, 854-858.

- Krämer, W. & Güttler, A. (2008). On comparing the accuracy of default predictions in the rating industry. *Empirical Economics*, 34, 343-356.
- Krämer, W. & Neumärker, S. (2016). Comparing the accuracy of default predictions in the rating industry for different sets of obligors. *Economics Letters*, 145, 48-51.
- Krämer, W. & Posch, P. N. (2015). *Partial orderings of default predictions* (Bericht Nr. 31). SFB 823 Discussion Paper. Zugriff am 03.09.2018 auf [https://eldorado.tu-dortmund.de/bitstream/2003/34200/1/DP_3115_SFB823_Kr%
c3%a4mer_Posch.pdf](https://eldorado.tu-dortmund.de/bitstream/2003/34200/1/DP_3115_SFB823_Kr%c3%a4mer_Posch.pdf).
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Journal of the American Statistical Association*, 9 (70), 209-219.
- Mason, S. J. (2008). Understanding forecast verification statistics. *Meteorological Applications*, 15 (1), 31-40.
- McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42 (9), 654-655.
- McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making*, 9, 190-195.
- Moody's. (2017). *Form NRSRO exhibit 1: Credit ratings performance measurement statistics*. Zugriff am 05.04.2017 auf [https://www.moody's.com/Pages/
reg001002.aspx](https://www.moody's.com/Pages/reg001002.aspx).
- Murphy, A. H. (1973). Hedging and skill scores for probability forecasts. *Applied Meteorology*, 12, 215-223.

- Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116, 2417-2424.
- Parry, M., Dawid, A. P. & Lauritzen, S. (2012). Proper local scoring rules. *The Annals of Statistics*, 40 (1), 561-592.
- Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: John Wiley & Sons.
- Rothschild, M. & Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, 2, 225-243.
- Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66 (336), 783-801.
- Silvapulle, M. J. & Sen, P. K. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. Hoboken: John Wiley & Sons.
- Standard & Poor's. (2017). *S&P global ratings performance for exhibit 1 form NRS-RO*. Zugriff am 05.04.2017 auf https://www.standardandpoors.com/en_US/web/guest/regulatory/form-nrsro.
- Stylianou, M. & Flournoy, N. (2002). Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*, 58, 171-177.
- U.S. Securities and Exchange Commission. (2017). *2017 summary report of commission staff's examinations of each nationally recognized statistical rating organization*. Washington D.C. Zugriff am 13.08.2018 auf <https://www.sec.gov/ocr/reports-and-publications/special-studies/nrsro-summary-report-2017.pdf>.

- Vardeman, S. & Meeden, G. (1983). Calibration, sufficiency, and domination considerations for bayesian probability assessors. *Journal of the American Statistical Association*, 78 (384), 808-816.
- Vazza, D. & Kraemer, N. W. (2017). *Default, transition, and recovery: 2016 annual global corporate default study and rating transitions* (Bericht). Standard & Poor's Financial Services LLC. Zugriff am 13.08.2018 auf <https://www.spratings.com/documents/20184/774196/2016+Annual+Global+Corporate+Default+Study+And+Rating+Transitions.pdf/2ddcf9dd-3b82-4151-9dab-8e3fc70a7035>.
- Wilks, D. (2006). *Statistical methods in the atmospheric sciences* (2. Aufl.). Burlington: Elsevier.
- Zadrozny, B. & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. *Proceedings of the eighth International Conference on Knowledge Discovery and Data Mining*, 694-699. Zugriff am 03.09.2018 auf https://www.researchgate.net/profile/Charles_Elkan/publication/2571315_Transforming_Classifier_Scores_into_Accurate_Multiclass_Probability_Estimates/links/0fcfd509ae852a8bb9000000.pdf.

Eidesstattliche Versicherung

Neumärker, Simon

Name, Vorname

133800

Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

Qualitätsvergleiche kalibrierter Wahrscheinlichkeitsprognosen mit Anwendung auf die internationale Ratingindustrie

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Dortmund, den 4. September 2018

Ort, Datum

Unterschrift

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden (§63 Abs. 5 Hochschulgesetz - HG -). Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft. Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z. B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Dortmund, den 4. September 2018

Ort, Datum

Unterschrift