

Kreditscoring unter Stichprobenselektion

Diplomarbeit

zur Erlangung des Grades eines Diplom-Kaufmanns

an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

vorgelegt von

Thomas Parnitzke

(Matrikel-Nr. 175089)

Prüfer: PD Dr. Marlene Müller

Berlin, 30. Juli 2003

Inhaltsverzeichnis

1. Einleitung	8
2. Problemstellung	10
2.1. Kreditscoringverfahren	10
2.2. Stichprobenselektion	11
2.3. Theoretische Darstellung	13
2.4. Leistungsmessung und Vergleich	14
2.4.1. Kontingenztafel	14
2.4.2. Trennschärfe	15
2.4.3. Lorenzkurve	16
3. Lösungsansätze	20
3.1. Veränderung des Kreditvergabeverhaltens	20
3.2. Re-Klassifizierung	21
3.3. Neugewichtung	21
3.4. Extrapolation	23
3.5. Heckman-Ansatz	24
3.6. Bivariater Probit-Ansatz	26
3.7. Bänder	27
3.7.1. Identifikation von Stichprobenanteilen	27
3.7.2. Bänder für Separationsmaße	29
3.8. Literaturüberblick	32
4. Praktische Anwendung	35
4.1. Simulierte Daten	35
4.1.1. Erstes Scoring	36
4.1.2. Methoden zur Minderung des Selektionseffektes	38
4.2. Datensatz aus Fahrmeir und Hamerle (1984)	44

4.2.1. Erstes Scoring	45
4.2.2. Methoden zur Minderung des Selektionseffektes	47
4.3. Der Selektionseffekt und seine Einflussgrößen	53
4.3.1. Simulierte Daten - Klassifikationsergebnis	54
4.3.2. Simulierte Daten - Trennschärfe	60
5. Zusammenfassung	65
A. Anhang Logit- und Probit-Modelle	67
B. Anhang Variablenbeschreibung	71

Tabellenverzeichnis

2.1. Beispiel Kontingenztabelle	14
3.1. Neugewichtung	22
3.2. Neugewichtung, Variablendefinition	22
4.1. Ergebnis simulierte Daten (erste Scorefunktion)	37
4.2. Ergebnis simulierte Daten (volle Stichprobe)	38
4.3. Ergebnis simulierte Daten (selektierte Stichprobe)	38
4.4. Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert 1	40
4.5. Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert 2	40
4.6. Ergebnis simulierte Daten, Re-Klassifizierung 1	40
4.7. Ergebnis simulierte Daten, Re-Klassifizierung 2	41
4.8. Ergebnis simulierte Daten, Neugewichtung	42
4.9. Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 1	43
4.10. Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 2	43
4.11. Ergebnis simulierte Daten, Extrapolation	44
4.12. Variablenbezeichnung des Datensatzes aus Fahrmeir und Hamer- le (1984)	45
4.13. Ergebnis empirische Daten (erste Scorefunktion)	46
4.14. Ergebnis empirische Daten (volle Stichprobe)	47
4.15. Ergebnis empirische Daten (selektierte Stichprobe)	47
4.16. Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert 1	48
4.17. Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert 2	49
4.18. Ergebnis empirische Daten, Re-Klassifizierung	49
4.19. Ergebnis empirische Daten, Neugewichtung	49

4.20. Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 1	50
4.21. Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 2	51
4.22. Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 3	51
4.23. Ergebnis empirische Daten, Extrapolation 1	52
4.24. Ergebnis empirische Daten, Extrapolation 2	52
4.25. Ergebnis empirische Daten, Vergleich	53
4.26. Selektionseffekt und Stichprobengröße, Mittelwerte der Abweichung	55
4.27. Selektionseffekt und Veränderung der ersten Scorefunktion	59
B.1. Variablenbeschreibung x_4	71
B.2. Variablenbeschreibung x_5	72
B.3. Variablenbeschreibung x_6	72
B.4. Variablenbeschreibung x_7	72

Abbildungsverzeichnis

2.1. Verteilungen und Kreditvergabe	12
2.2. Konzept der Trennschärfe	15
2.3. Lorenzkurve	17
2.4. „Receiver-Operating“-Charakteristik	19
3.1. Extrapolation	24
4.1. Kreditstatus und Score (simulierte Daten)	37
4.2. Kreditvergabe mit weichem Trennwert	39
4.3. Re-Klassifizierung	41
4.4. Kreditstatus und Score (empirisch)	46
4.5. Selektionseffekt und Stichprobengröße	54
4.6. Selektionseffekt und Trennwert c	56
4.7. Selektionseffekt und Veränderung der Ausfallhäufigkeit	57
4.8. Selektionseffekt und variable Ausfallhäufigkeit	58
4.9. Selektionseffekt und Veränderung der ersten Scorefunktion	60
4.10. Selektionseffekt und Trennschärfe	61
4.11. Selektionseffekt und Trennschärfe (Veränderung der Stichproben- größe)	62
4.12. Selektionseffekt und Trennschärfe (Veränderung des Trennwertes c)	63
4.13. Selektionseffekt und Trennschärfe (Veränderung der Ausfallhäufig- keit)	64
A.1. Logistische Verteilung gegen Normalverteilung 1	69
A.2. Logistische Verteilung gegen Normalverteilung 2	70

Symbolverzeichnis

x	Variable
X	Vektor von Variablen (x_1, \dots, x_n)
$f(\cdot)$	Dichtefunktion
$F(\cdot)$	Verteilungsfunktion
$\phi(\cdot)$	Dichtefunktion der Standardnormalverteilung
$\Phi(\cdot)$	Verteilungsfunktion der Standardnormalverteilung
s	Schnittpunkt der Verteilungen
U	Überschneidungsbereich
T	Trennschärfe
μ	Erwartungswert
σ^2	Varianz
cov	Kovarianz
ρ	Korrelation
ε, v	Störterme
$\lambda(\cdot)$	inverse Mills-Rate
N	Stichprobengröße
Y	Kreditstatus ($Y = y$ mit $y = 0, 1$)
A	Status der Kreditvergabe ($A = a$ mit $a = 0, 1$)
S	Scorewert
c	Trennwert
π	Rückzahlungswahrscheinlichkeit
$1 - \pi$	Ausfallwahrscheinlichkeit
η	Scorefunktion
β	Modellparameter
$p(S)$	aus Scorewerten bestimmte Rückzahlungswahrscheinlichkeit
t	Zeitperiode
e	Eulersche Zahl (2, 718...)
pi	Ludolfsche Zahl (3, 141...)

1. Einleitung

Bei der Vergabe von Krediten kommt der Nutzung von Kredit-scoringverfahren eine große Bedeutung zu. Der Begriff Kredit-scoring beschreibt dabei statistische Methoden, die für die Klassifikation von Kreditbewerbern in gute und schlechte Risikoklassen verwendet werden (Hand und Henley, 1997). Wurden früher Kreditentscheidungen allein auf den Erfahrungswerten von Bankmitarbeitern gefällt, ist die Kreditvergabe auf Basis determinierter Scoringregeln heute weit verbreitet. Erstmals wurde ein solches Kredit-scoringssystem von Durand (1941) mit der Verwendung einer Diskriminanzanalyse zur Unterscheidung von guten und schlechten Krediten vorgeschlagen (siehe Thomas, 2000). Ein entscheidender Grund für den sich im Folgenden verbreitenden Einsatz von Scoringverfahren war die durch Scoringssysteme verbesserte Unterscheidung von guten und schlechten Risiken. Als weitere begünstigende Faktoren sind die sich schnell entwickelnde Computertechnologie und die sich ergebende Kosten- und Zeitersparnis zu nennen.

Ein natürliches Interesse der kreditvergebenden Institution sollte es sein, die vorhandenen Kredit-scoringssysteme stetig zu verbessern, nicht zuletzt deshalb, weil einmal gewonnene Risikoparameter durch stetige Veränderungen der „Umwelt“ schnell veralten. Neben dem Hauptziel, die Verluste und Gewinnausfälle aus fehlklassifizierten Kreditbewerbern zu verringern, existieren weitere Gründe für die Verbesserung bestehender Verfahren. Hier ist vor allem die neue Eigenkapitalvereinbarung des Basler Ausschuss für Bankenaufsicht (Basel II) zu nennen. Sie geht von der einheitlichen Eigenkapitalunterlegung von risikobehafteten Vermögenspositionen ab, hin zu einer differenzierten, risikoabhängigen Unterlegung dieser Positionen. Für die kreditvergebende Institution ist es dabei von Vorteil, wenn sie zur Risikobestimmung interne Rating- bzw. Scoringmodelle verwenden kann. Ein Vorteil liegt darin begründet, dass alle eigenen Risiken bewertet werden können. Bei der alleinigen Verwendung externer Ratings werden hingegen viele Positionen unbewertet bleiben, die eine erhöhte Eigenkapitalunterlegung, verbunden mit höheren Kapitalkosten verlangen. Für die Verwendung eigener Modelle werden

durch Basel II jedoch bestimmte Mindestanforderungen postuliert. Zu den Anforderungen an das System gehört, dass es plausibel und intuitiv sein muss, eine bedeutende Differenzierung des Risikos erreicht, im Durchschnitt akkurate Ergebnisse erzielt und keine erkennbaren Verzerrungen aufweist.¹

Für die Bildung bzw. Verbesserung von Kreditscoringmodellen wird im Allgemeinen auf Vergangenheitsdaten, die registrierten Eingangsdaten und die Beobachtung über Ausfall oder Rückzahlung von Krediten zurückgegriffen. Diese Beobachtungen stehen jedoch nur für die Kreditbewerber zur Verfügung, denen auch wirklich ein Kredit gewährt wurde. Für die Kreditbewerber deren Kreditantrag abgelehnt wurde, sind diese Informationen nicht verfügbar. Aufgrund des Kreditvergabeverfahrens liegt eine nicht zufällige Stichprobe vor, die mit der zufälligen Stichprobe der Kreditbewerber nicht übereinstimmt. Deshalb wird es schwierig oder unmöglich, aus der beobachteten Stichprobe die Parameter einer Kreditscoringfunktion zu bestimmen, die für alle Kreditbewerber geeignet ist. Das hier geschilderte Problem ist in der Literatur unter dem Begriff der Stichprobenselektion bekannt.

Ziel dieser Arbeit soll es sein, Ansätze für die Minderung der durch Stichprobenselektion bedingten Verzerrungen vorzustellen und deren Wirkung an praktischen Beispielen zu demonstrieren. Weiterhin sollen Einflussfaktoren und ihre Auswirkung auf den Selektionseffekt untersucht und dargestellt werden. In Kapitel 2 wird das Problem der Stichprobenselektion näher beschrieben und es werden Methoden, die der Identifikation und Messung des Effektes dienen können, vorgestellt. Im dritten Kapitel erfolgt eine Darstellung der in der Literatur bekannten Lösungsansätze und eine Übersicht über die mit diesen Methoden erzielten Ergebnisse. Im folgenden Kapitel 4 werden die vorgestellten Ansätze an praktischen Beispielen getestet. In diesem Zusammenhang kommen zwei Modifikationen bereits bekannter Methoden zur Anwendung. Bestandteil dieses Kapitels ist ebenfalls die Untersuchung des Selektionseffektes und seiner Faktoren. In Kapitel 5 werden die erreichten Ergebnisse zusammengefasst.

¹Basel Committee on Banking Supervision (2003), S. 72 „The rating definitions and criteria must be both plausible and intuitive and must result in a meaningful differentiation of risk.“; S. 73 „The model must be accurate on average across the range of borrowers or facilities to which the bank is exposed and there must be no known material biases.“

2. Problemstellung

2.1. Kredit scoringverfahren

Unter dem Begriff Kredit scoring werden statistische Methoden zusammengefasst, die Kreditbewerber in gute und schlechte Risikoklassen unterteilen. Ein Kredit scoring weist im Allgemeinen Bewerbern um einen Kredit auf Basis statistischer Erkenntnisse einen Punktwert oder Score S zu. Dieser Score wird aufgrund verschiedener Eingangsvariablen (x_1, \dots, x_n) wie zum Beispiel Einkommen, monatliche Ausgaben, Vermögen, Beschäftigungsdauer berechnet und ist somit von ihnen bestimmt $S(x_1, \dots, x_n)$. Der ermittelte Scorewert sollte dabei im relationalen Zusammenhang zur Ausfallwahrscheinlichkeit $1 - \pi$ stehen. Es soll weiterhin davon ausgegangen werden, dass kleine Scores zu hohen und große Scores zu niedrigen Ausfallwahrscheinlichkeiten im Verhältnis stehen. Es liegt im Ermessen des Kreditors, im Rahmen seines Risiko-, Gewinn- und Kostenkontextes, einen Trennwert c festzulegen, der über Kreditvergabe oder Kreditverweigerung entscheidet.

Kredit scoring systeme kommen vor allem dann zum Einsatz, wenn bei gleicher Konditionierung viele ähnliche Risiken bewertet werden sollen. Beispiele dafür sind Debitsalden von Kreditkarten, Kleinkredite oder Hypothekendarlehen. Bei nicht homogenen Krediten (Großkredite, unterschiedliche Konditionierung) kommt auch heute noch eine Entscheidung auf Basis des Erfahrungsschatzes der Bankmitarbeiter zustande.

Für das Scoring von Krediten sind verschiedene Verfahren bekannt wie Punktsysteme, lineare Diskrimanzanalyse, quadratische Diskriminanzanalyse, Logit- bzw. Probit-Modelle, neuronale Netze, Entscheidungsbäume oder Expertensysteme. Im Rahmen dieser Arbeit werden vor allem Logit- und Probit-Modelle angewendet.

2.2. Stichprobenselektion

Der hier betrachtete Prozess besteht aus zwei relevanten Elementen, der Kreditvergabe und dem Kreditstatus. Die Kreditvergabe hängt dabei, wie im vorigen Abschnitt beschrieben, im Allgemeinen vom Kredit scoring, dem dabei erzielten Scorewert S und dem Trennwert c ab.

$$A = \begin{cases} 0 & \text{Kredit nicht vergeben } (S < c) \\ 1 & \text{Kredit vergeben } (S \geq c) \end{cases}$$

Nach der Kreditvergabe und Ablauf einer Zeitperiode t erhält man für die Gruppe, die einen Kredit erhalten hat ($A = 1$), zwei mögliche Ereignisse:

$$Y = \begin{cases} 0 & \text{Kredit nicht zurückgezahlt} \\ 1 & \text{Kredit zurückgezahlt} \end{cases}$$

Der Schritt der Kreditvergabe kann als Stichprobenselektion betrachtet werden. Die Stichprobe der Kreditbewerber wird durch das Verfahren von Kredit scoring und Kreditvergabe in eine reduzierte Stichprobe verwandelt. Diese Stichprobe kann, aufgrund der durch determinierte Regeln bezweckten Sortierung, der Gruppe der Bewerber um einen Kredit nicht mehr entsprechen. Ein zufälliges Fehlen einzelner Ausprägungen wäre dagegen weniger problematisch, da die Verteilungen und Wahrscheinlichkeiten der Stichprobe unbeeinflusst bleiben (siehe auch Greene, 1998).

Will man aus dem Kreditstatus Y und den Eingangsvariablen X ein neues Kredit scoringmodell schätzen, so stehen nur die akzeptierten Fälle (Gruppe $A = 1$) zur Verfügung. Die Parameter des neuen Modells sind im Vergleich zu einem Modell auf Basis der gesamten Stichprobe, aufgrund der verminderten Informationsmenge und der beschriebenen, durch das Scoring bewirkten Sortierung, verzerrt. Die verzerrten Parameter werden bei der Bewertung (Scoring) neuer Fälle, gegenüber den Parameteren eines Modells auf Basis der vollen Stichprobe, ebenso nur verzerrte Ergebnisse erzielen.

Dieser Sachverhalt wird auch in Abbildung 2.1 verdeutlicht. In dieser Abbildung sind die Verteilungen der Scorewerte in Abhängigkeit des Kreditstatus $Y = 0$ und $Y = 1$ dargestellt. Eine Trennung in abgelehnte und angenommene Fälle durch den Trennwert c , führt im rechten Teil der Abbildung zur verminderten Stichprobe. Logische Konsequenz ist, dass es durch die Stichprobenselektion zu einer Verschiebung der Mittelwerte und zu einer Verringerung der Varianz kommt. Eine veränderte Dichte ist für die beschnittenen Verteilungen ebenfalls zu verzeichnen. Es handelt sich hier um eine bedingte Beschneidung ($S \geq c$) der Verteilungen („incidental truncation“, vergleiche Greene, 1993).

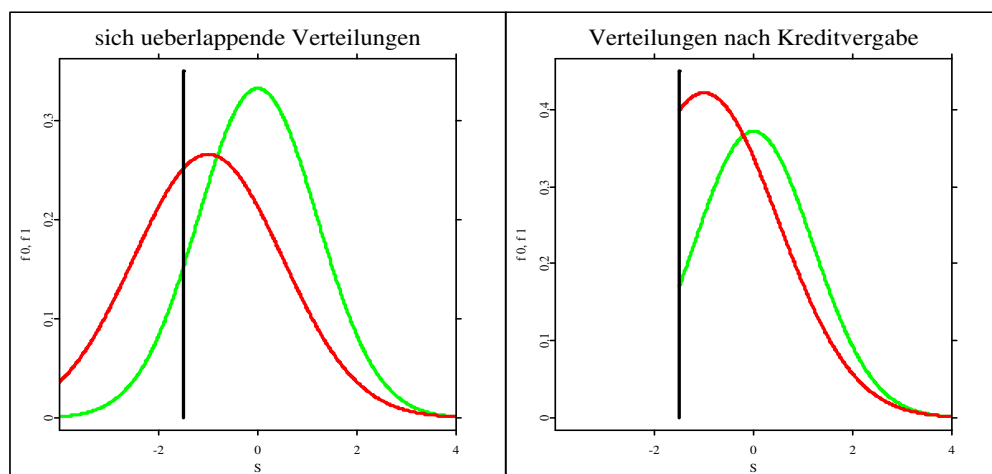


Abbildung 2.1.: Verteilung von S gegeben $Y = 0$ (rot) und $Y = 1$ (grün) mit Trennwert c (schwarz)

Aufgrund der verbreiteten Nutzung in der Literatur soll an dieser Stelle die Klassifizierung für fehlende Ausprägungen einer Variablen von Little und Rubin (1987) vorgestellt werden. Es existieren die Eingangsvariablen X und die Ausprägungen des Kreditstatus Y . Für das Fehlen der Ausprägung von Y gibt es folgende relevante Fälle:

- MCAR („missing completely at random“) Die Ausprägungen von Y fehlen zufällig. Das Fehlen bestimmter Beobachtungen ist dabei unabhängig von X und Y . Die Wahrscheinlichkeit in die Gruppe $A = 1$ aufgenommen zu werden, ist für alle Fälle gleich.

- MAR („missing at random“) Das Fehlen der Ausprägungen von Y und die Wahrscheinlichkeit für die Aufnahme in die Gruppe $A = 1$ hängen von X ab. Die Wahrscheinlichkeiten der beiden Gruppen bleiben unverändert.

$$P(Y = 1|A = 1) = P(Y = 1|A = 0)$$

- MNAR („missing not at random“) Das Fehlen der Ausprägungen von Y hängt von X und Y ab. Dies ist der beschriebene Fall der Stichprobenselektion, der durch die Verwendung von Kredit scoringmodellen, geschätzt aus Y und X bewirkt wird. Die Wahrscheinlichkeit zwischen den Gruppen werden durch die Selektion verändert.

$$P(Y = 1|A = 1) \neq P(Y = 1|A = 0)$$

2.3. Theoretische Darstellung

In diesem Abschnitt soll eine theoretische Erklärung für die in Abbildung 2.1 dargestellten Unterschiede der Verteilungen, welche durch ein Kreditvergabeverfahren bedingt waren, erfolgen. Kraft, Kroisandt und Müller (2002) demonstrieren, dass sich die Flächen der Verteilungen von S , gegeben $Y = y$ mit $y = 0, 1$, vor und nach einer Trennung am Trennwert c , in angenommene und abgelehnte Fälle unterscheiden. Nach erfolgter Kreditvergabe sind nur Informationen über die Scorewerte $\tilde{S} = (S|A = 1)$ und die Wahrscheinlichkeiten $P(\tilde{Y} = y)$ der akzeptierten Fälle verfügbar.

$$\begin{aligned} P(\tilde{S}_y \leq s) &= \frac{P(\tilde{S} \leq s, \tilde{Y} = y)}{P(\tilde{Y} = y)} = \frac{P(S \leq s, Y = y|A = 1)}{P(Y = y|A = 1)} \\ &= \frac{P(S \leq s, Y = y)}{P(S \geq c, Y = y)} \quad \text{wenn } s \geq c \end{aligned}$$

Aus $P(S_y \leq s) = P(S \leq s|Y = y) = P(S \leq s, Y = y)/P(Y = y)$ folgt

$$P(\tilde{S}_y \leq s) = \frac{P(S_y \leq s)P(Y = y)}{P(S \geq c)P(Y = y)} = \frac{P(S_y \leq s)}{P(S_y \geq c)}$$

was zeigt, dass

$$\tilde{F}_y(s) = \frac{F_y(s)}{1 - F_y(c)} \quad (2.1)$$

gilt. Wird Formel (2.1) abgeleitet, erhält man die durch Stichprobenselektion veränderte Dichte der Verteilungen. Auf Basis der verwendeten Definition lässt sich die Dichte für alle Werte von S größer c angeben.

$$\tilde{f}_y(S) = \frac{f_y(S)}{1 - F_y(c)} \quad \text{für alle } S \geq c \quad (2.2)$$

Es konnte somit gezeigt werden, dass sich die Verteilungen und Dichten aufgrund der Kreditvergabe und der damit verbundenen Beschneidung der Verteilungen in Abhängigkeit vom Trennwert c unterscheiden.

2.4. Leistungsmessung und Vergleich

Die Methoden zur Bewertung der Leistungsfähigkeit eines Kredit Scoringsystems lassen sich in Zählmaße und Separationsmaße unterscheiden (Hand und Henley, 1997). Im Rahmen dieser Arbeit sollen die vorgestellten Methoden primär zur Identifikation und Messung des Stichprobenselektionseffektes dienen.

2.4.1. Kontingenztabelle

Mit Hilfe einer 2 x 2 Kontingenztabelle (Zählmaß) lässt sich der Erfolg bei der Kreditvergabe im Zusammenhang mit dem Kreditstatus veranschaulichen. Dabei

	$A = 0$	$A = 1$
$Y = 0$	n_1	n_2
$Y = 1$	n_3	n_4

Tabelle 2.1.: Beispiel Kontingenztabelle

handelt es sich bei den Beobachtungen auf der Hauptdiagonalen n_1 und n_4 um richtig klassifizierte Fälle. Das sind abgelehnte Kredite, die bei Vergabe eines Kredits

wirklich ausgefallen wären und angenommene Kreditbewerber, die ihren Kredit zurückgezahlt haben. Die Elemente der Nebendiagonalen n_2 und n_3 sind falsch klassifizierte Fälle. Kreditanträge, die zurückgewiesen wurden, welche aber nicht ausgefallen wären, sind in n_3 zusammengefasst. Sie stellen entgangene Gewinne für das Unternehmen dar. Mit n_2 wird die Zahl derjenigen Bewerber spezifiziert, die einen Kredit erhalten, ihn aber nicht zurückgezahlt haben. Durch diese Kredite entstehen dem Unternehmen reale Verluste.

Die 2 x 2 Kontingenztabelle ist für die Praxis eher von geringem Nutzen, da nur die Werte von n_2 und n_4 beobachtet werden können. Für die in dieser Arbeit durchgeführten Simulationen (Y ist für $A = 0$ und $A = 1$ wahrnehmbar), können sie einen guten Eindruck über den Zuordnungserfolg verschiedener Ansätze liefern.

2.4.2. Trennschärfe

Das in der Arbeit von Kraft, Kroisandt und Müller (2002) vorgestellte Konzept der Trennschärfe bezieht sich auf die Größe des Überschneidungsbereiches der Verteilungen von S , gegeben den Ausprägungen $Y = 0$ und $Y = 1$.

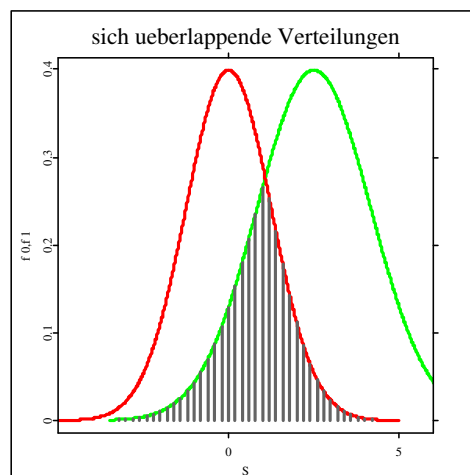


Abbildung 2.2.: Konzept der Trennschärfe - Überschneidungsbereich U (grau)

Der Überschneidungsbereich U kann dabei Werte zwischen 0 (perfekte Trennung)

und 1 (Verteilungen sind identisch) annehmen. Die Trennschärfe ist durch

$$T = 1 - U \quad T \in [0, 1]$$

definiert. Größere Werte von T sind mit einer besseren Trennung und niedrige Werte mit einer schlechten Trennung der beiden Verteilungen, auf Basis des zugrunde liegenden Scores, verbunden. Unter Normalverteilungsannahme ist der Wert von U und damit die Größe T über

$$U = F_1(s) + 1 - F_0(s) \quad (2.3)$$

bestimmbar. Für den einfachsten Fall, der Varianzgleichheit ($\sigma_0^2 = \sigma_1^2$) beider Verteilungen und der damit verbundenen Existenz eines Schnittpunktes s (für $\mu_0 \neq \mu_1$), berechnet sich dieser aus

$$s = \frac{\mu_0 + \mu_1}{2}.$$

Für den Fall, dass keine Aussagen über die Verteilungsform treffbar sind, ist die nichtparametrische Variante von U für beliebige Verteilungsformen und Anzahl von Schnittpunkten

$$U = \int \min\{f_0(s), f_1(s)\} ds \quad (2.4)$$

mit Hilfe von Kerndichteschätzern oder Histogrammen bestimmbar. Für monotone Zusammenhänge zwischen Score und Ausfallwahrscheinlichkeiten und der Existenz eines Schnittpunktes kann eine Variante dieser Definition durch

$$U = \min_s \{F_1(s) + 1 - F_0(s)\} \quad (2.5)$$

dargestellt werden (Kraft, Kroisandt und Müller, 2002).

2.4.3. Lorenzkurve

Die Lorenzkurve kann die Qualität der von Scorefunktionen erreichten Separation darstellen. Auf der in Abbildung 2.3 dargestellten Abszisse sind die nach Scorewerten, von schlecht nach gut, sortierten Kreditbewerber abgebildet. Auf der Ordinate ist der Anteil der schlechten Kredite dargestellt. Die ideale Score-

funktion sortiert die Kreditbewerber so, dass die n % schlechten Kredite den n % geringsten Scorewerten entsprechen. Für diesen Idealfall würde die Lorenzkurve der optimalen Kurve (grüne Linie) folgen. Im Normalfall wird die Scorefunktion gute und schlechte Kreditbewerber nicht perfekt separieren, somit wird die Lorenzkurve unter der optimalen Kurve liegen. Erreicht die Scorefunktion keine Separation, also die Kreditbewerber werden zufällig den Scorewerten zugeordnet, entspricht die Lorenzkurve der Diagonalen. Die Koordinaten der Lorenzkurve L in Abhängigkeit von den Scorewerten S können wie folgt dargestellt werden:

$$\{L_x(S), L_y(S)\} = \{F(S), F(S|Y = 0)\} \quad S \in (-\infty, \infty). \quad (2.6)$$

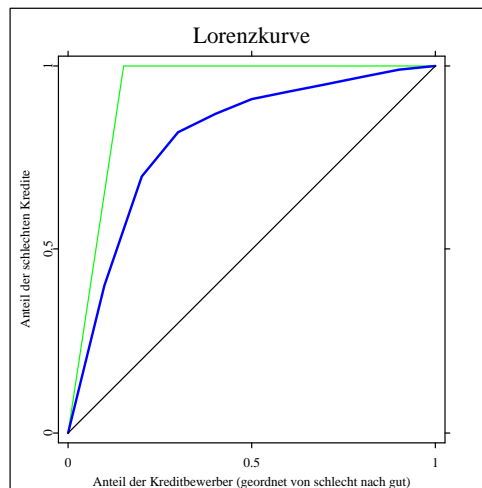


Abbildung 2.3.: Lorenzkurve

Gini-Koeffizient

Der Gini-Koeffizient G als quantitatives Maß basiert auf dem Konzept der Lorenzkurve. Er entspricht der zweifachen Fläche zwischen der Lorenzkurve und der dargestellten Diagonalen.

$$G = 1 - 2 \int_0^1 F_0^{-1}(S) dF(S) \quad (2.7)$$

Die Inverse $F_0^{-1}(S)$ von $F_0(S)$ wird dafür über die Spanne von $F(S)$ integriert. Der zweifache Wert dieser Fläche wird nun von eins abgezogen, um den Wert des Gini-Koeffizienten zu errechnen. Zur praktischen Berechnung des Gini-Koeffizienten bietet sich auch die Bestimmung der Trapezflächen unter der Lorenzkurve (Polygonzug, gegeben durch die Koordinatenpaare $\{F(S_j), F_0(S_j)\}$) an (vergleiche Neubauer, 1994).

$$G = \sum_{i=1}^n \{[F_0(S_i) + F_0(S_{i-1})][F(S_i) - F(S_{i-1})]\} - 1 \quad (2.8)$$

Größere Werte des Gini-Koeffizienten entsprechen dabei einer besseren Separation und umgekehrt.

„Accuracy-Ratio“

Die „Accuracy-Ratio“ AR setzt den für eine Scorefunktion erreichten Gini-Koeffizient zum optimalen Gini-Koeffizient ins Verhältnis. Der optimale Gini-Koeffizient entspricht, für die hier dargestellte Lorenzkurve, dem Anteil der guten Kreditbewerber an der Gesamtheit der Kreditbewerber.

$$AR = \frac{G}{G_{opt}} = \frac{G}{P(Y = 1)} \quad (2.9)$$

Der errechnete Wert AR kann zum Vergleich von Scorefunktionen genutzt werden. Für vorgestellte Konzepte von Lorenzkurve, Gini-Koeffizient und „Accuracy-Ratio“ im Kreditscoringkontext vergleiche Kraft, Kroisandt und Müller (2002).

Lorenzkurve 2

Mit dem Prinzip der Lorenzkurve können auch andere Zusammenhänge repräsentiert werden. Die im Folgenden vorgestellte Lorenzkurve, die auch unter dem Namen „Receiver-Operating“-Charakteristik bekannt ist, stellt auf der Abszisse den Anteil der schlechten akzeptierten Fälle und auf der Ordinate den Anteil der guten akzeptierten Fälle dar. Ein ideales Kreditscoringsystem würde erst die Gesamtheit aller guten Fälle akzeptieren, bevor es den ersten schlechten Fall akzeptiert. In der graphischen Darstellung (Abbildung 2.4) bedeutet dies, dass die Lorenzkurve erst der linken vertikalen Achse und dann der horizontalen Achse

folgen würde (grüne Linie). Für den Fall, dass das Scoringssystem keine Unterscheidungskraft besitzt, folgt die Kurve der Diagonalen (Hand und Henley, 1997). Der bekannte Gini-Koeffizient und die „Accuracy-Ratio“ können analog berechnet und verwendet werden.

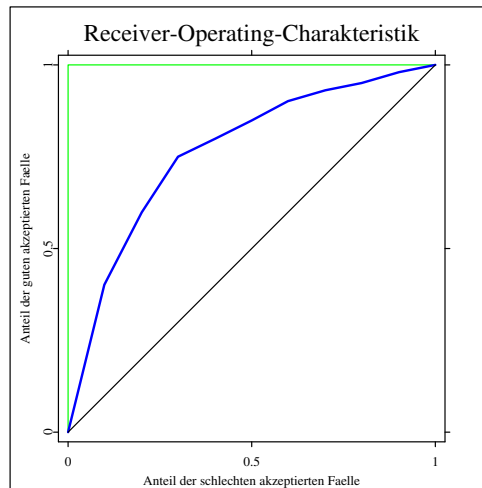


Abbildung 2.4.: „Receiver-Operating“-Charakteristik

3. Lösungsansätze

Die Modellparameter der selektierten Stichprobe können wie erwähnt, zu Verzerrungen bei der Klassifizierung zukünftiger Kreditbewerber führen. Ziel dieses Abschnittes soll es sein, Methoden vorzustellen, die zur Minderung dieser Verzerrung beitragen können. In der Literatur sind diese Methoden teilweise unter Folgerung für die abgelehnten Kreditbewerber („reject inference“) bekannt.

3.1. Veränderung des Kreditvergabeverhaltens

Ein in der Literatur bekannter Vorschlag ist, für eine kurze Zeit alle Kreditanträge zu bewilligen, um somit ein unverzerrtes Bild der Wirklichkeit für die Schätzung eines Kreditscoringmodells zu erhalten. Da diese Möglichkeit für die meisten Kredite mit sehr hohen finanziellen Risiken verbunden ist, scheint sie in der Praxis kaum durchführbar zu sein. Rosenberg und Gleit (1994) berichten in diesem Zusammenhang von einer Versandhausfirma, die allen Kunden einen geringen Kreditbetrag gewährte und mit den gesteigerten Verkaufsumsätzen entstandene Kreditverluste wieder ausgleichen konnte.

Hand (2002) schlägt für die Verbesserung von Kreditscoringmodellen eine weiche Schwelle für die Akzeptanz von Krediten vor. Risiken wird auf Basis ihres Scores S eine Wahrscheinlichkeit $p(S)$ zugewiesen, mit der sie akzeptiert oder abgelehnt werden. Somit erhalten auch Kreditbewerber mit einem hohen Ausfallrisiko, die bisher sicher dem Bereich $A = 0$ zugeordnet waren, eine wenn auch geringe Chance auf einen Kredit. Erhofft wird hierbei eine Verbesserung der Stichprobe und damit eine Minderung der auf Stichprobenselektion beruhenden Effekte. Zu beachten bleibt auch hier, dass die Menge der zusätzlichen Fälle in einem Kostenkontext betrachtet werden muss.

Eine Möglichkeit, die Vergabe von zusätzlichen Krediten und die damit verbundenen Risiken zu umgehen, wird von Ash und Meester (2002) vorgestellt. Ihr

Ansatz sieht die Informationsbeschaffung über das Ausfallverhalten eigener abgelehnter Kreditbewerber, die aber bei anderen Institutionen einen Kredit erhalten haben, vor. Diese Methode fordert eine gewisse Trefferrate bei der Beschaffung der gewünschten Informationen und ist natürlich nicht frei von Kosten. Ein möglicher verzerrender Einfluss auf die erhaltenen Informationen besteht darin, dass die Kreditbewerber bei den anderen Institutionen unter veränderten Konditionen (z.B. geringerer Kreditbetrag) den gewünschten Kredit erhalten haben.

3.2. Re-Klassifizierung

Bei der Methode der Re-Klassifizierung wird den durch die Kreditvergabe abgelehnten Fällen ein Kreditstatus Y zugeteilt. Die Kreditbewerber mit den schlechtesten Scores bzw. Eigenschaften erhalten dabei den Status $Y = 0$ (Kredit nicht zurückgezahlt). Diese Fälle werden dann zusammen mit der Gruppe $A = 1$ zur Schätzung eines neuen Scoring-Modells verwendet. In der Realität ist die Annahme $P(Y = 0|X) = 1$ für einen bestimmten Bereich der Fälle unrealistisch und wird zu Verzerrungen führen (Ash und Meester, 2002).

Eine Abwandlung der Re-Klassifizierung ist die Simulation der Ausprägung von Y für die abgelehnten Kreditbewerber. Als Wert für die Ausfallwahrscheinlichkeit wird dabei die beobachtete Ausfallrate der Stichprobe $A = 1$ verwendet. Die Simulation des Kreditstatus für die Gruppe $A = 0$ erfolgt auf Basis dieser Rate. Anschließend werden die so neu gewonnenen Fälle mit den Stichprobenelementen der Gruppe $A = 1$ zur Modellschätzung herangezogen. Diese Methode wird zu verzerrten Ergebnissen führen, da die Ausfallwahrscheinlichkeiten der Gruppen $A = 0$ und $A = 1$ aufgrund des verwendeten Scoringssystems nicht identisch sein werden.

3.3. Neugewichtung

Unter dem Begriff der Neugewichtung werden Methoden beschrieben, die durch Gewichtung bestimmter Stichprobenbestandteile eine Verbesserung für die Modellschätzung erzielen sollen.

Bei der von Ash und Meester (2002) beschriebenen Methode liegen die beob-

achteten Fälle sortiert in Bändern vor. Diese Bänder repräsentieren bestimmte Risikoklassen und enthalten jeweils eine bestimmte Anzahl von akzeptierten und abgelehnten Fällen. Es wird davon ausgegangen, dass das Ausfallverhalten in einem Band für die abgelehnten und akzeptierten Fälle identisch ist. Zur Umsetzung dieses Gedankens gehen die akzeptierten Fälle mit einem Gewicht in die Schätzung des neuen Scoringmodells ein. Das Bandgewicht entspricht dabei der Summe aus abgelehnten und akzeptierten Fällen, geteilt durch die Anzahl der abgelehnten Fälle des jeweiligen Bandes. Durch diese Art der Gewichtung wird der Anteil der abgelehnten Fälle berücksichtigt und dem Gedanken gleicher Ausfallwahrscheinlichkeiten Rechnung getragen (siehe Tabelle 3.1). Diese Methode wird ebenfalls zu verzerrten Ergebnissen führen, da sich die Ausfallwahrscheinlichkeiten für angenommene und abgelehnte Fälle in der Regel unterscheiden werden.

Band	gut	schlecht	gewährt	abgelehnt	Band-
b	$Y = 1$	$Y = 0$	$A = 1$	$A = 0$	gewicht
1	$y1_1$	$y0_1$	$a1_1 = y1_1 + y0_1$	$a0_1$	$(a1_1 + a0_1)/a1_1$
2	$y1_2$	$y0_2$	$a1_2 = y1_2 + y0_2$	$a0_2$	$(a1_2 + a0_2)/a1_2$
·	·	·	·	·	·
b	$y1_b$	$y0_b$	$a1_b = y1_b + y0_b$	$a0_b$	$(a1_b + a0_b)/a1_b$
·	·	·	·	·	·
n	$y1_n$	$y0_n$	$a1_n = y1_n + y0_n$	$a0_n$	$(a1_n + a0_n)/a1_n$

Tabelle 3.1.: Neugewichtung, vergleiche Crook und Banasik (2002), Variablendefinition siehe Tabelle 3.2

$y1_b$	Anzahl der gewährten guten Kredite pro Band
$y0_b$	Anzahl der gewährten schlechten Kredite pro Band
$a1_b = y1_b + y0_b$	Anzahl der gewährten Kredite pro Band
$a0_b$	Anzahl der der abgelehnten Kredite pro Band
$(a1_b + a0_b)/a1_b$	Bandgewicht

Tabelle 3.2.: Neugewichtung, Variablendefinition zur Tabelle 3.1

Die von Crook und Banasik (2002) vorgeschlagene Variante der Neugewichtung sieht die Gewichtung eines jeden akzeptierten Falles A_i ($i = 1 \dots m$) vor. Die

Scorewerte S stehen im monotonen Verhältnis zur Wahrscheinlichkeit der Kreditvergabe $p(S) \in [0, 1]$. Als Gewicht für die akzeptierten Fälle wird die Inverse $1/p(S)$ dieser Wahrscheinlichkeit verwendet. Akzeptierte Fälle mit sehr geringer Annahmewahrscheinlichkeit erhalten so ein sehr hohes Gewicht, während Kreditbewerber mit einem sehr hohen Wert von $p(S)$ nur ein geringeres Gewicht erhalten. Mit den so gewichteten Fällen erfolgt die Schätzung neuer Modellparameter.

3.4. Extrapolation

Unter dem Begriff der Extrapolation können, wie für die Neugewichtung, mehrere Möglichkeiten der Modifikation und Durchführung zusammengefasst werden. Im Unterschied zur Re-Klassifizierung und zur Neugewichtung werden für den Bereich der abgelehnten Fälle Werte angenommen, die auf logischen Vermutungen und statistischen Erkenntnissen beruhen (siehe Ash und Meester, 2002).

Die von Ash und Meester (2002) sowie von Crook und Banasik (2002) betrachtete Form der Extrapolation basiert auf den a posteriori Ausfallwahrscheinlichkeiten der akzeptierten Fälle. Aus dem Verlauf dieser empirischen Wahrscheinlichkeiten werden entsprechende Wahrscheinlichkeiten für die abgelehnten Kreditbewerber extrapoliert. In Abbildung 3.1 wird diese Idee für die Rückzahlungswahrscheinlichkeiten graphisch veranschaulicht. Für die extrapolierten Wahrscheinlichkeiten erfolgt die Festlegung eines Trennwertes, der die abgelehnten Fälle als gut bzw. schlecht klassifiziert. Beide Gruppen $A = 0$ und $A = 1$ werden nun zur Berechnung eines neuen Modells herangezogen. Dieses Vorgehen entspricht weitestgehend der Re-Klassifizierung der schlechtesten Kreditbewerber und wird nicht das gewünschte Ziel, die Minderung der Verzerrung, erreichen können.

Für die vorgestellte Methode der Extrapolation wird folgende Modifikation vorgeschlagen: Anstelle der Festlegung eines festen Trennwertes, wird für die extrapolierten Ausfallwahrscheinlichkeiten der Kreditstatus Y simuliert. Die abgelehnten Fälle und deren simulierte Werte von Y können nun mit den akzeptierten Fällen zur Berechnung des neuen Modells verwendet werden.

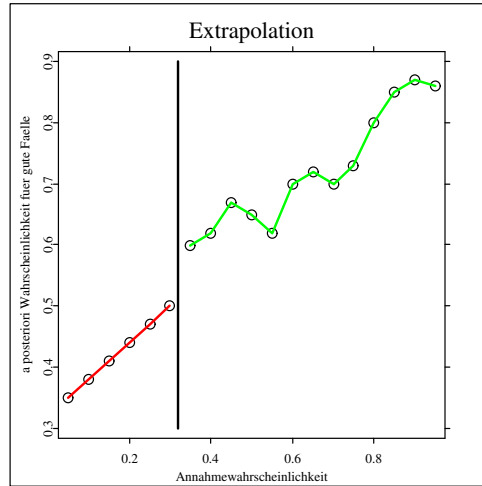


Abbildung 3.1.: Extrapolation

3.5. Heckman-Ansatz

Die Verwendung des Heckman-Ansatzes im Zusammenhang mit dem Problem der Stichprobenselektion wird von Ash und Meester (2002) vorgeschlagen. Der Hauptgedanke dieses Ansatzes besteht in der Bestimmung einer zusätzlichen Variablen, die den Selektionsmechanismus (Kreditscoring) widerspiegelt und deren Verwendung in der Modellbildung (hier Regression). Nach Heckman (1979) müssten Kreditvergabe und Kreditstatus als zwei normalverteilte Variablen aufgefasst werden:

$$\begin{aligned}
 Y &= \beta_1 X_1 + \varepsilon_1 \\
 A &= \beta_2 X_2 + \varepsilon_2.
 \end{aligned}
 \tag{3.1}$$

Die Wahrnehmbarkeit des Kreditstatus Y ist dabei abhängig vom verwendeten Selektionsverfahren. Man erhält aus der Selektion die Gruppe $A = 1$ (Y sichtbar für $S \geq 0$) und $A = 0$ (Y unsichtbar für $S < 0$). Es folgt

$$\begin{aligned}
 E(\varepsilon_1 | X_1, \text{Selektionsverfahren}) &= E(\varepsilon_1 | X_1, S \geq 0) \\
 &= E(\varepsilon_1 | X_1, \varepsilon_2 \geq -\beta_2 X_2).
 \end{aligned}$$

Für die Unabhängigkeit von ε_1 und ε_2 , also das zufällige Fehlen der Ausprägung von Y , ist das bedingte Mittel von ε_1 null. Im Allgemeinen wird dies nicht der

Fall sein, da die Ausprägungen von Y nicht zufällig fehlen. Somit ergibt sich die Regressionsfunktion

$$E(Y|X_1, S \geq 0) = \beta_1 X_1 + E(\varepsilon_1 | \varepsilon_2 \geq -\beta_2 X_2). \quad (3.2)$$

Führt man auf Basis der selektierten Stichprobe eine Regression durch, wird nur der erste Regressor aus Formel (3.2) berücksichtigt. Der zweite Teil der Formel wird vernachlässigt. Das Problem der Stichprobenselektion stellt sich so nach Heckman (1979) als einfaches Problem fallengelassener Variablen dar. Um diesen Fehler auszugleichen wird

$$E(\varepsilon_1 | \varepsilon_2 \geq -\beta_2 X_2)$$

durch

$$\frac{\sigma_{12}}{\sqrt{\sigma_{22}}} \lambda$$

ersetzt, wobei

$$\lambda = \frac{\phi(Z)}{1 - \Phi(Z)} = \frac{\phi(Z)}{\Phi(-Z)} \quad \text{für } S \geq c \quad (3.3)$$

der inversen Mills-Rate oder auch Hazard-Rate entspricht. Die inverse Mills-Rate kann dabei als Maß für die Menge der fallengelassenen Information bezüglich der abhängigen Variablen betrachtet werden. Das neue Modell ergibt sich wie folgt:

$$Y = \beta_1 X_1 + \frac{\sigma_{12}}{\sqrt{\sigma_{22}}} \lambda + v. \quad (3.4)$$

In dem von Heckman (1979) vorgeschlagenen zweistufigen Verfahren werden in der ersten Stufe die Parameter β_2 und somit die Ausprägungen der Variablen $Z = \beta_2 X_2 / \sqrt{\sigma_{22}}$ mit Hilfe eines Probit-Modells für die gesamte Stichprobe geschätzt. In der zweiten Stufe wird die auf Z beruhende inverse Mills-Rate als Regressor verwendet. Das beschriebene Verfahren setzt die Normalverteilung der Zufallsvariablen voraus. Diese Voraussetzung kann für die qualitativen Ausprägungen von Y nicht eingehalten werden. Der beschriebene Heckman-Ansatz wurde hauptsächlich für die Bestimmung unverzerrter Werte kontinuierlicher Variablen verwendet. Er könnte trotz verletzter Annahmen auch für Variablen mit qualitativer Ausprägung nützlich sein. Die vorgestellte Idee kann über die Verwendung ei-

ner zusätzlichen, den Selektionsmechanismus repräsentierenden Variablen, bei der Modellbildung einfließen.

3.6. Bivariater Probit-Ansatz

Boyes, Hoffman und Low (1989) nutzten für die Bestimmung der Ausfallwahrscheinlichkeiten ein zensiertes bivariates Probit-Modell. Dieses Modell, welches zuerst von Poirier (1980) vorgestellt und von Meng und Schmidt (1985) zusammenfassend mit anderen Modellen dargestellt wurde, fasst das Problem der Stichprobenselektion als einen Fall partieller Unbeobachtbarkeit auf. Ähnlich wie beim Heckman-Ansatz kommt es zu einer Berücksichtigung des zugrunde liegenden Selektionsmechanismus. Der durchgeführten Selektion wird durch Verwendung bivariater Normalverteilungen (Kreditvergabe und Kreditstatus) Rechnung getragen. Der Boyes, Hofman und Low (1989) zur Verfügung stehende Datensatz entsprach nicht der ursprünglichen Stichprobe aus akzeptierten und abgelehnten Kreditbewerbern (auswahlbasierte Stichprobe). Zur Behandlung dieser Verzerrung wurde eine Gewichtung auf Basis der von Manski und Lerman (1977) vorgeschlagenen WESML-Technik („weighted exogeneous sampling maximum likelihood“) durchgeführt. Das Modell für die Bestimmung der Parameter $\theta = (\beta_1, \beta_2, \rho)$ stellt sich wie folgt dar:

$$A = \beta_1 X + \varepsilon_1 \quad A = \begin{cases} 1 & \text{Kredit vergeben} \\ 0 & \text{nicht vergeben} \end{cases}$$

$$Y = \beta_2 X + \varepsilon_2 \quad Y = \begin{cases} 1 & \text{Kredit ausgefallen} \\ 0 & \text{zurückgeführt} \end{cases}$$

$$\begin{aligned} \ln L(\beta_1, \beta_2, \rho) &= \sum_i^n \{w_1 A_i Y_i \ln F(\beta_1 X_i, \beta_2 X_i; \rho) \\ &\quad + w_2 A_i (1 - Y_i) \ln [F(\beta_1 X_i) - F(\beta_1 X_i, \beta_2 X_i; \rho)] \\ &\quad + w_3 (1 - A_i) \ln F(-\beta_1 X_i)\}. \end{aligned} \quad (3.5)$$

Die Gewichte $w_{1,\dots,3}$ sind dabei wie folgt spezifiziert:

$$w_1 = Q_1/H_1$$

mit $Q_1 \equiv$ Populationsanteil von $A = 1$ und $Y = 1$, $H_1 \equiv$ Stichprobenanteil von $A = 1$ und $Y = 1$,

$$w_2 = Q_2/H_2$$

mit $Q_2 \equiv$ Populationsanteil von $A = 1$ und $Y = 0$, $H_2 \equiv$ Stichprobenanteil von $A = 1$ und $Y = 0$,

$$w_3 = Q_3/H_3$$

mit $Q_3 \equiv$ Populationsanteil von $A = 0$, $H_3 \equiv$ Stichprobenanteil von $A = 0$.

Der Populationsanteil entspricht dabei den Verhältnissen vor der auswahlbasierten Bildung der vorliegenden Stichprobe. Schätzungen für dieses Modell erhält man durch Maximierung der Likelihoodfunktion mit Hinblick auf die Parameter $\theta = (\beta_1, \beta_2, \rho)$.

3.7. Bänder

3.7.1. Identifikation von Stichprobenanteilen

Dieser Abschnitt beschreibt die von Horowitz und Manski (1998) vorgestellte Möglichkeit zur Identifikation von Stichprobenanteilen für den Fall fehlender Ergebnisausprägungen in empirischen Studien. Das Fehlen der Ergebnisausprägungen entspricht der Situation der fehlenden Information über den Kreditstatus der abgelehnten Kreditbewerber. Die Bestimmung eines bestimmten Stichprobenanteils $P(Y = y)$ mit $y = 0, 1$ folgt aus

$$E[y] = E[y|A = 1]P(A = 1) + E[y|A = 0]P(A = 0). \quad (3.6)$$

Der Erwartungswert des Anteils an guten bzw. schlechten Fällen $E[y]$ ergibt sich dabei aus den gewichteten Anteilen der einzelnen Gruppen. Der Anteil für die abgelehnten Kreditbewerber $E[y|A = 0]$ ist nicht wahrnehmbar, liegt aber innerhalb

bestimmter Grenzen $[K_0, K_1]$. Aus dieser Annahme und aus Formel (3.6) ergeben sich die Bänder für $E[y]$:

$$\begin{aligned} E[y|A = 1]P(A = 1) + K_0P(A = 0) &\leq E[y] \\ &\leq E[y|A = 1]P(A = 1) + K_1P(A = 0). \end{aligned} \quad (3.7)$$

Da der Anteil an $Y = 1$ bzw. $Y = 0$ für $A = 0$ garantiert im Bereich $[0, 1]$ liegt, ergibt sich folgender vereinfachter Ausdruck:

$$E[y|A = 1]P(A = 1) \leq E[y] \leq E[y|A = 1]P(A = 1) + P(A = 0). \quad (3.8)$$

Für bekannte geringere Spannen von $[K_0, K_1]$ können entsprechend engere Grenzen für den entsprechenden Stichprobenanteil bestimmt werden.

Zur Handhabung der Stichprobenvariation gegenüber der Realität wurden folgende Bonferroni- Konfidenzintervalle vorgeschlagen. Schreibt man die Formel (3.7) (hier für $Y = 1$) als

$$P(Y = 1|A = 1) \leq P(Y = 1) \leq 1 - P(Y = 0|A = 1), \quad (3.9)$$

so sind die asymptotischen Standardfehler für die untere (S_u) und obere Grenze (S_o) durch

$$\begin{aligned} S_u &= \{P(Y = 1|A = 1)[1 - P(Y = 1|A = 1)]/N\}^{1/2}, \\ S_o &= \{P(Y = 0|A = 1)[1 - P(Y = 0|A = 1)]/N\}^{1/2} \end{aligned} \quad (3.10)$$

gegeben. Für ein vorgegebenes Sicherheitsniveau α kann man nun die Konfidenzintervalle für das untere Band

$$P(Y = 1|A = 1) \pm \Phi^{-1}(\alpha/4)S_u$$

und das obere Band

$$[1 - P(Y = 0|A = 1)] \pm \Phi^{-1}(\alpha/4)S_o$$

bestimmen.

3.7.2. Bänder für Seperationsmaße

Kraft, Kroisandt und Müller (2002) zeigen in ihrer Arbeit, dass auch die Methoden der Leistungsmessung wie Trennschärfe und „Accuracy-Ratio“ von der durch Stichprobenselektion bewirkten Verzerrung beeinflusst werden. Zur Identifikation der wahren, unverzerrten Werte stellen sie Bänder vor. Dem in Abschnitt 2.3 dargestellten Ergebnis aus Formel (2.1)

$$\tilde{F}_y(s) = \frac{F_y(s)}{1 - F_y(c)}$$

folgend, werden der Überschneidungsbereich U und damit die Trennschärfe T ebenfalls von der Beschneidung der Verteilungen aufgrund der Kreditvergabe beeinflusst. Der Überschneidungsbereich U

$$U = F_1(s) + 1 - F_0(s)$$

verändert sich unter Anwendung von Formel (2.1) für die selektierte Stichprobe ($A = 1$) zu

$$\tilde{U} = \frac{F_1(s)}{1 - F_1(c)} + 1 - \frac{F_0(s)}{1 - F_0(c)}.$$

Aus der selektierten Stichprobe ist nur \tilde{T} und nicht der unverzerrte Wert von T bestimmbar. Unter Normalverteilungsannahme für die beiden Verteilungen von S gegeben Y ergibt sich die Bestimmung des Wertes von T aus den folgenden Ausführungen. Das unverzerrte erste und zweite Moment der Verteilungen lässt sich über die Lösung des folgenden Gleichungssystems

$$E(S_y|S_y \geq c) = \mu_y + \sigma_y \lambda(\alpha_y) \quad (3.11)$$

$$Var(S_y|S_y \geq c) = \sigma_y^2 \{1 - \lambda(\alpha_y)[\lambda(\alpha_y) - \alpha_y]\} \quad (3.12)$$

bestimmen, wobei durch

$$\lambda(\alpha) = \frac{\phi(\alpha)}{\Phi(-\alpha)} \quad \text{mit} \quad \alpha_y = \frac{c - \mu_y}{\sigma_y}$$

die inverse Mills-Rate verwendet wird. Die Werte von μ_y und σ_y^2 repräsentieren die Momente der unbeschnittenen Verteilungen. Mit den bestimmaren empirischen Momenten $E(S_y|S_y \geq c)$ und $Var(S_y|S_y \geq c)$ der selektierten Stichprobe kann man das Gleichungssystem [Formel (3.11) und (3.12)] lösen und somit den Erwartungswert $\mu_y = E(S_y)$ und die Varianz $\sigma_y^2 = Var(S_y)$ der unbeschnittenen Verteilung bestimmen (vergleiche Greene, 1993, Theorem 22.2). Mit Hilfe der so berechneten Momente lassen sich Schätzungen der unverzerrten Dichte- und Verteilungsfunktion und auch der unverzerrte Wert der Trennschärfe ermitteln.

Für den in der Praxis realistischeren Fall unbekannter Verteilungsformen schlagen Kraft, Kroisandt und Müller (2002) die Verwendung von Bändern für die Identifikation der unverzerrten Trennschärfe vor. Es werden dafür die folgenden Annahmen getroffen:

$$\begin{aligned} F_y(s) &= P(S \leq s|Y = y) \\ &= P(S \leq s, A = 1|Y = y) + P(S \leq s, A = 0|Y = y) \\ &= P(S \leq s|A = 1, Y = y)P(A = 1|Y = y) + P(S \leq s, A = 0|Y = y) \\ &= \tilde{F}_y(s)P(A = 1|Y = y) + P(S \leq s, A = 0|Y = y). \end{aligned} \quad (3.13)$$

Die obere Grenze des Wertes von $F_y(s)$ wird aus der Annahme abgeleitet, dass die Anzahl der Werte, für die $S \leq s \cap A = 0$ gilt, maximal der Gruppengröße von $A = 0$ entsprechen kann. Nutzt man dies, erhält man als obere Grenze für $F_y(s)$

$$\begin{aligned} F_y(s) &\leq \tilde{F}_y(s)P(A = 1|Y = y) + P(A = 0|Y = y) \\ &\leq 1 - P(A = 1|Y = y)[1 - \tilde{F}_y(s)]. \end{aligned} \quad (3.14)$$

Eine untere Grenze kann durch das Fallenlassen des zweiten Terms (keine Elemente für die $S \leq s \cap A = 0$ gilt) von Formel (3.13) bestimmt werden:

$$F_y(s) \geq \tilde{F}_y(s)P(A = 1|Y = y). \quad (3.15)$$

Die in den Formeln (3.14) und (3.15) enthaltene Wahrscheinlichkeit von $A = 1$ bedingt den Ausprägungen von Y [$P(A = 1|Y = y)$] ist nicht direkt bestimmbar, stattdessen wird die Spannweite für diese Wahrscheinlichkeit ermittelt. Für die weiteren Darstellungen wird folgende Notation eingeführt:

$$\alpha_y = P(A = 1|Y = y).$$

Die durch die Formeln (3.14) und (3.15) dargestellten Grenzen kann man wie folgt

$$\alpha_y \tilde{F}_y(s) \leq F_y(s) \leq 1 - \alpha_y + \alpha_y \tilde{F}_y(s) \quad (3.16)$$

schreiben. Aus $P(Y = y) = P(A = 1, Y = y) + P(A = 0, Y = y)$ wird

$$P(A = 1, Y = y) \leq P(Y = y) \leq P(A = 1, Y = y) + P(A = 0)$$

gefolgert. Von

$$\begin{aligned} \alpha_y = P(A = 0|Y = y) &= \frac{P(Y = y|A = 1)P(A = 1)}{P(Y = y)} \\ &= \frac{P(\tilde{Y} = y)P(A = 1)}{P(Y = y)} \end{aligned} \quad (3.17)$$

folgt

$$\alpha_y^* \leq \alpha_y \leq 1 \quad \text{mit} \quad \alpha_y^* = \frac{P(\tilde{Y} = y)P(A = 1)}{P(\tilde{Y} = y)P(A = 1) + P(A = 0)}. \quad (3.18)$$

Gleich der im vorangegangenen Kapitel vorgestellten Methode wird durch den Ausdruck $P(\tilde{Y} = y)P(A = 1) + P(A = 0)$ der wahre Stichprobenanteil $P(Y = y)$, durch eine obere Grenze determiniert. Aus der Kombination von Formel (3.16)

und Formel (3.18) erzielt man

$$\begin{aligned}\alpha_1^* \tilde{F}_1(s) + \alpha_0^* [1 - \tilde{F}_0(s)] &\leq F_1(s) + 1 - F_0(s) \\ &\leq 2 - \alpha_1^* [1 - \tilde{F}_1(s)] - \alpha_0^* \tilde{F}_0(s).\end{aligned}\quad (3.19)$$

Diesem Resultat folgend, erhält man nun die Bänder für den Trennwert $T = 1 - U$, die durch

$$\begin{aligned}1 - \min_s \{2 - \alpha_1^* [1 - \tilde{F}_1(s)] - \alpha_0^* \tilde{F}_0(s)\} &\leq T \\ &\leq 1 - \min_s \{\alpha_1^* \tilde{F}_1(s) + \alpha_0^* [1 - \tilde{F}_0(s)]\}\end{aligned}\quad (3.20)$$

gegeben sind. Für den Fall, dass alle Kreditbewerber einen Kredit erhalten ($\alpha_0 = \alpha_1 = 1$), reduziert sich der Ausdruck (3.20) zu

$$T = 1 - \min_s \{F_1(s) + 1 - F_0(s)\},$$

dem bereits im Abschnitt 2.4.2 aufgezeigten Ergebnis. Neben den hier dargestellten Bändern werden von Kraft, Kroisandt und Müller (2002) für die Lorenzkurve und die „Accuracy Ratio“ ebenfalls Bänder vorgestellt.

3.8. Literaturüberblick

Am Ende dieses Kapitels sollen Ergebnisse, die mit Methoden zur Minderung der Stichprobenselektion bereits erreicht wurden, zusammenfassend dargestellt werden.

Boyes, Hoffman und Low (1989) nutzten für die Bestimmung der Ausfallwahrscheinlichkeiten mit gleichzeitiger Behandlung des Stichprobenselektionseffektes das in Abschnitt 3.6 beschriebene bivariate Probit-Modell. Eine Aussage über den Einfluss der verwendeten Methode auf das erzielte Ergebnis wurde in ihrer Arbeit nicht getroffen. Der Grund dafür ist, dass der Fokus des Aufsatzes auf Gewinnmaximierung und nicht auf korrekter Bestimmung der Ausfallwahrscheinlichkeiten liegt. Erwähnenswert ist weiterhin, dass eine mögliche Heckman-Korrektur von den Autoren aufgrund der qualitativen, abhängigen Variablen abgelehnt wurde.

Greene (1998) nutzte für das Problem der Stichprobenselektion bei der Bestim-

mung der Ausfallwahrscheinlichkeiten ebenfalls bivariate Probit-Spezifikationen. Die von Boyes, Hoffman und Low (1989) benutzte Auslegung des Problems, partielle Unbeobachtbarkeit, unterscheidet sich dabei von der durch Greene (1998) verwendeten Interpretation der Zensierung. Die im Ergebnis erzielten Log-Likelihood-Funktionen sind nach Greene, trotz unterschiedlicher Interpretationen für beide Arbeiten, identisch. Gleich der Arbeit von Boyes, Hoffman und Low (1989) erfolgte kein Ergebnisvergleich von angewendeter Methode und unverzerrtem Modell (Modell der gesamten Bewerber). Verglichen wurde eine Spezifikation mit Berücksichtigung der Stichprobenselektion und ein Modell ohne deren Berücksichtigung. Beide Modelle erreichen eine geeignete Sortierung der Fälle. Das bedingte Modell (berücksichtigt Stichprobenselektion) nimmt eine höhere Ausfallrate für die gesamte Population an, weil es den letztlich abgelehnten Kreditbewerbern eine höhere Ausfallwahrscheinlichkeit zuordnet. Außerdem kann es die Fälle der Stichprobe, die aktuell ausfallen werden, etwas schärfer als das unbedingte Modell unterscheiden.

Feelders (2000) geht bei seinen Untersuchungen der Stichprobenselektion von dem beschriebenen MAR-Fall („missing at random“) für die Werte von Y aus. Wie bekannt, wird hierfür die Annahme $P(Y = 1|X, A = 1) = P(Y = 1|X, A = 0)$, gleiche Wahrscheinlichkeit für akzeptierte und abgelehnte Kreditbewerber, verwendet. Es erscheint aber unlogisch einer Gruppe von Kreditbewerbern, welche die gleiche Ausfallwahrscheinlichkeit wie die Restpopulation besitzt, den gewünschten Kredit zu verweigern. Die von Feelders (2000) durchgeführte ungewichtete Sortierung in gute und schlechte Fälle, mittels $x_1 + x_2 \geq c$, bewirkt dabei die beschriebenen gleichen Wahrscheinlichkeiten für beide Gruppen $A = 0$ und $A = 1$. Die Verwendung des EM-Algorithmus (siehe auch Dempster, Laird und Rubin, 1977), im Zusammenhang mit quadratischer Diskriminanzanalyse, erreichte für die hier verwendeten simulierten Daten gegenüber der Verwendung einfacher logistischer Regression eine leichte Ergebnisverbesserung.

Crook und Banasik (2002) untersuchten die Leistungsfähigkeit von Neugewichtung und Extrapolation. Die Vergleiche wurden gegen ein Modell für alle Kreditbewerber und ein Modell nur für akzeptierte Kreditbewerber durchgeführt. Für den Leistungsvergleich kam die vorgestellte „Receiver-Operating“-Charakteristik zur Anwendung. Die Methode der Neugewichtung konnte keine Ergebnisverbesserung

erzielen. Bei einem hohen Anteil abgelehnter Fälle verschlechterte sich das für die Neugewichtung erzielte Ergebnis. Die verwendete Extrapolation der Ausfallwahrscheinlichkeiten und die anschließende Unterteilung in gut und schlecht anhand eines Trennwertes ist nach Crook und Banasik (2002) nutzlos und harmlos. Eine wichtige Bemerkung ist, dass die nützliche Implementierung von „reject-inference“ Methoden, die akkurate Schätzung des Anteils guter bzw. schlechter Fälle für die Population aller Kreditbewerber voraussetzt.

Neben den eigenen erzielten Ergebnissen fassen Crook und Banasik (2002) auch die Resultate anderer Arbeiten zusammen. So fanden Banasik, Crook und Thomas (2001), dass die Anwendung von bivariaten Probit nur eine minimale Verbesserung gegenüber einem Modell, das nur auf den akzeptierten Fällen basiert, erreicht. Meester (2000) betrachtete zwei verschiedene Methoden der Extrapolation. Damit konnten zwar Ergebnisverbesserungen erzielt werden, diese blieben aber hinter den Erfolgen eines Modells der gesamten Stichprobe zurück.

Ash und Meester (2002) zeigen anhand eines empirischen Beispiels, dass die Verwendung von zusätzlichen Informationen, wie in Abschnitt 3.1 beschrieben, zu einem verbesserten Ergebnis gegenüber der alleinigen Verwendung der akzeptierten Fälle führen kann.

4. Praktische Anwendung

Im folgenden Kapitel soll an praktischen Beispielen der durch Stichprobenselektion bewirkte verzerrende Effekt demonstriert werden. Es ist wichtig zu bemerken, dass die in diesem Kapitel verwendeten Beispiele so konstruiert wurden, dass der Kreditstatus der abgelehnten Kreditbewerber beobachtbar ist. Im Kapitel 3 beschriebene Methoden zur Minderung der Verzerrung kommen zur Anwendung und werden in ihren erzielten Ergebnissen verglichen. Für den Ergebnisvergleich kommen vor allem die in Abschnitt 2.4.1 beschriebenen Kontingenztabellen zum Einsatz. Weiterhin soll eine Identifikation und Untersuchung von Faktoren, die einen Einfluss auf Größe und Verhalten des Selektionseffektes haben, erfolgen.

4.1. Simulierte Daten

Die Untersuchungen dieses Abschnittes basieren auf generierten normalverteilten Daten. Jeder der erzeugten Datensätze enthält 20000 Fälle und wurde in eine Trainingsstichprobe ($N = 10000$) und eine Teststichprobe ($N = 10000$) unterteilt. Ein Datensatz besteht aus vier Variablen x_j . Die Werte von x_1 repräsentieren den wahren Score und die damit verbundenen wahren Ausfallwahrscheinlichkeiten bzw. Rückzahlungswahrscheinlichkeiten der einzelnen Fälle. Die hier verwendete und allgemein bekannte Idee besteht darin, dass die Größen wahrer Score und wahre Ausfallwahrscheinlichkeit von einer sehr großen Anzahl Faktoren beeinflusst werden. Viele dieser Einflussgrößen, wie z. B. Charaktereigenschaften, sind außerdem nicht oder nur schwer wahrnehmbar bzw. messbar. Da der wahre Score für die kreditvergebende Institution nicht bestimmbar ist, werden die Werte von x_1 auch hier als unsichtbar angenommen. Die Variablen $x_{2,\dots,4}$ sollen dagegen typische für ein Scoring beobachtete Größen darstellen, die mit dem wahren Score x_1 im Zusammenhang stehen. Dieser Zusammenhang wird für die hier simulierten Daten

mit Hilfe einer Kovarianzmatrix dargestellt.

$$\text{cov} = \begin{pmatrix} 1 & 0.6 & -0.7 & 0.2 \\ & 1 & -0.6 & 0.3 \\ & & 1 & -0.7 \\ & & & 1 \end{pmatrix} \quad (4.1)$$

Für die Trainingsdaten wurden die wahren Rückzahlungswahrscheinlichkeiten π_{i1} mit der logistischen Verteilungsfunktion

$$\pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

für die Werte des wahren Scores x_1 erzeugt ($\eta_{i1} = x_{i1}$, siehe Anhang A). Mit Hilfe eines Zufallsgenerators wurde auf Basis dieser Wahrscheinlichkeiten der Kreditstatus Y nach Ablauf des Zeitraumes t simuliert. Der Generator war so eingestellt, dass ca. 75 % (genau 76,6 %) der Fälle dem Zustand $Y = 1$ zugeordnet wurden. Die restlichen Fälle erhielten die Ausprägung $Y = 0$.

4.1.1. Erstes Scoring

Auf Basis von $x_{2,\dots,4}$ wurde eine erste naive Scorefunktion

$$\eta_{2,\dots,4} = -0.2x_2 - 0.8x_3 + 0.6x_4 \quad (4.2)$$

zur Klassifizierung der Daten gebildet. Aus den so erzielten Scorewerten erfolgte die Berechnung der aus den sichtbaren Variablen bestimmaren Rückzahlungswahrscheinlichkeiten

$$p(S_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

Aufgrund der Wahrscheinlichkeiten $p(S_i)$ wurde mit dem Trennwert c die Gesamtstichprobe in abgelehnte Kredite $A = 0$ und gewährte Kredite $A = 1$ unterteilt. Die Stichprobenvariation bewirkt, trotz eines festen Wertes von c , variable Größen der Gruppe $A = 1$. Die Festlegung des Trennwertes erfolgte so, dass im Schnitt an ca. 75 % der Kreditbewerber ein Kredit vergeben wurde. Abbildung 4.1 stellt beispielhaft die Verteilung der Scores, bedingt den Ausprägungen von Y , für die

gesamte und die selektierten Stichprobe dar.

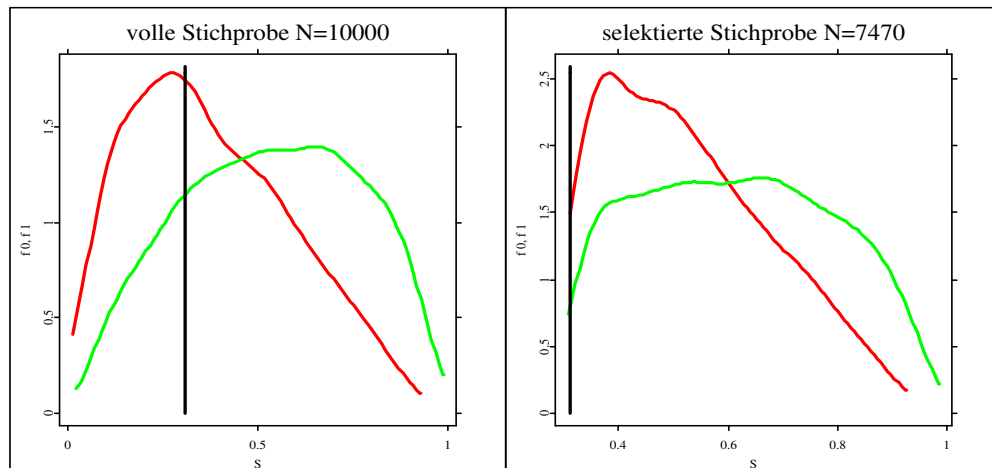


Abbildung 4.1.: Beispiel für simulierten Datensatz $Y = 0$ (rot), $Y = 1$ (grün) und Trennwert c (schwarz)

Die mit der ersten Scorefunktion [Formel (4.2)] ermittelten Scores und Wahrscheinlichkeiten $p(S)$ erreichen auf Basis des Trennwertes c einen Klassifikationserfolg von 71,8 %. Es handelt sich dabei um richtig zugeordnete Fälle für die $A = 0 \cap Y = 0$ oder $A = 1 \cap Y = 1$ gilt (Tabelle 4.1).

	$A = 0$	$A = 1$
$Y = 0$	10,22 %	13,12 %
$Y = 1$	15,08 %	61,58 %

Tabelle 4.1.: Ergebnis der ersten Scorefunktion mit $A =$ Kreditvergabe und $Y =$ Kreditstatus

Auf Basis der binären Größe Y und den zugeordneten Variablen $x_{2,\dots,4}$ wurden im Folgenden die Parameter β_j eines logistischen Modells geschätzt. Die Schätzung wurde dabei sowohl für die gesamte Stichprobe (alle Kreditbewerber erhalten Kredit) als auch für die reduzierte Stichprobe $A = 1$ durchgeführt.

Die geschätzten Parameter kommen nun für die Bewertung der Teststichprobe zur Anwendung. Die mit der logistischen Verteilungsfunktion, den Parametern β_j

und Werten von $x_{2,\dots,4}$ der Teststichprobe erzeugten Scores werden zum Sortieren der Prüfstichprobe verwendet. Die Ausprägungen des Kreditstatus wurden der Methodik für die Trainingsdaten folgend, generiert und den Scores zugeordnet. Zum Vergleich der geschätzten Modelle wurden die 25 % niedrigsten Scores in die Gruppe $A = 0$ eingeordnet, während höhere Werte von S in die Gruppe $A = 1$ gelangten. Die Durchführung der beschriebenen Trainings- und Testprozedur erfolgte dabei hundert Mal. Die Mittelwerte der Zuordnungserfolge sind in den Tabellen 4.2 und 4.3 dargestellt.

	$A = 0$	$A = 1$
$Y = 0$	14,54 %	8,84 %
$Y = 1$	10,46 %	66,16 %

Tabelle 4.2.: Scoring und Kreditstatus für die volle Stichprobe

	$A = 0$	$A = 1$
$Y = 0$	13,34 %	10,04 %
$Y = 1$	11,66 %	64,96 %

Tabelle 4.3.: Scoring und Kreditstatus für die selektierte Stichprobe

Die beiden Tabellen zeigen den Selektionseffekt für die simulierten Daten. Das Modell, dessen Parameter β_j auf den Informationen der gesamten Trainingsstichprobe beruht, konnte 80,7 % aller Fälle der Teststichprobe richtig klassifizieren. Das Modell, das auf Basis der reduzierten Stichprobe errechnet wurde, erreicht nur eine Genauigkeit von 78,3 %. Der Zuordnungserfolg für die beiden Modelle unterscheidet sich aufgrund der durch Stichprobenselektion verursachten Informationsverluste um 2,4 Prozentpunkte. Die Klassifikationserfolge der beiden Stichproben werden als Maßstab für die nun angewendeten Methoden zur Minderung des Selektionseffektes verwendet.

4.1.2. Methoden zur Minderung des Selektionseffektes

Veränderung des Kreditvergabebehavior

Durch die Verwendung aller Fälle zur Schätzung der Modellparameter erhält man, wie gezeigt, vom Selektionseffekt unbeeinflusste Parameter. Die Vergabe von Kre-

diten an alle Bewerber ist aber im Allgemeinen aufgrund des hohen finanziellen Risikos nicht durchführbar.

Der im Abschnitt 3.1 vorgestellte Ansatz von Hand (2002), der die Aufnahme von Fällen mittels weichem Trennwert vorsieht, scheint dagegen für die praktische Anwendung geeigneter zu sein. Bei den folgenden Berechnungen wurde der feste Trennwert c beibehalten. Für die Fälle, die bisher in der Trainingstichprobe eindeutig der Gruppe $A = 0$ zugeordnet waren, wurde die weiche Annaheschwelle von Hand (2002) eingeführt. Als Wahrscheinlichkeit für die Annahme oder Ablehnung eines Falles kam die mit Hilfe der logistischen Verteilungsfunktion und der ersten Scorefunktion ermittelten Rückzahlungswahrscheinlichkeiten $p(S)$ zur Anwendung. Ein Zufallsgenerator bestimmte auf Basis dieser Wahrscheinlichkeiten die Zugehörigkeit der Fälle zu einer der Gruppen $A = 0$ oder $A = 1$. Die als $A = 1$ klassifizierten Fälle ergeben zusammen mit der auf Basis von c bestimmten Stichprobe die neue Gruppe $A = 1$. Aufgrund dieser Methode erhalten ca. 19,5 % der bisher abgelehnten Kreditbewerber zusätzlich einen Kredit. Die Auswirkungen auf die selektierte Stichprobe werden in Abbildung 4.2 demonstriert.

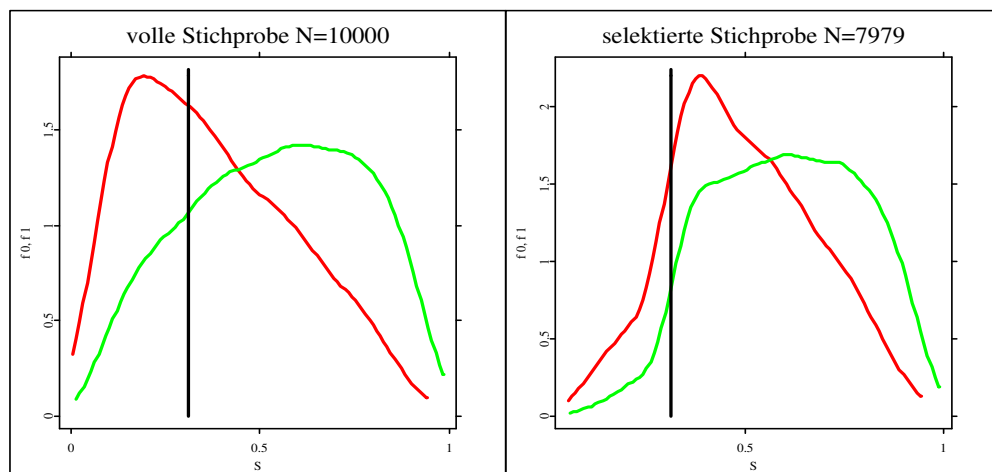


Abbildung 4.2.: Kreditvergabe mit weichem Trennwert

Für die Teststichprobe konnte mit dem Modell der vergrößerten selektierten Stichprobe das in der Tabelle 4.4 gezeigte Ergebnis erreicht werden. Mit der Umklassifizierung auf Basis der Wahrscheinlichkeiten $p(S)$ wurden 79,12 % aller Fälle richtig zugeordnet. Dieses Modell weist eine Verbesserung des Zuordnungserfolges

	$A = 0$	$A = 1$
$Y = 0$	13,74 %	9,62 %
$Y = 1$	11,26 %	65,38 %

Tabelle 4.4.: Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert 1

von 0,82 Prozentpunkten gegenüber dem einfachen selektierten Modell auf. Eine Erhöhung bzw. Absenkung des zusätzlich aufgenommenen Stichprobenanteils auf Basis der Wahrscheinlichkeiten führt zur Steigerung bzw. Minderung des Zuordnungserfolges. Tabelle 4.5 zeigt beispielhaft die Verwendung von 1,95 % der bisher abgelehnten Kredite. Es werden mit einem Klassifizierungserfolg von 78,48 %, gegenüber dem selektierten Modell (78,3 %), nur leicht verbesserte Ergebnisse erreicht.

	$A = 0$	$A = 1$
$Y = 0$	13,42 %	9,94 %
$Y = 1$	11,58 %	65,06 %

Tabelle 4.5.: Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert 2

Re-Klassifizierung

Bei dieser Methode wird den Kreditbewerbern mit den schlechtesten Scorewerten der Status $Y = 0$ zugeordnet. Die zusätzlichen Fälle werden zusammen mit den Werten der selektierten Stichprobe zur Schätzung des Modells verwendet.

	$A = 0$	$A = 1$
$Y = 0$	13,31 %	9,97 %
$Y = 1$	11,69 %	65,03 %

Tabelle 4.6.: Ergebnis simulierte Daten, Re-Klassifizierung (1 %)

Tabelle 4.6 zeigt das Ergebnis der Zuordnung zu $Y = 0$ für die Kreditbewerber mit den 1 % niedrigsten Scores. Das Ergebnis der Teststichprobe weist einen Zuordnungserfolg von 78,34 % auf. Ohne diese Daten wurde ein Wert von 78,3 %

erreicht. Ein deutlicher Unterschied konnte somit nicht erzielt werden.

	$A = 0$	$A = 1$
$Y = 0$	13,24 %	10,12 %
$Y = 1$	11,76 %	64,88 %

Tabelle 4.7.: Ergebnis simulierte Daten, Re-Klassifizierung (5 %)

In Tabelle 4.7 ist dargestellt, dass sich die Ergebnisse mit Erhöhung des re-klassifizierten Stichprobenanteils verschlechtern (78,12 %). Dieses Modell fällt noch hinter das Modell auf Basis der Stichprobe $A = 1$ zurück. Der rechte Teil

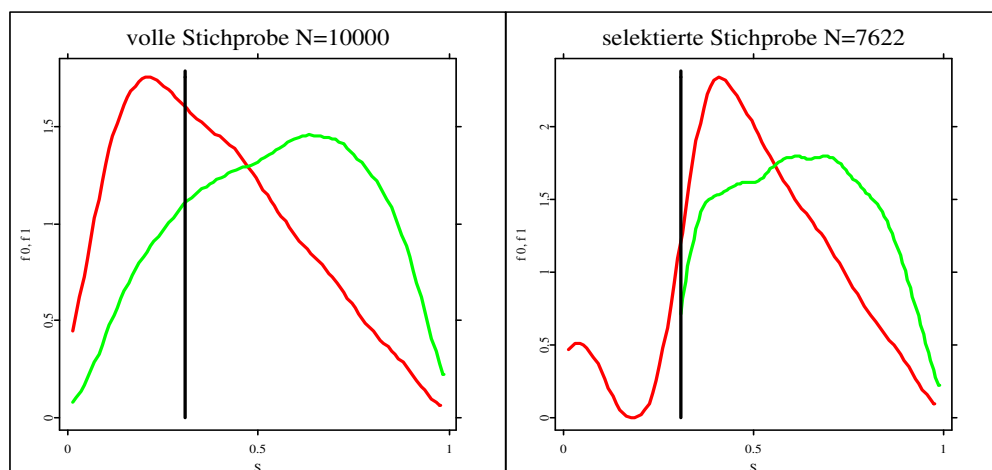


Abbildung 4.3.: Re-Klassifizierung

von Abbildung 4.3 zeigt den durch die Re-Klassifizierung hinzugewonnenen Bereich außerhalb der selektierten Stichprobe. Im linken Teil dieser Abbildung wird deutlich, warum der Ansatz der Re-Klassifizierung nicht erfolgreich ist. In dem Bereich, in dem der Kreditstatus der Fälle als $Y = 0$ angenommen wird, befindet sich noch ein erheblicher Anteil an Kreditbewerbern, die in Wirklichkeit nicht ausgefallenen wären. Diese Methode führt zu Verzerrungen, da die Annahme, dass die Gruppe der Bewerber mit den niedrigsten Scores ausgefallenen Krediten entspricht, nicht mit der Realität der simulierten Daten übereinstimmt.

Neugewichtung

Für die Neugewichtung kam die von Crook und Banasik (2002) vorgeschlagene Methode der Gewichtung zur Anwendung. Als Annahmewahrscheinlichkeiten wurden die aus der logistischen Verteilung gewonnenen Rückzahlungswahrscheinlichkeiten $p(S)$ verwendet. Alle akzeptierten Fälle erhielten die Inverse dieser Wahrscheinlichkeiten $1/p(S)$ als Gewicht. Die so gewichteten Fälle gingen in die Schätzung der neuen Modellparameter ein. Für die Teststichprobe wurden mit den Spezifikationen dieses Modells folgende Ergebnisse erreicht (Tabelle 4.8).

	$A = 0$	$A = 1$
$Y = 0$	13,86 %	9,49 %
$Y = 1$	11,14 %	65,51 %

Tabelle 4.8.: Ergebnis simulierte Daten, Neugewichtung

Mit dieser Art der Neugewichtung konnten im Schnitt 79,37 % aller Fälle der Teststichprobe richtig klassifiziert werden. Gegenüber der einfachen selektierten Stichprobe ergab das eine Verbesserung von 1,07 Prozentpunkten.

Kreditvergabe mit weichem Trennwert und Neugewichtung

Durch die Kreditvergabe mit weichem Trennwert werden Kredite an Kreditbewerber vergeben, die unter normalen Umständen keinen Kredit erhalten hätten. Die zusätzliche Information für den Bereich $A = 0$ verbessert das neue Selektionsmodell. Die im letzten Abschnitt behandelte Methode der Neugewichtung mit den inversen Wahrscheinlichkeiten $1/p(S)$ gibt den angenommenen Kreditbewerbern, die gerade noch akzeptiert wurden, ein höheres Gewicht und konnte die Effekte der Stichprobenselektion ebenfalls mildern.

In diesem Abschnitt sollen beide Methoden kombiniert zur Anwendung kommen. Auf Basis der Werte von $p(S)$ werden für den Bereich $A = 0$ zusätzliche Kredite angenommen. Für die Parameterschätzung erhalten sie die Inverse dieser Wahrscheinlichkeit $1/p(S)$ als Gewicht. Den Elementen der Stichprobe $A = 1$ wird ein Gewicht von 1 zugeteilt, da sie aufgrund des Trennwertes c sicher in der selektierten Stichprobe vorhanden sind. Für das erste Modell erhielten, aufgrund gleicher

Anwendung der weichen Aufnahmeschwelle, ca. 19.5 % aller sonst abgelehnten Bewerber einen Kredit. Tabelle 4.9 zeigt das Ergebnis für die Teststichprobe, das

	$A = 0$	$A = 1$
$Y = 0$	14,49 %	8,83 %
$Y = 1$	10,51 %	66,17 %

Tabelle 4.9.: Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 1

mit der Kombination beider Methoden erreicht wurde. Es konnte ein Klassifizierungserfolg von 80,66 % erzielt werden. Das Ergebnis ist dabei nahezu identisch mit dem vom Modell der gesamten Stichprobe erreichten Zuordnungserfolg von 80,7 %. Mit dem Ziel der Risikominimierung, also weniger Fälle aus dem Bereich $A = 0$ aufzunehmen, wurde im Folgenden durch multiplikative Veränderung die Chance der Aufnahme beeinflusst. Im gleichen Maße erfolgte auch die Veränderung der Gewichtung. So wurden im Extremfall nur noch 10 % der bisherigen Datenmenge, also 1,95 % der Fälle der Gruppe $A = 0$, zusätzlich aufgenommen. Die Gewichte der zusätzlichen Datenmenge sind entsprechend verzehnfacht worden (siehe Tabelle 4.10).

	$A = 0$	$A = 1$
$Y = 0$	14,30 %	9,02 %
$Y = 1$	10,70 %	65,98 %

Tabelle 4.10.: Ergebnis simulierte Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 2

Trotz Verringerung der zusätzlichen Datenmenge konnten immer noch 80,28 % aller Fälle richtig klassifiziert werden. Mit Verwendung dieser Methode erhält man hier eine Möglichkeit, den Stichprobenselektionseffekt weitgehend auszugleichen, ohne gleichzeitig zu hohe Risiken einzugehen.

Extrapolation

In diesem Abschnitt ist die Methode der Extrapolation verbunden mit der Simulation des Kreditstatus verwendet worden. Von der vorgeschlagenen Extrapolation

mit Festlegung des Status von $Y = 1$ über und $Y = 0$ unter einem bestimmten Trennwert für die Gruppe $A = 0$ wird abgesehen, da dieses Vorgehen einer Re-Klassifizierung entspricht, die bei den hier vorliegenden Verteilungen vermutlich keine Erfolge erreichen kann.

Die auf Basis der Annahmewahrscheinlichkeit sortierten akzeptierten Fälle sind in acht Bänder unterteilt worden. Für jedes Band erfolgte die Bestimmung des arithmetischen Mittels der Annahmewahrscheinlichkeiten. Die ermittelten a posteriori Wahrscheinlichkeiten für gute Fälle des jeweiligen Bandes wurden den jeweiligen Mittelwerten zugeordnet. Diese Wertepaare dienten zur Ermittlung der Koeffizienten einer linearen Regression. Mit diesen Koeffizienten erfolgte für die abgelehnten Kreditbewerber die Extrapolation der Wahrscheinlichkeit guter Fälle. Ein Zufallsgenerator ordnete nun auf Basis der extrapolierten Wahrscheinlichkeiten den abgelehnten Fällen einen Kreditstatus zu. Die so gewonnenen Beobachtungen und die Gruppe $A = 1$ wurden zur Schätzung eines neuen Modells verwendet.

	$A = 0$	$A = 1$
$Y = 0$	14,47 %	8,99 %
$Y = 1$	10,53 %	66,01 %

Tabelle 4.11.: Ergebnis simulierte Daten, Extrapolation

Die hier beschriebene Extrapolation erreicht ein Ergebnis von 80,48 % richtig klassifizierter Fälle. Sie schneidet damit vergleichbar positiv wie die beschriebene Methode des weichen Trennwertes mit zusätzlicher Neugewichtung ab, ohne dass zusätzliche Fälle aus der Gruppe der abgelehnten Kreditbewerber aufgenommen wurden. Zu bedenken bleibt, dass bei den hier verwendeten normalverteilten Daten, der Verteilungsverlauf des Anteils guter Fälle relativ monoton ist, was eine gute Voraussetzung für die angewendete lineare Regression ist.

4.2. Datensatz aus Fahrmeir und Hamerle (1984)

Der aus Fahrmeir und Hamerle (1984) entnommene Datensatz einer süddeutschen Großbank enthält Beobachtungen über den Kreditstatus von 1000 Kreditbewerbern. Er fasst die Merkmalsausprägungen X von 700 guten Krediten und 300 schlechten Krediten. Es handelt sich bei dem Datensatz um eine geschichtete

Stichprobe, da ihm zusätzlich schlechte Kredite weit über den üblichen Anteil von drei bis sieben Prozent hinzugefügt wurden. Der Datensatz enthält 20 Merkmale x_j . Zur Vereinfachung sind für die Modellbildung die sieben in Tabelle 4.12 dargestellten Variablen ausgewählt worden. Mit Hilfe eines Zufallsgenerators wurde

x_1	Darlehenshöhe in DM
x_2	Laufzeit des Kredits in Monaten
x_3	Alter des Antragstellers in Jahren
x_4	bisherige Zahlungsmoral
x_5	weitere Ratenkredite
x_6	Sparkonto oder Wertpapiere
x_7	Wohnung

Tabelle 4.12.: Variablenbezeichnung des Datensatzes aus Fahrmeir und Hamerle (1984), für eine nähere Beschreibung siehe Anhang B

der Datensatz in zwei Teile geteilt. Der dabei erzeugte Trainingsdatensatz besteht aus ca. 600 Beobachtungen und soll zur Modellbildung verwendet werden. Der Testdatensatz enthält die Verbleibenden der 1000 Fälle. Die Durchführung der Simulation erfolgte jeweils hundertmal.

Es sollen wie bei den simulierten Daten der Effekt der Stichprobenselektion aufgezeigt und Methoden zur Minderung der Verzerrung angewendet werden. Bei dem vorliegenden Datensatz handelt es sich um eine bereits selektierte Stichprobe, die durch Scoring und Kreditentscheidung zustande gekommen ist. Da der Datensatz einen relativ hohen Anteil an schlechten Krediten besitzt, wird er als Abbildung der Realität verwendet. Die Annahme, dass es sich bei den 1000 Fällen um eine zufällige Stichprobe der Kreditbewerber handelt, lässt die Möglichkeit zu, ein eigenes Scoring und eine eigene Stichprobenselektion durchzuführen. Der Kreditstatus aller Kreditbewerber ist bekannt, deshalb können ähnliche Untersuchungen wie bei den simulierten Daten durchgeführt werden.

4.2.1. Erstes Scoring

Die Parameter der ersten Scorefunktion wurden einmalig mit einem Logit-Modell aus einer zufällig gezogenen Stichprobe bestimmt, die 250 Elemente der Gesamt-

stichprobe umfasst. Für die Trainingsstichprobe sind mit diesen Parametern der zugehörige Score und die Rückzahlungswahrscheinlichkeit $p(S)$ berechnet worden. Die 150 niedrigsten Scores (ca. 25 % der Fälle) wurden als abgelehnt $A = 0$ deklariert. Alle Fälle mit höheren Scorewerten erhielten den Status eines angenommenen Kreditbewerbers ($A = 1$).

	$A = 0$	$A = 1$
$Y = 0$	11,63 %	18,49 %
$Y = 1$	13,37 %	56,51 %

Tabelle 4.13.: Ergebnis der ersten Scorefunktion

Die Scorefunktion und die Trennung nach den ersten 150 Werten ordnete 68,14 % aller Fälle richtig ein (Tabelle 4.13). In Abbildung 4.4 werden die Verteilungsfunktionen des Scores, gegeben $Y = 0$ und $Y = 1$, für die volle Stichprobe und die selektierte Stichprobe dargestellt.

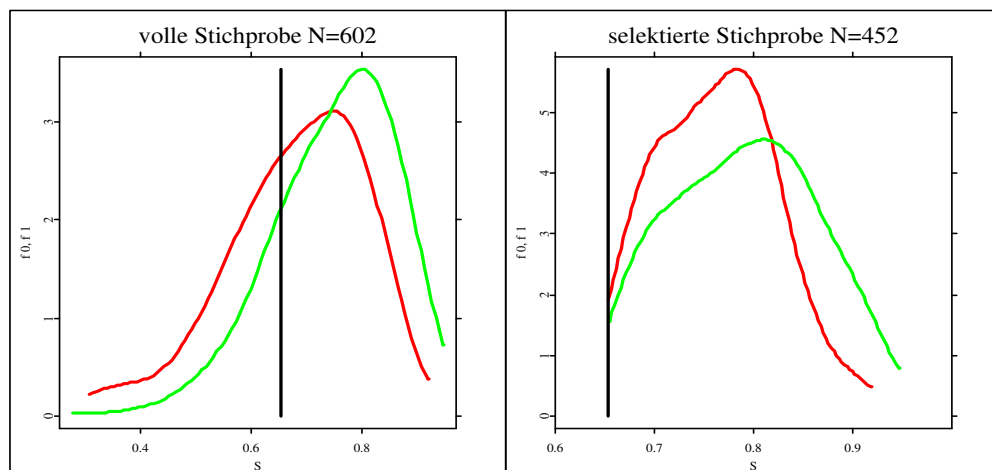


Abbildung 4.4.: empirische Verteilungen von S , gegeben den Ausprägungen von Y

Die Informationen der vollen Stichprobe (ca. 600 Elemente) und der selektierten Stichprobe (ca. 450 Elemente) liefern die Basis für die Berechnung von zwei neuen Logit-Spezifikationen. Die so erhaltenen Parameter dienen zum Scoring der Teststichprobe. Die Trennung erfolgte ebenfalls nach ca. 25 % (entspricht 100 Fällen)

der niedrigsten Scorewerte. Das erzielte Klassifikationsergebnis ist in den Tabellen 4.14 und 4.15 dargestellt. Die aus der vollen Stichprobe gewonnene Scoringfunk-

	$A = 0$	$A = 1$
$Y = 0$	11,82 %	17,81 %
$Y = 1$	13,18 %	57,19 %

Tabelle 4.14.: Scoring und Kreditstatus für die volle Stichprobe

	$A = 0$	$A = 1$
$Y = 0$	11,33 %	18,30 %
$Y = 1$	13,67 %	56,70 %

Tabelle 4.15.: Scoring und Kreditstatus für die selektierte Stichprobe

tion erreichte für die Teststichprobe im Schnitt 69,01 % richtige Klassifizierungen. Die Scoringfunktion, die aus der selektierten Stichprobe ermittelt wurde, erzielte 68,03 % korrekte Zuordnungen. Der Unterschied zwischen den beiden Modellen, der auf Stichprobenselektion zurückzuführen ist, beträgt 0,98 Prozentpunkte. Es bleibt hier zu beachten, dass die Verschlechterung des Klassifikationserfolges nicht monoton für alle der 100 Simulationen vorliegt und nur im Durchschnitt identifizierbar ist. Für die Zelle $A = 1$ und $Y = 1$ erreichte das auf der selektierten Stichprobe beruhende Modell 23 mal bessere Ergebnisse und elf mal das gleiche Ergebnis wie das Modell auf Basis der vollen Stichprobe. Ursachen für diesen Sachverhalt werden im Abschnitt 4.3 näher untersucht.

4.2.2. Methoden zur Minderung des Selektionseffektes

Bei den durchgeführten hundert Simulationsläufen kam es zu Abweichungen für die erzielten Zuordnungserfolge. Die festgestellten Unterschiede für den Zuordnungserfolg zwischen selektierter und voller Stichprobe wichen von dem oben dargestellten Wert (0,98) teilweise um 0,3 Prozentpunkte ab. Die von den Methoden erreichten prozentualen Verbesserungen blieben dabei relativ stabil. Um den Einfluss der verschiedenen Methoden besser und klarer darzustellen, werden für jede Methode zusätzlich die Klassifikationserfolge für das Modell der vollen und der selektierten Stichprobe angegeben. Ebenfalls wurden bestimmte Variationen der

Methoden im gleichem Simulationslauf angewendet. Die erreichten Ergebnisse der angewandten Methoden sollten somit besser einzuordnen sein. Der Klassifikationserfolg ($\sum \text{Hd.} \equiv$ Summe der Hauptdiagonalelemente) wurde für die Modelle der vollen und der selektierten Stichprobe und für die verwendete Methode in jeder Tabelle mit angegeben.

Kreditvergabe mit weichem Trennwert

Die mit der ersten Scorefunktion erreichten Rückzahlungswahrscheinlichkeiten $p(S)$ wurden als Ausgangspunkt für die wahrheitsbedingte Aufnahme in die selektierte Stichprobe verwendet. Aufgrund der Verteilung der Scorewahrscheinlichkeiten erhielten ca. 60% der in $A = 0$ enthaltenen Bewerber zusätzlich zur Gruppe $A = 1$ einen Kredit.

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,83 %	18,23 %	11,38 %	18,68 %	11,67 %	18,39 %
$Y = 1$	13,02 %	56,92 %	13,47 %	56,47 %	13,18 %	56,76 %
$\sum \text{Hd.}$	68,75 %		67,85 %		68,43 %	

Tabelle 4.16.: Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert 1

Wie Tabelle 4.16 zeigt, steigert die erhöhte Kreditgewährung den Anteil der richtig klassifizierten Fälle für die Teststichprobe. Eine Erhöhung des Anteils akzeptierter Fälle um ca. 60 % der in Gruppe $A = 0$ enthaltenen Fälle ist aufgrund des hohen Risikos nicht die erwünschte Lösung des Problems. Tabelle 4.17 zeigt die Ergebnisse für ein Fünftel dieser Datenmenge gleich 12% der Gruppe $A = 0$. Die Aufnahme von 12 % zusätzlicher bisher abgelehnter Fälle bringt nur eine leichte Verbesserung gegenüber der reinen selektierten Stichprobe. Wie bei den simulierten Daten stehen auch bei diesem Datensatz die zusätzliche Datenmenge und die Minderung des Selektionseffektes im relationalen Zusammenhang.

Re-Klassifizierung

Für die Methode der Re-Klassifizierung erhielten 10 % der schlechtesten Scores der Gruppe $A = 0$ den Kreditstatus $Y = 0$. Die zusätzlichen Fälle wurden mit der

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	12,08 %	17,90 %	11,58 %	18,40 %	11,65 %	18,34 %
$Y = 1$	12,89 %	57,13 %	13,39 %	56,63 %	13,32 %	56,69 %
\sum Hd.	69,21 %		68,21 %		68,34 %	

Tabelle 4.17.: Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert 2

Gruppe der akzeptierten Kreditbewerber für die Bildung eines Modells verwendet. Mit der Re-Klassifizierung konnte im Gegensatz zu den simulierten Daten eine

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,97 %	17,91 %	11,48 %	18,4 %	11,74 %	18,14 %
$Y = 1$	13,02 %	57,10 %	13,52 %	56,60 %	13,26 %	56,86 %
\sum Hd.	69,07 %		68,08 %		68,60 %	

Tabelle 4.18.: Ergebnis empirische Daten, Re-Klassifizierung

deutliche Minderung des Selektionseffektes erreicht werden.

Neugewichtung

Gleich den simulierten Daten ist die Inverse der durch das Logit-Modell gewonnenen Wahrscheinlichkeiten $p(S)$ zur Neugewichtung der selektierten Stichprobe verwendet worden. Im Unterschied zu den simulierten Daten, konnte die Neu-

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,98 %	17,80 %	11,46 %	18,32 %	11,46 %	18,32 %
$Y = 1$	13,03 %	57,19 %	13,56 %	56,66 %	13,55 %	56,67 %
\sum Hd.	69,17 %		68,12 %		68,13 %	

Tabelle 4.19.: Ergebnis empirische Daten, Neugewichtung

gewichtung für diesen empirischen Datensatz keine Minderung des Stichproben-selektionseffektes erzielen (Tabelle 4.19). Eine mögliche Erklärung ist, dass die

akzeptierten Fälle nahe am Trennwert ein geringeres Gewicht [hohe Werte von $p(S)$], als die akzeptierten Fälle nahe c bei den simulierten Daten erhalten. Eine andere Begründung wäre, dass die höhere Gewichtung der Fälle am Trennwert nichts hilft, da diese Fälle den abgelehnten Fällen weniger entsprechen.

Kreditvergabe mit weichem Trennwert und Neugewichtung

Bei dieser Methode gingen, wie bei den simulierten Daten, die zusätzlich aus $A = 0$ aufgenommenen Fälle mit der Inversen der Wahrscheinlichkeit $p(S)$ und die Fälle der Gruppe $A = 1$ mit dem Gewicht 1 in die Schätzung des neuen Logit-Modells ein. Die hier beschriebene Methode wurde im selben Simulationslauf wie die Kreditvergabe mit weichem Trennwert angewendet, um die Ergebnisse vergleichbar zu machen. Tabelle 4.20 stellt das Ergebnis für 60 % zusätzlicher Elemente aus

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,83 %	18,23 %	11,38 %	18,68 %	11,74 %	18,32 %
$Y = 1$	13,02 %	56,92 %	13,47 %	56,47 %	13,11 %	56,83 %
\sum Hd.	68,75 %		67,85 %		68,57 %	

Tabelle 4.20.: Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 1

$A = 0$ mit Neugewichtung dar. Es zeigt sich, dass der Zuordnungserfolg durch die Gewichtung leicht verbessert werden kann und das Ergebnis des Modells auf Basis der vollen Stichprobe fast erreicht wird. Trotz eines weitaus höheren Anteils verwendeter Daten sind die erzielten Ergebnisse jedoch nicht so überzeugend wie bei den simulierten Daten. Die Verwendung von 12 % zusätzlicher Daten und die Nutzung des entsprechend mit fünf multiplizierten Gewichtsvektors erzielten die in Tabelle 4.21 dargestellten Klassifikationsergebnisse. Die verwendete Methodik, welche bei den simulierten Daten relativ gut funktionierte, weist hier ein schlechteres Ergebnis auf als die alleinige Nutzung der selektierten Stichprobe. Die Ursache dafür ist in der größeren Ausprägungsnähe der Variablen für die simulierten Daten zu sehen. Bei zusätzlicher Selektion eines Bewerbers entspricht dieser vielen anderen, die in der „Nähe“ dieses Falls liegen. Weiterhin wirkt die große Bandbreite (Anzahl) an abgelehnten Fällen stabilisierend. Die auf Grund zufälliger Selektion erhaltenen zusätzlichen Kreditbewerber werden gleichmäßiger verteilt sein. Bei

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,84 %	17,73 %	11,36 %	18,21 %	11,09 %	18,48 %
$Y = 1$	13,19 %	57,24 %	13,67 %	56,76 %	13,94 %	56,49 %
\sum Hd.		69,08 %		68,12 %		67,58 %

Tabelle 4.21.: Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 2

den hier verwendeten empirischen Daten ist der Bereich der Scorewerte, für den Kreditanfragen abgelehnt werden, sehr klein. Er umfasst, wie bereits angeführt, nur 150 Werte. Bei ca. 18 zusätzlichen Fällen (12 % der Gruppe $A = 0$) werden diese auf Grund des kleinen Bereichs von $A = 0$ und der geringeren Ausprägungsnähe der Variablen nicht den Nachbarfällen entsprechen. Die geringe Stichprobengröße kann zusätzlich zu einer nicht gleichmäßigen Verteilung der zufällig ausgewählten Kreditbewerber führen. So ist eine weitere Verzerrung und Verschlechterung des Ergebnisses erklärbar. Tabelle 4.22 zeigt die Nutzung der 12 % zusätzlichen Daten

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	12,08 %	17,90 %	11,58 %	18,40 %	11,70 %	18,29 %
$Y = 1$	12,89 %	57,13 %	13,39 %	56,63 %	13,27 %	56,74 %
\sum Hd.		69,21 %		68,21 %		68,44 %

Tabelle 4.22.: Ergebnis empirische Daten, Kreditvergabe mit weichem Trennwert und Neugewichtung 3

mit der einfachen Gewichtung, die verbesserte Ergebnisse erreicht. Dies zeigt erneut, dass die Multiplikation der Gewichte mit fünf nicht die wahren Eigenschaften der Daten aufzeigen kann und die Ergebnisse nur zusätzlich verzerrt.

Extrapolation

Für die Extrapolation erfolgte eine Unterteilung der selektierten Stichprobe $A = 1$ in mehrere Bänder. Das arithmetische Mittel der Wahrscheinlichkeiten $p(S)$ der in den Bändern enthaltenen Fälle wurde gebildet. Den Mittelwerten ist anschließend

die relative Häufigkeit guter Fälle des jeweiligen Bandes zugeordnet worden. Die so erhaltenen Wertepaare bildeten die Grundlage für die Berechnung der Parameter einer linearen Regression, mit denen die Wahrscheinlichkeit guter Fälle im Bereich $A = 0$ extrapoliert wurde. Die extrapolierten Wahrscheinlichkeiten dienten als Grundlage für die Simulation des Status Y für die abgelehnten Kreditbewerber. Die abgelehnten Fälle konnten so mit der Stichprobe $A = 1$ zur Modellbildung beitragen. Die erreichten Ergebnisse für die Teststichprobe sind in Tabelle 4.23 dargestellt. Der hier erzielte Klassifikationserfolg ist nicht so überzeugend wie bei

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	12,14 %	17,89 %	11,63 %	18,40 %	11,89 %	18,14 %
$Y = 1$	12,89 %	57,08 %	13,41 %	56,56 %	13,15 %	56,82 %
\sum Hd.	69,22 %		68,19 %		68,71 %	

Tabelle 4.23.: Ergebnis empirische Daten, Extrapolation 1

den simulierten Daten. Eine Ursache mag die schlechtere Anpassung der linearen Regression an die vorliegende Anteilsverteilung guter Fälle sein. Eine weitere Fehlerquelle ist möglicherweise der Simulationsvorgang des Kreditstatus im Zusammenhang mit der geringen Fallanzahl des Bereiches $A = 0$. Zum Ausgleich von Verzerrungen erfolgte die Simulation des Kreditstatus dreimal. Die jeweils erhaltenen Parameter des Logit-Modells wurden gemittelt und erst dann für die Teststichprobe verwendet. Tabelle 4.24 zeigt für die verwendete Anpassung eine Verbesserung des Klassifikationserfolges.

	voll		selektiert		Methode	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	12,14 %	17,89 %	11,63 %	18,40 %	12,00 %	18,03 %
$Y = 1$	12,89 %	57,08 %	13,41 %	56,56 %	13,04 %	56,93 %
\sum Hd.	69,22 %		68,19 %		68,93 %	

Tabelle 4.24.: Ergebnis empirische Daten, Extrapolation 2

Vergleich Kreditvergabe mit Neugewichtung und Extrapolation

Da die hier erreichten Ergebnisse nicht so gut vergleichbar waren wie bei den simulierten Daten (sehr stabil), werden in diesem Abschnitt die zwei am besten funktionierenden Methoden, Kreditvergabe mit weichem Trennwert und die Extrapolation nochmals an einem Datensatz verglichen. Die beiden Methoden unter-

	voll		selektiert	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,96 %	18,36 %	11,41 %	18,92 %
$Y = 1$	13,05 %	56,63 %	13,60 %	56,07 %
\sum Hd.	68,59 %		67,48 %	
	Methode 1		Methode 2	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11,81 %	18,51 %	11,77 %	18,55 %
$Y = 1$	13,20 %	56,48 %	13,24 %	56,44 %
\sum Hd.	68,29 %		68,21 %	

Tabelle 4.25.: Ergebnis empirische Daten, Vergleich Kreditvergabe mit weichem Trennwert (Methode 1) und Extrapolation (Methode 2)

scheiden sich kaum in ihren Ergebnissen. Da die Extrapolation keine zusätzlichen Fälle und somit Risiken für das erreichte Ergebnis aufnimmt, ist diese Methode der Kreditvergabe mit weichem Trennwert und Neugewichtung vorzuziehen.

4.3. Der Selektionseffekt und seine Einflussgrößen

Bei den durchgeführten Untersuchungen ist aufgefallen, dass das Modell der selektierten Stichprobe durchaus bessere Zuordnungserfolge für die Teststichprobe erzielte, als das Modell, was mit Hilfe der vollen Stichprobe geschätzt wurde (siehe Abschnitt 4.2.1). Ähnliche Effekte konnten bei Separationsmaßen wie der Trennschärfe festgestellt werden. Die Trennschärfe der selektierten Stichprobe kann unter Umständen auch Werte annehmen, die größer sind, als die der vollen Stichprobe. Ursachen und Einflussgrößen für diese Beobachtung sollen in diesem Abschnitt ebenfalls untersucht werden.

4.3.1. Simulierte Daten - Klassifikationsergebnis

Der Ausgangspunkt für die folgenden Untersuchungen ist das in Kapitel 4.1 vorgestellte Simulationsmodell. Die Trainings- und die Teststichprobe enthalten jeweils 1000 Fälle, von denen ca. 25 % schlechte Kredite sind. Der Trennwert c wurde so gewählt, dass 75 % aller Bewerber einen Kredit erhielten. Die volle Stichprobe und die selektierte Stichprobe dienen zur Schätzung eines Logit-Modells mit dessen Parametern die Bewertung der Teststichprobe erfolgte. Die 25 % schlechtesten Scores der Teststichprobe werden dabei als abgelehnt $A = 0$ betrachtet. Der Vergleich des für die Teststichprobe erreichten Klassifikationserfolges fand mittels Differenzbildung statt. Die Simulation wurde hundertmal durchgeführt. In den folgenden Abbildungen sind die Simulationsanzahl und die sortierten Differenzen dargestellt. Negative Differenzen bedeuten, dass das Modell der selektierten Stichprobe bessere Zuordnungserfolge als das Modell der vollen Stichprobe erreicht.

Stichprobengröße

Für die folgende Untersuchung ist die Größe der Stichprobe verändert worden, während alle anderen Einflussfaktoren konstant blieben. Die Größenänderung erfolgte für Trainings- und Teststichprobe simultan. Bei kleinen Stichprobengrößen

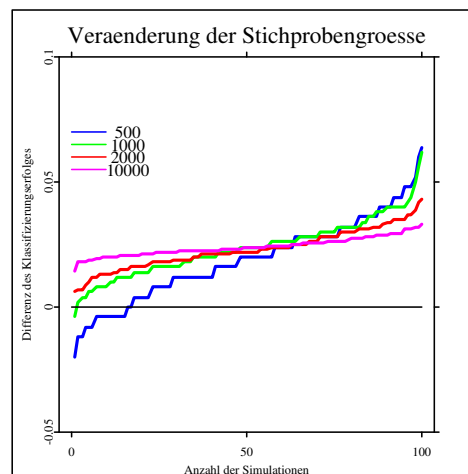


Abbildung 4.5.: Selektionseffekt und Stichprobengröße

(500 und 1000) kommt es häufiger vor, dass das aus der selektierten Stichprobe geschätzte Modell bessere Ergebnisse bei der Teststichprobe erreicht, als das

Modell der vollen Stichprobe. Deutlich wird, dass mit steigender Fallanzahl die Varianz der Differenzen des Klassifizierungserfolges sinkt. Die Mittelwerte der Differenzen scheinen dagegen gleich zu bleiben. Die in Tabelle 4.26 dargestellten Mittelwerte bestätigen dies bis auf eine Ausnahme. Für den Simulationslauf mit 500 Fällen liegt der beobachtete Mittelwert weit unter dem der anderen Simulationsläufe. Eine Erklärung dafür ist, dass bei geringer Fallanzahl die Span-

Stichprobengröße	Mittelwert der Abweichung
500	1,92
1000	2,38
2000	2,31
10000	2,41

Tabelle 4.26.: Selektionseffekt und Stichprobengröße, Mittelwerte der Abweichung in Prozentpunkten

ne der möglichen Variablenausprägungen nicht mehr ausgenutzt (ausgeleuchtet) wird. Die auf der selektierten Stichprobe geschätzten Modellparameter können nun für die Fälle der Teststichprobe zufällig besser oder schlechter als die Parameter des vollen Modells sein. Bei großen Fallanzahlen (erhöhte Ausnutzung der Ausprägungspanne der Variablen) kann dieser Effekt nicht mehr so stark auftreten und die Differenz des Klassifizierungserfolges stabilisiert sich. Zusammenfassend kann gesagt werden: Bei großer Fallanzahl ist die Varianz für die Differenz des Klassifizierungserfolges sehr gering. Es muss, trotz möglicher Stichprobenvariation, mit einem stabilen Verzerrungseffekt gerechnet werden. Bei einer geringen Anzahl von Fällen ist der Verzerrungseffekt weniger stabil (steigende Varianz) und das Modell der selektierten Stichprobe kann bei der Klassifizierung der Teststichprobe durchaus besser abschneiden. Die hier angeführte Begründung kann auch die beobachtete bessere Klassifikation für das Modell der selektierten Stichprobe bei den auf den Daten von Fahrmeir und Hamerle (1984) durchgeführten Simulationen erklären. Die Parameter der selektierten Stichproben sind aufgrund kleiner Stichprobengrößen zufällig besser für die Bewertung der Teststichprobe geeignet als die Parameter der vollen Stichprobe.

Trennwert c

In diesem Abschnitt erfolgte für die Trainingsstichprobe eine Variation des Trennwertes c . Die Trennung für die Teststichprobe nach den 25 % niedrigsten Scorerwerten wurde beibehalten. Alle anderen Faktoren blieben unverändert. Abbildung 4.6 zeigt für die Teststichprobe die Differenzen des Klassifikationserfolges in Abhängigkeit der in der Trainingsstichprobe zugelassenen Kreditanzahl. Werden in

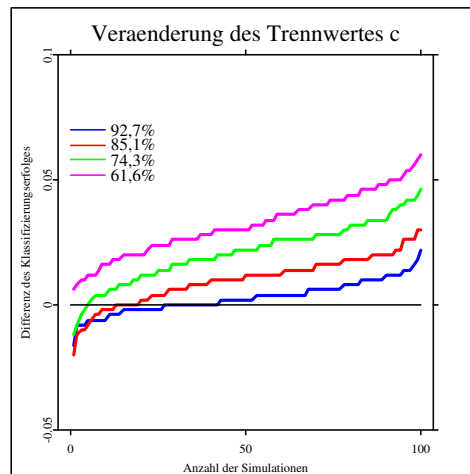


Abbildung 4.6.: Selektionseffekt und Trennwert c

der Trainingsstichprobe mehr Kredite genehmigt, sinkt der Unterschied des Klassifikationserfolges. Bei Akzeptanz aller Kreditbewerber entstehen bedingt keine Differenzen. Mit Erhöhung der Anzahl abgelehnter Kredite steigt der beobachtete Unterschied an. Die aufgetretenen Schwankungen sind auf die Menge der verwendeten Fälle (Informationen), die in der selektierten Trainingsstichprobe zur Verfügung stehen, zurückzuführen. Eine größere Anzahl Fälle bedeutet die Minderung der Differenz des Zuordnungserfolges, eine geringere Fallanzahl bewirkt das Gegenteil.

Ausfallhäufigkeit

An dieser Stelle ist die Ausfallhäufigkeit verändert worden. Der Zufallsgenerator wurde dabei so angepasst, dass auf Basis derselben wahren Scores mehr oder weniger Fälle ausfielen. Die beschriebene Anpassung wurde für Trainings- und

Teststichprobe gleichermaßen durchgeführt. Abbildung 4.7 zeigt die Differenzen des Klassifikationserfolges in Abhängigkeit des Anteils guter Kreditbewerber. Für

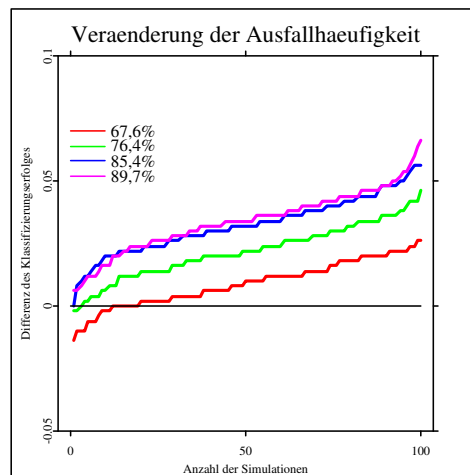


Abbildung 4.7.: Selektionseffekt und Veränderung der Ausfallhäufigkeit

diese Daten zeigt sich, dass bei einem geringeren Anteil guter Fälle die Differenzen des Zuordnungserfolges sinken und bei steigendem Anteil anwachsen. Die Varianz der Differenzen bleibt davon unberührt. Eine Erklärung für die beobachteten Unterschiede ist, dass bei kleineren Anteilen guter Fälle und gegebenem Trennwert für die Modellbildung mehr Informationen bezüglich ausgefallener Kredite zur Verfügung stehen. Bei geringen Ausfallhäufigkeiten und gegebenem c werden nur sehr wenig schlechte Kredite in die selektierte Stichprobe gelangen. Die mit dieser Stichprobe gebildeten Modelle erreichen für die Teststichprobe nur stark verzerrte Ergebnisse.

Im Zusammenhang mit der Festlegung des Trennwertes scheint die Ausfallhäufigkeit für die Modellbildung in der Praxis ein erhebliches Problem darzustellen. Stehen aufgrund des Trennwertes nur Stichproben mit einem geringen Anteil schlechter Kredite zur Verfügung, werden die neu bestimmten Modellparameter für die Population aller Bewerber, im Vergleich zu Modellparametern der vollen Stichprobe, eine stark verschlechterte Performance erzielen.

Die beschriebene Veränderung der Ausfallhäufigkeit kann auch zur Darstellung der Veränderung ökonomischer Verhältnisse verwendet werden. Das bedeutet, dass

Modelle unter guten ökonomischen Verhältnissen geschätzt und unter schlechten Verhältnissen getestet werden und umgekehrt. Abbildung 4.8 zeigt, dass eine

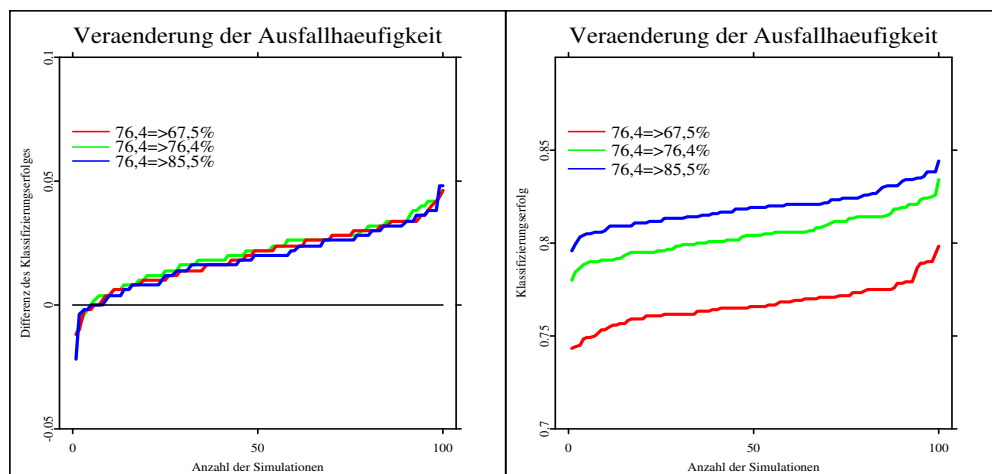


Abbildung 4.8.: Selektionseffekt und variable Ausfallhäufigkeit

Veränderung der Ausfallhäufigkeit (dargestellt ist der Anteil der guten Fälle) über die Zeit keinen Einfluss auf die Differenz des Klassifikationserfolgs hat. Trotz allem existiert ein verzerrender Einfluss, der auf das Modell der vollen Stichprobe und das Modell der selektierten Stichprobe gleichzeitig wirkt. Im rechten Teil von Abbildung 4.8 ist dies für das Modell der vollen Stichprobe dargestellt. Sinkt die Häufigkeit guter Fälle bei sonst konstanten Parametern, verringert sich der Klassifikationserfolg und umgekehrt. Neben dem in dieser Arbeit diskutierten Problem der Verzerrung aufgrund von Stichprobenselektion existieren also auch andere Faktoren, die auf die Leistungsfähigkeit von Kreditscoringmodellen einen Einfluss haben.

Veränderung der ersten Scorefunktion

Im Folgenden soll der Einfluss der ersten Scorefunktion auf den Selektionseffekt untersucht werden. Dabei wurden die in Tabelle 4.27 dargestellten Scorefunktionen verwendet. Da sich für die Scorefunktionen unterschiedliche Wertebereiche von $p(S)$ ergaben und der festgelegte Trennwert c die Stichprobe in unterschiedliche Größen unterteilte, wurde für die Trainingsstichprobe eine einheitliche Trennung nach den 25 % niedrigsten Scores festgelegt.

Score- funktion	Spezifikation	Anteil $Y = 1$ in $A = 0$	Anteil $Y = 1$ in $A = 1$
Score 1	$\eta = 0.2x_2 - 0.8x_3 - 0.6x_4$	42,6 %	87,6 %
Score 2	zufällige Zuordnung	76,0 %	76,6 %
Score 3	$\eta = -0.2x_2 - 0.8x_3 + 0.6x_4$	58,9 %	82,3 %
Score 4	$\eta = -0.4x_2 - 0.4x_3 + 0.6x_4$	69,9 %	78,5 %

Tabelle 4.27.: Selektionseffekt und Veränderung der ersten Scorefunktion

Die Funktion Score 3 ist die stets verwendete und in Kapitel 4.1.1 vorgestellte Scorefunktion. Score 2 entspricht einer zufälligen Zuordnung der einzelnen Fälle zu den Gruppen $A = 0$ und $A = 1$. Aufgrund der Art der Zuteilung ist keine Ergebnisverzerrung zwischen dem Modell der selektierten und der vollen Stichprobe zu erwarten. Das bestätigte sich auch in der Praxis (siehe Abbildung 4.9). Die Parameter von Score 1 entsprechen annähernd den aus einer vollen Stichprobe erreichbaren unverzerrten Parametern. Trotz der Stichprobenselektion unterliegen die geschätzten Modellparameter kaum einer Verzerrung. Mit Score 4 ist eine Scorefunktion dargestellt, für die der Klassifikationserfolg der Teststichprobe noch verzerrter ist als die primär benutzte Funktion Score 3. Die Scorefunktion erreicht dabei für die Trainingsstichprobe immer noch einen besseren Zuordnungserfolg als die zufällige Zuordnung (Score 2). Eine weitere Ursache für den auf Stichprobenselektion beruhenden Effekt liegt für diese Daten in den verzerrten Parametern der ersten Scorefunktion. Eine erste Scorefunktion mit wahren Parametern, führt für das Modell der selektierten Stichprobe bei der Bewertung der Teststichprobe zu fast unverzerrten Ergebnissen. Eine Erklärung für die fehlende Verzerrung kann sein, dass die unverzerrte selektierte Stichprobe und die normalverteilten Daten es möglich machen, ebenfalls unverzerrte Parameter aus dieser Stichprobe zu schätzen, die für die Teststichprobe gleiche Ergebnisse wie für das Modell der vollen Stichprobe erreichen.

Um zu zeigen, dass das unverzerrte Ergebnis aufgrund korrekter Parameter in der ersten Scorefunktion nicht für jede Datenstruktur erreichbar ist, wurden für den Datensatz aus Fahrmeir und Hamerle (1984) ähnliche Untersuchungen durchgeführt. Für jede Simulation erfolgte die Bestimmung der ersten Scorefunktion aus

allen Fällen der Trainingsstichprobe. Die Parameter sind jetzt ebenso unverzerrt wie bei den simulierten Daten. Die weitere Simulationsdurchführung blieb unverändert. Für die erste Scorefunktion konnte ein Klassifikationserfolg von 69,9 %, gegenüber den erzielten 68,14 % der ersten Scorefunktion aus Abschnitt 4.2.1, erreicht werden. Für die Teststichprobe wurde mit dem Modell der vollen Stichprobe (Aufgrund der getroffenen Annahme, entspricht dieses Modell der ersten Scorefunktion) 69,13 % und mit dem Modell der selektierten Stichprobe ein Klassifikationserfolg von 68,1 % erreicht. Im Unterschied zu den simulierten Daten konnte der Selektionseffekt nicht durch eine unverzerrte erste Scorefunktion verringert werden. An dieser Stelle spielen für die Entstehung des Selektionseffektes auch andere Faktoren, wie z. B. fehlende Informationen, eine Rolle.

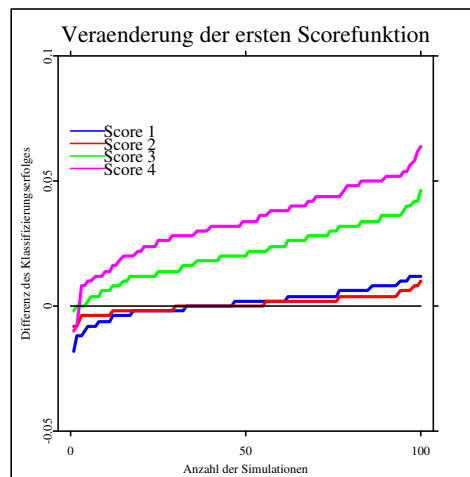


Abbildung 4.9.: Selektionseffekt und Veränderung der ersten Scorefunktion

4.3.2. Simulierte Daten - Trennschärfe

Die folgenden Darstellungen basieren auf den von Kraft, Kroisandt und Müller (2002) durchgeführten Simulationen. Sie zeigen den Verzerrungseffekt für das Maß der Trennschärfe, der auch schon theoretisch aufgezeigt wurde, für die Verteilungen von S , gegeben $Y = 0$ und $Y = 1$. Der Verzerrungseffekt, der bisher anhand des unterschiedlichen Klassifikationserfolges in der Teststichprobe dargestellt wurde, wird hier anhand der Verteilungen vor und nach der Stichprobenselektion (Kreditvergabe) nachgewiesen. Ziel des Abschnittes soll es sein, auch für das Maß der

Trennschärfe den Einfluss verschiedener Faktoren auf den Verzerrungseffekt zu untersuchen und eventuelle Unterschiede aufzuzeigen.

Die Scorewerte der Kreditbewerber wurden aus einer Normalverteilung mit den Parametern $\mu = -1.5$ und $\sigma^2 = 2.25$ gewonnen. Die Bestimmung der zugehörigen Wahrscheinlichkeiten erfolgte wieder mit Hilfe einer Logit-Spezifikation

$$p(S) = \frac{e^S}{1 + e^S}.$$

Anhand dieser Wahrscheinlichkeiten ist mit einem vorher festgelegten Trennwert c die Gruppe der Kreditbewerber in abgelehnte und angenommene Fälle unterteilt worden. Mit den Werten von $p(S)$ erfolgte ebenfalls die Simulation des Kreditstatus. Der Wert der Trennschärfe wurde für die gesamte Stichprobe und für die selektierte Stichprobe ($A = 1$) berechnet. Die so beschriebene Simulation ist hundertmal durchgeführt worden. Die Bestimmung der Scorewerte aus der definierten Normalverteilung und die Trennung aufgrund des Wertes von c fand dabei für jede Simulation neu statt. Als Ausgangspunkt wurden ähnlich, wie in den vorigen Abschnitten, 1000 Fälle mit ca. 25 % Ausfällen und einer Trennung nach den 25 % niedrigsten Scores verwendet. Die Trennschärfe für die volle (grün) und die selektierte Stichprobe (rot) ist in Abbildung 4.10 dargestellt.

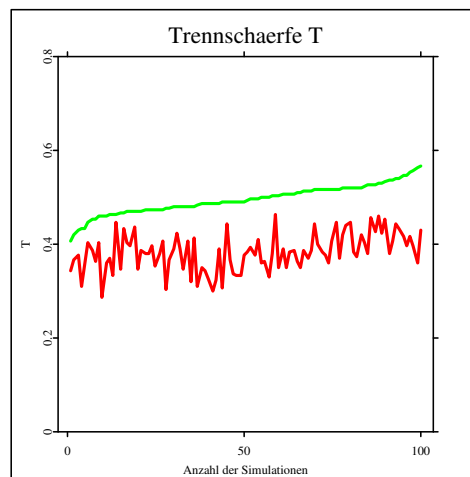


Abbildung 4.10.: Selektionseffekt und Trennschärfe

Im Folgenden ist die Beeinflussung der Stichprobengröße, des Trennwertes c und der Ausfallhäufigkeit dargestellt. Zur besseren Verdeutlichung des Stichprobenselektionseffektes ist analog zur Vorgehensweise des vorangegangenen Abschnittes die Differenz der Trennschärfe gebildet und sortiert dargestellt worden.

Stichprobengröße

Die Berechnung der Trennschärfedifferenzen erfolgte für Stichproben mit 250, 500, 1000 (Standard), 2000 und 10000 Fällen. Die erzielten Ergebnisse der durchgeführten Simulationen sind in Abbildung 4.11 dargestellt. Wie bei den zuvor

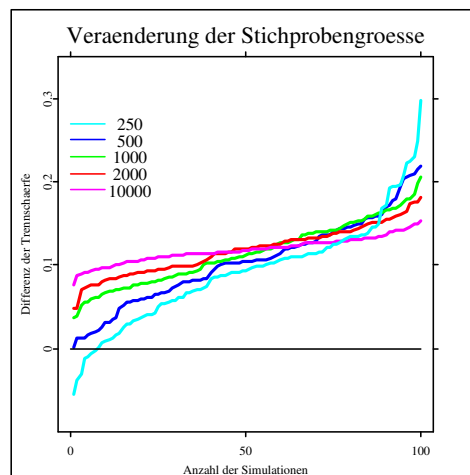


Abbildung 4.11.: Selektionseffekt und Trennschärfe (Veränderung der Stichprobengröße)

durchgeführten Simulationen konnte für eine steigende Anzahl von Fällen festgestellt werden, dass sich die Varianz der Differenzen mindert. Mit dem Selektionseffekt ist bei großen Fallanzahlen sicherer und in ähnlicher Ausprägung zu rechnen. Für kleine Fallanzahlen vermindert sich unter einer sehr großen Ergebnisstreuung der Mittelwert der Trennschärfedifferenzen.

Trennwert c

Durch die Festlegung des Trennwertes c wurden unterschiedliche Größen der selektierten Stichprobe erzeugt (siehe Abbildung 4.12). Eine Vergrößerung der se-

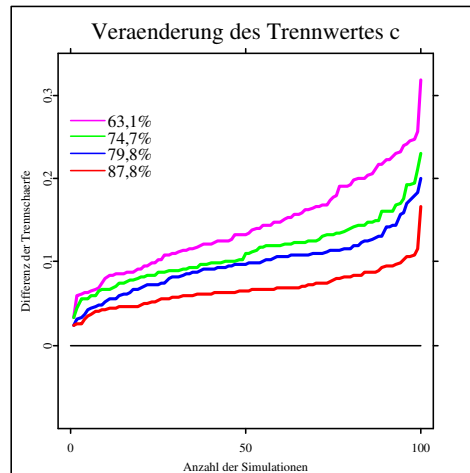


Abbildung 4.12.: Selektionseffekt und Trennschärfe (Veränderung des Trennwertes c)

lektierten Stichprobe, hin zur vollen Stichprobe, bewirkt ebenfalls eine Minderung der Verzerrung, die mit einer Erhöhung der verfügbaren Menge an Information zu erklären ist.

Ausfallhäufigkeit

In Abbildung 4.13 werden die Differenzen der Trennschärfe in Abhängigkeit des Anteils guter Fälle dargestellt. Sinkt dieser Anteil, steigt die Größe der Verzerrung. Bei Vergrößerung des Anteils guter Fälle, fallen die Trennschärfedifferenzen. Die Kurve für den Simulationslauf mit 89,7 % guten Fällen soll zeigen, dass ab einem bestimmten Anteilswert die Differenzen nicht mehr parallel verlaufen. Die Streuung der Differenzen steigt sehr stark an.

Für den Klassifikationserfolg und die Trennschärfe wurden gleichzeitig das Verhalten bei unterschiedlichen Stichprobengrößen, die Anpassung des Trennwertes und die Veränderung der Ausfallhäufigkeit untersucht. Allgemein lässt sich sagen, dass bei großen Stichproben die auf Verzerrung beruhenden Differenzen deutlich und mit sehr geringen Schwankungen hervortreten. Eine durch den Trennwert bewirkte Vergrößerung der selektierten Stichprobe, hin zur vollen Stichprobengröße, vermindert folgerichtig den Verzerrungseffekt. Für die Veränderung der Ausfallhäufigkeit konnten nur konträre Bewegungen festgestellt werden. Die Ur-

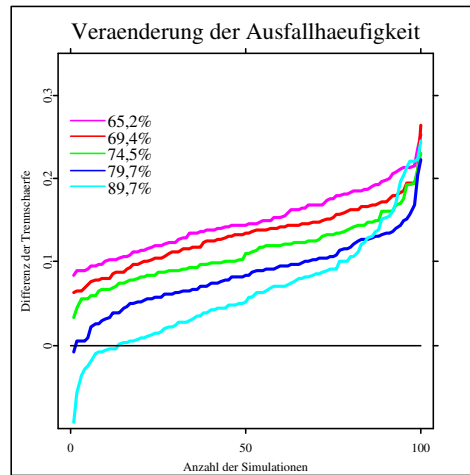


Abbildung 4.13.: Selektionseffekt und Trennschaerfe (Veränderung der Ausfallhäufigkeit)

sache für diese Beobachtung kann in den unterschiedlichen Herleitungen bzw. Definitionen von Klassifikationserfolg und Trennschaerfe liegen.

5. Zusammenfassung

In dieser Arbeit ist das Problem der Stichprobenselektion im Zusammenhang mit der Bildung und Entwicklung von Kredit Scoringsystemen untersucht worden. Es wurden Methoden zur Minderung des Selektionseffektes vorgestellt und an praktischen Beispielen angewendet. Die einfache Neugewichtung und Re-Klassifizierung erscheinen für die Minderung der bewirkten Verzerrung ungeeignet. Diese Methoden werden ohne genauere Kenntnisse über das Ausfallverhalten der abgelehnten Kreditbewerber keine sicheren Verbesserungen erzielen können. So erreichten sie, für die verwendeten Beispiele, eher zufällig Ergebnissteigerungen. Die Vergabe zusätzlicher Kredite, mittels wahrscheinlichkeitsbedingter Aufnahmeschwelle, konnte im Verhältnis zur zusätzlichen Datenmenge eine Minderung der Verzerrung erreichen. Mittels der durchgeführten Kombination von weicher Annahmeschwelle und Gewichtung ist das erzielte Ergebnis weiter verbessert worden. Trotz der Verbesserungen dürfen hier die zusätzlich aufgenommenen Risiken und die damit verbundenen Kosten nicht außer Acht gelassen werden. Es stellt sich die Frage, ob die Verbesserung des Klassifikationserfolges, die höhere Risikoaufnahme rechtfertigt. Die Extrapolation der empirischen Ausfallwahrscheinlichkeiten, mit vorgeschlagener Simulation des Kreditstatus, erreichte ebenfalls gute Erfolge bei der Minderung des Stichprobenselektionseffektes. Bei dieser Methode sind keine zusätzlichen Fälle und damit verbundene Risiken erforderlich. Für verbesserte Ergebnisse muss mit den extrapolierten Wahrscheinlichkeiten eine gute Approximation der wahren Wahrscheinlichkeiten erzielt werden. Bei der Anwendung dieser Methoden in der Praxis ergeben sich verschiedene Probleme: Der Kreditstatus der abgelehnten Kreditbewerber ist nicht wahrnehmbar. Dieses Problem wurde hier durch entsprechende Bildung bzw. Verwendung der Daten umgangen. Die Spanne für mögliche Verbesserungen ist sehr klein. Für den in der Praxis nur partiell erkennbaren Klassifizierungserfolg ist sie entsprechend geringer.

Die Untersuchung von Einflussfaktoren und deren Wirkung auf den Selektionseffekt ergab folgende Ergebnisse: Bei in der Praxis relevanten großen Kreditportfo-

lios ist, abhängig von anderen Faktoren, bei der Modellbildung sicher mit einer Verzerrung aufgrund von Stichprobenselektion zu rechnen. Vergrößert man die selektierte Stichprobe durch Veränderung des Trennwertes c , erhält man mehr Informationen über die Population der Kreditbewerber und kann die Verzerrung mindern. Es konnte weiterhin festgestellt werden, dass durch hohe Trennwerte c bewirkte realitätsnahe geringe Ausfallhäufigkeiten, der selektierten Stichprobe, einen negativen Einfluss (große Verzerrung) auf die Bildung neuer Modelle haben können. Die durch Veränderung der ersten Scorefunktion erzielten Ergebnisse lassen folgende Schlüsse zu: Bereits verzerrte Parameter können zu einer falschen Sortierung und somit wieder zu verzerrten Modellparametern führen. Die Struktur des Datensatzes und der Variablenausprägungen spielt ebenfalls eine hier nicht näher untersuchte Rolle. Für die aufgrund von Stichprobenselektion verzerrten Werte der Trennschärfe konnten für die Stichprobengröße und die zusätzliche Datenmenge ähnliche Ergebnisse wie für den Klassifikationserfolg ermittelt werden.

Der Wechsel der ökonomischen Verhältnisse im Zeitablauf bewirkte keine zusätzliche Verzerrung zwischen dem Modellergebnis, der vollen und der selektierten Stichprobe. Diese Veränderung hatte jedoch einen Einfluss auf den absoluten Klassifikationserfolg der Modelle. Es zeigt sich, dass neben den in dieser Arbeit untersuchten Verzerrungen, aufgrund von Stichprobenselektion, auch andere Faktoren auf den Erfolg von Kredit Scoringsystemen einwirken. Neben dem Wandel ökonomischer Verhältnisse wären hier beispielsweise das Übergehen des Scoringsystems durch Bankmitarbeiter und Veränderungen in der Kreditbewerberpopulation, bedingt durch Konditionsänderungen etc. (siehe auch Phillips und Yezer, 1996) zu nennen. Obwohl durch Stichprobenselektion prozentual gesehen nur geringe Verschlechterungen des Klassifikationserfolges bewirkt werden, können diese, auch mit der Wirkung anderer Faktoren, einen erheblichen Einfluss auf das finanzielle Ergebnis eines Kreditportfolios haben. Mit dem Ziel der Gewinnmaximierung und aufgrund des durch Basel II postulierten Anspruchs, erkennbare Verzerrungen zu vermeiden, sollten ergebnisbeeinflussende Faktoren, wie die Stichprobenselektion von kreditvergebenden Institutionen nicht vernachlässigt sondern mittels geeigneter Methoden bei der Modellbildung berücksichtigt werden.

A. Logit- und Probit-Modelle

Bei den hier vorliegenden Daten ist die abhängige Variable Y (Kreditrückzahlung) binär.

$$Y = \begin{cases} 1 & \text{Kredit zurückgezahlt} \\ 0 & \text{Kredit nicht zurückgezahlt} \end{cases}$$

Dem Fall $Y = 1$ wird die Wahrscheinlichkeit π und der Ausprägung $Y = 0$ die Wahrscheinlichkeit $1 - \pi$ zugeordnet. Die Wahrscheinlichkeitsverteilung von Y ergibt sich aus:

$$f(y, \pi) = P(Y = y) = \pi^y(1 - \pi)^{1-y}, \quad y = 0, 1.$$

Die so beschriebene Verteilung besitzt den Erwartungswert $E(Y) = \pi$ und die Varianz $Var(Y) = \pi(1 - \pi)$. Im Kontext des Kreditscorings sind die Ausfall- bzw. Rückzahlungswahrscheinlichkeiten eines jeden Kreditbewerbers von Interesse. Die Binomialverteilung der Fälle in Abhängigkeit der erklärenden Variablen X lässt sich wie folgt schreiben:

$$P(Y_i = y_i | X_i) = f(y_i, \pi_i | X_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, \quad y_i = 0, 1.$$

Für die Darstellung des Zusammenhangs zwischen den Wahrscheinlichkeiten π und den Variablen X ist eine Bestimmung der Parameter β_j mittels linearer Regression ungeeignet. Ein wichtiger Grund dafür sind die nicht normal sondern binomial verteilten Störterme ε . Geeigneter für die Darstellung des Zusammenhangs erscheinen die Spezifikationen des Logit- oder des Probit-Modells, die im Folgenden näher beschrieben werden.

Das Logit-Modell ist durch

$$\pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} = \frac{1}{1 + e^{-\eta_i}} \quad (\text{A.1})$$

gegeben und stellt den Zusammenhang zwischen der Wahrscheinlichkeit π_i und der Funktion $\eta_i = X_i^\top \beta_j$ her. Löst man Formel (A.1) nach e^{η_i} auf, erhält man:

$$e^{\eta_i} = \frac{\pi_i}{1 - \pi_i}. \quad (\text{A.2})$$

Der Bruch

$$\frac{\pi_i}{1 - \pi_i}$$

stellt dabei das Verhältnis der Wahrscheinlichkeit des Erfolgs zur Wahrscheinlichkeit des Misserfolgs dar. Im Allgemeinen wird dieser Bruch auch als odds (Chance des Erfolgs) bezeichnet. Durch Logarithmierung des Ausdrucks (A.2) erhält man die gesuchte Link-Funktion des Logit-Modelles

$$\eta_i = \log \left(\frac{\pi_i}{1 - \pi_i} \right). \quad (\text{A.3})$$

Das Logit-Modell ist somit durch

$$\eta_i = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = X_i^\top \beta = \sum_j x_{ij} \beta_j \quad (\text{A.4})$$

spezifiziert.

Für das Probit-Modell wird der Zusammenhang zwischen π und η durch

$$\pi_i = \Phi(\eta_i) \quad (\text{A.5})$$

dargestellt. Die Funktion Φ entspricht der Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$. Die Linkfunktion des Probit-Modells ergibt sich aus den Werten der inversen Verteilungsfunktion Φ^{-1} für die Wahrscheinlichkeiten π_i . Dar-

aus folgt das Probit-Modell:

$$\eta_i = \Phi^{-1}(\pi_i) = X_i^\top \beta = \sum_j x_{ij} \beta_j \quad (\text{A.6})$$

Für die Schätzung der Parameter β_j wird die Maximum-Likelihood-Methode verwendet. Die Ableitungen nach den Parametern führt zu einem nichtlinearen Gleichungssystem, welches mit Hilfe iterativer Verfahren (z.B. Newton-Raphson-Algorithmus) gelöst werden kann.

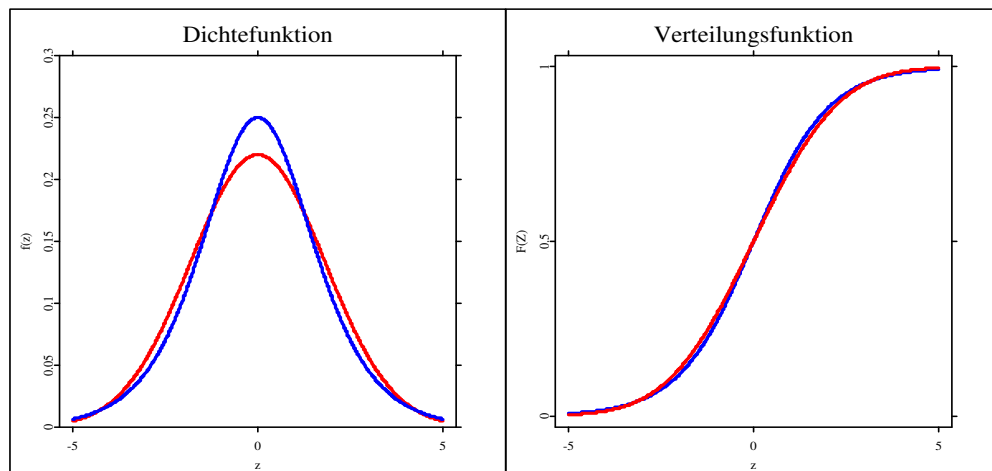


Abbildung A.1.: Logistische Verteilung (blau) gegen Normalverteilung $N(0, \pi^2/3)$ (rot)

Die für das Logit- und Probit-Modell erreichten Schätzergebnisse sind sehr ähnlich. Das soll durch die gemeinsame Darstellung von logistischer Verteilung und Normalverteilung in Abbildung A.1 aufgezeigt werden. Da die Varianz der logistischen Verteilung $\pi^2/3$ beträgt, wird für die Darstellung eine Normalverteilung mit der gleichen Varianz verwendet. Beide Verteilungen unterscheiden sich leicht an ihren Enden (Abbildung A.2). Die Dichte der logistischen Verteilung liegt hier über der Dichte der Normalverteilung. (vergleiche Rönz, 2001 sowie Hosmer und Lemeshow, 1989).

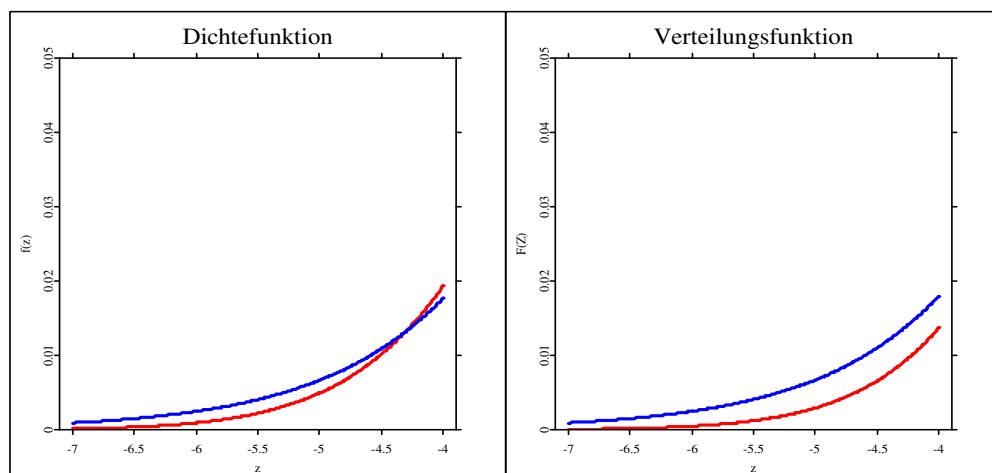


Abbildung A.2.: Logistische Verteilung (blau) gegen Normalverteilung $N(0, \pi^2/3)$ (rot) (linkes Ende von Dichte und Verteilung)

B. Variablenbeschreibung des Datensatzes aus Fahrmeir und Hamerle (1984)

In diesem Abschnitt werden die für das Scoring verwendeten Variablen des Fahrmeir und Hamerle (1984) entnommenen Datensatzes näher beschrieben. Da Variablen mit kategorialen Ausprägungen vorliegen und diese nicht direkt in einer logistischen Regression verwendet werden können, wurden sie auf binäre Variable umcodiert. Soweit eine Umwandlung erfolgte, ist sie in Form einer Tabelle näher beschrieben.

- Variable x_1 : Darlehenshöhe in DM ($250 \leq x_1 \leq 18424$)
- Variable x_2 : Laufzeit des Kredites in Monaten ($4 \leq x_2 \leq 72$)
- Variable x_3 : Alter des Antragstellers in Jahren ($19 \leq x_3 \leq 75$)
- Variable x_4 : bisherige Zahlungsmoral (Tabelle B.1)

Ausprägung alt		Ausprägung neu	
1	keine Kredite bisher/ alle Kredite ordnungsgemäß zurückgezahlt	0	keine Probleme
2	bei der Bank frühere Kredite einwandfrei abgewickelt		
3	bei der Bank noch bestehende Kredite bisher einwandfrei		
4	früher zögernde Kreditführung	1	Probleme
5	kritisches Konto/ es bestehen weitere Kredite bei anderer Bank		

Tabelle B.1.: Variablenbeschreibung x_4

- Variable x_5 : weiter Ratenkredite (Tabelle B.2)

Ausprägung alt		Ausprägung neu	
1	bei einer anderen Bank	0	weitere Ratenkredite
2	bei einem Kaufhaus		
3	keine	1	keine weiteren Ratenkredite

Tabelle B.2.: Variablenbeschreibung x_5

- Variable x_6 : Sparkonto oder Wertpapiere (Tabelle B.3)

Ausprägung alt		Ausprägung neu	
5	nicht ermittelbar / kein Sparkonto	0	nicht ermittelbar, geringes Guthaben
1	< 100 DM		
2	$100 \leq \dots < 500$ DM	1	Guthaben ab 100 DM
3	$500 \leq \dots < 1000$ DM		
4	≥ 1000 DM		

Tabelle B.3.: Variablenbeschreibung x_6

- Variable x_7 : Wohnung (Tabelle B.4)

Ausprägung alt		Ausprägung neu	
1	Mietwohnung	0	Mietwohnung
2	Eigentumswohnung	1	Eigentum, kostenlos
3	kostenlose Wohnung		

Tabelle B.4.: Variablenbeschreibung x_7

Literaturverzeichnis

- Ash, D. and S. Meester. (2002). *Best practices in reject inferencing*, Presentation at credit risk modelling and decisioning conference, Wharton Financial Institutions Center, Philadelphia, <http://fic.wharton.upenn.edu/fic/ash.pdf>.
- Basel Committee on Banking Supervision. (2003). *The new basel capital accord*, Bank for International Settlements, <http://www.bis.org>.
- Banasik, J., J. Crook, and L. C. Thomas. (2001). *Sample selection bias in credit scoring models*, Working Paper Series No. 01/5, Credit Research Centre, University of Edingburgh.
- Boyes, W. J., D. L. Hoffman, and S. A. Low. (1989). *An econometric analysis of the bank credit scoring problem*, Journal of Econometrics **40**, 3–14.
- Crook, J. and J. Banasik. (2002). *Does reject inference really improve the performance of application scoring models?*, Working Paper Series No. 02/3, www.crc.ems.ed.ac.uk/workingpapers/workingpaper02-3.pdf.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. (1977). *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society Series B **39**, 1–38.
- Durand, D. (1941). *Risk elements in consumer installment financing*, National Bureau of Economic Research, New York.
- Fahrmeir, L., A. Hamerle. (1984). *Multivariate statistische Verfahren*, Walter de Gruyter, Berlin.
- Feelders, A. J. (2000). *Credit scoring and reject inference with mixture models*, International Journal of Intelligent Systems in Accounting, Finance & Management **9**, 1–8.
- Greene, W. H. (1993). *Econometric Analysis*, 2nd ed., Prentice Hall.

- Greene, W. H. (1998). *Sample selection in credit-scoring models*, Japan and the World Economy **10**, 299–316.
- Hand, D. J. and W. E. Henley. (1997). *Statistical classification methods in consumer credit scoring: a review*, Journal of the Royal Statistical Society Series A **160**, 523–541.
- Hand, D. J. (2002). *Measurement and prediction models in consumer credit*, Presentation at credit risk modelling and decisioning conference, Wharton Financial Institutions Center, Philadelphia, <http://fic.wharton.upenn.edu/fic/hand.pdf>.
- Heckman, J. (1979). *Sample selection bias as a specification error*, Econometrica **47**, 153–161.
- Horowitz, J. L. and C. F. Manski. (1998). *Censoring of outcomes and regressors due to survey nonresponse: Identification using weights and imputations*, Journal of Econometrics **84**, 37–58.
- Hosmer, D. W. and S. Lemeshow. (1989). *Applied logistic regression*, John Wiley & Sons, Inc., New York.
- Kraft, H., G. Kroisandt, and M. Müller. (2002). *Assessing the discriminatory power of credit scores*, Diskussionspapier, Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM) und Humboldt-Universität zu Berlin.
- Little, R. J. A. and D. B. Rubin. (1987). *Statistical analysis with missing data*, John Wiley & Sons, Inc., New York.
- Manski, C. F. and S. R. Lerman. (1977). *The estimation of choice probabilities from choice based samples*, Econometrica **45**, 1977–1988.
- Meester, S. (2000). *Reject inference for credit scoring model development using extrapolation*, Mimeo, New Jersey: CIT Group.
- Meng, C. L. and P. Schmidt. (1985). *On the cost of partial observability in the bivariate probit model*, International Economic Review **26**, 71–85.
- Neubauer, W. (1994). *Statistische Methoden*, Franz Vahlen, München.
- Phillips, R. F. and A. M. Yezer. (1996). *Self-selection and tests for bias and risk in mortgage lending: can you price the mortgage if you don't know the process?*, The Journal of Real Estate Research **11**, 87–102.
- Poirier, D. J. (1980). *Partial observability in bivariate probit models*, Journal of Econometrics **12**, 210–217.

- Rönz, B. (2001). *Generalisierte lineare Modelle*, Skript zur Vorlesung, Humboldt-Universität zu Berlin.
- Rosenberg, E. and A. Gleit. (1994). *Quantitative Methods in credit management: a survey*, Operations Research **42**, 589–613.
- Thomas, L. C. (2000). *A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers*, International Journal of Forecasting **16**, 149–172.

Erklärung zur Urheberschaft

Hiermit erkläre ich, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Thomas Parnitzke

Berlin, 30. Juli 2003