

Kalibrierung von Ratingsystemen – eine erste Analyse

Luise Breinlinger,
Evgenia Glogova,
Andreas Höger¹)

I Einleitung

Im Gesamtprozess des Aufbaus bzw. der fortlaufenden Wartung eines Ratingsystems stellt sich zwangsläufig auch die Frage nach einer adäquaten Kalibrierung des Systems. Unter anderem als Folge der Umsetzung der neuen Basler Eigenkapitalvereinbarung (Basel II) wird dieser Problemstellung in näherer Zukunft zunehmende Aufmerksamkeit gewidmet werden. Basis der vorliegenden Untersuchung bilden von der Auskunftsei Creditreform zur Verfügung gestellte Bonitätsdaten im Umfang von rund 10.000 Datensätzen für jedes der Jahre 1996 bis 2001. Im Rahmen einer Erstbeschäftigung mit der Kalibrierung beschränkt sich unsere Arbeit auf statische Methoden, das heißt, die Schätzungen der Ausfallwahrscheinlichkeiten beruhen auf jeweils einjährigen Migrationen, und auch die Einteilung der Ratingklassen bezieht sich im Prinzip auf jeweils einen Zeitpunkt. Die durch intertemporale Betrachtung entstehende Dynamik findet vorerst nur in konzeptionellen Überlegungen ihren Niederschlag. Insgesamt liegt der Fokus der Arbeit auf methodischen Aspekten. Von den Resultaten der Analysen ist zu erwähnen, dass bei statischer Betrachtungsweise die Ceteris-paribus-Erhöhung der Anzahl an Ratingklassen bei den verwendeten Kalibrierungsmethoden das Eigenmittelerfordernis sinken lässt. Spätestens jedoch bei intertemporaler Betrachtungsweise wird der maximal möglichen Anzahl an Ratingklassen durch die Forderung nach Monotonie in der

Struktur der Ausfallwahrscheinlichkeiten eine natürliche Grenze gesetzt. Die Notwendigkeit der intertemporalen Modellierung (sowie einer ausreichend langen Datenhistorie) wird auch durch die Sensitivität der Eigenmittel in Abhängigkeit von schwankenden Ausfallraten untermauert.

Der vorliegende Beitrag ist folgendermaßen gegliedert: Abschnitt 2 skizziert die Datenbasis. Die Kapitel 3.1 und 3.2 sind den empirischen Kalibrierungsanalysen gewidmet: zum einen, wenn Ausfallwahrscheinlichkeiten auf Basis relativer Häufigkeiten geschätzt werden (Kapitel 3.1), zum anderen, wenn Ausfallwahrscheinlichkeiten mittels logistischer Regression geschätzt werden (Kapitel 3.2). Integrative Überlegungen bilden den Inhalt von Kapitel 3.3, während Abschnitt 4 mit einer Zusammenfassung und einem Ausblick die Arbeit abschließt.

2 Datenbasis

Die Creditreform stellte der Oesterreichischen Nationalbank (OeNB) zweimal Daten zur Verfügung: Einmal für den Zeitraum von 31. Dezember 2000 bis 31. Dezember 2001 und einmal für den Zeitraum von 31. Dezember 1996 bis 31. Dezember 2001 (Auswertungen zu den Stichtagen am jeweiligen Jahresende). Die beiden Datensätze wurden zu unterschiedlichen Zeitpunkten (im Juli bzw. im August 2002) aus der Datenbank gezogen und sind für den gemeinsamen Zeitraum, 31. Dezember 2000 bis 31. Dezember 2001, nicht direkt miteinander vergleichbar: Stichproben-

¹ Die Autoren sind Mitarbeiter der Abteilung für Bankenanalyse und -revision der Oesterreichischen Nationalbank. Die in diesem Beitrag geäußerten Meinungen sind jene der Autoren und sollten nicht als Sicht der Oesterreichischen Nationalbank interpretiert werden.

Für Kommentare und Diskussionen danken wir Walter Schwaiger, für die Unterstützung im Rahmen der Datenaufbereitung Gerhard Fiam und Wolfgang Schüller.

artige Überprüfungen haben ergeben, dass sich zwischen den beiden Ziehungen aus der Datenbank Scores und andere Werte (zum Teil auch rückwirkend) geändert haben; es gab Neueinträge, Löschungen etc. Schließlich wurden die Daten der Creditreform um Daten der Großkreditevidenz (GKE) der OeNB erweitert.

Die für die Studie verwendeten Datensätze weisen die folgende Struktur auf:

Datensatz 1: 9.752 Beobachtungen, unter anderem mit den Merkmalen Firmenbuchnummer, Postleitzahl (Region), Branchen-Code (Zuordnung der Creditreform), Umsatz zum 31. Dezember 2001, Bonitätsindex (Score) der Creditreform zum 31. Dezember 2000 und zum 31. Dezember 2001, GKE-Daten (vor allem Kreditausnutzung) zum 31. Dezember 2001. Von diesen 9.752 zum 31. Dezember 2000 solventen Unternehmen waren im Lauf des Jahres 2001 196 Ausfälle zu verzeichnen, was eine durchschnittliche Ausfallwahrscheinlichkeit von rund 2% für dieses Jahr ergibt.

Datensatz 2: 10.273 Beobachtungen mit denselben Merkmalen wie bei Datensatz 1 für den Zeitraum von 31. Dezember 1996 bis 31. Dezember 2001. Die Umsatzzahlen dieses Datensatzes bestehen aus Zahlen zum 31. Dezember 2000 und zum 31. Dezember 2001.

Wesentlicher Bestandteil beider Datensätze ist der Bonitätsindex der Creditreform. Dieser Score gibt die Bonitätsbeurteilung eines Unternehmens durch die Creditreform an. Der Wert kann dabei zwischen 100 (beste Bonität) und 600 (schlechteste Bonität, Ausfall) liegen und setzt sich aus 15 Kriterien zusammen. Diese Kriterien umfassen unter anderem Zahlungsweise, Krediturteil, Unternehmensentwicklung, Auftragslage

und Lage der Branche. Zahlungsweise und Krediturteil bilden mit einer Gewichtung von zusammen rund 50% den Schwerpunkt des Bonitätsindex.

3 Durchgeführte Untersuchungen und Ergebnisse

3.1 Klassenbildung auf Basis relativer Ausfallhäufigkeiten

Unter Klassenbildung auf Basis relativer Ausfallhäufigkeiten (im Folgenden kurz als Häufigkeitsanalyse bezeichnet) wird die Konstruktion der Ratingklassen anhand von Zählvorgängen verstanden, die direkt auf dem Bonitätsindex (Score) basieren. Dabei werden die Unternehmen dem Score nach größenmäßig geordnet und danach in eine bestimmte Anzahl von Klassen (z. B. Klasse 1: Score 100 bis 200 etc.) eingeteilt. Eine Möglichkeit hierbei liegt darin, die Anzahl an Unternehmen je Klasse annähernd konstant zu halten (Unternehmensgleichverteilung). Die Operationalisierung erfolgt z. B. durch folgende Vorgabe: Bei zehn Ratingklassen sind pro Ratingklasse 10% der Unternehmen zuzuteilen, bei fünf Ratingklassen 20% etc. Diese Vorgehensweise wird im Folgenden aufgegriffen. Eine andere Variante ist jene, die in der Arbeit von Lawrenz und Schwaiger (2002), dem Ausgangspunkt der vorliegenden Studie, verwendet wurde: Für jede Ratingklasse ist hierbei vorzugeben, welcher Anteil an den Gesamtausfällen der jeweiligen Ratingklasse zuzuordnen ist (vorgegebenes Aufklärungsprofil). Als Schätzer der Ausfallwahrscheinlichkeit dient bei beiden Methoden die relative Häufigkeit der Ausfälle der betreffenden Klasse.

Es wurden die folgenden Untersuchungen durchgeführt, wobei jeweils (das gilt für alle durchgeführten Berechnungen) auf Basis des Formel-

werks vom Oktober 2002 gerechnet wurde: Foundation-IRB-Ansatz mit 45% LGD („loss given default“) und einem Abschlag für kleine und mittlere Unternehmen (KMUs) in Abhängigkeit vom Umsatz des Unternehmens, nachdem die Datenbasis daraufhin entsprechend gefiltert worden war. Es wurde die Kreditausnutzung gemäß GKE als Proxy für die tatsächliche Kreditanspruchnahme verwendet, und es wurden die Ausfallwahrscheinlichkeiten (in Ermangelung einer ausreichenden Datenbasis) entgegen der Basler Eigenkapitalvereinbarung (siehe Kapitel 3.3) jeweils auf Basis eines nur einjährigen Beobachtungszeitraums ermittelt.

Berechnung 1: Berechnung des Eigenmittelerfordernisses und der PD-Struktur (der Struktur der Ausfallwahrscheinlichkeiten – „probabilities of default“ – eines Ratingsystems) bei Variation der Zahl der Ratingklassen unter sonst möglichst unveränderten Bedingungen: Ziel ist die Analyse der Auswirkungen einer systematischen Variation der Zahl der Ratingklassen auf die beiden genannten Größen Eigenmittelerfordernis und PD-Struktur. Verwendet wird Datensatz 1.

Berechnung 2: Berechnung von Eigenmittelerfordernis und Prüfung der PD-Struktur im Zeitablauf anhand von Datensatz 2: Auch hier erfolgt die Berechnung der Ausfallwahrscheinlichkeiten auf Basis eines jeweils einjährigen Beobachtungszeitraums.

Um Vergleichbarkeit zu gewährleisten wurde die Methode der Unternehmensgleichverteilung gewählt: Die Variation der Zahl der Ratingklassen wurde dabei also unter der Bedingung durchgeführt, dass annähernd gleich viele Unternehmen in jeder Klasse vorkommen. Bei Verwendung eines vorgegeben Aufklärungsprofils zur Klassenbildung (anstelle der Un-

ternehmensgleichverteilung) müsste in einem ersten Schritt hingegen erst definiert werden, wie eine Ceteris-paribus-Variation der Anzahl der Ratingklassen überhaupt konsistent durchgeführt werden kann.

Zu Berechnung 1: Es wurden das Eigenmittelerfordernis und die PD-Struktur für 5, 7, 10, 12 und 15 Klassen untersucht. Zum einen zeigt sich, dass das Eigenmittelerfordernis mit zunehmender Zahl von Ratingklassen kontinuierlich sinkt, nämlich von 6'18% (5 Klassen) auf 5'88% (15 Klassen).

Zum anderen ergibt sich, dass die PD-Struktur nur bei einem System von fünf Klassen monoton steigend ist (die Ausfallwahrscheinlichkeiten steigen von der besten zur schlechtesten Klasse monoton an), ab einer Zahl von sieben oder mehr Klassen aber keine Monotonie mehr zu beobachten ist. Das bedeutet, dass die Zahl der Ratingklassen ceteris paribus nicht beliebig gesteigert werden kann, ohne bestimmte wünschenswerte Eigenschaften der PD-Struktur, wie z. B. die Monotonie, zu verlieren.

Zu Berechnung 2: Anhand von Datensatz 2 wurde überprüft, wie sich Eigenmittel und PD-Struktur im Zeitablauf verhalten. Es wurden jeweils sieben Klassen gebildet, und zwar wiederum mit annähernd gleich vielen Unternehmen je Klasse. Aus der Gesamtzahl an Beobachtungen wurden für jedes Jahr individuell die relevanten Datensätze gefiltert, also Unternehmen, die zu Beginn des untersuchten Jahres insolvent waren, wurden jeweils eliminiert etc. Pro Jahr stehen daher unterschiedlich viele Datensätze zur Verfügung: von 6.137 Beobachtungen für 1996/97 steigend auf bis zu 9.419 Beobachtungen für 2000/01. Die durchschnittliche Ausfallwahrscheinlichkeit steigt

Tabelle 1

Eigenmittelerfordernis im Zeitablauf					
	1996/97	1997/98	1998/99	1999/2000	2000/01
	<i>Anzahl</i>				
Unternehmen	6.176	6.883	7.653	8.527	9.419
	<i>in %</i>				
Eigenmittelerfordernis	4'68	5'32	5'40	5'29	5'86
Ausfallrate	1'16	1'23	1'33	1'37	1'93

Quelle: Creditreform, OeNB, eigene Berechnungen.

im gesamten Beobachtungszeitraum von 1'16% im Jahr 1996/97 auf 1'93% im Jahr 2000/01¹).

Da für die Berechnung des Eigenmittelerfordernisses im Zeitablauf keine Umsatzzahlen von der Creditreform für die Jahre 1997, 1998 und 1999 vorhanden waren, wurde der KMU-Abschlag auf Basis der Umsatzzahlen von 2001 konstant über alle Jahre hinweg gerechnet. Als Schätzwerte für Kreditinanspruchnahmen wurden wiederum GKE-Daten (Kreditausnutzung) für die jeweiligen Jahre herangezogen. Für das Eigenmittelerfordernis im Zeitablauf ergibt sich das folgende Bild (siehe Tabelle 1).

Anzumerken ist, dass die Entwicklung des Eigenmittelerfordernisses den wesentlichen Anstieg der durchschnittlichen Ausfallraten im Beobachtungszeitraum abbildet: In den Jahren 1997/98 bis 1999/2000 sind Ausfallraten zwischen 1'2 und 1'4% zu verzeichnen. Das Eigenmittelerfordernis bewegt sich in diesem Zeitraum zwischen rund 5'3 und 5'4%. Im Vergleich dazu ist die Ausfallrate im Jahr 2000/01 wesentlich höher, nämlich 1'93%, und auch das Eigenmittelerfordernis ist mit 5'86% entsprechend höher als in den Jahren zuvor. Gleiches gilt mit umgekehrten Vorzeichen für das Jahr 1996/97 im Vergleich zu den folgenden Jahren oder ergibt sich aus einem direkten

Vergleich der Jahre 1996/97 und 2000/01, wobei die Zahlen für 1996/97 auf Grund des geringeren Umfangs der Stichprobe mit Vorsicht zu interpretieren sind.

Das durchschnittliche Eigenmittelerfordernis über die fünf beobachteten Jahre beläuft sich auf 5'31%, bei einer über diesen Zeitraum gerechneten durchschnittlichen Ausfallrate von 1'41%. Wenn man das Jahr 1996/97 für diese Berechnungen weglässt, beläuft sich das durchschnittliche Eigenmittelerfordernis auf 5'47% bei einer durchschnittlichen Ausfallrate von 1'47%.

Zu den Bewegungen der PD-Struktur bei Anwendung der Häufigkeitsanalyse ist anzumerken, dass die Form im Zeitablauf schwankt. Einheitlich ist nur, dass die Ausfallwahrscheinlichkeiten ab Klasse 5 in allen Jahren monoton steigend sind. In den besseren Klassen ist keine im Zeitablauf durchgängige Struktur zu verzeichnen. Als Schlussfolgerung daraus lässt sich sagen, dass mit der Methode der Häufigkeitsanalyse die Eigenschaft der Monotonie einer PD-Struktur, gerechnet auf Basis eines einjährigen Beobachtungszeitraums, zwar einmal erzeugt werden kann, aber ceteris paribus im Zeitablauf wieder verloren gehen wird.

Dies gilt sowohl für die unter Berechnung 1 und Berechnung 2 gewählte Methode der Unternehmens-

¹ Die Abweichungen der Datensätze 1 und 2 betreffend das Jahr 2000/01 ergeben sich wie angeführt aus den unterschiedlichen Abfragezeiträumen aus der Creditreform-Datenbank.

gleichverteilung als auch bei Verwendung eines vorgegebenen Aufklärungsprofils: Eine im Zeitablauf durchgehende Monotonie wird auch bei Vorgabe des Aufklärungsprofils nur dann zu erreichen sein, wenn das Aufklärungsprofil jedes Jahr in Abhängigkeit von der Datenlage geändert wird, nicht jedoch mit tatsächlich als konstant vorgegebenen Aufklärungsschritten. Letzteres würde aber auch ein stetes Ändern der (Score-) Grenzen der Ratingklassen erfordern, was wiederum das erratische Wandern mancher Unternehmen durch verschiedene Ratingklassen verstärken würde, und zwar unabhängig von der spezifischen wirtschaftlichen Lage des betreffenden Unternehmens.

3.2 Logistische Regression

Bei der logistischen Regression wird die Abhängigkeit zwischen einer linearen Kombination von Einflussfaktoren $\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ und der abhängigen Variablen, welche nur zwei Werte annehmen kann (Ausfall/kein Ausfall), modelliert. Weiters ist es wichtig, dass die logistische Funktion den Index $\beta'x$ auf das Intervall $[0,1]$ abbildet und somit jedem Score-Wert, ausgedrückt durch $\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$, einen Wert, der als Ausfallwahrscheinlichkeit interpretiert werden kann, zuweist (n ist die Anzahl der Unternehmen, x ist der Vektor der unabhängigen Variablen, β ist der Vektor der Parameter). Der Index $\beta'x$ kann als Maß für die Bonität eines Kreditnehmers dienen.

Für eine nicht direkt beobachtbare Variable y^* gilt:

$$y^* = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

ε sei logistisch verteilt, wobei der Mittelwert 0 und die Varianz 1 der Verteilung keine einschränkenden Annah-

men sind. Die Verteilungsfunktion der logistischen Verteilung ist:

$$\Lambda(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}$$

Folgendes können wir direkt beobachten:

$$y = 1 \text{ wenn } y^* > 0 \\ y = 0 \text{ wenn } y^* < 0$$

Die Wahrscheinlichkeit, dass $y = 1$ ist somit

$$P(y = 1) = P(y^* > 0) = \\ = P(\beta_0 + \beta_1 x_1 + \dots + \varepsilon > 0) = \\ = P(\varepsilon > -\beta_0 - \beta_1 x_1 - \dots - \beta_n x_n) = \\ = P(\varepsilon < \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n) = \\ = \Lambda(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)$$

da die logistische Verteilung symmetrisch ist.

Für die Schätzung des Koeffizientenvektors $\boldsymbol{\beta}$ wird die Maximum-Likelihood-Optimierungsmethode verwendet. Bei dieser Methode wird die Wahrscheinlichkeit (L), dass das geschätzte Modell die beobachteten y -Werte reproduziert, maximiert. Die logistische Funktion Λ gibt die Ausfallwahrscheinlichkeit an (siehe weiter oben), während $1 - \Lambda$ die Überlebenswahrscheinlichkeit charakterisiert.

$$MaxL = P(Y_1 = y_1, \dots, Y_n = y_n) = \\ = \prod_{y_i=0} [1 - \Lambda(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)] \\ \prod_{y_i=1} \Lambda(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)$$

Die oben beschriebene logistische Regression wird eingesetzt, um jedem Unternehmen eine Ausfallwahrscheinlichkeit zuzuordnen, Klassen zu bilden und das Eigenmittelerfordernis für Kredite an österreichische KMUs zu berechnen, wobei die Vorgehensweise in vier Unterpunkten wie folgt zusammengefasst ist:

3.2.1 Auswahl der erklärenden Variablen
anhand von Datensatz 1 im Sinne der
Maximierung der Robustheit und des
Informationsgehalts des Modells

Modellspezifikation (MS) 1: Als Erklärende wurden eine Konstante und der Bonitätsindex aufgenommen. Danach wurde untersucht, ob sich beim Einbeziehen anderer uns zur Verfügung stehender Variablen mit potenziellem Einfluss auf die Bonität eines Kreditnehmers die Qualität des Modells erhöht.

Modellspezifikation 2: Konstante, Bonitätsindex und ln(Umsatz 2001) als Approximation für die Größe des Unternehmens.

Modellspezifikation 3: Konstante, Bonitätsindex und anhand von Dummy-Variablen die Bundeslandzugehörigkeit.

Modellspezifikation 4: Konstante, Bonitätsindex, ln(Umsatz 2001) und anhand von Dummy-Variablen die Bundeslandzugehörigkeit.

Die Branchenzugehörigkeit wurde auch anhand von Dummy-Variablen und Datensatz 2 getestet. Durch den Einsatz der gleichen Maximum-Likelihood-Optimierungsmethode konnte zu keiner Lösung für die Parameter gelangt werden. Ein Grund dafür ist,

dass Creditreform den Effekt der Branchenzugehörigkeit als Risikokennzahl im Scoring-Modell schon berücksichtigt hat, womit das explizite Einbeziehen von Branchen-Dummies überflüssig wird und zu einer schlechten Modellspezifikation führt.

Anhand der t-Statistik, welche angibt, ob ein Koeffizient signifikant verschieden von 0 ist („Trennkraft“), und ihrem p-Wert (die Wahrscheinlichkeit diesen t-Wert zu beobachten), wird ersichtlich, dass nicht nur der Score sondern auch ln(Umsatz 2001) einen hohen Informationsgehalt besitzt. Die gemeinsame Signifikanz der Bundesländer-Dummies wurde mit Hilfe des Wald-Tests (Greene, 1993) geprüft: die χ^2 -Statistik ergibt 1'12, wobei der kritische Wert der χ^2 -Verteilung mit acht Freiheitsgraden auf dem 95-Prozent-Signifikanzniveau 15'50 beträgt, und damit die Hypothese, dass die Koeffizienten der Bundesländer-Dummies alle gemeinsam 0 sind, nicht verworfen werden kann. Die Bundesländereffekte sind von Creditreform zwar nicht im Bonitätsindex berücksichtigt (eine gewisse Signifikanz ist deshalb bei den obigen Schätzergebnissen zu beobachten), eine bundesländerspezifische Modell-

Tabelle 2

**Schätzergebnisse von Datensatz 1: Koeffizienten mit den t-Werten
und (in Klammern) p-Werten**

	Konstante	Score 2000	ln(Umsatz 2001)	Wien	Nieder- österreich	Tirol	Ober- österreich	Salzburg	Vorarlberg	Steiermark	Burgenland	Eigenmittel- erfordernis in %
MS 1	- 9'85 - 24'84 (0'0)	0'02 17'2 (0'0)	6'04
MS 2	14'23 - 9'63 (0'0)	0'02 17'14 (0'0)	0'26 3'12 (0'002)	6'50
MS 3	10'38 - 20'43 (0'0)	0'02 16'83 (0'0)	0'73 2'34 (0'02)	0'60 1'78 (0'08)	0'68 2'00 (0'05)	0'66 2'00 (0'05)	0'25 0'71 (0'48)	- 0'41 - 1'09 (0'28)	1'00 2'35 (0'02)	0'13 0'39 (0'70)	..	5'94
MS 4	14'27 - 9'37 (0'0)	0'02 16'7 (0'0)	0'23 2'75 (0'006)	0'67 2'13 (0'03)	0'58 1'70 (0'09)	0'69 2'02 (0'04)	0'64 1'95 (0'05)	0'25 0'70 (0'48)	- 0'39 - 1'03 (0'30)	1'00 2'34 (0'02)	0'11 0'33 (0'74)	6'27

Quelle: Creditreform, OeNB, eigene Berechnungen.

spezifikation bedarf aber noch weiterführender Überlegungen.

Um in weiterer Folge Gütemaße sinnvoll anwenden zu können, muss zuerst die Robustheit des Schätzmodells gewährleistet sein. Die meisten Probleme bezüglich der Robustheit eines Logit-Modells bereitet die Heteroskedastie, welche zur Inkonsistenz der geschätzten Koeffizienten führt (das bedeutet, dass die Präzision, mit welcher der Parameter geschätzt wird, mit zunehmender Stichprobengröße sinkt). Der statistische Test von Davidson und MacKinnon (1993) wurde angewandt um die Hypothese H_0 von Homoskedastie zu testen. Die Ergebnisse daraus zeigen, dass für Modellspezifikationen 1 und 3, H_0 nicht verworfen werden kann: im ersten Fall ist die χ^2 -Statistik 0'08 und im zweiten 14'15, wobei der kritische Wert der χ^2 -Verteilung mit neun Freiheitsgraden auf dem 10-Prozent-Signifikanzniveau 14'68 beträgt. Bei den Modellspezifikationen 2 und 4 kann Heteroskedastie nur auf dem 10-Prozent-Signifikanzniveau verworfen werden.

Somit ergibt sich, dass das beste Modell betreffend Robustheit und Signifikanz der erklärenden Variablen von Modellspezifikation 1 dargestellt wird.

Die Gütetests hingegen treffen keine Aussagen über die Modellanpassung, sondern lediglich über den Informationsgehalt, der in den erklärenden Variablen steckt. Implementiert wurde das Gütemaß von McKelvey und Zavoina (1975):

$$R_{MZ}^2 = \frac{\sum_{i=1}^n (\hat{y}_i^* - \tilde{y}_i^*)^2}{\left(\sum_{i=1}^n (\hat{y}_i^* - \tilde{y}_i^*)^2 - n \right)}$$

wobei n die Anzahl von Beobachtungen, \hat{y}_i^* den geschätzten Wert von y_i^* und \tilde{y}_i^* den entsprechenden Mittel-

wert repräsentieren. Laut Testergebnissen (R_{MZ}^2 beträgt 59%) ist Modellspezifikation 4 jene mit dem größten Informationsgehalt.

Ein anderes Gütemaß ist der Gini-Koeffizient, welcher mit der Hilfe von Gini-Kurven berechnet wird. Zur Ermittlung der Gini-Kurve werden die Unternehmen zunächst nach ihrem Risikogehalt, das heißt nach ihrem Bonitätsindex, geordnet. Für jeden Anteil x der risikoreichsten Unternehmen wird der Anteil der ausgefallenen Unternehmen $y(x)$ bestimmt. Der Gini-Koeffizient ist definiert als die Fläche zwischen der Gini-Kurve des Modells und der Gini-Kurve eines Zufallsmodells (für das Zufallsmodell gilt $y(x) = x$) dividiert durch die Fläche zwischen der Gini-Kurve eines perfekten Modells und der Gini-Kurve eines Zufallsmodells. Je höher also der Wert des Gini-Koeffizienten ist, desto höher ist die Fähigkeit des Modells zur Diskriminierung der Unternehmen in Bezug auf ihre Bonität. Die Gini-Koeffizienten variieren von 62'08% für Modell 1 bis 66'56% für Modell 3.

Auffällig ist, dass die Modelle, welche ln(Umsatz 2001) als Erklärende beinhalten, schlechter spezifiziert sind, aber einen höheren Informationsgehalt aufweisen. Auch aus den Schätzergebnissen der Modellspezifikation 1 (siehe Tabelle 3) ist ersichtlich, dass der durchschnittliche Umsatz pro Ratingklasse grundsätzlich sinkt, wenn sich die Ausfallwahrscheinlichkeit erhöht. Diese Resultate deuten darauf hin, dass entweder die Unternehmensgröße einen Einfluss auf die Ausfallwahrscheinlichkeit hat – diese ist aber in dem Creditreform-Scoring schon berücksichtigt worden – oder aber möglicherweise andere Größenindikatoren wie etwa die Zahl der Beschäftigten hier geeigneter wären.

Tabelle 3

Ergebnisse anhand von Modellspezifikation 1 und Datensatz 1

	Ausfallwahrscheinlichkeit in %	Ausfallquote	Durchschnittlicher Umsatz in EUR	Durchschnittliche Ausnutzung
Ratingklasse 1	0'303	0'359	12.369.335	661.056
Ratingklasse 2	0'454	0'574	10.978.920	447.701
Ratingklasse 3	0'553	0'431	655.008	225.184
Ratingklasse 4	0'737	0'790	582.990	285.106
Ratingklasse 5	1'298	0'790	767.034	343.341
Ratingklasse 6	2'323	2'728	527.660	198.157
Ratingklasse 7	8'402	8'399	394.803	173.116

Quelle: Creditreform, OeNB, eigene Berechnungen.

3.2.2 Kalibrierung und Klassenbildung

Unsere Kalibrierungsannahme ist, dass unsere Stichprobe das gleiche Ausfallniveau hat wie die Grundgesamtheit, das heißt, dass sie ausreichend groß und repräsentativ ist.

Die Klassenbildung beim logistischen Ansatz erfolgt wiederum unter der Vorgabe, jeder Ratingklasse annähernd die gleiche Anzahl von Unternehmen zuzuweisen. Unser Ziel war nämlich die Berechnung des Eigenmittelerfordernisses entsprechend den Basler Anforderungen für Monotonie sowie für eine ausreichende Anzahl von Unternehmen pro Klasse, aber eben jeweils immer nur für ein Jahr. Das heißt, es wurde nicht unter Verfolgung des Ziels zeitlicher Stabilität (und damit verbundener Homogenität der Ratingklassen) kalibriert.

Die geschätzten Ausfallwahrscheinlichkeiten im Gegensatz zu den beobachteten relativen Ausfallhäufigkeiten (Ausfallquoten) sind Tabelle 3 zu entnehmen. Eine interessante Beobachtung stellen die in den Spalten vier und fünf enthaltenen sinkenden durchschnittlichen Um-

sätze bzw. Ausnutzungen mit abnehmender Bonität dar.

3.2.3 Variation der Anzahl von Ratingklassen

Für die in Kapitel 3.2.1 definierte Modellspezifikation 1 wurde das Eigenmittelerfordernis für eine verschiedene Anzahl von Klassen – 5, 7, 10, 12 und 15 Klassen – anhand von Datensatz 1 berechnet. Die Klassen wurden gebildet, indem jeder Klasse gleich viele Unternehmen zugeordnet wurden. Zu beobachten ist, dass das Eigenmittelerfordernis von 6'08% im Falle der fünf Ratingklassen auf 5'96% bei 15 Ratingklassen kontinuierlich sinkt.

3.2.4 Fünf-Jahres-Entwicklung

Analog zu Berechnung 2 im Ansatz der Häufigkeitsanalyse wurde anhand von Datensatz 2 überprüft, wie sich Eigenmittel und PD-Struktur im Zeitablauf verhalten.

Das Eigenmittelerfordernis schwankt zwischen 5'29% im Jahr 1997 und 6'12% im Jahr 2001, wobei der Durchschnittswert über die fünf Jahre 5'54% beträgt.

Tabelle 4

Fünf-Jahres-Entwicklung – Schätzergebnisse anhand von Datensatz 2

	1996/97	1997/98	1998/99	1999/2000	2000/01
Eigenmittelerfordernis in %	5'29	5'66	5'44	5'19	6'12

Quelle: Creditreform, OeNB, eigene Berechnungen.

3.3 Integrative Überlegungen

3.3.1 Relevante Mindestanforderungen von Basel II und deren Implikationen für die vorliegende Arbeit

Als umfassende Prinzipien hinter den IRB-Mindestanforderungen wird die Gewährleistung folgender Punkte seitens der Rating- und Risikoschätzsysteme gesehen:

- eine fundierte Beurteilung der Schuldner- sowie der Transaktionscharakteristika
- eine aussagekräftige Risikodifferenzierung
- hinreichend genaue und konsistente quantitative Risikoschätzungen

3.3.1.1 Ratingstruktur

Im gegenwärtigen Kontext explizit Bezug nehmend auf Forderungen gegenüber Unternehmen, Banken und Staaten muss eine Bank sowohl bei ihrem Schuldner-rating als auch bei ihrem transaktionsbezogenen Rating eine aussagekräftige Verteilung der Forderungen über die Stufen hinweg (ohne übermäßige Konzentrationen) aufweisen – ausreichend um eine entsprechende Risikodifferenzierung zu gewährleisten. Zur Sicherstellung, dass diese Vorgabe erreicht wird, muss eine Bank mindestens sieben Schuldnerstufen für nicht ausgefallene und eine für ausgefallene Schuldner aufweisen. Die Aufsichtsbehörde kann (bei Banken mit Schuldner heterogener Schuldnerqualität) eine stärkere Differenzierung verlangen.

Banken, deren Kreditportfolios auf ein ausgewähltes Marktsegment und eine Bandbreite an Ausfallrisiko konzentriert sind, müssen innerhalb dieser Bandbreite genügend viele Stufen aufweisen, um übermäßige Schuldnerkonzentrationen in einer bestimmten Stufe zu vermeiden. Signifikante Konzentrationen innerhalb einer ein-

zelnen Stufe müssen durch stichhaltige empirische Evidenz dahingehend unterstützt werden, dass die Stufe ein einigermaßen enges PD-Band abdeckt.

Der Erörterung wert ist in diesem Zusammenhang die Basler Vorgabe von mindestens sieben Ratingstufen für nicht ausgefallene Schuldner. Wenn wir diese Mindestanzahl mit den in dem vorliegenden Beitrag erhaltenen Ergebnissen zur Ratingklassenbildung in Verbindung mit der Monotonie in der PD-Struktur vergleichen, wird ersichtlich, dass die Creditreform-Daten für die Klassenbildung auf Basis relativer Ausfallhäufigkeiten bei gleicher Anzahl von Unternehmen je Klasse nur dann eine monoton steigende PD-Struktur aufweisen, wenn auf fünf Ratingklassen abgestellt wird. Dabei ist anzunehmen, dass eine monoton steigende PD-Struktur mit einer – wie in dieser Arbeit gewählt – auf jeweils nur einen Zeitpunkt bezogenen Modellierung eher zu erzeugen sein sollte als mit Hilfe einer dynamischen Betrachtung. Diese für die Creditreform-Daten gemachte Beobachtung kann in folgedessen wohl als Beispiel dafür gesehen werden, dass sich die Anzahl der wählbaren und zu besetzenden Ratingklassen als Funktion der Diskriminierungsfähigkeit des Ratingsystems darstellt und nicht davon losgelöst behandelt werden kann. Der Grad der Auffächerung der Ratingklassen ist durch die Diskriminierungsfähigkeit des Ratingsystems begrenzt.

3.3.1.2 Evaluationshorizont und Stressszenarios

Obwohl der für die PD-Schätzung verwendete Zeithorizont ein Jahr beträgt, müssen Banken für die Ratingzuweisung einen längeren Zeithorizont verwenden. Die PD-Schät-

zungen müssen ein langfristiger Durchschnitt der realisierten Ein-Jahres-Ausfallraten für die Schuldner der betreffenden Stufe sein. Eine Bank darf einen einfachen Durchschnitt der PD-Schätzungen für einzelne Schuldner in einer gegebenen Schuldnerklasse verwenden. Unabhängig davon, ob eine Bank externe, interne oder gepoolte Datenquellen oder eine Kombination der drei verwendet, muss die Länge der für die PD-Schätzungen zu Grunde liegenden Beobachtungsreihe für mindestens eine Quelle mindestens fünf Jahre betragen. Wenn die verfügbare Beobachtungsperiode für eine der drei Quellen einen längeren Zeitraum umfasst und diese Daten relevant sind, muss auf diese längere Periode zurückgegriffen werden. Um ungerechtfertigten Optimismus zu vermeiden, muss die Bank zu ihren Schätzungen eine konservative Spanne hinzuaddieren, deren Ausmaß von der wahrscheinlichen Bandbreite der Schätzfehler abhängt.

Gegeben die Schwierigkeiten bei der Prognose zukünftiger Ereignisse sowie deren Wirkung auf die finanzielle Situation bestimmter Schuldner, muss die Bank eine konservative Perspektive in Bezug auf die prognostizierte Information einnehmen. Wenn ferner nur eine beschränkte Datenbasis verfügbar ist, hat die Bank einen konservativen Bias in ihre Analysen einfließen zu lassen.

Bankinterne Einschätzungen der Leistungsfähigkeit der eigenen Ratingssysteme müssen auf einer langen Datenhistorie basieren und eine Bandbreite fundamentalwirtschaftlicher Bedingungen abdecken – idealerweise einen oder mehrere gesamte Konjunkturzyklen. Ein Schuldner-rating muss die Einschätzung der Bank in Bezug auf Fähigkeit und Bereitschaft

des Schuldners, seine Verpflichtungen trotz ungünstigem wirtschaftlichem Umfeld oder unerwartet eintretender Ereignisse vertragsgerecht zu erfüllen widerspiegeln.

Eine Bank kann dieses Erfordernis erfüllen, indem sie ihre Ratingzuteilungen auf spezifische adäquate Stressszenarios begründet. Alternativ kann die Bank diese Anforderung durch entsprechende Berücksichtigung jener Schuldnercharakteristika, die die Verletzbarkeit des Schuldners durch ungünstige wirtschaftliche Bedingungen sowie unvorhergesehene Ereignisse widerspiegeln, ohne explizite Spezifikation eines Stressszenarios, erfüllen. Die Bandbreite wirtschaftlicher Bedingungen, die bei der Beurteilung einfließen, muss sowohl den gegenwärtigen Bedingungen als auch jenen Bedingungen, die über einen Konjunkturzyklus hinweg innerhalb einer bestimmten Branche bzw. geografischen Region voraussichtlich auftreten werden, entsprechen.

Die Bank muss über einen geregelten Ablauf der Modellvalidierung verfügen, welcher die Überwachung der Modellperformance und -stabilität sowie Vergleiche des Modell-Outputs mit tatsächlichen Realisierungen einschließt. In der Modellvalidierung müssen Out-of-Time- und Out-of-Sample-Tests enthalten sein. Ferner sind jene Umstände anzuführen, unter welchen das Modell nicht wirksam arbeitet.

Da nicht dynamische (im Sinne von intertemporalen) PD-Schätzungen, sondern vielmehr die Analyse jeweils auf einen Zeitpunkt bezogener Aufnahmen Gegenstand der vorliegenden Arbeit war, wurden die diesbezüglichen Regelungen der neuen Basler Eigenkapitalvereinbarung nicht schlagend. Dennoch sei erwähnt, dass die für die PD-Schätzung vorgeschrie-

benen langfristigen Durchschnitte der Ein-Jahres-Ausfallraten nicht mit der Klasseneinteilung nach dem Prinzip eines vorgegebenen Aufklärungsprofils vereinbar sind. Der Unterschied zwischen beiden Ansätzen manifestiert sich im Wesentlichen darin, dass die Basler Regelungen durch die angestrebte Einbeziehung der Ausfallraten über den Konjunkturzyklus hinweg der prozyklischen PD-Entwicklung (und damit einhergehend einer konjunkturell schwankenden Entwicklung des Eigenmittelerfordernisses) entgegenwirken. Dahingegen wird bei vorgegebenem Aufklärungsprofil aus zweierlei Gründen keine im Zeitablauf konstante PD-Struktur sowohl für die Ratingklassen selbst als auch für die einzelnen Unternehmen erzeugt: Einerseits schwanken die PDs der einzelnen Ratingklassen in Abhängigkeit davon, wie viele Kreditnehmer mit welchen realisierten Ein-Jahres-Ausfallraten einer Klasse zugeordnet werden, wobei die Zuordnung, basierend auf der Vorgabe der Reihung durch das Scoring, nach fixierten Schwellenwerten für die kumulierten Ausfallraten erfolgt. Andererseits können Unternehmen im Zeitablauf zwischen angrenzenden Ratingklassen fluktuieren, da die Klassenbildung eben auf die kumulierten Ausfallraten abstellt.

Es sei an dieser Stelle betont, dass die Problemstellung der Modellvalidierung nicht Gegenstand der vorliegenden Untersuchung war und auf Grund der dafür nicht ausreichenden Datenbasis auch nicht hätte sein können. Zweifellos wird der Validierung von Ratingsystemen jedoch bei Entwicklung wie Implementierung eine entscheidende Rolle zukommen.

In diese Richtung weiterführend ist auch die Vorgabe zu sehen, dass eine IRB-Bank für die Bemessung

der Kapitaladäquanz jedenfalls einwandfreie Stresstestprozesse zur Verfügung haben muss. Stresstests haben jene möglichen Ereignisse oder künftigen Änderungen wirtschaftlicher Rahmenbedingungen zu identifizieren, die nachteilige Effekte auf das Kreditexposure der Bank haben könnten, und müssen auch einschätzen, wie gut die Bank solchen Änderungen standhalten könnte.

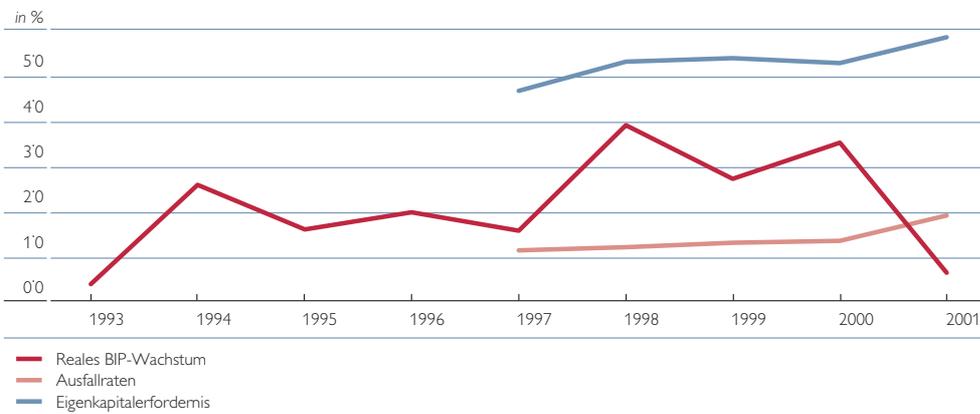
3.3.2 Die makroökonomische Einbettung der Daten

Grafik 1 setzt die Ausfallraten der verwendeten Creditreform-Daten sowohl in Relation zum berechneten Eigenmittelerfordernis, wie es sich nach der Methode der Häufigkeitsanalyse bei Unternehmensgleichverteilung ergibt, als auch zum jährlichen realen BIP-Wachstum. Zur Illustration wurde der Zeitraum von 1993 bis 2001 gewählt, um die Daten in den konjunkturellen Verlauf eingebettet darstellen zu können. Dabei fällt ins Auge, dass die reale Wachstumsreduktion im Ausmaß von rund 3 Prozentpunkten für den Zeitraum von 2000 bis 2001 von einem Anstieg der Ausfallraten um etwa $\frac{1}{2}$ Prozentpunkt begleitet wird, was seinerseits eine entsprechende Erhöhung im berechneten Eigenmittelerfordernis nach sich zieht.

Im Sinne des weiter oben Ausgeführten ist die illustrierte Entwicklung des Eigenmittelerfordernisses aber nicht als Indiz für Prozyklizität zu werten, da die PD-Schätzungen auf Ein-Jahres-Ausfallraten beruhen und somit die von der Basler Eigenkapitalvereinbarung in der aktuellen Fassung für die intertemporale Modellierung vorgeschriebene Glättung über den Konjunkturzyklus hinweg nicht zum Tragen kommt. Aus der Grafik geht jedoch dennoch

Grafik 1

Ausfallraten im Konjunkturzyklus



Quelle: Creditreform, OeNB, eigene Berechnungen.

anschaulich hervor, dass es für den Zeitraum von 2000 bis 2001 ohne diese Glättung zu prozyklischen Effekten in dem Sinne kommen würde, dass die Reduktion des realen BIP-Wachstums über die erhöhten Ausfallraten zu einem gestiegenen Eigenmittelerfordernis führen würde.

4 Zusammenfassung

Die Berechnungen zeigen die Sensitivität des jährlichen Eigenmittelerfordernisses in Abhängigkeit von der jeweiligen durchschnittlichen jährlichen Ausfallrate: Steigt die Anzahl der Ausfälle erheblich, wie im Vergleich des Jahres 2000/01 mit den Jahren zuvor, resultiert das auch in einem deutlich gestiegenen Eigenmittelerfordernis. Um einer unerwünschten prozyklischen Wirkung vorzubeugen, wird daher vom aktuellen Basler Regelwerk vorgeschrieben, die Ausfallwahrscheinlichkeiten je Ratingklasse unter Zugrunde-Legung einer mehrjährigen Datenbasis (zumindest fünf Jahre) zu rechnen. Da die Methode der Häufigkeitsanalyse jedenfalls nicht dazu geeignet ist, im Zeitablauf gesehen stabile sowie monotone PD-Strukturen zu erzeugen, müssen andere Verfahren

zum Einsatz kommen. Vielversprechend in diesem Zusammenhang erscheint die Anwendung der logistischen Regression.

Die Struktur der Ausfallwahrscheinlichkeiten eines Ratingsystems (PD-Struktur) zum einen bei einer Variation der Zahl der Ratingklassen, zum anderen bei einer Prüfung im Zeitablauf zeigt das folgende Bild: Die maximale Zahl der Ratingklassen hängt von der Struktur der zu Grunde liegenden Daten und von der Güte (Trennfähigkeit) des Ratings ab. Je mehr repräsentative Daten in ausreichender Qualität zur Verfügung stehen und je trennschärfer das Rating ist, desto mehr Ratingklassen mit einer monoton steigenden PD-Struktur können gebildet werden.

Arbeitsschritte, die sich im Anschluss an die in der vorliegenden Studie durchgeführten Untersuchungen ergeben würden bzw. nahe lägen, sind

1. die Kalibrierung auf Basis einer längeren Zeitreihe,
2. generell die Validierung von Ratingsystemen sowie
3. die dynamische (intertemporale) Modellierung der PD-Struktur.

Von den drei genannten Punkten ist jedenfalls der große Bereich der Validierung als komplexe Aufgabe anzusehen. Obwohl sich eine vom Basler Ausschuss für Bankenaufsicht eingesetzte Validierungsarbeitsgruppe mit dem Thema beschäftigt, ist anzunehmen, dass auf nationaler Ebene erhebliche Ermessensspielräume zu füllen sein werden.

Literaturverzeichnis

- Davidson, R. und J. G. MacKinnon. 1993.** Estimation and Inference in Econometrics. Oxford: Oxford University Press.
- Greene, W. H. 1993.** Econometric Analysis. New York: Prentice-Hall International.
- Kaiser, U. und A. Szczesny. 2002.** Logit- und Probit-Modelle für Kreditrisiken. In: Schroeder, M. (Hrsg.). Finanzmarkt-Ökonometrie – Basistechniken, Fortgeschrittene Verfahren, Prognosemodelle. Stuttgart: Schäffer-Poeschel.
- Lawrenz, J. und W. S. A. Schwaiger. 2002.** Basel II: Quantitative Impact Study für Österreich (QIS 2.5). In: BankArchiv – Zeitschrift für das gesamte Bank- und Börsenwesen 50. 77–89.
- McKelvey, R. und W. Zavoina. 1975.** A Statistical Model for the Analysis of Ordinal Level Dependent Variables. In: Journal of Mathematical Sociology 4. 103–120.