

**Im «Zeitalter» von Basel II setzen Banken zunehmend statistisch abgestützte Ratingmodelle ein. Diese Ausrichtung auf quantitative Methoden entspricht einem langfristigen Trend im Kreditrisikomanagement. An einem Praxisbeispiel wird gezeigt, wie solche Modelle gebildet werden, welche Kriterien sich für die Ausfallprognose eignen und wie diese zu kombinieren sind.**

---

CHRISTIAN MEIER

---

# RATINGMODELLE BAUEN UND VALIDIEREN

## Ein Ansatz für Schweizer KMU

### 1. MOTIVIERUNG

Während es früher genügte, Schuldner über sogenannte Expertenmodelle mehr intuitiv in Ratingkategorien einzuteilen, sind die Ansprüche an die Messung von Kreditrisiken heute wesentlich höher. Sinkende Zinsmargen sowie die gestiegenen Anforderungen von Basel II an bankinterne Ratingmodelle haben diesen Prozess wesentlich beschleunigt.

Statistische Ansätze führen zu exakteren Ausfallprognosen, weil sie einen optimalen Nutzen aus dem vorhandenen Datenmaterial ziehen. Ohne quantitative Methoden ist eine zuverlässige Aussage schwierig, ob ein Kriterium in einem Modell überhaupt signifikant ist und wie es im Zusammenspiel mit anderen Faktoren zu gewichten ist.

Eine der grössten Herausforderungen für Banken und die Bankenaufsicht ist es, die Performance eines Ratingmodells zu validieren [1]. Auch dies ist ohne eine konsistente Datenbasis und quantitative Ansätze praktisch unmöglich. Kleinere und mittlere Banken gleichen darum ihren Grössennachteil in bezug auf die erforderliche Datenmenge und das Know-how zunehmend über Kooperationen und das Pooling von Kreditdaten aus.

### 2. KREDIT-DATENPOOL

Die Stichprobe stammt aus einem Datenpool, dem gegenwärtig 13 Schweizer Banken angeschlossen sind. In diesem Kontext werden die eingesetzten Ratingmodelle jährlich validiert und aktualisiert. Für die vorliegende Untersuchung wurden rund 2200 Jahresabschlüsse von guten und ausgefallenen Firmenkunden aus dem Zeitraum 1998–2004 benutzt, die von den Banken nach einem einheitlichen Raster erfasst

wurden. Eine Firma gilt als «ausgefallen», wenn bei ihr im Sinne von Basel II ein Ausfallereignis aufgetreten ist (d. h. die Forderung ist mehr als 90 Tage überfällig oder sie wird von der Bank als gefährdet bezeichnet). Dabei wird derjenige Abschluss herangezogen, welcher der Bank mindestens ein Jahr vor dem Ausfallereignis vorgelegen hat. Die Unternehmen weisen Umsätze grösser als CHF 200 000 auf, dürfen aber nicht börsenkotiert sein. Die Stichprobe ist nach Branchen und Rechtsformen gestreut, enthält aber keine Finanzdienstleister, Immobilienfirmen und öffentlich-rechtliche Körperschaften. Bei Firmen mit Konzernabhängigkeit werden nur die konsolidierten Abschlüsse miteinbezogen. Alle Abschlüsse haben einen mehrschichtigen Datenfilter durchlaufen, der unplausible Erfassungen eliminiert.

### 3. FINANZKENNZAHLEN

In der Literatur findet man sehr viele Kennzahlen, die in mindestens einem empirischen Modell vertreten sind. Von zahlreichen Kennzahlen wie z. B. dem Eigenfinanzierungsgrad lassen sich zudem Varianten bilden. Soll man beim Eigenfinanzierungsgrad z. B. nur das ausgewiesene Eigenkapital berücksichtigen oder auch die nachrangigen Darlehen? Und sollten langfristige Darlehen an Aktionäre vom Eigenkapital abgezogen werden? Bereits die Auswahl trennscharfer Kennzahlen gestaltet sich somit zu einem zentralen Schritt bei der Modellierung von Ausfallrisiken. Insgesamt haben wir in diese Untersuchung 110 Finanzkennzahlen miteinbezogen.

Die Prognosekraft der Kennzahlen wurde zunächst mit der *Receiver Operating Characteristic (ROC)* [2] getestet. Die ROC misst die Fähigkeit eines Kriteriums bzw. eines Modells, zwischen Firmen zu differenzieren, die später ausfallen und solchen, die zahlungsfähig bleiben. *Abbildung 1* bringt eine Auswahl von Kennzahlen in eine Rangliste, wobei sich die Trennfähigkeit im Performancemass *AUROC (Area under ROC)* zusammenfassen lässt. Ein AUROC von 1 steht für ein Kriterium, das perfekt zwischen guten und schlechten Schuldnern trennt. Ein AUROC von 0.5 bedeutet, dass man die Bonität ebensogut mit einem Zufallsexperiment bestimmen könnte. In der Realität liegen die sinnvollen Kriterien irgendwo zwischen diesen Extremen. Ein AUROC von 0.82 beim Liquiditätsgrad 2 entspricht z. B. der Wahrscheinlichkeit, mit der ein



CHRISTIAN MEIER,  
DR. OEC. PUBL.,  
MANAGING DIRECTOR,  
RISK CONSULTING  
GROUP, ZÜRICH  
CH.MEIER@RCG.CH

Abbildung 1: **TRENNFÄHIGKEIT VON FINANZKENNZAHLEN**

Finanzkennzahl	AUROC <sup>1</sup>
Liquiditätsgrad 2	0.82
Kfr. FK (bereinigt)/Nettoumsatz x 360	0.81
Liquiditätsgrad 1	0.79
Selbstfinanzierungsgrad	0.78
Eigenfinanzierungsgrad (bereinigt)	0.77
EBITDA/Finanzaufwand	0.74
Cash-flow/FK-Anzahlungen	0.74
Kurzfristiges FK/Bilanzsumme	0.73
Zinsdeckungsgrad	0.73
Eigenkapitalrendite	0.72
Unternehmensgewinn/Nettoumsatz	0.72
Earnings Before Taxes/Aktiven	0.71
Cash-Flow/Nettoumsatz	0.68
Anlagendeckungsgrad 2	0.67
Gesamtkapitalrendite	0.65
Lagerdauer	0.65
EBITDA/Nettoumsatz	0.62
Kapitalumschlag	0.57
Verschuldungsfaktor	0.53

<sup>1</sup> AUROC: Area under ROC

zufällig selektierter ausgefallener Kunde anhand dieser Kennzahl schlechter klassifiziert wird als ein zufällig ausgewählter guter Kunde.

Die Liste zeigt, dass Kennzahlen aus den Dimensionen Liquidität, Finanzierung und Vermögenslage grundsätzlich etwas besser abschneiden als die klassischen Rentabilitätskennzahlen. Betrachtet man die Kennzahlen genauer, wird klar, dass bei den zugrundeliegenden Jahresabschlusspositionen der Bewertungsspielraum (Window Dressing) von oben nach unten tendenziell zunimmt. Gewisse Kriterien wie z. B. der Kapitalumschlag sind zudem branchenabhängig, was ihre Verwendbarkeit in einem branchenübergreifenden Modell einschränkt. Finanzkennzahlen mit einem AUROC < 0.55 werden nicht weiterverfolgt.

Bei den meisten Kennzahlen beinhalten Werte ab einem bestimmten Niveau nur noch wenig Informationen bezüglich eines potentiellen Ausfalls. *Abbildung 2* zeigt beispielsweise, mit welchem Ausfallrisiko ein bestimmter Liquiditätsgrad 2 verbunden ist. Ab einem Wert von 200% verändert sich das Ausfallrisiko praktisch nicht mehr. In einem solchen Fall wird bei einer Kennzahl an der richtigen Stelle ein «Cap» gesetzt. Dies verhindert, dass Ausreisser bei der folgenden Schätzung der Modellparameter zu Verzerrungen führen und Extremwerte bei der Berechnung des Scorewertes einer Firma einen übermässigen Einfluss ausüben. In gleicher Weise werden auch alle anderen Kennzahlen transformiert.

**4. SCORINGMODELL «QUANTITATIV»**

In einem nächsten Schritt werden die skalierten Kennzahlen in einem multivariaten Zusammenhang untersucht, um die Prognosefähigkeit verschiedener Kennzahlenkombinationen zu bestimmen. Der Praktiker ist oft versucht, zu viele Kriterien in einem Modell zu berücksichtigen, um möglichst alle finanziellen Aspekte eines Unternehmens abzudecken. Es ist jedoch eine statistische Erkenntnis, dass sich mit zusätzlichen Kriterien zwar der Erklärungsgehalt eines Modells bezüglich der verwendeten Stichprobe erhöhen lässt, dieses aber an Prognosekraft verliert, wenn es ausserhalb der Stichprobe angewendet wird. Bei der Kriterienauswahl sollte man sich deshalb vom Grundsatz «weniger ist mehr» leiten lassen.

Als Modellierungsmethode wurde die logistische Regression gewählt. Sie bestimmt für ein 0/1-Ereignis, mit welchen Einflussgrössen sich die beiden Gruppen «solvente» und «ausgefallene» Kunden besonders gut unterscheiden lassen. Logitmodelle haben gegenüber der sonst noch verwendeten Diskriminanzanalyse den Vorteil, dass sie mit weniger Annahmen bezüglich der Kennzahlenverteilung auskommen.

Allerdings ergeben sich nur schon mit 20 Kriterien mehr als 1 Million mögliche Kennzahlenkombinationen. Ausgehend von den univariat trennfähigsten Kriterien wird deshalb versucht, in einem schrittweisen Prozess dasjenige Modell zu finden, welches am besten zwischen guten und ausgefallenen Kunden unterscheidet [3]. Als Mass für die Trennkraft eines Modells verwenden wir wiederum den AUROC-Wert. Die zusätzliche Performanceverbesserung sinkt jedoch relativ schnell, je mehr Kennzahlen in ein Modell integriert werden. Die meisten statistisch abgestützten Ratingmodelle haben deshalb ein Set von 6–10 Kennzahlen.

Die acht Faktoren in unserem Ansatz zeigt *Abbildung 3*, wobei bestimmte Kennzahlen mit einer nichtlinearen Transformation in das Modell eingehen [4]. Dieses deckt alle wichtigen finanziellen Dimensionen einer Firma ab (Liquidität, Finanzierung, Vermögenslage, Rentabilität, Umschlag). Die Ertragssituation ist dabei mit dem Unternehmensgewinn, dem Cash-Flow sowie dem (betrieblichen) EBITDA über verschiedene Aggregate im Modell vertreten.

Die Modellperformance lässt sich signifikant steigern, wenn bei *Gewerbekunden* (Umsatz < CHF 1,5 Mio.) der Liquiditäts-

Abbildung 2: **LIQUIDITÄT UND AUSFALLRISIKO**

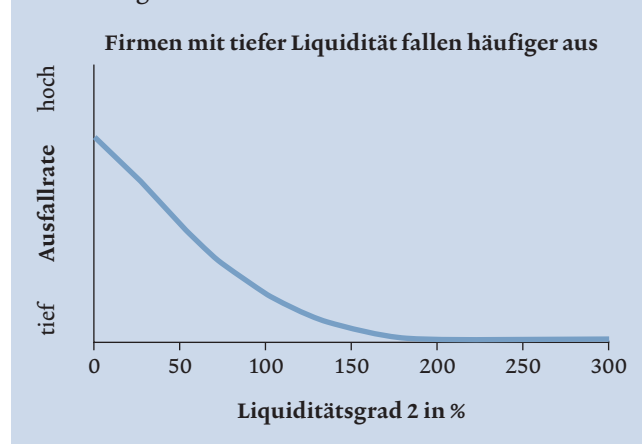


Abbildung 3: **MODELL MIT FINANZIELLEN FAKTOREN**

Kennzahl	Bedeutung
Liquiditätsgrad 1 ( <i>Unternehmen</i> ) oder 2 ( <i>Gewerbe</i> )	↑
Kapitalbindung (Kfr. FK x 360/Umsatz)	↑
Eigenfinanzierungsgrad	↑
Cash-Flow/Fremdkapital – Anzahlungen	→
EBITDA/Finanzaufwand	→
Umsatzrentabilität	↓
Lagerdauer	↓
Personengesellschaft	↓

grad 2, bei grösseren *Unternehmen* hingegen der Liquiditätsgrad 1 genommen wird. Praktisch arbeiten die Banken in diesen beiden Kundensegmenten daher mit zwei separaten Modellen, die unterschiedliche Kennzahlengewichte aufweisen. Die relative Bedeutung der Kriterien in Abbildung 3 gibt deshalb nur eine Tendenz für einen durchschnittlichen Kunden an. Der AUROC bei Gewerbekunden beträgt 0.88, derjenige bei Unternehmen 0.92. Ausfallprognosen beim Kleingewerbe sind somit etwas schwieriger als bei grösseren Firmen.

*Personengesellschaften* weisen gegenüber *Kapitalgesellschaften* bei den meisten Kennzahlen etwas schlechtere Werte auf, ohne dass die Ausfallrate wesentlich höher ist. Der Indikator «Personengesellschaft» wird als binäre Variable (0/1) model-

liert und «korrigiert» diese strukturellen Unterschiede, die sonst insbesondere bei Freiberuflern (Ärzten, Anwälten usw.) oft zu einer Überschätzung des Risikos führen. Relevant ist das Kriterium vorwiegend bei Gewerbekunden.

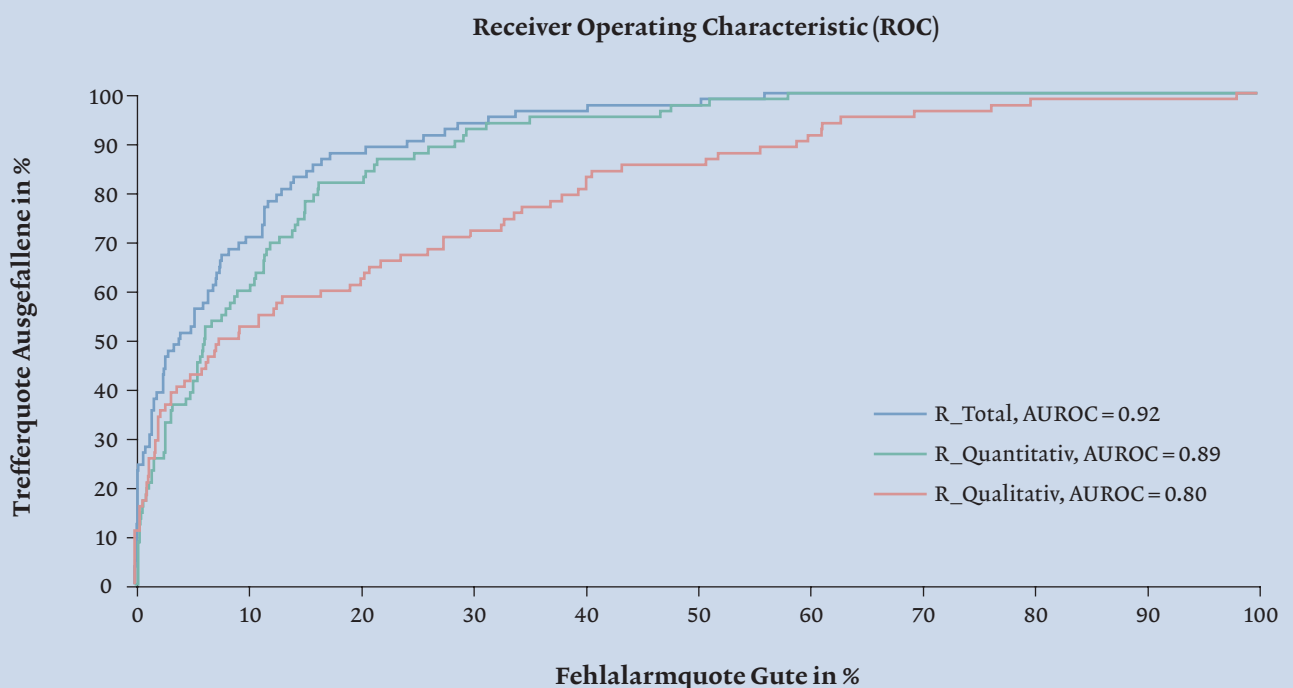
Die *Lagerdauer* gehört einzeln zwar nicht zu den trennfähigsten Kennzahlen, bringt im multivariaten Kontext aber noch wertvolle Aspekte ein (Warenaufwand, Lager), die in den übrigen Kennzahlen nicht enthalten sind. Dies führt zu einer leichten Verbesserung und Abrundung des Modells.

Ein gutes Modell widerspiegelt oft nicht das ökonomische Maximum, sondern das praktische Optimum. So hat das Feedback von den am Datenpool beteiligten Banken oder

Abbildung 4: **TRENNFÄHIGKEIT QUALITATIVER FAKTOREN**

Kriterium	AUROC
Informationspolitik	0.73
Unternehmensstrategie	0.72
Managementqualität	0.69
Rechnungswesen	0.66
Marktattraktivität	0.63
Abhängigkeiten	0.63
Wettbewerbsvorteile	0.61
Controlling	0.59
Firmenalter	0.55
Nachfolgeregelung	0.53

Abbildung 5: **MESSUNG DER MODELLPERFORMANCE**



die Suche nach einer guten Kennzahlen-Diversifikation den Modellbau wesentlich beeinflusst. Das Modell enthält daher auch keine Kriterien, die mit falschen Vorzeichen eingehen (je besser die Kennzahl, desto höher das Risiko) oder die stark miteinander korreliert sind. Dies würde nur zu einer Überanpassung (overfitting) des Modells an die Stichprobe führen. Selbstverständlich sind alle Variablen sowie das Modell als Ganzes statistisch signifikant.

### 5. SCORINGMODELL «QUALITATIV»

Führt die Berücksichtigung weicher Kriterien wie «Managementqualität» oder «Wettbewerbsvorteile» zu einer Verbesserung des Modells? In der Frühphase der Ratingmodelle wurde oft argumentiert, Finanzkennzahlen würden lediglich die Vergangenheit abbilden und eine zukunftsorientierte Bonitätsprognose sei nur mit qualitativen Faktoren möglich. Der Optimismus, mit einer subjektiven Unternehmensanalyse Ausfälle besser prognostizieren zu können, ist heute einer realistischeren Einschätzung gewichen. Da uns genügend lange Zeitreihen mit qualitativen Merkmalen vorliegen, werden wir die Frage nach deren Gewichtung wiederum statistisch beantworten.

Die im Datenpool vertretenen Banken beurteilen verschiedene qualitative Faktoren auf einer Skala von 1–4 nach einem einheitlichen Schema. Die Ausformulierung der Kriterien hat sich seit Ende 2001 kaum mehr verändert, so dass heute ein konsistenter Datensatz vorliegt. In die Stichprobe sind dabei rund 1300 qualitative Bewertungen eingeflossen. *Abbildung 4* zeigt die univariate Trennkraft ausgewählter Kriterien.

Es fällt auf, dass die Trennfähigkeit der qualitativen Kriterien tendenziell tiefer ist als diejenige der Finanzkennzahlen. Für einen Aussenstehenden ist es ja oft schwierig, die Kernkompetenzen und Wettbewerbsvorteile einer Firma in einem volatilen Marktumfeld richtig zu beurteilen. Vor dieser Herausforderung ist die Bewertung der Kreditsachbearbeiter sicher positiv zu werten, übertrifft sie eine reine Zufallsbeurteilung doch bei weitem. Da die Kombinationsmöglichkeiten hier wesentlich geringer sind, werden alle Faktoren in Betracht gezogen, die einen AUROC grösser als 0.5 aufweisen. Ohne die Details hier zu zeigen, schätzen wir mit diesen qualitativen Merkmalen erneut ein Logitmodell.

*Abbildung 5* visualisiert die Performance dieses Modells (R\_Qualitativ) und vergleicht sie mit dem Kennzahlenmodell (R\_Quantitativ) anhand sogenannter ROC-Kurven. Diese gehören heute zum Standardrepertoire bei der Validierung von Ratingmodellen und werden wie folgt konstruiert: Angenommen, eine Bank lege in ihrem Ratingmodell einen *Trennwert* fest. Schuldner mit einem höheren Scorewert teilt sie in die Kategorie der «Guten» ein, solche mit einem tieferen Scorewert bezeichnet sie als potentielle «Ausfälle». Je konservativer der Trennwert und somit die Kreditpolitik ist, desto grösser ist der Anteil ausgefallener Schuldner, der richtig klassifiziert wird (Trefferquote Ausgefällene). Dies hat aber andererseits zur Folge, dass die Bank auch sehr viele gute

Kunden fälschlicherweise als potentielle Ausfälle bezeichnet (Fehlalarmrate Gute). Je nach Trennwert ergibt sich jeweils ein anderes Verhältnis von Trefferquote zu Fehlalarmrate. Die ROC-Kurve ist nun nichts anderes als eine Generalisierung dieses Verhältnisses über alle möglichen Trennwerte. Die Fläche unterhalb der Kurve entspricht dem AUROC des jeweiligen Modells. Je weiter oben links eine ROC-Kurve liegt, desto besser ist das Modell. In unserem Beispiel ist die Überlegenheit der Finanzkennzahlen klar, auch wenn die weichen Faktoren bei sehr schlechten Kunden einen lokalen Vorteil aufweisen.

### 6. RATING UND AUSFALLWAHRSCHEINLICHKEIT

Die beiden Teilmodelle liefern für jeden Kunden je einen Scorewert. Um die Modelltransparenz zu erhöhen und die Rating-

---

*«Der Nutzen qualitativer Faktoren für die Ausfallprognose ist eindeutig belegt.»*

kommunikation gegenüber dem Kunden zu erleichtern, werden die Scores zu kontinuierlichen Teilratings auf eine Skala von 1–12 umgerechnet [5]. Mit diesen zwei Teilratings lässt sich nun über ein neues Logitmodell die relative Bedeutung der quantitativen und qualitativen Faktoren bestimmen. Die ROC-Kurve des kombinierten Modells (R\_Total) in *Abbildung 5* zeigt, dass dieses einem reinen Finanzkennzahlenmodell in jeder Hinsicht überlegen ist (die Kurve liegt nie unter dem quantitativen Modell). Der Nutzen qualitativer Faktoren für die Ausfallprognose ist somit eindeutig belegt. Klar ist aber auch, dass die Gewichtung der quantitativen Faktoren wesentlich höher ausfallen muss.

Im Kreditrisikomanagement interessiert neben dem Rating eines Kunden vor allem seine Ausfallwahrscheinlichkeit. Zu diesem Zweck werden die Ratings abschliessend so mit Ausfallwahrscheinlichkeit kalibriert, dass die Anzahl Ausfälle pro Ratingkategorie den historischen Ausfallraten in den Kreditportfolios der Banken entsprechen.

### 7. FAZIT

Die ROC ist heute die populärste Technik zur Validierung von Ratingmodellen, weil ihre Resultate auch für einen Praktiker intuitiv nachvollziehbar sind. An einem für Schweizer Banken gebauten Modell wurde gezeigt, wie sich die ROC flexibel sowohl zur Beurteilung einzelner Bonitätskriterien als auch zum Vergleich ganzer Modelle einsetzen lässt. Mit Hilfe des Logitansatzes gelingt es dabei, sowohl weiche als auch harte Firmenmerkmale so zu kombinieren, dass das resultierende Ratingmodell betriebswirtschaftlich überzeugt und mit möglichst wenigen Kriterien eine hohe Prognosekraft erzielt. ■

**Anmerkungen:** 1) In «Studies on the Validation of Internal Rating Systems», Working Paper No. 14, Februar 2005, widmet das Basler Komitee der Thematik mehr als 100 Seiten. 2) Vgl. z. B. Engelmann, B./Hayden, E./Tasche, D.: Testing rating accuracy, in: RISK, January 2003, S. 82–86. 3) Um die Komplexität bei der Modellauswahl zu reduzieren,

verwenden wir das «Forward/Backward»-Selektionsverfahren aus dem Standardwerk von Hosmer, D./Lemeshow, S.: Applied Logistic Regression, S. 116–128, Wiley & Sons, 2000. 4) Nachdem das beste lineare Modell gefunden ist, implementieren wir die Methode fraktionaler Polynome von Royston und Altman. Damit lässt sich die korrekte

Beziehung von Variablen modellieren, die einen nichtlinearen Bezug zum Ausfallrisiko (genauer: zum Logit-Wert) aufweisen. 5) Zur Bestimmung der Ratingintervalle verwenden wir den K-Means-Algorithmus, eine spezielle Methode der Cluster-Analyse.

## RÉSUMÉ

### Mettre en œuvre des modèles de notation et les valider

A la suite de Bâle II, la tendance à quantifier les risques de crédit s'est imposée et a conduit à un renforcement de l'appui statistique des modèles de notation. Le modèle de notation présenté ici est basé sur un pool de données auquel ont adhéré 13 banques suisses. La présente étude s'appuie sur environ 2200 états financiers de PME solvables ou non. Dans une première étape, les auteurs examinent 110 ratios financiers sous l'angle de la qualité du pronostic d'un défaut sur prêt. La capacité de séparation entre entreprises solvables et insolubles se mesure au moyen de la courbe ROC (Receiver Operating Characteristic), intégrée à la mesure de performance AUROC (Area under ROC). La courbe ROC est actuellement la technique de validation de modèles de notation la plus répandue, car ses résultats peuvent être retracés de manière intuitive.

La liste de ratios séparables montre que les ratios des domaines de la liquidité, du financement et de l'état du patrimoine font meilleure figure que les ratios de rentabilité classiques. Dans la majorité des ratios, les valeurs ne contiennent plus, à partir d'un certain niveau, que peu d'informations relatives à un défaut potentiel, raison pour la-

quelle elles sont transformées en conséquence.

Dans une seconde étape, on définit une combinaison de ratios grâce à laquelle les clients solvables se démarquent le mieux des clients insolubles. Pour la sélection et la pondération des ratios, on utilise la méthode statistique du modèle logit et pour mesurer la performance du modèle, le critère AUROC. La capacité prédictive peut être augmentée si l'on choisit pour les plus petites entreprises un modèle légèrement différent (p. ex. degré de liquidité 2 au lieu de 1) avec des pondérations de ratios différentes de celles des plus grandes entreprises. Les deux modèles contiennent des ratios provenant de toutes les dimensions financières, les défauts sur prêt dans les grandes entreprises étant plus aisés à prévoir que dans les petites entreprises.

Le pool de données contient également des données historiques avec l'évaluation de facteurs qualitatifs tels la qualité du management ou les avantages compétitifs d'une entreprise. La procédure à suivre avec ces critères est la même que pour les ratios financiers, c'est-à-dire que l'on commence par tester leur capacité de séparation individuelle

et que l'on forme ensuite un modèle d'évaluation du crédit «qualitatif». Au niveau des divers facteurs souples déjà, on constate qu'ils ont tendance à pouvoir moins bien se démarquer. La performance du modèle d'évaluation du crédit qualitatif et quantitatif peut très bien être visualisée à l'aide des courbes ROC. Il apparaît alors clairement que les critères financiers l'emportent sur les critères qualitatifs.

Reste à savoir comment combiner les facteurs quantitatifs et qualitatifs. Afin que les résultats restent transparents pour les banques concernées, la valeur critère des deux modèles partiels est calculée selon une échelle continue de 1 à 12. Le résultat est deux notations partielles, lesquelles sont pondérées à leur tour en un modèle logit, la pondération du modèle quantitatif étant manifestement plus élevée. Toutefois, la prise en compte de facteurs qualitatifs génère un modèle de notation qui se démarque nettement mieux par rapport à une approche pure de ratios financiers. Pour conclure, les notations sont calibrées à l'aide de probabilités de défaut qui correspondent aux taux de défaut historiques du pool de données. CM/AFB