

Discussion Paper

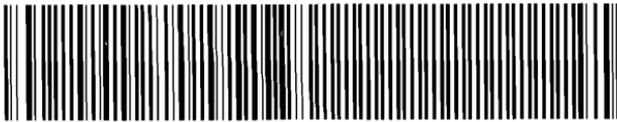
Discussion Paper No. 96-27

Prognose von Insolvenzwahrscheinlichkeiten mit Hilfe logistischer neuronaler Netzwerke

– Eine Untersuchung von kleinen und mittleren Unternehmen –

Ulrich Anders
Andrea Szczesny

W 636 (96.27)



16. SEP. 1996 Weltwirtschaft

W 636 (96.27) mit 95 Sig 5a

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

International Finance Series

Prognose von Insolvenzwahrscheinlichkeiten mit Hilfe logistischer neuronaler Netzwerke

– Eine Untersuchung von kleinen und mittleren Unternehmen –

Ulrich Anders & Andrea Szczesny

Zentrum für Europäische Wirtschaftsforschung (ZEW)

Tel: 0621/1235-141 bzw. -143	Kaiserring 14-16
Fax: 0621/1235-223	Postfach 10 34 43
Email: anders@zew.de bzw. szczesny@zew.de	68034 Mannheim

Mannheim, den 13. August 1996

Zusammenfassung

In diesem Artikel prognostizieren wir Insolvenzwahrscheinlichkeiten von kleinen und mittleren Unternehmen, die bislang noch keine hinreichende Beachtung gefunden haben, obwohl diese Unternehmen besonders insolvenzgefährdet sind. Die Schwierigkeit bei der Insolvenzprognose von kleinen und mittleren Unternehmen besteht darin, daß meist nur wenige Bilanzdaten verfügbar sind, so daß qualitative Angaben zu den Unternehmen hinzugezogen werden müssen. Wir zeigen, daß trotz dieser Einschränkungen gute Insolvenzprognosen möglich sind. Als Verfahren verwenden wir die logistische Regression und neuronale Netzwerke. Zur Modellselektion nutzen wir bei beiden Verfahren statistische Inferenztechniken. Für die logistische Regression ist dies ein Standard, für neuronale Netzwerke ist es dagegen neu. Mit Hilfe der Inferenztechniken gelangen wir zu sparsam parametrisierten Modellen und vermeiden ein *Overfitting* der Daten. Abschließend werden die Wirkungszusammenhänge in den resultierenden Modellen analysiert. Mit Hilfe des neuronalen Netzwerkmodells können die Nichtlinearitäten in den Wirkungszusammenhängen abgebildet werden, wodurch wesentlich differenziertere Aussagen über das Insolvenzrisiko in Abhängigkeit bestimmter Merkmalsausprägungen getroffen werden können als bei der logistischen Regression.

Wir sind Daniel Schwamm für seine wertvolle Unterstützung zu großem Dank verpflichtet. Die Daten für diese Studie wurden dem ZEW freundlicherweise vom Verband der Vereine Creditreform zur Verfügung gestellt und uns von Dietmar Harhoff, Konrad Stahl und Michael Woywode in aufbereiteter Form überlassen. Wir danken Olaf Korn und Michael Schröder für hilfreiche Anmerkungen.

1 Einleitung

In den letzten Jahren ist die Zahl der Unternehmensinsolvenzen in Deutschland deutlich gestiegen.¹ Die hohen Insolvenzraten führen zu erhöhten Wertberichtigungen in der Kreditwirtschaft. In vielen Fällen reagieren die Kreditgeber auf ihre gestiegenen Kreditausfallrisiken mit einer generell vorsichtigeren Kreditvergabe und einer Reduzierung des Kreditvolumens. Jedoch greift diese Reaktion nicht an die Wurzel des Problems. Zur Verminderung des Kreditrisikos wäre vielmehr eine exaktere Risikoerfassung als bisher wünschenswert. Die Verwendung eines Verfahrens, das eine kundenindividuelle Ermittlung der Insolvenzwahrscheinlichkeit ermöglicht und somit eine entsprechende Konditionengestaltung zuläßt, ist daher eine viel geeignetere Vorgehensweise.

Während in den meisten Arbeiten zur Insolvenzprognose größere veröffentlichungspflichtige Unternehmen betrachtet werden, liegt das Augenmerk in dieser Untersuchung auf kleinen und mittleren Unternehmen. Eine fundierte Untersuchung zum Insolvenzrisikos dieser Unternehmen ist uns derzeit nicht bekannt. Dies ist um so erstaunlicher, als gerade die Gruppe kleiner und mittlerer Unternehmen mit einem hohen Insolvenzrisiko behaftet sind. Eine Untersuchung der Creditreform (1996) zeigt, daß 95% aller Unternehmensinsolvenzen kleinere und mittlere Betriebe betreffen.

Die Schwierigkeit der Insolvenzprognose von kleinen und mittleren Unternehmen besteht darin, daß diese häufig nicht publizitätspflichtig sind, ihre Bilanz nur in verkürzter Form aufstellen oder ihrer Veröffentlichungspflicht sehr verspätet oder gar nicht nachkommen. Aus diesem Grund liegen Bilanzen zum Kreditvergabezeitpunkt meist gar nicht oder nur unvollständig vor. Der Kreditgeber muß sich daher häufig mit wenigen quantitativen und einigen qualitativen Angaben zum Unternehmen zufrieden geben.

Ein Verfahren, welches für die Prognose von Insolvenzwahrscheinlichkeiten von kleinen und mittleren Unternehmen geeignet ist, muß daher vor allem den Einsatz von qualitativen Daten erlauben. In den meisten Arbeiten zur Insolvenzanalyse wurde die multivariate lineare Diskriminanzanalyse eingesetzt.² Diese Methode hat jedoch den Nachteil, daß sie sich theoretisch nur für normalverteilte, nicht jedoch für qualitative Merkmalsausprägungen eignet.

Aus diesem Grund verwenden wir in dieser Untersuchung die logistische Regression. Dieses Verfahren ist robust gegen die Verwendung von nicht normalverteilten Merkmalen (Press/Wilson, 1978). Mit Hilfe der logistischen Regression soll die Wahrscheinlichkeit modelliert werden, mit der ein Unternehmen in Abhängigkeit bestimmter Merkmale zahlungsunfähig wird. Leider erlaubt das einfache logistische Regressionsmodell lediglich eine lineare Modellierung zwischen den Unternehmensmerkmalen und dem Solvenz/Insolvenz-Verhältnis. Tatsächlich beeinflussen manche Merkmale das Solvenz/Insolvenz-Verhältnis

¹Das Statistische Bundesamt nennt für das Jahr 1995 den Rekord von 22.244 Konkursen und Vergleichen, fast 19 Prozent mehr als im Jahr 1994. Die von den Gläubigern bezifferte Schadenssumme beläuft sich auf 33 Milliarden DM.

²Hüls (1995), Rehkugler/Poddig (1993), Bleier (1989) oder Hausschildt (1988) verwenden multivariate Diskriminanzanalysen zur Jahresabschlußanalyse von Unternehmen. Keysberg (1989), Feulner (1980) oder Häußler (1981) führen mit Hilfe der multivariaten linearen Diskriminanzanalyse eine Kreditwürdigkeitsanalyse für das Privatkundengeschäft von Banken durch.

jedoch nichtlinear. Eine solche Nichtlinearität ließe sich im logistischen Regressionsmodell durch eine nichtlineare Transformation der exogenen Merkmale abbilden. Da jedoch die genaue Form der Nichtlinearität nicht bekannt ist, ist die Wahl einer geeigneten Transformation ausgesprochen schwierig. Zudem ist es nicht sinnvoll, Indikatorvariablen zu transformieren. Aus diesem Grund setzen wir neuronale Netzwerke ein, die eine nichtlineare Modellierung des Solvenz/Insolvenz-Verhältnisses erlauben, ohne daß im voraus Annahmen über die Form der abzubildenden Nichtlinearität getroffen werden müssen. Darüber hinaus können neuronale Netzwerke gegenseitige Abhängigkeiten in den Variablen abbilden, wodurch auch Indikatorvariablen nichtlinear in das Modell eingehen können.

Die Insolvenzanalyse mit Hilfe neuronaler Netze ist bislang mit einer Reihe von Defiziten behaftet. Diese liegen vor allem darin begründet, daß neuronale Netze nicht als statistische Analyseverfahren erkannt und entsprechend eingesetzt wurden. Als Folge davon werden die Auswahl von Variablen und die Spezifikation einer geeigneten Netzwerkarchitektur mittels einer eher heuristischen Vorgehensweise durchgeführt. Darüber hinaus wurden in einigen Untersuchungen viel zu große Netzwerke auf der Basis eines viel zu kleinen Datensatzes aufgestellt.³

Die in dieser Studie verwendeten Daten stammen aus dem Mannheimer Unternehmenspanel (MUP). Das Panel umfaßt etwa 13.000 westdeutsche Unternehmen, deren Daten seit September 1989 gesammelt werden. Neben Merkmalen zur Struktur und zur ökonomischen Situation des Unternehmens (z.B. Anzahl der Mitarbeiter im Betrieb, Alter des Unternehmens, Branche, Rechtsform, vergangene Umsätze) stellt das MUP Informationen zur Person des Unternehmers (z.B. Ausbildung, Alter, Familienstand, Geschlecht) bereit. Da bei kleinen und mittleren Unternehmen eine besonders enge Bindung zwischen der Person des Unternehmers und seinem Unternehmen besteht,⁴ sind solche Faktoren für die Analyse ausgesprochen wichtig. Für diese Studie wurden dem MUP 302 kleine und mittlere Unternehmen mit maximal 500 Mitarbeitern entnommen. Von diesen Unternehmen wurde die Hälfte im Zeitraum von 1990 bis 1993 zahlungsunfähig.

Der vorliegende Artikel ist wie folgt organisiert: In Abschnitt 2 wird die logistische Regression und anschließend die Theorie neuronaler Netzwerke unter einer statistischen Perspektive dargestellt. In Abschnitt 3 erfolgt eine Beschreibung der verwendeten Daten. Abschnitt 4 gibt die Ergebnisse der Untersuchung wieder, die in Abschnitt 5 zusammengefaßt wird.

2 Klassifikationsverfahren

Unter Klassifikationsverfahren versteht man Verfahren, die dazu verwendet werden, Objekte verschiedenen Klassen zuzuordnen. Dabei ist die tatsächliche Klassenzugehörigkeit der einzelnen Objekte unbekannt, so daß die Klassifizierung allein aufgrund der Merkmale der Objekte vorgenommen werden muß. Das am häufigsten eingesetzte Klassifikationsverfahren ist die multivariate Diskriminanzanalyse (MDA). Ihr Einsatz ist jedoch nur dann

³In einigen Untersuchungen wurden sogar Datensätze verwendet, die weniger als fünfzig insolvente Unternehmen besitzen. Vgl. Hauschildt (1988), Odom/Sharda (1990) oder Bischoff/Bleile/Graalfs (1992).

⁴Vgl. Endress (1979).

zulässig, wenn die Merkmale einer Klasse annähernd normalverteilt sind. Die Voraussetzung der Normalverteilung wird insbesondere bei der Verwendung von qualitativen Merkmalen, die für die Insolvenzanalyse herangezogen werden müssen, verletzt.

Es bieten sich zwei Alternativen: die logistische Regression, die nach Press/Wilson (1978) robust gegen die Verwendung von nicht normalverteilten Merkmalen ist, und nichtparametrische Verfahren. Nichtparametrische Verfahren, wie z.B. Kerndichteschätzer oder die Methode der nächsten Nachbarn, eignen sich für beliebige Klassenverteilungen, da sie keine Annahmen über diese treffen. Bei diesen Verfahren lassen sich jedoch keine statistischen Inferenztechniken verwenden, wodurch die Auswahl eines geeigneten Modells sowie die Selektion von relevanten Merkmalen erheblich erschwert wird. Im Gegensatz dazu ermöglicht die logistische Regression die Anwendung von Inferenztechniken für die Modellselektion. Für das hier gegebene Problem ist die logistische Regression also wesentlich besser geeignet als nichtparametrische Verfahren.

Eine Restriktion der einfachen logistischen Regression besteht in der Annahme, daß der zu modellierende Zusammenhang linear ist. Einige theoretische Ansätze deuten jedoch auf einen nichtlinearen Einfluß einiger Unternehmensmerkmale auf die Insolvenzwahrscheinlichkeit hin.⁵ Nichtlinearitäten lassen sich im logistischen Regressionsmodell lediglich durch eine geeignete Transformation der Merkmalsausprägungen abbilden. Da die genaue Form der Nichtlinearität jedoch unbekannt ist, ist die Wahl einer geeigneten Transformation ausgesprochen schwierig. Aus diesem Grund setzen wir neuronale Netzwerke ein. Mit einer bestimmten Architektur lassen sich diese als eine Verallgemeinerung der logistischen Regression interpretieren. Ihr Vorteil besteht darin, daß nichtlineare Zusammenhänge modelliert werden können, über deren funktionale Form man im voraus keine Annahmen treffen muß. Zudem sind neuronale Netzwerke in der Lage, gegenseitige Abhängigkeiten in den Variablen abzubilden. Dadurch können auch nichtlineare Einflüsse in den Indikatorvariablen modelliert werden, die durch Transformationen dieser Variablen nicht abgebildet werden können. Die Transformation einer Indikatorvariable ergibt nämlich wieder nur eine Indikatorvariable. Ein nichtlineare Einfluß einer Indikatorvariable bedeutet, daß diese in Abhängigkeit der anderen Merkmalsausprägungen einer Beobachtung verschieden stark im Modell wirkt. Im folgenden werden das logistische Regressionsmodell und neuronale Netzwerke dargestellt.

2.1 Die logistische Regression

Es wird angenommen, daß zwischen der endogenen Variable y und den Merkmalsausprägungen der zu klassifizierenden Objekte $X = [x_0, x_1, \dots, x_I]$ der folgenden Zusammenhang existiert:

$$y = f(X) + \varepsilon \quad (1)$$

Die endogene Variable y nimmt die Werte Eins oder Null an. Die Funktion $f(X)$ entspricht der konditionalen Erwartung $E[y|X]$. Für die Schätzung der konditionalen Erwartung ver-

⁵Vgl. Aldrich/Auster (1986) und Freeman/Carrol/Hannan (1983).

wendet man die logistische Funktion $\pi(z) = (1 + e^{-z})^{-1}$, deren Wertebereich im Intervall $]0;1[$ liegt. Mit der logistischen Funktion werden die Wahrscheinlichkeiten geschätzt, mit denen ein Objekt einer bestimmten Klasse zugeordnet werden kann. Eine ausführliche Begründung für die Verwendung der logistischen Funktion findet sich u.a. bei Cox (1970). Die lineare logistische Regressionsfunktion ergibt sich damit durch:

$$f(X, \alpha) = \pi(X\alpha) = \frac{1}{1 + e^{-X\alpha}}. \quad (2)$$

Dabei wird ein Objekt, das die Merkmale X besitzt, der Klasse Eins mit der Wahrscheinlichkeit $\pi(X\alpha)$ zugeordnet und der Klasse Null mit Wahrscheinlichkeit $(1 - \pi(X\alpha))$. Das logistische Regressionsmodell kann durch Transformation in das Logit-Modell überführt werden.⁶ Ein Logit λ_t ergibt sich durch:

$$\lambda_t = \ln \left(\frac{\pi(X_t\alpha)}{1 - \pi(X_t\alpha)} \right) = X_t\alpha. \quad (3)$$

Der Fehler der Regression in Gleichung (1) wird mit ε bezeichnet. Im Fall einer dichotom verteilten endogenen Variable y kann ε zwei Werte annehmen. Der Wert $y = 1$ wird mit Wahrscheinlichkeit $\pi(X\alpha)$ angenommen und ε erhält den Wert $\varepsilon = 1 - \pi(X\alpha)$. Der Wert $y = 0$ wird mit Wahrscheinlichkeit $(1 - \pi(X\alpha))$ angenommen und ε erhält den Wert $\varepsilon = -\pi(X\alpha)$. Die Fehler ε unterliegen also einer Binomialverteilung mit Mittelwert Null und Varianz $\pi(X\alpha) \cdot (1 - \pi(X\alpha))$.

Zur Schätzung der Parameter α wird die *Maximum Likelihood*-Methode verwendet. Im Fall von identischen und unabhängig binomialverteilten Fehlern wird dabei folgende *Log-likelihood*-Funktion maximiert:

$$\ln L(\alpha) = \sum_{t=1}^T y_t \ln \pi(X_t\alpha) + (1 - y_t) \ln(1 - \pi(X_t\alpha)), \quad (4)$$

Die kleinste-Quadrate-Methode sollte in diesem Zusammenhang nicht benutzt werden, da sie sich aus der Annahme normalverteilter Fehler herleitet, und damit zu einer ineffizienten Parameterschätzung führen würde.⁷

⁶Vgl. Ronning (1991).

⁷Vgl. Greene (1993).

2.2 Logistische neuronale Netzwerke

Nur wenigen Verfahren ist in den letzten Jahren so viel Aufmerksamkeit zuteil geworden wie neuronalen Netzwerken. Neuronale Netzwerke haben ihre Leistungsfähigkeit in den verschiedensten Anwendungsgebieten unter Beweis gestellt,⁸ und sie wurden bereits mehrfach für die Insolvenzanalyse eingesetzt.⁹ In den meisten dieser Anwendungen wurden neuronale Netzwerke jedoch nicht als statistische Verfahren begriffen. Tatsächlich sind neuronale Netzwerke und insbesondere die in den meisten Fällen verwendeten *Multi Layer Perceptron*-Netzwerke (MLP) nichts anderes als eine sehr flexible Klasse von statistischen Regressionsverfahren.¹⁰ Als solche bieten sie die Möglichkeit zu einer fundierten statistischen Analyse. Mittels einer statistischen Analyse lassen sich beispielsweise geeignete Netzwerkarchitekturen bestimmen, die Signifikanz von einzelnen Variablen testen, das sogenannte *Overfitting* vermeiden sowie bestimmte Datenprobleme erkennen.

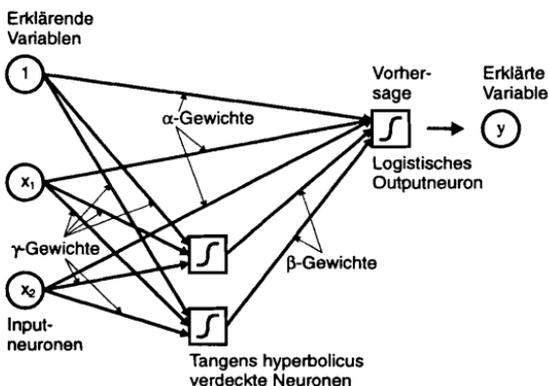


Abbildung 1: MLP-Netzwerk mit drei Schichten.

In diesem Beitrag werden ausschließlich MLP-Netzwerke der Form betrachtet, wie sie in Abbildung 1 dargestellt ist. Netzwerke dieser Form stellen eine nichtlineare Verallgemeinerung der linearen logistischen Regression dar.¹¹ Jedes neuronale Netzwerk läßt sich durch eine Funktion der erklärenden Variablen $X = [x_0, x_1, \dots, x_I]$ und der Netzwerkgewichte $w = (\alpha', \beta', \gamma')$ beschreiben. x_0 ist konstant und wird für alle Beobachtungen als $x_0 \equiv 1$ definiert. I bezeichnet die Anzahl der nichtkonstanten erklärenden Variablen, H

⁸Z.B. Anders/Korn/Schmitt (1996) oder Rehkugler/Zimmermann (1994).

⁹Baetge(1994), Baetge u.a. (1994), Krause (1993), Pytlík (1995) und Rehkugler/Poddig (1992) analysieren mit Hilfe neuronaler Netzwerke Jahresabschlüsse. Rehkugler/Schmidt-von Rhein (1993) und Schumann (1992) verwenden neuronale Netze zur Kreditwürdigkeitsanalyse im Privatkundengeschäft.

¹⁰Vgl. Sarle (1994).

¹¹Vgl. McLachlan (1992).

die Anzahl der verwendeten verdeckten Neuronen. Für das in Abbildung 1 dargestellte Netzwerk ergibt sich damit die folgende funktionale Form $f(X, w)$:

$$f(X, w) = \pi \left(X\alpha + \sum_{h=1}^H \beta_h g \left(\sum_{i=0}^I \gamma_{hi} x_i \right) \right) \quad (5)$$

Die Funktion $g(\cdot)$ stellt die sogenannte Aktivierungsfunktion eines Neurons dar und wird für diese Arbeit als die tangens hyperbolicus Funktion gewählt. Die Funktion $\pi(\cdot)$ ist die in Gleichung (2) definierte logistische Funktion.

Die Schätzung der Netzwerkgewichte w erfolgt für das gegebene Problem wiederum mit der *Maximum Likelihood*-Methode, wobei die Zielfunktion der logistischen Regression (4) durch folgende Zielfunktion ersetzt wird:

$$\ln L(w) = \sum_{t=1}^T y_t \ln f(X_t, w) + (1 - y_t) \ln(1 - f(X_t, w)). \quad (6)$$

Wie oben dargestellt, führt die Verwendung der kleinsten-Quadrate-Methode bei nicht normalverteilten Fehlertermen zu einer ineffizienten Parameterschätzung.¹² Trotzdem wird diese Methode im Kontext der Insolvenzanalyse für die Bestimmung der Netzwerkgewichte häufig eingesetzt.¹³

Die Verallgemeinerung der linearen logistischen Regression durch das neuronale Netzwerk läßt sich durch Nebeneinanderstellung der Gleichungen (2) und (5) erkennen. Das zu Gleichung (3) korrespondierende Logit λ_t ergibt sich jetzt nichtlinear durch

$$\lambda_t = \ln \left(\frac{f(X_t, w)}{1 - f(X_t, w)} \right) = X\alpha + \sum_{h=1}^H \beta_h g \left(\sum_{i=0}^I \gamma_{hi} x_{i,t} \right). \quad (7)$$

Der Vorteil der Verwendung von neuronalen Netzwerken besteht darin, keine explizite Annahme über die funktionale Form der konditionalen Erwartung $f(X)$ aus Gleichung (1) treffen zu müssen. Neuronale Netzwerke sind nämlich in der Lage, jede Funktion bis zu einem beliebig hohen Grad an Genauigkeit zu approximieren, vorausgesetzt sie sind hinreichend groß dimensioniert (Hornik/Stinchcombe/White, 1989). Eine Schwierigkeit besteht jedoch darin, die für ein Problem geeigneteste Netzwerkarchitektur zu finden. Diese sollte gerade groß genug sein, die Funktion $f(X)$ abzubilden, jedoch nicht größer. Ein solches Netzwerk weist keinen systematischen Fehler auf und neigt nicht zum *Overfitting* der Daten.¹⁴

¹²Es wird häufig fälschlicherweise angenommen (z.B. Goede/Weinrich, 1996), daß neuronale Netzwerke dann Verwendung finden sollten, wenn man keine Annahme über die Verteilung der Störterme machen möchte.

¹³Vgl. Bischoff/Bleile/Graalfs (1992), Krause (1993), Pytlik (1995).

¹⁴Vgl. Anders (1995).

Die Suche nach einem geeigneten Netzwerkmodell ist ein iterativer Prozeß, in dem eine Reihe von verschiedenen statistischen Verfahren behilflich sein können.¹⁵ Zu den wichtigsten Verfahren gehören Signifikanztests für die Parameter eines Modells, mit deren Hilfe man sowohl relevante Inputvariablen als auch die notwendige Anzahl verdeckter Neuronen identifizieren kann.

Um ein gutes Klassifikationsmodell zu erhalten, ist die Auswahl geeigneter Merkmale von besonderer Bedeutung. Nur wenn die selektierten Merkmale — eventuell auch im Zusammenspiel mit anderen Merkmalen — eine gewisse Klassenzugehörigkeit aufweisen, läßt sich ein gutes Klassifikationsmodell aufstellen. Man muß daher feststellen, ob ein Merkmal tatsächlich signifikant zu einem Modell beiträgt.

Dazu können u.a. *Likelihood Ratio*-Tests (LR-Tests) und Wald-Tests verwendet werden. Mit einem LR-Test lassen sich zwei Modelle mit einer verschiedenen großen Parameterzahl vergleichen. Insbesondere ist es sinnvoll, die geschätzten *Loglikelihoods* von zwei Modellen miteinander zu vergleichen, von denen das eine Modell zusätzliche Parameter besitzt. Damit läßt sich feststellen, ob die zusätzlichen Parameter das ursprüngliche Modell wesentlich verbessern. Mit einem Wald-Test läßt sich u.a. testen, ob ein Parameter signifikant von Null verschieden ist. In der logistischen Regression ist die Anwendung von Tests ein Standard.¹⁶ Bei neuronalen Netzwerken ist die Anwendung von Hypothesentests dagegen kompliziert und wird daher meist unterlassen.

Statistische Inferenz in neuronalen Netzwerken wurde von White (1989a,b) entwickelt. Er hat bewiesen, daß die Parameter eines neuronalen Netzwerks mit Hilfe von *Maximum Likelihood*-Verfahren konsistent geschätzt werden können, wenn diese bis auf symmetrische Lösungen eindeutig bestimmbar sind. Die geschätzten Parameter \hat{w} sind asymptotisch normalverteilt mit Mittelwert w und Kovarianz $\frac{1}{T}C$, wobei T die Anzahl der Beobachtungen bezeichnet. Als Schätzung der Kovarianzmatrix $\frac{1}{T}C$ kann u.a. die Inverse der Fisher'schen Informationsmatrix $E[\nabla^2 \ln L_t]$ verwendet werden.¹⁷ Dabei steht $\ln L_t$ für die *Loglikelihood Contribution* und $\nabla^2 \ln L_t$ für die Hessematrix der *Loglikelihood Contribution*.

Im Prinzip reicht dieses Wissen aus, um Hypothesentests durchzuführen. Vor der Anwendung der Tests muß allerdings sichergestellt sein, daß die Parameter — bis auf symmetrische Lösungen — tatsächlich eindeutig sind. Dies ist nicht der Fall, wenn sich in einem Netzwerk irrelevante verdeckte Neuronen befinden. Hat das β -Gewicht eines irrelevanten verdeckten Neurons nämlich theoretisch den Wert Null, können die γ -Gewichte, die in dieses verdeckte Neuron führen, jeden beliebigen Wert annehmen und sind somit nicht mehr eindeutig bestimmbar. Das Problem besteht darin, daß die Verteilung im Fall von nicht eindeutigen Gewichten nicht mehr multivariat normal ist, sondern einer viel allgemeineren Klasse von gemischten Gauß-Verteilungen angehört (Phillips, 1989). Damit ist eine auf der Normalverteilung beruhende Inferenz nicht mehr möglich.

Alle irrelevanten verdeckten Neuronen müssen also zunächst aus dem entsprechenden Netzwerk entfernt werden. Die Identifikation von irrelevanten verdeckten Neuronen ist jedoch äußerst schwierig. Deshalb schlagen Anders/Korn (1996) vor, nur dann eine weiteres

¹⁵Vgl. Anders (1995).

¹⁶Vgl. Hosmer/Lemeshow (1989).

¹⁷Vgl. Greene (1993).

verdecktes Neuron in ein neuronales Netzwerk aufzunehmen, wenn dieses signifikant zur Erklärung der endogenen Variable beiträgt. Dafür lassen sich Methoden verwenden, wie sie von White (1989c) und Teräsvirta/Lin/Granger (1993) entwickelt wurden.¹⁸

Unter der Voraussetzung, daß sich keine irrelevanten verdeckten Neuronen in dem Netzwerk befinden, lassen sich nun die Gewichte eines neuronalen Netzwerks mittels eines Wald-Tests auf ihre Signifikanz überprüfen.¹⁹ Wie erste empirische Studien zeigen (Anders/Korn/Schmitt, 1996), hat sich die Anwendung von Inferenztechniken als eine sehr leistungsfähige Methode herausgestellt, geeignete Netzwerkarchitekturen zu konstruieren und *Overfitting* zu vermeiden.

2.3 Qualität der Modelle

Die naheliegendste Methode, die Qualität eines Modells zu evaluieren, besteht darin, die Klassifikationsergebnisse auszuwerten. Wünschenswert sind hohe Trefferquoten, wobei geschätzte Wahrscheinlichkeiten, die über 0,5 liegen, der Klasse Eins (insolvent), geschätzte Wahrscheinlichkeiten, die unter 0,5 liegen, der Klasse Null (solvent) zugeordnet werden. Die Trefferquote macht jedoch keine Aussage über die Wahrscheinlichkeit, mit der ein Unternehmen als insolvent bzw. solvent prognostiziert wird. Dennoch werden in vielen Untersuchungen ausschließlich Trefferquoten zur Evaluierung der spezifizierten Modelle verwendet.²⁰ Beispielsweise hat ein Modell eine Trefferquote von 100%, wenn es für alle Objekte der Klasse Eins eine Wahrscheinlichkeit von 0,51 und für alle Objekte der Klasse Null eine Wahrscheinlichkeit von 0,49 liefert. Trotz der 100-prozentigen Trefferquote wäre einem Kreditgeber mit solch einem Modell im Fall der Insolvenzprognose nicht sehr geholfen, denn für eine Kreditentscheidung ist es letztendlich unerheblich, ob ein Unternehmen mit einer Wahrscheinlichkeit von 0,51 oder 0,49 insolvent wird. Aus diesem Grund werden Unternehmen, für die eine Insolvenzwahrscheinlichkeit im Bereich von 0,5 prognostiziert wird, von den Kreditgebern häufig einer weitergehenden Überprüfung unterzogen. Dieses Vorgehen verursacht im Kreditinstitut Kosten, die mit Hilfe eines trennschärferen Modells reduziert werden können. Wünschenswert sind daher Modelle, die die Insolvenz bzw. Solvenz eines Unternehmens mit jeweils hohen Wahrscheinlichkeiten $\hat{\pi}_t$ bzw. $(1 - \hat{\pi}_t)$ (z.B. größer als 0,7) voraussagen.

Bei der Evaluierung der Modelle ist es deshalb interessant, die beiden Häufigkeitsverteilungen der geschätzten Wahrscheinlichkeiten für die solventen bzw. insolventen Unternehmen nebeneinander zu stellen. Je näher die prognostizierten Insolvenzwahrscheinlichkeiten der solventen Unternehmen an der Wahrscheinlichkeit von Null und je näher die prognostizierten Insolvenzwahrscheinlichkeiten der insolventen Unternehmen an der Wahrscheinlichkeit von Eins liegt, desto besser ist die Treffsicherheit des Modells.

¹⁸Eine ausführliche Beschreibung dieser Methoden findet sich bei Anders/Korn (1996).

¹⁹Die genaue Berechnung des Wald-Tests für neuronale Netzwerke wird bei Anders (1996b) dargestellt.

²⁰Z.B. Baetge et al. (1994) und Pytlik (1995).

Die Treffsicherheit drückt sich u.a. in der *Deviance* D des Modells aus.²¹ Je näher ihr Wert an Null liegt, desto größer ist die Trennschärfe.

$$D = \sum_{t=1}^T y_t \ln \left(\frac{\hat{\pi}_t}{y_t} \right) + (1 - y_t) \ln \left(\frac{1 - \hat{\pi}_t}{1 - y_t} \right). \quad (8)$$

Ein weiteres Maß, die Qualität der logistischen Regression zu bestimmen, ist der *Goodness of Fit*-Wert Z^2 . Er hat eine ähnliche Bedeutung wie das Bestimmtheitsmaß R^2 der linearen Regression und berechnet sich durch:

$$Z^2 = \sum_{t=1}^T \frac{(y_t - \hat{\pi}_t)^2}{\hat{\pi}_t(1 - \hat{\pi}_t)}. \quad (9)$$

Je besser der *Goodness of Fit*, desto näher liegt der Wert Z^2 an Null. Daneben gibt es eine Reihe von Rangkorrelationskoeffizienten, die die geschätzten Wahrscheinlichkeiten $\hat{\pi}(\cdot)$ zu den entsprechenden Klassenzugehörigkeiten in Beziehung setzen. Diese sind u.a. Somers' D , Godman-Kruskals *Gamma* und Kendalls *Tau-a*.

3 Datenbeschreibung

3.1 Datenbasis

Grundlage der durchgeführten Untersuchungen ist das Mannheimer Unternehmenspanel (MUP) des Zentrums für Europäische Wirtschaftsforschung (ZEW). Das MUP ist eine Datenbank, die auf den Daten des Verbandes der Vereine Creditreform (VVC), Deutschlands größter Kreditauskunftei, beruht. Das MUP enthält umfangreiche Daten zu ca. 15.000 Unternehmen verschiedener Größe und Rechtsform und bildet damit einen Querschnitt durch die westdeutsche Unternehmenspopulation, der in bezug auf die Arbeitsstättenzählung 1987 repräsentativ ist. Die Daten des MUP werden seit fünf Jahren gesammelt und regelmäßig aktualisiert, so daß Entwicklungen einzelner Unternehmen nachvollzogen werden können. Im Fall von unklaren Unternehmensschicksalen hat das ZEW rund 4.000 Telefonate geführt, um die Daten zu vervollständigen.

Das MUP beinhaltet unter anderem Informationen zum Alter und der Beschäftigtenzahl von Unternehmen, zu den Beteiligungsstrukturen, der Vermögensstruktur und zur Person des Unternehmers. Marktaustritte wurden nach einem detaillierten System codiert.²²

Für diese Untersuchung wurden aus dem MUP 302 kleine und mittlere Unternehmen bis zu einer maximalen Größe von 500 Beschäftigten entnommen. Die Stichprobe umfaßt zur Hälfte Unternehmen, die im Zeitraum zwischen 1990 und 1993 zahlungsunfähig

²¹Die *Deviance* der logistischen Regression entspricht in etwa der *Sum of Squared Errors* einer linearen Regression.

²²Eine ausführliche Beschreibung des MUP findet sich bei Harhoff/Stahl/Woywode (1995).

geworden sind. Trotz der Größe des MUP-Datensatzes stellen insolvente Unternehmen einen Engpaßfaktor dar. Daher wurden die verfügbaren zahlungsunfähigen Unternehmen vollständig verwendet. Den 151 insolventen Unternehmen wurden 151 solvente Unternehmen gegenübergestellt, die zufällig gezogen wurden.

Die entnommenen Unternehmen stammen vorwiegend aus dem verarbeitenden Gewerbe, dem Bau und dem Handel. Ein kleinerer Teil der Betriebe gehört der Dienstleistungsbranche und dem Transportwesen an. Der Großteil der Unternehmen hat die Rechtsform der GmbH bzw. der GmbH & Co. KG. Der Rest verteilt sich auf KG, OHG und Einzelunternehmen. Gemeinnützige und öffentliche Unternehmen sowie freie Berufe wurden nicht berücksichtigt.

Die Mehrzahl der Unternehmen sind klein und haben maximal fünfzig Beschäftigte. Nur ein geringer Anteil der Betriebe im betrachteten Datensatz hat mehr als hundert Beschäftigte. Mehr als die Hälfte der Unternehmen im ausgewählten Datensatz sind jünger als 20 Jahre. Ein Viertel der Betriebe fällt in die Gruppe der Nachkriegsunternehmen mit einem Alter von bis zu fünfzig Jahren. Viele dieser Unternehmen werden noch vom Unternehmensgründer selbst geleitet.

Das MUP stellt außer den Angaben zu den Unternehmen noch Informationen zur Ausbildung und dem Alter der Unternehmer zur Verfügung. Im Falle mehrerer Anteilseigner wurden die persönlichen Daten des dominierenden Anteilseigners betrachtet. Rund die Hälfte der Unternehmer haben maximal eine abgeschlossene Ausbildung. Ein Viertel der Unternehmer haben eine Meisterprüfung abgelegt und ein Viertel verfügt über eine akademische Ausbildung.

4 Resultate der Studie

Für die Bildung der Modelle werden nur solche Kenngrößen kleiner und mittlerer Unternehmen betrachtet, die einem Kreditgeber bei der Prognose der Insolvenzwahrscheinlichkeit zur Verfügung stehen, wenn kein Jahresabschluß des kreditnachfragenden Unternehmens vorliegt. Dazu gehören Unternehmensalter, Mitarbeiterzahl, Unternehmensumsatz, Umsatzentwicklung, Rechtsform, Branche, Diversifikationsgrad des Unternehmens, Bildungsabschluß, Geschlecht und Alter des Unternehmers sowie das durchschnittliche Alter der Anteilseigner. Daten zur Ertragslage sowie zur Gewinn und Verlustrechnung wurden also nicht berücksichtigt.

Aus dem zur Verfügung stehenden Datensatz von 302 Beobachtungen wurden zum Zwecke der Validierung per Zufall 20 Prozent der Beobachtungen gezogen. Es ergab sich eine Trainingsmenge von 242 und eine Validierungsmenge von 60 Unternehmen, von denen jeweils die Hälfte der Unternehmen solvent bzw. insolvent waren.

4.1 Logistisches Regressionsmodell

Zunächst wurde ein logistisches Regressionsmodell aufgestellt. Die Selektion des Modells wurde schrittweise rückwärts durchgeführt, d.h. aus einem anfänglich zu großen Modell

wurden mit Hilfe von LR-Tests sukzessive alle Variablen entfernt, die nicht signifikant zur Erklärung der Unternehmensinsolvenzen beitragen. Als Signifikanzniveau wurde für die gesamte Untersuchung 5 Prozent festgelegt.

Als Resultat dieser Modellselektion wurden folgende Variablen identifiziert, die einen signifikanten Erklärungsbeitrag zur Insolvenzprognose von Unternehmen liefern.

- Das *Unternehmensalter* gemessen in Jahren.
- Die *Umsatzentwicklung*, die als Differenz der jeweils letzten Umsatzeintragungen errechnet wurde.
- Die Variable *beschränkte Haftung*, die den Wert Eins enthält, wenn die Gesellschafter nur beschränkt haften.
- Die Variablen der Branchenzugehörigkeit *verarbeitendes Gewerbe* und *Handel*.
- Die Indikatorvariable *Bildungsabschluß*, die den Wert Eins besitzt, wenn der Unternehmer maximal eine Lehre abgeschlossen hat und Null, falls eine Meisterprüfung abgelegt wurde oder der Unternehmer über einen akademische Abschluß verfügt.

Das resultierende logistische Regressionsmodell ist in Abbildung 2 dargestellt. Die zugehörigen Parameter zusammen mit ihren Standardabweichungen und der sich ergebenden Wald-Teststatistik finden sich in Tabelle 1.

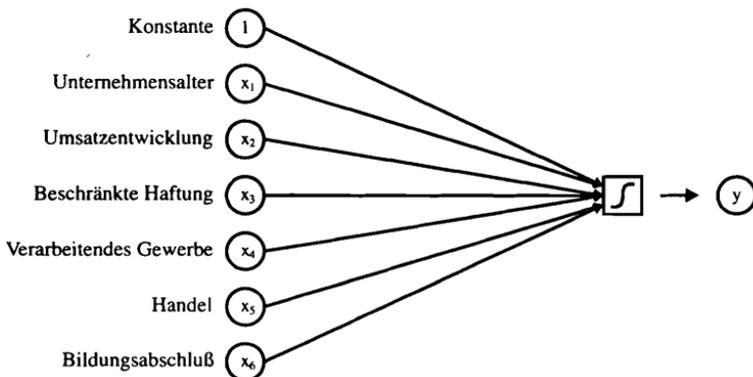


Abbildung 2: Geschätztes logistisches Regressionsmodell.

Aus der Tabelle 1 läßt sich ablesen, daß ein zunehmendes Unternehmensalter risikoreduzierend in das Modell eingeht; ebenso wirkt eine Umsatzsteigerung. Je größer also die Umsatzsteigerung in der Vergangenheit war, desto geringer ist das Insolvenzrisiko.

Eine Erhöhung des Risikos ergibt sich, wenn der Unternehmer maximal eine Lehre abgeschlossen hat sowie bei einer beschränkten Haftung der Gesellschafter. Eine Zugehörigkeit des Unternehmens zum Handel erniedrigt, eine Zugehörigkeit zum verarbeitenden Gewerbe dagegen erhöht das Insolvenzrisiko. Andere Brancheneinflüsse haben sich im Rahmen dieser Studie als nicht signifikant erwiesen.

Variable	Parameter	Wert	Std.abw.	Wald-Stat.	Signif.
Konstante	$\hat{\alpha}_0$	-1.5877	0.1466	117.24	0.0000
Unternehmensalter	$\hat{\alpha}_1$	-0.3724	0.0730	26.02	0.0000
Umsatzentwicklung	$\hat{\alpha}_2$	-0.5006	0.1005	24.81	0.0000
Beschränkte Haftung	$\hat{\alpha}_3$	1.6964	0.1402	146.47	0.0000
Verarbeitendes Gewerbe	$\hat{\alpha}_4$	0.3232	0.1548	4.36	0.0368
Handel	$\hat{\alpha}_5$	-0.5388	0.1736	9.64	0.0019
Bildungsabschluß	$\hat{\alpha}_6$	1.2381	0.1405	77.64	0.0000

Tabelle 1: Parametertests im logistischen Regressionsmodell.

4.2 Neuronales Netzwerkmodell

Als Ausgangsmodell für die Spezifikation des neuronalen Netzes wurde das logistische Regressionsmodell des vorhergehenden Abschnitts verwendet. Mittels der in Abschnitt 2.3 erwähnten LR-Tests und unter Zuhilfenahme der Methode von Teräsvirta/Lin/Granger (1993) läßt sich feststellen, ob das logistische Regressionsmodell nichtlineare Beziehungen unberücksichtigt läßt. Dabei testet man, ob ein zusätzliches verdecktes Neuron die Erklärung der endogenen Variable signifikant verbessert. In unserer Untersuchung war dies der Fall, so daß das logistische Regressionsmodell um ein verdecktes Neuron ergänzt wurde. Anschließend wurden alle nicht-signifikanten Parameter durch Anwendung des Wald-Tests identifiziert und sukzessive aus dem Netzwerk entfernt. Dabei wurde das Netzwerk jedesmal neu geschätzt, wenn ein Parameter entfernt wurde. Nach Entfernung aller nicht-signifikanten Parameter aus dem Netzwerk wurde mittels eines erneuten LR-Tests überprüft, ob noch weitere Nichtlinearitäten unberücksichtigt geblieben sind. Da dies nicht der Fall war, wurde kein weiteres verdecktes Neuron in das Netzwerk eingefügt. Das resultierende Netzwerk besitzt damit lediglich ein verdecktes Neuron (siehe Abbildung 3). Die Parameter des Netzwerks sind zusammen mit ihrer Wald-Teststatistik in Tabelle 2 dargestellt.

Im resultierenden neuronalen Netzwerkmodell geht das Unternehmensalter nichtlinear in das Netzwerk ein. Umsatzsteigerungen wirken sich linear risikovermindernd aus, eine Haftungsbeschränkung risikoverhöhend. Eine Zugehörigkeit des Unternehmens zum Handel oder zum verarbeitenden Gewerbe beeinflusst das Insolvenzrisiko nichtlinear. Der Bildungsabschluß des Unternehmers hat einen nichtlinearen Einfluß auf die Wahrscheinlichkeit einer Zahlungsunfähigkeit. Wie sich die einzelnen Merkmalsausprägungen auf die Insolvenzwahrscheinlichkeit eines Unternehmens im Detail auswirken, wird in Abschnitt 4.4 dargestellt.

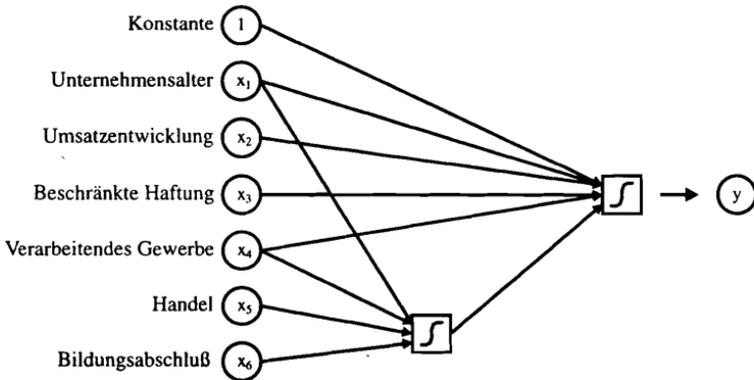


Abbildung 3: Geschätztes neuronales Netzwerkmodell.

Variable	Parameter	Wert	Std.abw.	Wald-Stat.	Signif.
Konstante	$\hat{\alpha}_0$	-1.8861	0.1553	147.59	0.0000
Unternehmensalter	$\hat{\alpha}_1$	0.6232	0.1763	12.49	0.0004
Umsatzentwicklung	$\hat{\alpha}_2$	-0.5179	0.1014	26.08	0.0000
Beschränkte Haftung	$\hat{\alpha}_3$	1.6411	0.1427	132.27	0.0000
Verarbeitendes Gewerbe	$\hat{\alpha}_4$	1.6393	0.5584	8.61	0.0033
	$\hat{\beta}$	2.5313	0.5240	23.34	0.0000
Unternehmensalter	$\hat{\gamma}_{1-1}$	-0.9849	0.2606	14.29	0.0002
Verarbeitendes Gewerbe	$\hat{\gamma}_{1-4}$	-0.7599	0.2772	7.52	0.0061
Handel	$\hat{\gamma}_{1-5}$	-0.4473	0.1646	7.38	0.0066
Bildungsabschluß	$\hat{\gamma}_{1-6}$	0.8518	0.2773	9.43	0.0021

Tabelle 2: Parameter und Teststatistiken im neuronalen Netzwerkmodell.

Das geschätzte Netzwerkmodell ist deutlich weniger komplex als die in bisherigen Studien verwendeten Modelle. Um eine zuverlässige Schätzung der Parameter eines Netzwerks zu erreichen, sollte die Anzahl der Beobachtungen die Anzahl der Parameter um mindestens Faktor zehn überschreiten (White, 1992). Man kann deshalb davon ausgehen, daß die in anderen Arbeiten konstruierten Netze viel zu groß sind.²³ In der Regel wurden die

²³Odom/Sharda (1990) schätzen ein Netz mit fünf Eingabeneuronen, fünf verdeckten Neuronen und einem Ausgabeneuron auf der Basis eines Datensatzes mit 74 Unternehmen, Bischoff/Bleile/Graalfs (1992) verwenden für ihre Untersuchung die Daten von 81 Unternehmen, um ein Netz mit 19 Eingabeneuronen zu schätzen und Baetge (1994) konstruiert Netze mit bis zu 73 Eingabeneuronen auf der Basis eines Datensatzes mit 224 Unternehmen

Eingabeneuronen vollständig mit den verdeckten Neuronen verknüpft. In manchen Untersuchungen wurde die Anzahl der verdeckten Neuronen darüber hinaus lediglich mit Hilfe von Daumenregeln bestimmt. Diese Vorgehensweisen führen zu einer unnötig hohen Komplexität der Netzwerkmodelle, die mit einiger Sicherheit viele statistisch nicht signifikante Parameter enthalten. Eine unnötig hohe Komplexität der Modelle macht die Interpretation der Ergebnisse schwierig und führt zu einer deutlichen Verschlechterung der Prognosevarianz.²⁴

4.3 Vergleich der beiden Modelle

Beide Modelle sind für die Insolvenzprognose geeignet: die logistische Regression erzielt gute, das neuronale Netzwerk sehr gute Resultate. Die Gütemaße der beiden Modelle sind in Tabelle 3 zusammengefaßt.

Gütemaße	Logistische Regression		Neuronales Netzwerk	
	in-sample	out-of-sample	in-sample	out-of-sample
Anzahl Beob.	240	60	240	60
TQ: Solvent	70,0	63,3	72,5	63,3
TQ: Insolvent	75,0	73,3	75,0	83,3
TQ: Gesamt	72,5	68,3	73,8	73,3
$\frac{1}{T}D$	-0,5532	-0,6094	-0,5236	-0,5731
$\frac{1}{T}Z^2$	1,05	1,17	0,956	1,17
Somers' <i>D</i>	0,59	0,47	0,62	0,55
Gamma	0,59	0,47	0,62	0,55
Tau-a	0,29	0,24	0,31	0,28

Tabelle 3: Gütemaße der geschätzten Modelle.

Um einen Vergleich zu bisherigen Untersuchungen im Bereich der Insolvenzprognose zu ermöglichen, betrachten wir zunächst die erzielten Trefferquoten (TQ) der logistischen Regression und des neuronalen Netzwerks. Die Gesamttrefferquote des neuronalen Netzwerks liegt *in-sample* bei 73,8 Prozent, *out-of-sample* werden 73,3 Prozent erreicht. Dabei werden insbesondere insolvente Unternehmen mit einer Trefferquote von 83,3 Prozent in fast allen Fällen richtig klassifiziert. Die Trefferquoten des neuronalen Netzwerks liegen in allen Fällen über denen der logistischen Regression, deren Gesamttrefferquote *in-sample* bei 72,5 Prozent und *out-of-sample* bei immerhin 68,3 Prozent liegt.

²⁴Vgl. Anders/Schwamm (1996).

Die *out-of-sample*-Trefferquoten anderer Untersuchungen, die Jahresabschlüsse von Unternehmen betrachten, liegen im Schnitt ebenfalls zwischen 70 und 80 Prozent. Mit Hilfe unserer Modelle erreicht man also ein vergleichbares Ergebnis, obwohl keinerlei Informationen zur Finanz-, Vermögens- und Ertragslage der Unternehmen genutzt wurden.²⁵ Würden solche Informationen hinzugezogen werden, sollte sich die ohnehin schon gute Qualität der Modelle dieser Studie noch weiter verbessern lassen.

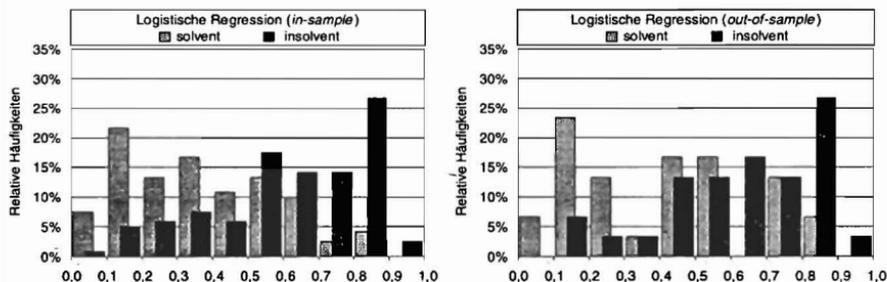


Abbildung 4: Häufigkeitsverteilung der prognostizierten Wahrscheinlichkeiten des logistischen Regressionsmodells.

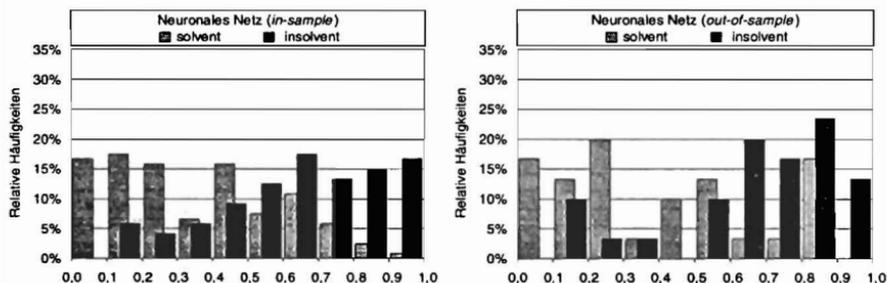


Abbildung 5: Häufigkeitsverteilung der prognostizierten Wahrscheinlichkeiten des neuronalen Netzwerkmodells.

Neben der Betrachtung der Trefferquoten soll die Qualität der Modelle dieser Untersuchung an ihrer Treffsicherheit bzw. Trennschärfe gemessen werden. Je besser die

²⁵Beispielsweise erreichen Odom/Sharda (1990) bei der Bilanzanalyse mit Hilfe der LDA eine Gesamttrefferquote von 74,5 Prozent, mit ihrem Netzwerkmodell eine Trefferquote von 81,8 Prozent. Pytlík (1995) erzielt 77,2 Prozent mit der LDA, seine neuronalen Netzwerkmodelle klassifizieren zwischen 76,3 und 88,8 Prozent der Unternehmen korrekt.

Trennschärfe, desto sicherer können Unternehmen der Klasse der solventen bzw. insolventen Unternehmen zugeordnet werden. In den folgenden Säulendiagrammen werden die *in-sample*- und *out-of-sample*-Häufigkeiten der prognostizierten Insolvenzwahrscheinlichkeiten sowohl für die logistische Regression als auch für das neuronale Netzwerk dargestellt. Dabei wird unterschieden, ob die Insolvenzprognosen für die solventen Unternehmen (graue Säulen) oder für die insolventen Unternehmen (schwarze Säulen) durchgeführt wurden.

Wie sich aus den Säulendiagrammen in Abbildung 4 und 5 ablesen läßt, erzeugt das Netzwerk Insolvenzprognosen, die sowohl *in-sample* als auch *out-of-sample* für insolvente Unternehmen deutlich näher an 1 und für solvente Unternehmen deutlich näher an 0 liegen. Dies läßt sich daran erkennen, daß sich die Masse der grauen Säulen beim neuronalen Netzwerk gegenüber der logistischen Regression nach links, die Masse der schwarzen Säulen nach rechts verschoben hat. Mit Hilfe des nichtlinearen neuronalen Netzwerks ist es also möglich, deutlich trennschärfere Modelle zu erzeugen als mit dem logistischen Regressionsmodell.

Die höhere Trennschärfe des neuronalen Netzwerks wird auch durch die durchschnittliche *Deviance* reflektiert, die möglichst nahe an Null liegen sollte. Das neuronale Netzwerkmodell liefert *in-sample* einen Wert von -0,5236 und *out-of-sample* einen Wert von -0,5731 und damit bessere Ergebnisse als die logistische Regression, die *in-sample* einen Wert von -0,5532 und *out-of-sample* einen Wert von -0,6094 erreicht. Gleiches gilt auch für den durchschnittlichen *Goodness of Fit*-Wert $\frac{1}{n}Z^2$. Diese Ergebnisse werden durch die Rangkorrelationskoeffizienten Somers' *D*, Godman-Kruskals *Gamma* und Kendalls *Tau-a* noch einmal bestätigt.

Abschließend läßt sich feststellen, daß das neuronale Netzwerk der logistischen Regression hinsichtlich aller Gütemaße sowohl *in-sample* als auch *out-of-sample* überlegen ist. Mit Hilfe des neuronalen Netzwerks ist es also gelungen, die bereits guten Resultate der logistischen Regression weiter zu verbessern.

4.4 Analyse und Interpretation der Wirkungszusammenhänge

Vergleicht man das logistische Regressionsmodell mit dem neuronalen Netzwerkmodell, so wird deutlich, in welchen Bereichen das Insolvenzrisiko durch die übliche Vorgehensweise unter- bzw. überschätzt wird. Die folgenden Grafiken veranschaulichen die Wirkung der verschiedenen Variablen auf die Prognose der Insolvenzwahrscheinlichkeiten.

Abbildungen 6a und 6b zeigen den Einfluß des Unternehmensalters und der Umsatzentwicklung auf die Insolvenzwahrscheinlichkeit, wenn die Gesellschafter unbeschränkt haften, das Unternehmen nicht dem Handel oder dem verarbeitenden Gewerbe angehört und der Unternehmer Meister oder Akademiker ist. Für das neuronale Netzwerk wird der nicht-lineare Einfluß des Unternehmensalters im Bereich sinkender Umsätze deutlich. In jungen Jahren ist das Risiko hoch, sinkt anschließend ab, um nach etwa dreißig Jahren wieder leicht anzusteigen. Dieses Phänomen wurde bereits in den empirischen Untersuchungen von Brüderl/Schüssler (1990) und Szczesny/Korn (1995) festgestellt und ist kompatibel mit den evolutionstheoretischen Hypothesen zum Unternehmensalter, der *Liability of*

Newness und der *Liability of Aging*.²⁶ Die *Liability of Newness*-Hypothese besagt, daß das Insolvenzzisiko mit steigendem Alter abnimmt. Es wird argumentiert, daß Unternehmen in jungen Jahren noch Legitimationsprobleme gegenüber ihrer Umwelt haben. Neugegründete Unternehmen haben es schwer, Unterstützung zu mobilisieren. Banken neigen dazu, vorrangig langjährige Kunden zu bedienen. Viele Jungunternehmer können sich daher nicht behaupten und scheiden aus dem Markt aus. Mit der *Liability of Aging*-Hypothese wird ausgedrückt, daß Unternehmen mit steigendem Alter zunehmend unflexibler werden, wodurch sich das Insolvenzzisiko der Unternehmen entsprechend erhöht.

Wie sich aus dem Vergleich der Abbildungen 6a und 6b der beiden Modelle ergibt, würde eine auf dem logistischen Regressionsmodell aufbauende Kreditvergabe das Insolvenzzisiko von Unternehmen kurz nach der Gründung unterschätzen, das Risiko von Unternehmen mit einem Alter von um die 30 Jahren jedoch überschätzen. Unternehmen dieses Alters würden in einer Konditionengestaltung, die auf dem logistischen Regressionsmodell beruhen würde, ungerechtfertigt benachteiligt werden.

Um den Einfluß der Indikatorvariablen beschränkte Haftung, Zugehörigkeit zum verarbeitenden Gewerbe und Bildungsabschluß des Unternehmers zu verdeutlichen, wird im Vergleich zu den Grafiken 6a und 6b jeweils eine Indikatorvariable in den Grafiken 6c bis 6h auf Eins gesetzt. Dabei ist insbesondere der nichtlineare Einfluß der Indikatorvariablen verarbeitendes Gewerbe und beschränkte Haftung interessant. Diese wirken in Abhängigkeit der anderen Merkmalsausprägungen verschieden stark.

Die Haftungsbeschränkung wirkt im neuronalen Netzwerk (Abbildung 3) genau wie im logistischen Regressionsmodell lediglich linear. Aus diesem Grund verschieben sich die Insolvenzwahrscheinlichkeiten in den Abbildungen 6c und 6d parallel zu den Insolvenzwahrscheinlichkeiten in den Abbildungen 6a und 6b. Das höhere Insolvenzzisiko bei einer Haftungsbeschränkung ist plausibel. Stiglitz und Weiß (1981) beschreiben, daß eine Haftungsbeschränkung der Eigentümer eines Unternehmens Anreize zur Verfolgung risikoreicherer Projekte liefert, da diese höhere Gewinne erwarten lassen.²⁷ Mit einem steigenden Risiko der Projekte wächst allerdings auch das Risiko eines Scheiterns und die Gefahr, zahlungsunfähig zu werden.

Die Zugehörigkeit eines Unternehmens zum verarbeitenden Gewerbe wirkt im neuronalen Netzwerkmodell dagegen nichtlinear. Dies wird im Vergleich der Abbildungen 6b und 6f offenbar. Während sich die Wahrscheinlichkeiten beim logistischen Regressionsmodell in den Abbildungen 6a und 6e lediglich parallel verschieben, erhöht sich bei Unternehmen des verarbeitenden Gewerbes in Abbildung 6f vor allem das Risiko älterer Unternehmen. In Klein- und Mittelbetrieben ist ein steigendes Unternehmensalter häufig gleichbedeutend mit einem steigenden Alter des Unternehmers. Steht kein geeigneter Nachfolger zur Verfügung oder hält der Unternehmer an seinen gewohnten Handlungsweisen fest, sinkt im Alter die Bereitschaft zu notwendigen Investitionen, Innovationen oder Anpassungen an veränderte Marktstrukturen.²⁸ Bei so einer Verhaltensweise sind insbesondere Unternehmen des verarbeitenden Gewerbes sehr stark insolvenzgefährdet, da das Überleben solcher Unternehmen üblicherweise von einer großen Modernität der Produktion abhängig ist.

²⁶Vgl. Aldrich/Auster (1986) und Freeman/Caroll/Hannan (1983).

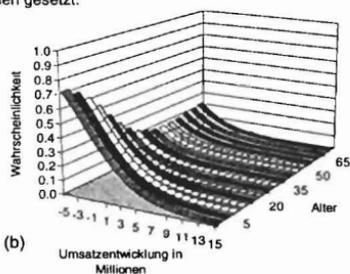
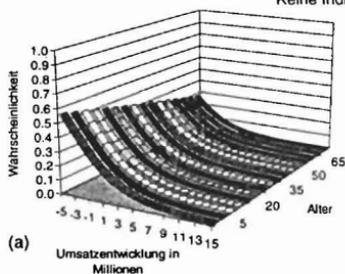
²⁷Vgl. dazu auch Harhoff/Stahl (1995)

²⁸Endress (1979)

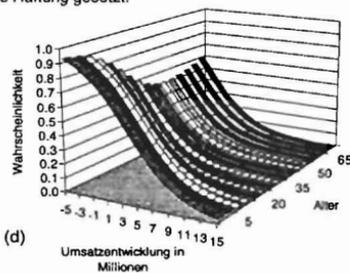
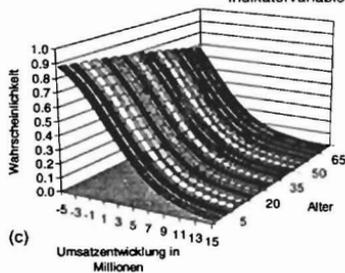
Logistische Regression

Neuronales Netzwerk

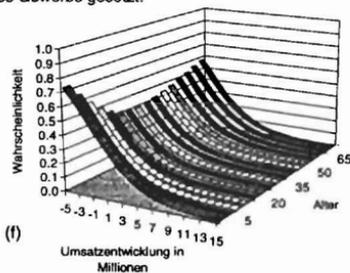
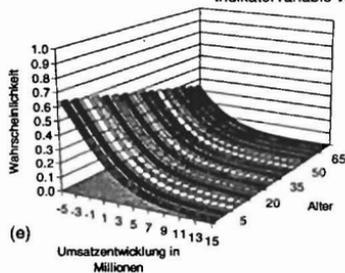
Keine Indikatorvariablen gesetzt:



Indikatorvariable *beschränkte Haftung* gesetzt:



Indikatorvariable *verarbeitendes Gewerbe* gesetzt:



Indikatorvariable *Bildungsabschluß* gesetzt:

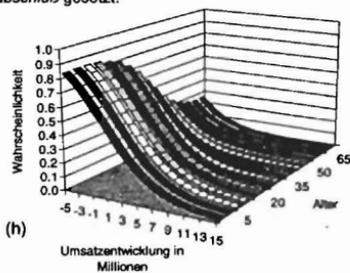
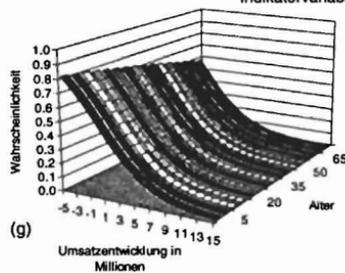


Abbildung 6: Insolvenzwahrscheinlichkeit in Abhängigkeit des Unternehmensalters, der Umsatzentwicklung und verschiedener Indikatorvariablen.

In den Abbildung 6g und 6h wird die Wirkung auf die Insolvenzwahrscheinlichkeit veranschaulicht, wenn der Unternehmer lediglich eine Lehre abgeschlossen hat. Während die Insolvenzwahrscheinlichkeit bei der logistischen Regression wiederum nur linear ansteigen kann, wird beim neuronale Netzwerk in Abbildung 6h im Vergleich zu Abbildung 6b deutlich, daß das Insolvenzrisiko insbesondere kurz nach der Gründung des Unternehmens sehr hoch ist. Bei älteren Unternehmen scheint es für die Bestimmung der Insolvenzwahrscheinlichkeit dagegen unerheblich zu sein, über welchen Bildungsabschluß der Unternehmer verfügt. Dieses Ergebnis ist naheliegend. In organisationssoziologischen Untersuchungen hat sich ergeben, daß die Produktivität eines Unternehmens mit einem höheren Humankapital steigt.²⁹ Da die Berufs- und Lebenserfahrung des Unternehmers mit zunehmendem Alter wächst, und damit das Humankapital in seinem Unternehmen zunimmt, kann der eigentliche Bildungsabschluß ab einem bestimmten Zeitpunkt vernachlässigt werden.

Insgesamt zeigt sich, daß sich mit Hilfe eines neuronalen Netzwerks sehr nuancierte Aussagen über die Wirkungszusammenhänge zwischen den Unternehmensmerkmalen und dem Insolvenzrisiko treffen lassen. Würden solche nichtlinearen Abhängigkeiten bei der Bewertung des Insolvenzrisikos berücksichtigt, sollten Kreditgeber zu einer erheblich differenzierteren Konditionengestaltung gelangen als dies mit linearen logistischen Modellen möglich ist.

5 Zusammenfassung

In dieser Untersuchung haben wir gezeigt, wie sich Insolvenzprognosen für kleine und mittlere Unternehmen durchführen lassen. Da kleine und mittlere Unternehmen entweder nicht publizitätspflichtig sind oder ihrer Publizitätspflicht häufig erst verspätet nachkommen, wurden in dieser Studie ausschließlich solche Daten verwendet, die einem Kreditgeber in der Regel zum Kreditvergabezeitpunkt vorliegen. Daten zur Ertragslage oder zur Gewinn- und Verlustrechnung wurden also nicht berücksichtigt. Trotz dieser Restriktion erreichen wir mit den eingesetzten statistischen Verfahren *out-of-sample*-Trefferquoten von über 70 Prozent.

Für die Insolvenzprognose wurden das logistische Regressionsmodell und neuronale Netzwerke eingesetzt. Im Gegensatz zu der in den meisten anderen Untersuchungen verwendeten multivariaten linearen Diskriminanzanalyse sind diese Verfahren auch bei der Auswertung qualitativer Daten sehr gut geeignet.

Beide Modelle liefern hervorragende Ergebnisse bei der Prognose von Insolvenzwahrscheinlichkeiten. Das lineare logistische Regressionsmodell erreicht *out-of-sample* eine Gesamttrefferquote von 68,3 Prozent, das neuronale Netzwerk eine Gesamttrefferquote von 73,3 Prozent. Im Vergleich zu der logistischen Regression konnte durch das neuronale Netzwerk aber nicht nur die Trefferquote, sondern vor allem auch die Trennschärfe der Prognose erheblich verbessert werden. Dadurch wurde die Sicherheit der Insolvenzprognose sowohl für insolvente als auch für solvente Unternehmen wesentlich erhöht.

²⁹Vgl. Brüderl/Preisendörfer/Ziegler (1992).

Würde die Entscheidung einer Kreditvergabe auf der Grundlage der in dieser Arbeit vorgestellten Modelle getroffen, kann mit Hilfe des neuronalen Netzwerks eine wesentlich differenzierter Konditionengestaltung erreicht werden als mit der logistischen Regression. Das Netzwerk ist nämlich im Gegensatz zur logistischen Regression in der Lage, auch nichtlineare Wirkungszusammenhänge zwischen den Merkmalsausprägungen und der Insolvenzwahrscheinlichkeit von Unternehmen herzustellen. Dabei hat das neuronale Netzwerk insbesondere die Möglichkeit, gegenseitige Abhängigkeiten unter den Merkmalsausprägungen abzubilden, wodurch die nichtlineare Modellierung von Indikatorvariablen gewährleistet ist. Es hat sich ergeben, daß das Insolvenzrisiko von Unternehmen in jungen Jahren groß ist, danach abnimmt und in höherem Alter wieder ansteigt. Der Bildungsabschluß des Unternehmers verliert mit zunehmendem Unternehmensalter an Bedeutung.

Zur Spezifikation beider Modelle wurden Inferenztechniken eingesetzt. Für die logistische Regression ist die Verwendung von Inferenztechniken Standard. Die Verwendung von Inferenztechniken in neuronalen Netzwerken ist jedoch neu. Wie sich durch die sehr gute *out-of-sample* Prognosequalität des neuronalen Netzwerks gezeigt hat, lassen sich Inferenztechniken in neuronalen Netzwerken erfolgreich dazu verwenden, bestehende Modelle zu verbessern und *Overfitting* zu vermeiden.

Literatur

- Aldrich, Howard E. / Auster, Ellen (1986): *Even Drafts Started Small: Liability of Age and Size and their Strategic Implications*. In: Aldrich H.E. (1986): *Population Perspectives on Organizations*. Uppsala, 29–60.
- Anders, Ulrich / Korn, Olaf / Schmitt, Christian (1996): *Improving the Pricing of Options — A Neural Network Approach*. ZEW Discussion Paper 96-04.
- Anders, Ulrich / Korn Olaf (1996): *Model Selection in Neural Networks*. ZEW Discussion Paper 96-21.
- Anders, Ulrich / Schwamm, Daniel (1996): *Prediction Intervals for Neural Networks*. ZEW Discussion Paper (bevorstehend).
- Anders, Ulrich (1995): *Neuronale Netzwerke in der Ökonometrie*. ZEW Discussion Paper 95-26.
- Anders, Ulrich (1996a): *Was neuronale Netze wirklich leisten*. Die Bank, 3, 162–165.
- Anders, Ulrich (1996b): *Statistische Grundlagen neuronaler Netzwerke*. In: Schröder, Michael (Hrsg.) (1988): *Quantitative Verfahren im Finanzbereich: Methoden und Anwendungen*. ZEW Wirtschaftsanalysen Bd. 5, Nomos Verlag.
- Baetge, Jörg et al. (1994): *Bonitätsbeurteilung von Jahresabschlüssen nach neuem Recht (HGB 1985) mit Künstlichen Neuronalen Netzen auf der Basis von Clusteranalysen*. Der Betrieb, 47/7, 337–343.
- Baetge, Jörg (1994): *Rationalisierung des Firmenkundengeschäfts auf der Basis empirisch-statistisch gewonnener Bilanzbonität*. vbo-Informationen, 1, 4–25.
- Bischoff, Rainer / Bleile, Clemens / Graafls, Jürgen (1992): *Der Einsatz Neuronaler Netze zur betriebswirtschaftlichen Kennzahlenanalyse*. Forschungsbericht der Fachhochschule Furtwangen herausgegeben von Zahradnik W., WEKA-Verlag.
- Bleier, Ernst (1984): *Unternehmensanalyse aus dem Jahresabschluß – Risikoklassifikation mit Hilfe von Diskriminanzfunktionen*. Service-Fachverlag.
- Brüderl, Josef / Schüssler, Rudolf (1990): *Organizational Mortality: the Liability of Newness and Adolescence*. Administrative Science Quarterly, 24/3, 530–547.
- Brüderl, Josef / Preisendörfer, Peter / Ziegler, Rudolf (1992): *Survival Chances of Newly Founded Business Organizations*. American Sociological Review, 57, 227–241.
- Cox, David R. (1970): *The Analysis of Binary Data*. Methuen.
- Endress, Ruth (1979): *Insolvenz- und Risikofaktoren im Mittel- und Kleinbetrieb*. RKW, Frankfurt am Main.

- Fahrmeir, Ludwig / Hamerle, Alfred (1984): *Multivariate statistische Verfahren*. Walter de Gruyter.
- Freeman, John / Carroll, Glenn R. / Hannan, Michael T. (1983): *The Liability of Newness: Age Dependence in Organizational Death Rates*. American Sociological Review 48, 692-710.
- Goede, Karl / Weinrich Günter (1996): *Bessere Kreditentscheidungen durch neuronale Netze?* Die Bank, 7/96, 420-423.
- Greene, William H. (1993): *Econometric Analysis*. Macmillan.
- Hand, D.J. (1981): *Discrimination and Classification*. John Wiley & Sons.
- Harhoff, Dietmar / Stahl, Konrad / Woywode, Michael (1994): *Growth and Exit of West German Firms*. ZEW Discussion Paper 95-15.
- Harhoff, Dietmar / Stahl, Konrad (1995): *Unternehmens- und Beschäftigungsdynamik in Westdeutschland zum Einfluß von Haftungsregeln und Eigentümerstruktur*. Ifo-Studien, 41/1, 17-50.
- Hauschildt, Jürgen (1988): *Krisendiagnose durch Bilanzanalyse*. Otto Schmidt.
- Hornik, Kurt / Stinchcombe, Max / White, Halbert (1989): *Multilayer Feedforward Networks are Universal Approximators*. Neural Networks, 2, 359-366.
- Hosmer, David W. / Lemeshow, Stanley (1989): *Applied Logistic Regression*. John Wiley & Sons.
- Hüls, Dagmar (1995): *Früherkennung insolvenzgefährdeter Unternehmen*. IDW-Verlag.
- Häußler, Walter M. (1981): *Punktbewertungen bei Kredit Scoringsystemen. Über Verfahren der Punktbewertung und Diskrimination mit Anwendung auf Kredit Scoringsysteme*. Fritz Knapp Verlag.
- Keysberg, Gerhard (1989): *Die Anwendung der Diskriminanzanalyse zur statistischen Kreditwürdigkeitsprüfung im Konsumentenkreditgeschäft*. Müller Botermann.
- Krause, Clemens (1995): *Kreditwürdigkeitsuntersuchung mit Neuronalen Netzen*. IDW-Verlag.
- McLachlan, Geoffrey J. (1992): *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons.
- Odom, Marcus D. / Sharda, Ramesh (1990): *A Neural Network Model for Bankruptcy Prediction*. IJCNN International Joint Conference on Neural Networks. SOS Printing, II, 163-167.
- Phillips, P.C.B. (1989): *Partially Identified Econometric Models*. Econometric Theory, 5, 181-240.

- Press, S. James / Wilson, Sandra (1978): *Choosing between Logistic Regression and Discriminant Analysis*. Journal of the American Statistical Association, 73, 699–705.
- Pytlik, Martin (1995): *Diskriminanzanalyse und Künstliche Neuronale Netze zur Klassifizierung von Jahresabschlüssen – Ein empirischer Vergleich*. Peter Lang Verlag.
- Rehkugler, Heinz / Poddig, Thorsten (1992): *Klassifikation von Jahresabschlüssen mittels Multilayer-Perceptrons – Erste Ergebnisse und weiterführende Fragestellungen*. Bamberger Betriebswirtschaftliche Beiträge, 87/1992.
- Rehkugler, Heinz / Schmidt-von Rhein, Andreas (1993): *Kreditwürdigkeitsanalyse und -prognose für Privatkundenkredite mittels statistischer Methoden und Künstlicher Neuronaler Netze – Eine empirisch-vergleichende Studie*. Bamberger Betriebswirtschaftliche Beiträge, 93/1993.
- Rehkugler, Heinz / Zimmermann, Hans Georg (1994): *Neuronale Netze in der Ökonomie*. Vahlen.
- Ronning, Gerd (1991): *Mikroökometrie*. Springer-Verlag.
- Sarle, Warren S. (1994): *Neural Networks and Statistical Models*. Proceedings of the Nineteenth Annual SAS Users Group International Conference.
- Schumann, Matthias / Lohrbach, Thomas / Bähns, Peter (1995): *Versuche zur Kreditwürdigkeitsprüfung mit Künstlichen Neuronalen Netzen*. Arbeitspapier der Georg-August-Universität Göttingen, 2/92.
- Szczesny, Andrea / Korn, Olaf (1995): *Insolvenzanalyse mit neuronalen Netzen*. ZEW-Newsletter, 2/96, 11-14.
- Teräsvirta, Timo / Lin, Chien-Fu / Granger, Clive W.J. (1993): *Power of the Neural Network Linearity Test*. Journal of Time Series Analysis, 14 (2), 209–220.
- White, Halbert (1989a): *Some Asymptotic Results for Learning in Single Hidden-Layer Feedforward Network Models*. Journal of the American Statistical Association, 84 (404), 1003–1013.
- White, Halbert (1989b): *Learning in Neural Networks: A Statistical Perspective*. Neural Computation, 1, 425–464.
- White, Halbert (1989c): *An Additional Hidden Unit Test for Neglected Non-linearity in Multilayer Feedforward Networks*. Proceedings of the International Joint Conference on Neural Networks, Washington, DC. SOS Printing, II, 451–455.
- White, Halbert (1992): *Artificial Neural Networks*. Blackwell.