

regio's. Dit betekent concreet dat de vraag naar woningen in bijv. West-Vlaanderen onder meer afhangt van de relatieve prijs van woningen in West-Vlaanderen ten opzichte van de prijs in Limburg. Dit lijkt weinig waarschijnlijk. In eerste instantie werd dan overwogen om de kruisprijs-effecten tussen niet-aangrenzende regio's te elimineren (door het opleggen van nulrestricties op de betrokken θ_{ij}). Uiteindelijk werd echter gekozen voor een nog eenvoudiger vorm, waarbij geen enkel paar goederen specifieke substituten of complementen vormt. Dit is de hypothese van onafhankelijkheid van preferenties ("preference independence"), die impliceert dat $\theta_{ij} = 0$ voor $i \neq j$. De vraagvergelijkingen krijgen dan de eenvoudiger vorm:

$$\bar{w}_{i,t} d \ln X_{i,t} = \theta_i DX_t + \phi \theta_i (d \ln P_{i,t} - DP_t) + \delta_i d \ln DG_t$$

($i = 1, \dots, 6$)

Paul Pompe *

Jan Bilderbeek *

Faillissementspredictie: een vergelijking tussen lineaire discriminantanalyse en neurale netwerken

Trefwoorden: faillissementspredictie; neurale netwerken; praktische toepassing

In dit artikel wordt een onderzoek beschreven met als doel het vergelijken van de prestaties van de methoden lineaire discriminantanalyse (LDA) en neurale netwerken bij het afleiden van een model om faillissementen te voorspellen. Beide methoden werden toegepast op een grote dataverzameling met jaarrekeningen afkomstig van lopende en gefailleerde Belgische ondernemingen. We gebruikten een passend onderzoeksontwerp teneinde de verschillen in presteren op significantie te kunnen toetsen. De methoden bleken opvallend gelijkwaardig te presteren. Alleen in het geval van een zeer geringe hoeveelheid trainingsdata deden de neurale netwerken het beter dan LDA. In het artikel wordt één model in een dusdanige vorm gepresenteerd dat gebruik van het model in de praktijk mogelijk is. Getoond wordt op welke verschillende wijzen het model eenvoudig kan worden toegepast. Ook wordt aangegeven hoe de prestaties van een model worden beïnvloed door de leeftijd en grootte van de te classificeren onderneming en door de macro-economische situatie.

* Universiteit Twente, Vakgroep Financieel Management en Bedrijfseconomie

1. Inleiding

In de afgelopen decennia is veel onderzoek verricht naar het door middel van statistische modellen voorspellen van faillissementen (bijv. Altman, 1968; Bilderbeek, 1977; Dimitras et al., 1996). Ook in België, het land waaruit de gegevens voor het beschreven onderzoek afkomstig zijn, zijn dergelijke modellen afgeleid (Ooghe et al., 1993; Ooghe et al., 1995). Een model voor het voorspellen van faillissementen heeft tot doel de relatie te beschrijven tussen een naderend faillissement en een aantal verklarende financiële ratio's. Deze ratio's kunnen berekend worden met informatie uit de jaarrekening van een onderneming. Het uiteindelijke doel is een instrument te verkrijgen om een faillissement vroegtijdig te kunnen signaleren. Vaak wordt bij het afleiden van een faillissementsmodel gebruikgemaakt van de statistische methode lineaire discriminantanalyse (LDA). Sinds enige jaren bestaat er aandacht voor een nieuwe methode: neurale netwerken (bijv. O'Leary, 1998; Barniv et al., 1997; Pompe en Feelders, 1997; Boritz en Kennedy, 1995; Altman et al., 1994; Fletcher en Goss, 1993). De met neurale netwerken bereikte resultaten zijn wisselend; veel onderzoekers rapporteren over betere faillissementsvoorspellingen door het gebruik van neurale netwerken, maar in andere onderzoeken zijn de prestaties niet beter dan die van de klassieke methoden zoals lineaire discriminantanalyse. Het doel van het onderzoek dat in dit artikel wordt beschreven, is het vergelijken van de prestaties van LDA en neurale netwerken bij het voorspellen van faillissementen in België. Beide methoden worden toegepast op een dataverzameling met jaarrekeningen afkomstig van lopende en gefailleerde Belgische ondernemingen. Het artikel is als volgt opgebouwd. In paragraaf 2 gaan we nader in op de verzamelde data en in paragraaf 3 worden de methoden LDA en neurale netwerken beschreven. Paragraaf 4 gaat over het onderzoeksontwerp en in paragraaf 5 worden de resultaten aan de orde gesteld. Na de presentatie van een model in paragraaf 6 en het plaatsen van twee kanttekeningen bij het gebruik van het model in paragraaf 7, sluiten we af met conclusies in paragraaf 8.

2. Data

Al langer dan tien jaar plaatst de centrale bank van België jaarrekeningen op cd-rom van ondernemingen die onderworpen zijn aan de open-

baarmaking van hun jaarrekening. Onder deze ondernemingen bevinden zich onder meer alle naamloze vennootschappen en alle besloten vennootschappen met beperkte aansprakelijkheid. In dit onderzoek is voor het verzamelen van data gebruikgemaakt van deze cd-roms. We konden beschikken over de cd-roms uit de jaren 1988 tot en met 1994. Op een cd-rom staan per onderneming de jaarrekeningen van enkele kalenderjaren. Alleen jaarrekeningen van vennootschappen uit de sector industrie (NACE-code 2, 3 of 4) zijn verzameld. Bij het afleiden van een faillissementsmodel wordt gebruikgemaakt van een steekproef met jaarrekeningen. De steekproef wordt getrokken uit de totale populatie. Na het afleiden van een model wordt voor het inschatten van de voorspellingskracht van het model een controlesteekproef gebruikt. In dit artikel wordt de steekproef aangeduid met trainingsverzameling en de controlesteekproef met testverzameling. Het afleiden van een model wordt trainen genoemd en bij het gebruik van de testverzameling wordt er gesproken van testen. In de trainings- en testverzameling bevinden zich twee klassen jaarrekeningen: de klasse "lopend" en de klasse "failliet". Jaarrekeningen van de eerste klasse zijn afkomstig van ondernemingen die niet binnen een zekere periode failliet zijn gegaan. Jaarrekeningen van de tweede klasse horen bij ondernemingen die een bepaald aantal jaren na het jaar waar de jaarrekening betrekking op heeft, zijn gefailleerd.¹ De jaarrekeningen waarvoor dit aantal jaren gelijk is aan i ($i = 1, 2, 3, \dots$) worden aangeduid met ig_i . We onderscheiden vijf groepen jaarrekeningen ig_i , met $i = 1, 2, 3, 4, 5$. Een jaarrekening in ig_1 is altijd de laatst gepubliceerde jaarrekening vóór het faillissement. Het jaar na het jaar waar de laatst gepubliceerde jaarrekening betrekking op heeft, wordt daarmee beschouwd als het jaar van failleren. Hoewel het formeel-juridische tijdstip van het faillissement niet in dat jaar hoeft te liggen, zien we het niet langer publiceren van jaarrekeningen door een onderneming als het feitelijke moment van failleren. Van deze keuze is niet altijd sprake in onderzoek naar faillissementspredictie; vaak wordt het jaar van failleren gelijkgesteld met het jaar van het juridische tijdstip van het faillissement. De tijdspanne tussen de sluitingsdatum van de laatst gepubliceerde jaarrekening en het formeel-juridische moment van het faillissement bleek bij ongeveer 90% van de onderzochte ondernemingen een half tot twee jaar te bedragen.

Wat betreft de klasse "failliet" zijn de hoeveelheden jaarrekeningen beschikbaar die vermeld staan in tabel 1. Het zijn bijna alle jaarrekeningen

¹ Deze ondernemingen hadden op de cd-roms de rechtstoestand "faillissement".

van deze klasse die op de cd-roms aanwezig waren. Onder het kopje "Jaar i " staat het aantal jaarrekeningen jg_i . Elke jaarrekening van de klasse "failliet" heeft betrekking op een van de volgende negen kalenderjaren: 1985, 1986, ..., 1992 of 1993. De hoeveelheden in de tabel nemen bij een hogere waarde voor i gestaag af. Hiervoor bestaan twee redenen. In de eerste plaats bevinden zich in de groepen jg_i met een hoge waarde voor i vanzelfsprekend geen jaarrekeningen van ondernemingen die te kort hebben bestaan om in deze groepen een jaarrekening te hebben. De belangrijkste reden is dat in het onderzoek alleen kennis bestaat over faillissementen die plaatsvonden in de genoemde tijdsperiode van negen jaar. Daarom zijn de enige jaarrekeningen in bijvoorbeeld jg_5 de jaarrekeningen die betrekking hebben op een kalenderjaar aan het begin van de tijdsperiode en die afkomstig zijn van ondernemingen die aan het einde van de tijdsperiode faillieerden. Wat betreft de klasse "lopend" wordt er gebruikgemaakt van jaarrekeningen die betrekking hebben op een van de volgende vier kalenderjaren: 1988, 1989, 1990 of 1991. Op zich kunnen uit één kalenderjaar meer dan voldoende data van deze klasse worden verkregen. Om een wat langere tijdsperiode te representeren, is echter voor vier kalenderjaren gekozen. We gebruiken 1266 aselect gekozen jaarrekeningen, van elk kalenderjaar ongeveer evenveel. De jaarrekeningen zijn afkomstig van 1266 verschillende ondernemingen.² Wat betreft de klasse "failliet" worden voor elk jaar opnieuw de prestaties van LDA en neurale netwerken met elkaar vergeleken. Voor zo'n vergelijking gebruiken we, naast de jaarrekeningen van de klasse "failliet" uit dat jaar, ook jaarrekeningen van de klasse "lopend". De laatstgenoemde jaarrekeningen zijn altijd de 1266 jaarrekeningen van de klasse "lopend". Wel houden we in elke vergelijking het aantal jaarrekeningen van beide klassen gelijk. Bijvoorbeeld: voor de vergelijking in jaar 4 gebruiken we de 600 jaarrekeningen van de klasse "failliet" en verwijderen we daarnaast aselect 666 jaarrekeningen van de klasse "lopend" uit de totale verzameling van 1266.

2 Elk van deze ondernemingen was op de cd-rom die de gekozen jaarrekening bevat zonder een specifieke rechtstoestand. Verder is bij elke onderneming gecontroleerd dat deze op een recentere cd-rom (tot en met de cd-rom van september 1994) nog steeds zonder specifieke rechtstoestand was.

Tabel 1.

	Failliet					Lopend
	Jaar 1	Jaar 2	Jaar 3	Jaar 4	Jaar 5	
Aantal	1266	956	758	600	446	1266

De aantallen jaarrekeningen in tabel 1 zijn tot stand gekomen na drie reducties in de data en vertegenwoordigen 87% van de originele dataset in het geval van klasse "failliet" en 94% van de originele dataset in het geval van klasse "lopend". De eerste reductie betrof het verwijderen van jaarrekeningen die betrekking hebben op een periode anders dan twaalf maanden (1% reductie voor beide klassen). Vervolgens werden de jaarrekeningen verwijderd met rekenkundige fouten (11% reductie voor klasse "failliet" en 3% reductie voor klasse "lopend"). Ten slotte werden jaarrekeningen verwijderd die niet voor elke gebruikte financiële ratio een waarde hadden (zie vierde voetnoot, 1% reductie voor klasse "failliet" en 2% voor klasse "lopend"). Een voorbeeld van een rekenkundige fout is dat het eigen vermogen plus het vreemd vermogen niet gelijk is aan het totaal vermogen. In de balans en de resultatenrekening is gekeken naar rekenkundige fouten in alle posten en in de toelichting alleen naar fouten in de posten die gebruikt worden bij het berekenen van de financiële ratio's. Het grote verschil tussen beide klassen wat betreft het percentage jaarrekeningen met rekenkundige fouten maakt duidelijk dat het wel of niet aanwezig zijn van fouten al iets zegt over de kans op een bepaalde klasse. In dit onderzoek is verder niets gedaan met deze constatering. Bij de vergelijking van LDA met neurale netwerken zal zowel de trainingsverzameling als de testverzameling altijd bestaan uit evenveel jaarrekeningen van beide klassen. Deze verhouding is niet representatief voor de verhouding tussen de klassen in de populatie. In werkelijkheid bedraagt de kans dat een onderneming binnen een aantal jaren failliet gaat slechts enkele procenten. Door rekening te houden met de kosten van misclassificatie kan de gehanteerde verhouding worden verantwoord.³

3 De kosten van het onjuist classificeren van een jaarrekening van de klasse "failliet" (c_f) worden over het algemeen als veel hoger beschouwd dan de kosten van het onjuist classificeren van een jaarrekening van de klasse "lopend" (c_l). De exacte verhouding tussen deze kosten (c_f/c_l) hangt af van het specifieke gebruik van het model. Voor dit onderzoek nemen we aan dat $c_f/c_l = p_l/p_f$ met p_i de proportie jaarrekeningen van de klasse " i " in de populatie. Aan de hand van de stelling van Bayes kan worden aangetoond dat deze aanname gelijk staat met een situatie waarbij beide klassen even vaak

In het onderzoek wordt geen bepaald jaarrekeningschema uitgesloten; zowel jaarrekeningen volgens het verkorte schema als jaarrekeningen volgens het volledige schema doen mee. Ondernemingen die publiceren volgens het verkorte schema zijn niet verplicht om de waarde van de omzet in de jaarrekening te vermelden. Als gevolg daarvan is de omzet maar in ongeveer 60% van de jaarrekeningen bekend. Om geen data te verliezen, is besloten om ratio's die voor hun berekening de waarde van de omzet nodig hebben, niet te gebruiken. Bij het uitrekenen van de ratio's wordt de berekeningswijze van Ooghe en Van Wymeersch (1994) gevolgd. In het onderzoek wordt elke jaarrekening beschreven door tien ratio's.⁴ De keuze van deze tien ratio's is gebaseerd op literatuur- en vooronderzoek. Het vooronderzoek bestond uit korte experimenten op gedeelten van de data en gaf een indicatie over welke ratio's het meest voorspellend waren. Hierbij is gebruikgemaakt van de stapsgewijze selectie van ratio's (minimalisatie van Wilks' lambda, zie SPSS-handleiding, 1993). Om een model te verkrijgen dat de alom onderscheiden dimensies van de financiële positie (rentabiliteit, liquiditeit, activiteit en solvabiliteit) van een onderneming beoordeelt, is ervoor gezorgd dat ra-

voorkomen in de populatie ($p_1 = p_f = 0,5$) en de beide soorten kosten gelijk zijn aan elkaar ($c_f = c_j$). Door in dit onderzoek van deze situatie uit te gaan, kunnen we een klas-senverhouding van 0,5:0,5 in de trainings- en testverzameling rechtvaardigen.

- 4 De berekening van de tien ratio's verloopt als volgt. De gebruikte codes zijn posten uit de balans, de resultatenrekening en de toelichting. Omwille van de vergelijkbaarheid worden ratio's bij beide soorten schema's op dezelfde wijze berekend (hoewel af en toe een verschillende code wordt gebruikt, bijvoorbeeld bij de berekening van de toegevoegde waarde). Bij verkort schema: $tw = 7061 + 6170$, $w = 8079 + 8279 + 8475 + 8089 + 8289 + 8485$. Bij volledig schema: $tw = 7074 - 60 - 61$, $w = 8071 + 8072 + 8073 + 8074 + 8271 + 8272 + 8273 + 8274 + 8275 + 8276 + 661 - 760 - 761$.
 Ratio 1 = $(7066 + 6670 - 65 - 9126 + 6314 + 6357 - 9125 + w) / 2058$.
 Ratio 2 = $(7066 + 6670 - 65 - 9126 + 780 + 680) / 2058$.
 Ratio 3 = absolute waarde van $((7067 + 6770) / 1015)$. De absoluut-tekens bij de derde ratio zijn het gevolg van het feit dat veel jaarrekeningen uit de jaren vlak vóór het faillissement een verlies combineren met een negatief eigen vermogen, wat vaak leidt tot een hoge waarde voor de ratio *winst na belastingen / eigen vermogen*.
 Ratio 4 = $(2958 - 29 - 4248 - 4923) / 2058$.
 Ratio 5 = $5458 / (2958 - 29)$.
 Ratio 6 = $tw / 2058$.
 Ratio 7 = $62 / tw$.
 Ratio 8 = $1015 / 2058$.
 Ratio 9 = $(13 + 140 + 141) / 2058$.
 Ratio 10 = $(7067 + 6770 - 780 - 680 + 6314 + 6357 - 9125 + w) / (16 + 1749)$.
 De enkele jaarrekeningen die voor ten minste één ratio een noemer gelijk aan 0 hebben, worden in het onderzoek uitgesloten. Vanzelfsprekend wordt niet uitgesloten een jaarrekening waarvan één van de ratio's de waarde 0 heeft. Dit laatste geldt bijvoorbeeld voor 6% van de 1266 jaarrekeningen van klasse "lopend". Bij deze jaarrekeningen zijn de personeelskosten 0 en dus heeft de zevende ratio de waarde 0.

tio's van iedere dimensie in de groep van tien ratio's aanwezig zijn. Het vooronderzoek gaf ook aan dat het wel of niet opnemen in de groep van ratio's van ratio's die voor de berekening de omzet nodig hebben, weinig of geen invloed had op de voorspellingskracht van een model. Dit betekent niet dat de omzetratio's individueel geen voorspellende waarde hebben, want die waarde hebben ze zeker. Binnen de groep van tien ratio's kon hun bijdrage aan de voorspellingskracht van een model echter worden overgenomen door andere ratio's.

Rentabiliteitsratio's

1. bruto resultaat / totaal vermogen
2. netto resultaat / totaal vermogen
3. | winst na belastingen / eigen vermogen |

Liquiditeitsratio's

4. netto bedrijfskapitaal / totaal vermogen
5. liquide middelen / vlottende activa

Activiteitsratio's

6. toegevoegde waarde / totaal vermogen
7. personeelskosten / toegevoegde waarde

Solvabiliteitsratio's

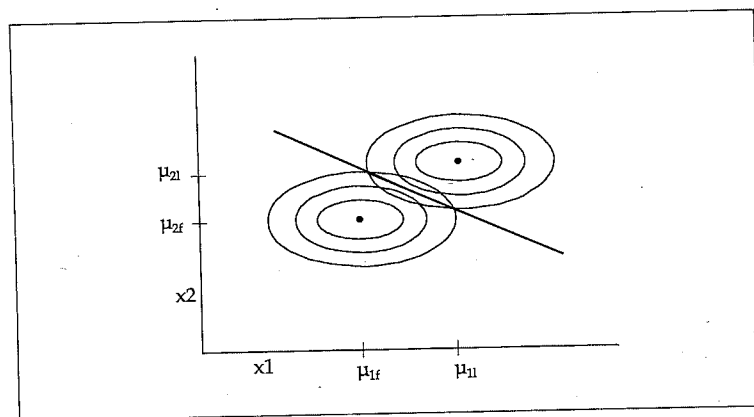
8. eigen vermogen / totaal vermogen
9. (reserves + overgedragen winst of verlies) / totaal vermogen
10. cashflow / vreemd vermogen

3. Methoden

Het verschil tussen LDA en neurale netwerken zal aan de hand van voorbeelden worden toegelicht. Voor een meer uitgebreide beschrijving van de methoden wordt verwezen naar bijv. James (1985) in het geval van LDA en naar bijv. Bishop (1995) en Haykin (1994) in het geval van neurale netwerken. Stel dat elke jaarrekening wordt beschreven door twee ratio's, x_1 en x_2 . De verdeling van jaarrekeningen kan nu worden aangegeven in een universum dat wordt opgespannen door de twee ratio's (figuur 1). LDA veronderstelt dat er bij de data sprake is van multivariaat normale verdelingen met gelijke covariantiematrices. Voor de jaarrekeningen in figuur 1 klopt deze veronderstelling, in de werkelijkheid echter voldoen jaarrekeningen over het algemeen niet of slechts in beperkte mate aan deze voorwaarden (zie bijv. Bilderbeek, 1977). Om de

vorm van de verdeling te kunnen aangeven, zijn in figuur 1 per klasse een aantal lijnen getekend. De lijnen verbinden punten waar de kans op voorkomen van jaarrekeningen van de betreffende klasse even groot is. Elke lijn vormt een ellips rond het punt met de grootste kans. Dit punt is het gemiddelde (μ_{1f} , μ_{2f}) in het geval van de klasse "failliet" en het gemiddelde (μ_{1l} , μ_{2l}) in het geval van de klasse "lopend". Naarmate een punt meer verwijderd is van het gemiddelde, neemt de kans op voorkomen af. Een model dat is afgeleid met een methode als LDA of neurale netwerken, is eigenlijk niet meer dan een opdeling van het universum in twee delen. Als een jaarrekening zich in het ene deel bevindt, dan beschouwt het model de jaarrekening als behorende tot de klasse "lopend"; zit een jaarrekening in het andere deel, dan wijst het model de jaarrekening toe aan de klasse "failliet". Zoals in voetnoot 3 is aangegeven, gaan we er in dit onderzoek van uit dat de populatie van jaarrekeningen uit evenveel jaarrekeningen van beide klassen bestaat en dat de kosten van het onjuist classificeren van een jaarrekening van de klasse "failliet" even hoog zijn als de kosten van het onjuist classificeren van een jaarrekening van de klasse "lopend". In die situatie deelt LDA het universum op door middel van de rechte lijn in figuur 1. Jaarrekeningen die onder de lijn liggen, worden als "failliet" geclassificeerd en jaarrekeningen boven de lijn als "lopend". Deze opdeling is optimaal, dat wil zeggen dat de totale kosten van onjuiste classificaties minimaal zijn. Er bestaat geen betere scheiding. LDA geeft een lineaire opdeling; in het geval van twee ratio's is de scheiding altijd een rechte lijn.

Figuur 1.

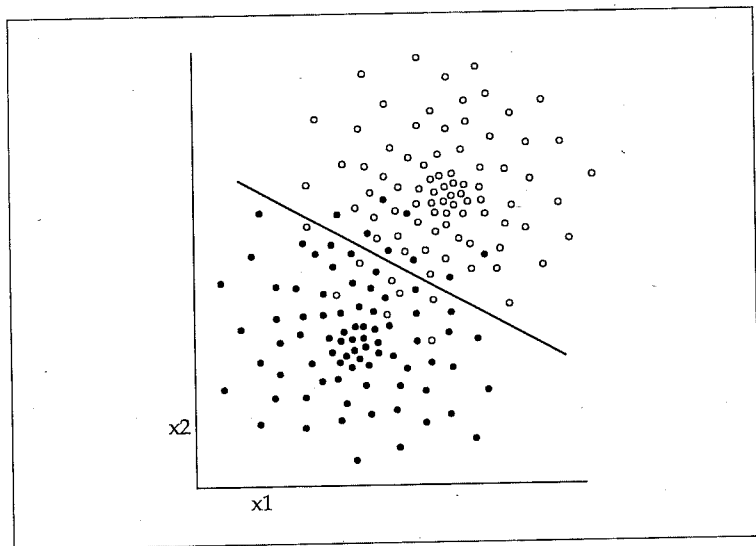


Een veelgebruikte vorm van het neurale netwerk in praktische toepassingen is het "feed-forward"-netwerk met één "hidden layer". Ook in ons onderzoek wordt dit netwerk toegepast.⁵ Het netwerk is tot veel meer scheidingen in staat dan alleen een lineaire opdeling. In principe kan iedere scheiding tussen de twee klassen met een willekeurige nauwkeurigheid worden benaderd. We bekijken een aantal scheidingen in figuur 2 tot en met 6. Een jaarrekening wordt beschreven door twee ratio's. In figuur 2 en 3 worden de voorbeelden uit eenzelfde trainingsverzameling getoond. Ook in figuur 4, 5 en 6 komen de voorbeelden uit eenzelfde verzameling. Zwarte rondjes zijn jaarrekeningen van de klasse "failliet" en witte rondjes zijn jaarrekeningen van de klasse "lopend". Stel dat in figuur 2 en 3 de voorbeelden afkomstig zijn van een populatie die in redelijke mate voldoet aan de voorwaarden die LDA stelt om een optimale scheiding voort te brengen (dat wil zeggen multivariaat normale verdelingen met gelijke covariantiematrices). De lineaire scheiding van LDA staat in figuur 2 en een veel betere scheiding dan deze bestaat niet. De scheiding in figuur 3 geeft een perfecte opdeling van de voorbeelden in de trainingsverzameling. Toch is deze scheiding minder goed dan de scheiding in figuur 2. Waar het namelijk om gaat is het generaliserend vermogen van een scheiding, dat wil zeggen hoe goed de scheiding is in het geval van de totale populatie. In figuur 3 heeft de opdeling zich te veel aangepast aan de trainingsdata. Toevalligheden in de trainingsverzameling, die niet in de populatie terug te vinden zijn, zijn wel verwerkt in de opdeling. De situatie in figuur 3 wordt aangeduid met de term "overfitting". Door het vermogen elke mogelijke scheiding te kunnen benaderen, zouden neurale netwerken zowel de scheiding in figuur 2 als de scheiding in figuur 3 kunnen realiseren. Aangezien de scheiding van figuur 2 nauwelijks verbeterd kan worden, heeft het gebruik van neurale netwerken in deze situatie geen toegevoegde waarde.

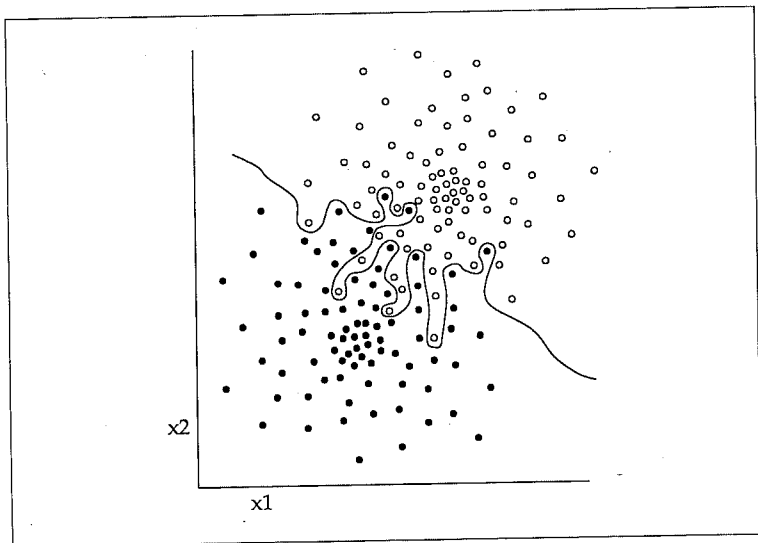
Alleen wanneer niet aan de voorwaarden van LDA wordt voldaan (zoals in de werkelijkheid bij jaarrekeningen het geval is), is het toepassen van neurale netwerken mogelijk zinvol. Ook in figuur 4, 5 en 6 gelden de voorwaarden niet. De scheiding in figuur 5 lijkt het best te zijn. Waarschijnlijk zal deze opdeling het ook goed doen op de gehele populatie. In figuur 6 is er sprake van "overfitting". De scheiding heeft zich te veel

5 In dit onderzoek is gebruikgemaakt van de functie *net* binnen S-Plus (Venables en Ripley, 1994). De "hidden units" en "output unit" hebben logistische functies en als "error"-functie wordt de "cross-entropy" gebruikt. De functie gebruikt de "quasi newton"-methode als optimalisatie-algoritme.

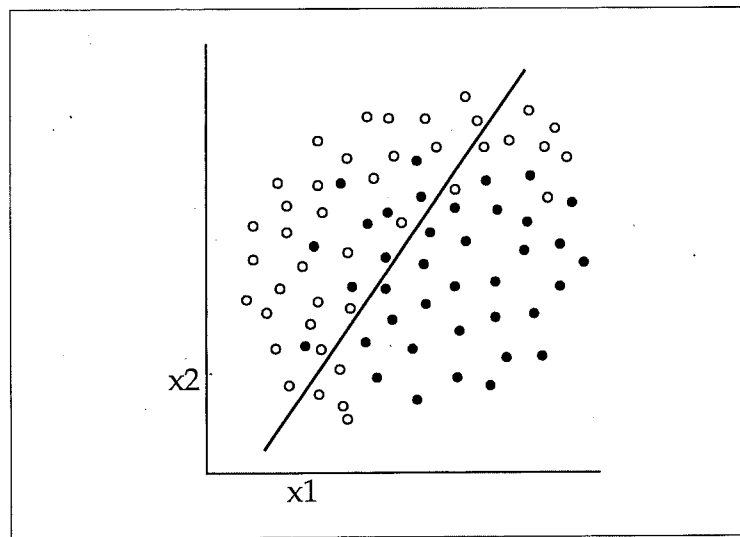
Figuur 2.



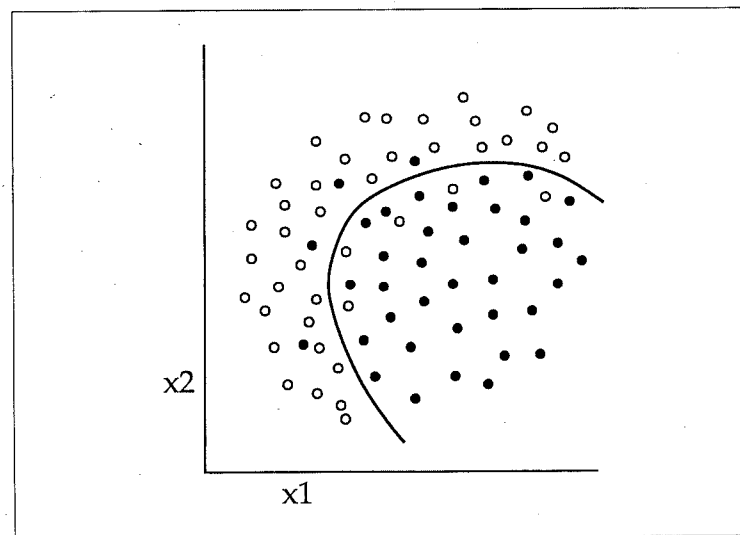
Figuur 3.



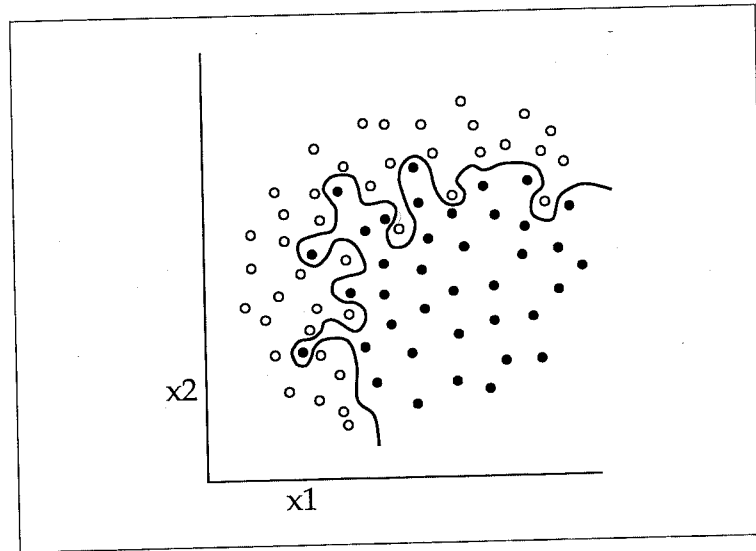
Figuur 4.



Figuur 5.



Figuur 6.



aangepast aan de trainingsdata. In figuur 4 heeft de scheiding zich te weinig aangepast aan de data. Er is sprake van "underfitting". Een simpele lineaire scheiding is onvoldoende voor een goede aanpassing. Stel dat LDA bij deze trainingsverzameling de scheiding in figuur 4 realiseert. Omdat LDA alleen een lineaire scheiding kan voortbrengen, is deze methode in deze situatie minder geschikt. Neurale netwerken kunnen alle drie de opdelingen realiseren.

4. Onderzoeksontwerp

Voordat een methode kan worden gebruikt, moet een aantal parameters van een waarde worden voorzien. De precieze instelling van de parameters heeft invloed op de vorm van het uiteindelijke model. De instelling moet dusdanig zijn dat het generaliserend vermogen van de scheiding zo goed mogelijk is. Op voorhand is echter moeilijk in te schatten wat een goede instelling van de parameters is. Daarom wordt een aantal instellingen geprobeerd en de beste instelling wordt gekozen. Bij LDA zijn er twee parameters waarvoor verschillende waarden worden gepro-

beerd. De eerste parameter is de combinatie van priorkansen (11 waarden) en de tweede parameter is de mate waarin jaarrekeningen met extreme ratiowaarden niet worden meegenomen tijdens het trainen (12 waarden).⁶ Bij LDA proberen we dus in totaal 11 (eerste parameter) \times 12 (tweede parameter) = 132 instellingen.

Bij neurale netwerken hanteren we twee versies. Bij de eerste versie van neurale netwerken, aangeduid met nn1, zijn er vier parameters die van een waarde moeten worden voorzien: het aantal "hidden units" (6 waarden), de mate van "weight decay" (3 waarden), de gewichtsverhouding tussen beide klassen in de "cross-entropy"-formule (7 waarden) en de startgewichten (2 waarden).⁷ We proberen dus in totaal $6 \times 3 \times 7 \times 2 = 252$ instellingen. De tweede versie (nn2) verschilt van de eerste in zoverre dat er veel minder parameterinstellingen worden geprobeerd, namelijk tien instellingen.⁸ Het proberen van minder parameterinstellingen betekent wel dat de kans bestaat dat instellingen die het meest geschikt zijn voor de data waarop wordt getraind, worden gemist; in dat geval wordt een slechter model verkregen. De vraag is dan ook interessant in welke mate de tweede versie tot slechtere modellen leidt dan de eerste. Het grote voordeel van nn2 is dat het bepalen van de beste instelling veel minder tijd in beslag neemt.

De gevolgde werkwijze in dit onderzoek is als volgt. Voor de vergelijking van de methoden in een bepaald jaar hebben we een verzameling met evenveel jaarrekeningen van beide klassen. Bijvoorbeeld: in het ge-

- 6 Wat betreft de eerste parameter kijken we naar de volgende elf combinaties (prior van "lopend", prior van "failliet"): $(0,40, 0,60)$, $(0,42, 0,58)$, $(0,44, 0,56)$, ..., $(0,60, 0,40)$. Bij de tweede parameter proberen we zowel het niet als het wel verwijderen van extremen. In het laatste geval is de procedure als volgt. Wanneer van ten minste één ratio waarmee een jaarrekening wordt beschreven de waarde meer dan i standaardafwijkingen afwijkt van het gemiddelde voor de betreffende klasse, dan wordt de jaarrekening verwijderd uit de trainingsverzameling. Voor de mate van verwijdering i proberen we elf waarden: $1, 1,5, 2, \dots, 6$.
- 7 De waarden die geprobeerd worden, zijn de volgende. Parameter 1: $0, 1, 2, 3, 4, 5$. Parameter 2: $0, 0,05, 0,2$. Bij parameter 3 de volgende combinaties (gewicht van "lopend", gewicht van "failliet"): $(1,2, 1)$, $(1,1, 1)$, $(1,05, 1)$, $(1, 1)$, $(1, 1,05)$, $(1, 1,1)$, $(1, 1,2)$. Ten slotte wordt geëxperimenteerd met twee sets startgewichten (hierbij krijgt elk gewicht in het netwerk een aselechte waarde in het interval $[-0,7, 0,7]$ toegewezen). Vanwege het gebruik van "weight decay" worden de ratiowaarden van alle jaarrekeningen herschaald. Na de herschaling ligt het grootste deel (90%) van de waarden voor een bepaalde ratio in het interval $[0,1]$.
- 8 De te proberen waarden zijn nu als volgt. Parameter 1: $0, 1, 2, 3, 4$. Parameter 2 heeft altijd waarde $0,2$. Parameter 3 heeft altijd waarde $(1, 1)$. Wel zijn er nog steeds twee sets startgewichten. Het aantal instellingen is dus nog maar $5 \times 1 \times 1 \times 2 = 10$.

val van jaar 4 zijn dat 600 jaarrekeningen van klasse "failliet" en 600 jaarrekeningen van klasse "lopend". Deze verzameling delen we aselekt op in twee gelijke delen, een trainings- en een testverzameling. We zorgen ervoor dat in ieder deel evenveel jaarrekeningen van beide klassen voorkomen. Vervolgens bepalen we door middel van "10-fold cross-validation" op de trainingsverzameling wat de beste parameterinstelling van elke methode is. Met de techniek "10-fold cross-validation" is het mogelijk de trainingsverzameling waarmee een model wordt getraind ook te gebruiken om een testresultaat van het model te verkrijgen. In het geval van LDA vindt 132 keer "10-fold cross-validation" plaats, iedere keer met een andere parameterinstelling. Bij nn1 en nn2 gaat het om respectievelijk 252 keer en 10 keer. Per methode beschouwen we de parameterinstelling met het hoogste "cross-validation"-resultaat als de beste instelling. Het hoogste "cross-validation"-resultaat is echter optimistisch gebiased, dat wil zeggen: het geeft een te positieve indruk van de prestatie van de methode. We trainen daarom per methode een nieuw model op de hele trainingsverzameling en met de beste parameterinstelling. Dit model wordt getest op de testverzameling. Per methode hebben we nu een niet-gebiased testresultaat en deze drie testresultaten zijn geschikt om met elkaar te worden vergeleken. Verschillen in de testresultaten worden door middel van de McNemar-toets onderzocht op significantie. Om de betrouwbaarheid van de onderzoeksresultaten te vergroten, hebben we besloten om voor de vergelijking in elk jaar het beschreven experiment tien keer uit te voeren. De belangrijke verandering in elke uitvoering ten opzichte van de negen andere uitvoeringen is dat de totale dataverzameling opnieuw aselekt wordt opgedeeld in een trainings- en een testverzameling.

5. Resultaten

De belangrijkste resultaten zijn te vinden in tabel 2. In de tweede tot en met vierde kolom staan de gemiddelde testresultaten. Bijvoorbeeld: tijdens elk van de tien uitvoeringen van de vergelijking in jaar 1 werd één model van LDA getest met de testverzameling, namelijk het model getraind op de hele trainingsverzameling en met de beste parameterinstelling. Het percentage 77,2% in de linkerbovenhoek van de tabel is het gemiddelde van die tien testresultaten. Een testresultaat is het percentage jaarrekeningen in de testverzameling dat juist werd geclassificeerd. In kolom 5, 8 en 11 staat respectievelijk het verschil in de gemiddelde test-score van LDA en nn1, van LDA en nn2 en van nn1 en nn2. In kolom 6

staan drie waarden ($A : B : C$) en in kolom 7 twee waarden ($D : E$). Waarde A , B en C is het aantal maal tijdens de tien uitvoeringen van het experiment dat het model van LDA respectievelijk een beter, eenzelfde en een slechter testresultaat behaalde dan het model van nn1. In kolom 7 worden alleen de significante verschillen geteld. Waarde D en E staat voor het aantal keren dat het model van LDA respectievelijk een significant beter en een significant slechter testresultaat behaalde dan het model van nn1. Bijvoorbeeld: in jaar 5 had het model van LDA in vier van de tien uitvoeringen een betere testscore dan het model van nn1 en in zes uitvoeringen een slechtere. Tijdens een van de laatstgenoemde zes uitvoeringen was er sprake van een significant verschil. Bij het toetsen op significantie is gekozen voor een onbetrouwbaarheid van 0,05 (tweezijdige toets). Wanneer het verschil in testresultaat van twee modellen volgens de McNemar-toets significant is, dan betekent dit het volgende. Met 5% onbetrouwbaarheid is het aantal jaarrekeningen in de gehele populatie dat het ene model juist classificeert niet gelijk aan het aantal jaarrekeningen in de gehele populatie dat het andere model juist classificeert. Het model met het hoogste testresultaat is dus zeer waarschijnlijk beter dan het andere model. Voor kolom 9 en 10 geldt een soortgelijk verhaal als voor kolom 6 en 7, alleen zijn de methoden nu LDA en nn2. Uit de tabel volgt dat de drie methoden opvallend gelijkwaardig hebben gepresteerd. Zo is in kolom 5, 8 en 11 te zien dat het verschil in de gemiddelde testscore van twee methoden bijna altijd minder dan 1% bedraagt. Kolom 6 en 9 geven geen aanleiding om te concluderen dat de prestaties van LDA duidelijk beter of slechter zijn dan de twee versies van neurale netwerken. In kolom 11 valt op dat de verschillen tussen de twee versies van neurale netwerken beperkt blijven.

Ook is nog gekeken naar het verschil in presteren van de methoden in het geval dat er weinig trainingsdata zijn. In grote lijnen is dezelfde procedure gevolgd als eerder beschreven. Veranderingen zijn de volgende. De trainings- en de testverzameling zijn niet meer even groot. De jaarrekeningen van elke klasse worden aselekt opgedeeld in twee ongelijke delen. Het totaal minus 200 komt in de testverzameling terecht. Hoeveel van de 200 overgebleven jaarrekeningen worden gebruikt in de trainingsverzameling, hangt af van het precieze experiment dat wordt gedaan. Per jaar wordt gekeken naar negen verschillende trainingsverzamelingen, ieder van een andere grootte. De verzamelingen bevatten 10, 15, 20, 25, 30, 40, 50, 100 of 200 jaarrekeningen van iedere klasse. Omdat in tabel 2 de resultaten van nn1 en nn2 niet veel van elkaar verschillen, wordt nu alleen gebruikgemaakt van nn2. In tabel 3 staan de resultaten

Tabel 2.

Jaar	LDA	Nn1	Nn2	LDA-Nn1	> = <	Mc Nemar	LDA-Nn2	> = <	Mc Nemar	Nn1-Nn2
1	77,2	76,9	76,5	0,3	5:0:5	0:0	0,7	5:0:5	3:2	0,4
2	72,8	72,3	72,6	0,5	5:1:4	0:0	0,2	5:0:5	1:0	-0,3
3	69,5	69,4	69,5	0,1	6:1:3	0:0	0	5:1:4	0:0	-0,1
4	65,8	65,8	65,7	0	5:0:5	0:0	0,1	6:0:4	0:0	0,1
5	65,5	65,9	66,6	-0,4	4:0:6	0:1	-1,1	3:0:7	0:0	-0,7

vermeld voor jaar 1 en 3. Uit de tabel volgt dat bij heel weinig trainingsdata de neurale netwerken beter in staat waren een goed model af te leveren. Vanaf ongeveer een trainingshoeveelheid van 30 jaarrekeningen per klasse en richting lagere hoeveelheden wordt het verschil in prestaties van beide methoden steeds groter.

Tabel 3.

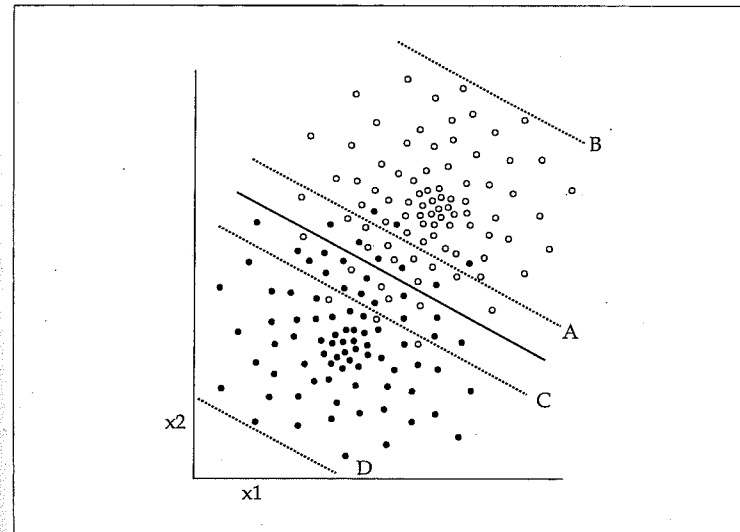
	Jaar 1					Jaar 3				
	LDA	Nn2	LDA-Nn2	> = <	Mc Nemar	LDA	Nn2	LDA-Nn2	> = <	Mc Nemar
10	61,2	70,8	-9,6	1:0:9	0:8	56,4	62,5	-6,1	0:0:10	0:7
15	66,0	72,7	-6,7	0:0:10	0:8	57,0	62,6	-5,6	1:0:9	0:8
20	69,6	73,0	-3,4	2:0:8	0:7	60,5	64,2	-3,7	2:0:8	1:6
25	70,7	73,1	-2,4	3:0:7	2:4	62,5	64,6	-2,1	2:0:8	2:5
30	70,0	72,9	-2,9	1:0:9	1:7	63,2	64,9	-1,7	3:0:7	1:6
40	72,6	74,2	-1,6	3:0:7	1:4	65,2	66,5	-1,3	2:0:8	1:2
50	74,4	74,7	-0,3	6:0:4	2:3	65,3	67,1	-1,8	2:0:8	1:5
100	75,4	75,6	-0,2	5:0:5	2:2	67,7	68,3	-0,6	3:0:7	0:1
200	76,3	76,4	-0,1	5:0:5	1:1	68,3	68,2	0,1	5:0:5	0:0

6. Een praktische toepassing

In deze paragraaf wordt één model in een dusdanige vorm gepresenteerd dat toepassing van het model in de praktijk mogelijk is. In het onderzoek presteerden LDA en neurale netwerken gelijkwaardig. We geven daarom alleen een model van LDA, aangezien modellen van LDA minder complex zijn dan modellen van neurale netwerken. In figuur 2 vinden we een voorbeeld van een model van LDA. De lijn in deze figuur

is $M = 5,7$, waarbij M gelijk is aan $0,54 x_1 + x_2$. De waarde 5,7 noemen we de grenswaarde. Het model in de figuur bestaat hieruit dat een jaarrekening met ratiowaarden x_1 en x_2 die boven de lijn ligt (ofwel $M > 5,7$) gerekend wordt tot klasse "lopend" en dat een jaarrekening die zich onder de lijn bevindt (ofwel $M < 5,7$) gerekend wordt tot klasse "failliet". Het is mogelijk om het model aan te passen (bijvoorbeeld op grond van een bepaald gewenst kredietbeleid) door een andere grenswaarde te kiezen. Het model geeft dan andere classificatieresultaten voor de beide klassen. Het vergroten van de grenswaarde van 5,7 tot bijvoorbeeld 6,5 heeft tot gevolg dat de lijn die de klassen scheidt hoger komt te liggen. In figuur 7 geeft stippellijn A de nieuwe ligging aan. Meteen is duidelijk dat meer jaarrekeningen van klasse "failliet" juist worden geclassificeerd, ten koste van meer jaarrekeningen van klasse "lopend" die niet juist worden geclassificeerd. Hoe groter de grenswaarde wordt genomen, hoe hoger de lijn komt te liggen, hoe beter het classificatieresultaat voor klasse "failliet" en hoe slechter het resultaat voor klasse "lopend". Op een gegeven moment is de grenswaarde zo groot en ligt de bijbehorende lijn zo hoog dat alle jaarrekeningen van klasse "failliet" juist en alle jaarrekeningen van klasse "lopend" onjuist worden geclassificeerd. Dit is het geval bij stippellijn B, met bijbehorende grenswaarde 10,1. Bij het verkleinen van

Figuur 7.



de grenswaarde geldt een omgekeerd verhaal; het resultaat voor klasse "failliet" wordt slechter en het resultaat voor klasse "lopend" wordt beter. Stippellijn C en D geven de ligging van de scheidslijn bij respectievelijk grenswaarde 4,9 en 1,5.

Het model dat in de praktijk kan worden toegepast, komt uit jaar 2. De modellen uit dat jaar bleken een goede voorspellingskracht te bezitten, zowel vlak vóór het faillissement als enige jaren eerder. Om een zo goed mogelijk model te verkrijgen, bevat de trainingsverzameling, anders dan bij de vergelijking van LDA met neurale netwerken, alle 956 jaarrekeningen van klasse "failliet" (jaar 2) en 956 jaarrekeningen van klasse "lopend". Vanwege de vrij hoge correlatie tussen de eerste en tweede ratio is besloten de eerste ratio niet mee te nemen. Het model is getraind met de beste parameterinstelling (bepaald met "10-fold cross-validation") en ziet er als volgt uit. Stel, een jaarrekening heeft voor de negen ratio's de waarden x_2, x_3, \dots, x_{10} . We berekenen voor de jaarrekening de waarde van:

$$M = 0,42 x_2 + -0,07 x_3 + 0,46 x_4 + 5,26 x_5 + -0,67 x_6 + -0,91 x_7 + 2,27 x_8 + 0,20 x_9 + 1,06 x_{10}$$

De grenswaarde is 0,01. Het model rekent de jaarrekening tot de klasse "failliet" wanneer geldt dat $M < 0,01$ en tot de klasse "lopend" wanneer geldt dat $M > 0,01$.

In tabel 4 staan voor het model een groot aantal grenswaarden vermeld. De bijbehorende resultaten voor verschillende verzamelingen staan eraanast. Het eerste percentage is het resultaat voor klasse "lopend". De andere vijf percentages hebben betrekking op klasse "failliet", achtereenvolgens het resultaat voor jaar 1, jaar 2, jaar 3, jaar 4 en jaar 5. We hebben "cross-validation" (namelijk "leave-one-out") toegepast voor het berekenen van de percentages, waarbij alle jaarrekeningen uit tabel 1 bij de berekening werden gebruikt.⁹ Hoe groter de grenswaarde, hoe beter het

⁹ We zijn als volgt te werk gegaan. Per grenswaarde (anders gezegd per combinatie van priorkansen) worden de zes percentages vastgesteld. Bij iedere grenswaarde opnieuw worden alle jaarrekeningen uit tabel 1 geïnclassificeerd om de percentages te bepalen. Niet elke jaarrekening wordt op dezelfde wijze geïnclassificeerd. De jaarrekeningen uit de trainingsverzameling (956 van klasse "lopend" en 956 uit jaar 2) worden geïnclassificeerd door middel van "leave-one-out". Voor de jaarrekeningen van klasse "failliet" uit andere jaren is het volgende bedacht. In het geval dat tijdens de "leave-one-out"-procedure de uitgesloten jaarrekening van klasse "failliet" is, wordt niet alleen deze jaarrekening geïnclassificeerd, maar ook andere jaarrekeningen van de betreffende on-

der onderneming (die betrekking hebben op andere jaren dan jaar 2). Blijven over de jaarrekeningen van ondernemingen die geen jaarrekening uit jaar 2 hebben, maar wel uit andere jaren, en de 310 (= 1266 - 956) overige jaarrekeningen van klasse "lopend". Deze jaarrekeningen worden apart, buiten de "leave-one-out"-procedure om, door het model (met de betreffende grenswaarde) geïnclassificeerd.

Tabel 4.

Gr.w.	lop	fai 1	2	3	4	5	Gr.w.	lop	fai 1	2	3	4	5
9,30	1	100	99	99	100	100	-0,10	75	77	70	61	55	54
5,39	3	99	99	99	99	99	-0,21	78	74	67	55	50	49
4,53	5	99	98	99	98	98	-0,27	80	72	64	53	48	46
3,31	10	99	97	98	97	97	-0,37	82	68	58	48	42	41
2,69	15	98	97	97	97	97	-0,41	83	67	55	46	41	39
2,20	20	97	96	96	96	95	-0,50	85	64	50	41	37	34
1,86	25	97	95	95	94	93	-0,56	87	61	47	39	35	32
1,61	30	96	94	93	92	91	-0,64	88	58	44	36	31	28
1,37	35	95	93	92	90	88	-0,72	90	55	40	32	28	26
1,11	40	94	92	90	87	86	-0,81	91	51	35	29	24	22
0,90	45	92	90	89	84	83	-0,90	92	48	31	25	21	20
0,70	50	91	89	86	81	80	-0,99	93	45	29	22	18	17
0,52	55	89	86	82	77	77	-1,13	94	42	26	18	16	15
0,38	60	87	84	79	73	72	-1,29	95	38	23	17	14	12
0,21	65	85	80	74	67	67	-1,59	96	32	17	12	11	9
0,08	68	82	77	69	62	63	-2,14	97	24	10	8	7	4
0,03	70	81	76	66	61	60	-2,81	98	16	7	5	4	3
-0,04	73	79	74	63	59	57	-4,66	99	8	3	2	2	2

Ook kan met grijze gebieden worden gewerkt. Stel, men wil dat het model hoogstens 10% van de jaarrekeningen van klasse "failliet binnen twee jaar" en 15% van de jaarrekeningen van klasse "lopend" onjuist classificeert. We zoeken in de tabel de grenswaarde waarbij het gemiddelde van de score voor jaar 1 en jaar 2 90% is (dat is 0,70) en de grenswaarde waarbij de score voor klasse "lopend" 85% is (dat is -0,50). We definiëren vervolgens drie gebieden; gebied I is $-\infty, -0,50$, gebied II is $<-0,50, 0,70>$ en gebied III is $<0,70, \infty>$. Het aangepaste model is nu als volgt. Een jaarrekening met een waarde voor M in gebied I wordt als "failliet" geclassificeerd en een jaarrekening met een waarde voor M in gebied III als "lopend". Als de waarde voor M in gebied II ligt, wordt geen beslissing genomen. We vinden dan dat de jaarrekening niet duidelijk genoeg bij een van de twee klassen hoort, waardoor er te veel kans is dat we op basis van deze waarde een foute beslissing nemen. Nadere analyse van de onderhavige onderneming is in deze situatie geboden. Gebied II wordt het grijze gebied genoemd. We kunnen eenvoudig met de tabel narekenen hoe de jaarrekeningen van beide klassen verdeeld zijn over de drie gebieden. Bij grenswaarde 0,70 zijn de scores voor "lopend" en "failliet" 50% en 90%. Bij grenswaarde -0,50 zijn de percentages 85% en 57%. De jaarrekeningen van klasse "failliet" zijn als volgt verdeeld; in gebied I bevindt zich 57%, in gebied III $100 - 90 = 10\%$ en in gebied II $100 - 57 - 10 = 33\%$. En voor de jaarrekeningen van klasse "lopend" geldt: in gebied I is het percentage $100 - 85 = 15\%$, in gebied III 50% en in gebied II $100 - 15 - 50 = 35\%$. Wat betreft de klasse "failliet binnen twee jaar" classificeert het model dus 10% van de jaarrekeningen onjuist, 57% juist en bij 33% wordt geen directe beslissing op basis van het model genomen. Bij de klasse "lopend" zijn deze percentages achtereenvolgens 15%, 50% en 35%.

Het is altijd belangrijk om zich bewust te blijven van de werkelijke klassenverhouding in de totale populatie. Stel dat in werkelijkheid ongeveer 4% van de ondernemingen binnen vijf jaar failliet gaat. In dit voorbeeld maken we gebruik van het model met grenswaarde -0,27 en met een test-score van 80% voor "lopend" en 57% voor "failliet binnen vijf jaar". Met het model classificeren we 2000 aselect gekozen jaarrekeningen, waarvan er dus 80 (4% van 2000) van klasse "failliet" zijn en 1920 van klasse "lopend". 430 jaarrekeningen worden door het model als "failliet" geclassificeerd. Merk op dat hiervan slechts 46 stuks (57% van 80) daadwerkelijk bij deze klasse horen. De overige 384 jaarrekeningen (20% van 1920) horen eigenlijk bij de klasse "lopend". Dit voorbeeld laat zien dat als een jaarrekening door een model tot de klasse "failliet" wordt gere-

kend, de kans nog steeds heel klein is dat het werkelijk om een jaarrekening van deze klasse gaat (in dit voorbeeld is deze kans $(46 / 430) * 100\% = 11\%$).

Een model kan ook worden gebruikt om een onderneming een waardeering te geven. Een onderneming met een hoge waarde voor M wordt dan positiever gewaardeerd dan een onderneming met een lage M-waarde. Stel dat de jaarrekeningen van twee ondernemingen worden beoordeeld met het model. De eerste onderneming behaalt een M-waarde van 1,11 en de tweede onderneming een waarde van -0,72. In vergelijking met de jaarrekening van de eerste onderneming heeft 60% van de jaarrekeningen van klasse "lopend" en 90% van de jaarrekeningen van klasse "failliet binnen vijf jaar" een lagere M-waarde. Voor de tweede onderneming zijn deze percentages 10% en 36%. Of de onderneming in de toekomst blijft voortbestaan, is dus bij de tweede onderneming onzekerder dan bij de eerste.

7. Kanttekeningen bij het gebruik van het model

Bij de beschrijving in de vorige paragraaf van de wijze waarop een model in de praktijk kan worden toegepast, willen we twee kanttekeningen plaatsen. Ten eerste zijn de classificatiepercentages in tabel 4 tot stand gekomen met jaarrekeningen die op verschillende kalenderjaren betrekking hebben. Nader onderzoek bracht aan het licht dat voor een specifiek kalenderjaar de percentages in redelijke mate kunnen afwijken. Ten tweede zijn de percentages in tabel 4 vastgesteld met jaarrekeningen van een breed gedefinieerde groep ondernemingen. Het met een model classificeren van een specifieke categorie ondernemingen, bijvoorbeeld ondernemingen van een bepaalde grootte, kan de percentages doen verbeteren of verslechteren. We zullen de kanttekeningen kort met daarop betrekking hebbende onderzoeksresultaten ondersteunen.

Voor onderzoek naar de invloed van het kalenderjaar op de prestatie van een model hebben we gekeken naar de reeks van jaren 1988, 1989, 1990 en 1991. Dit is een interessante periode, omdat de groei van het bruto nationaal product een afname liet zien en het aantal faillissementen in de populatie waarvan in dit onderzoek gebruik wordt gemaakt (vennootschappen uit de sector industrie) enorm opliep. In 1988 is het aantal ondernemingen dat zich één jaar vóór het faillissement bevindt (ofwel

het aantal jaarrekeningen in jg_1) 117, in 1989 168, in 1990 201 en in 1991 260. We maken vier trainingsverzamelingen en vier testverzamelingen. De jaarrekeningen uit dezelfde verzameling hebben allemaal betrekking op hetzelfde kalenderjaar. Bij elk van de vier kalenderjaren hoort één trainingsverzameling en één testverzameling (zie het linkerdeel van tabel 5). In de trainingsverzameling van jaar 1988 zitten bijvoorbeeld 450 jaarrekeningen van klasse "lopend" en 125 jaarrekeningen van klasse "failliet". En voor de testverzameling van jaar 1988 zijn deze aantallen respectievelijk 450 en 109. De jaarrekeningen van klasse "failliet" hebben betrekking op het laatste of het op één na laatste jaar vóór het jaar van het faillissement (jaarrekeningen uit jg_1 en jg_2).¹⁰ We maken vier modellen; uit elke trainingsverzameling wordt een model afgeleid. Vervolgens testen we elk model met elk van de vier testverzamelingen. Zowel LDA als nn2 worden toegepast. In het rechterdeel van tabel 5 staan de testresultaten bij het gebruik van LDA, maar de resultaten van nn2 wijken daar niet veel van af.¹¹ Elk van de vier modellen laat hetzelfde beeld zien, een afnemend testresultaat naarmate het kalenderjaar van de testverzameling hoger wordt. Het verschil tussen het resultaat voor de testverzameling van 1988 en het resultaat voor de testverzameling van 1991 is ongeveer 6%.¹² De resultaten laten zien dat de prestaties van een model binnen enkele jaren in redelijke mate kunnen veranderen, in ieder geval gedurende een periode waarin het aantal faillissementen snel verandert. De classificatiepercentages bij de grenswaarden in tabel 4 moe-

- 10 We hebben ervoor gezorgd dat in elk van de acht verzamelingen jaarrekeningen uit jg_1 even vaak voorkomen als jaarrekeningen uit jg_2 . De jaarrekeningen zijn afkomstig uit de groep van 1266 (jaar 1) en 956 (jaar 2) jaarrekeningen uit tabel 1. De 3600 (= 4 * 900) jaarrekeningen van klasse "lopend" zijn de 1266 jaarrekeningen uit tabel 1 met daarnaast 2334 extra verzamelde jaarrekeningen.
- 11 Om een trainingsverzameling te verkrijgen waarin beide klassen even vaak voorkomen, worden de jaarrekeningen van klasse "failliet" dusdanig vaak gedupliceerd dat het aantal van 450 wordt bereikt. Toegepast wordt dezelfde werkwijze als beschreven in paragraaf 4, dat wil zeggen dat met "10-fold cross-validation" op de trainingsverzameling wordt bepaald wat de beste parameterinstelling van elke methode is. Het testresultaat is altijd gelijk aan: (het percentage voor klasse "lopend" + het percentage voor klasse "failliet") / 2. Ook nu wordt het experiment tien keer uitgevoerd, met steeds weer een nieuwe aselechte opdeling in een trainings- en een testverzameling bij elke uitvoering. De testresultaten in de tabel zijn de gemiddelde resultaten van de tien uitvoeringen van het experiment.
- 12 Het verschil tussen het resultaat voor de testverzameling van 1988 en het resultaat voor de testverzameling van 1991 is getoetst op significantie. Er is alleen gekeken naar de modellen van 1988 en 1991. In het geval van het model van 1988 was er in zeven van de tien uitvoeringen van het experiment een p-waarde van kleiner dan 0,1 (eenzijdige toets). Bij het model van 1991 was er in acht van de tien uitvoeringen een p-waarde van kleiner dan 0,1 (eenzijdige toets).

ten daarom met zorg worden beschouwd en zouden regelmatig opnieuw vastgesteld moeten worden met recente data. In Ooghe et al. (1995) wordt bij een Belgisch faillissementsmodel vastgesteld dat de prestaties juist wel redelijk stabiel waren gedurende een aantal jaren. Mogelijk betrof het een periode met weinig verandering in het aantal faillissementen. Interessant is dat in dit onderzoek een verslechtering van de prestaties van een model samengaat met een verslechtering van de economie en een toename van het aantal faillissementen. Het onderscheid tussen een onderneming die failliet gaat en een onderneming die blijft voortbestaan, lijkt in economisch goede tijden beter te maken. Waarschijnlijk is in slechte tijden de groep kwakkelende ondernemingen groter. Dit zijn de ondernemingen waarvan de afloop het meest onzeker is; ze zijn niet duidelijk gezond, maar stevenen ook niet met grote zekerheid af op een faillissement.

Tabel 5.

	1988	1989	1990	1991		Test			
	Lop, fai	Lop, fai	Lop, fai	Lop, fai		1988	1989	1990	1991
Train	450,125	450,125	450,125	450,125	Train	78	77	74	72
Test	450,109	450,183	450,277	450,265		78	77	74	71
Totaal	900,234	900,308	900,402	900,390		77	76	73	72
					1991	78	76	74	71

Inzake de tweede kanttekening hebben we de invloed van de leeftijd en de grootte van een onderneming op de prestatie van een model beschouwd. We beginnen met de invloed van de leeftijd. Er wordt gebruikgemaakt van 956 jaarrekeningen van klasse "failliet" (uit jg_2) en 3000 jaarrekeningen van klasse "lopend".¹³ Elke jaarrekening krijgt een bepaalde leeftijd mee, namelijk de leeftijd van de bijbehorende onderneming op het moment dat de jaarrekening gepubliceerd werd.¹⁴ We han-

- 13 De jaarrekeningen van klasse "failliet" zijn de 956 (jaar 2) jaarrekeningen uit tabel 1. De 3000 jaarrekeningen van klasse "lopend" zijn 3000 van de 3600 jaarrekeningen uit voetnoot 10.
- 14 Het oprichtingsjaar van een onderneming is geschat aan de hand van het getal dat gevormd wordt door het tweede en derde cijfer in het BTW-nummer. De schatting vindt als volgt plaats. Als het getal 12 is, dan is het oprichtingsjaar 1973 of 1972. Andere getallen: 13 dan 1974 of 1973, 14 dan 75 of 74, 15 dan 76 of 75, 16 dan 77 of 76, 17 t/m 18 dan 78 of 77, 19 dan 79 of 78, 20 dan 80 of 79, 21 dan 81 of 80, 22 dan 82 of 81, 23 t/m 24 dan 83 of 82, 25 dan 84 of 83, 26 t/m 27 dan 85 of 84, 28 t/m 29 dan 86 of 85, 30 t/m 31 dan 87

teren drie leeftijdscategorieën: 1 tot en met 4 jaar, 5 tot en met 13 jaar, en 14 jaar en ouder. Tabel 6 geeft aan hoe de jaarrekeningen over de categorieën verdeeld zijn. De trainingsverzameling bestaat uit de helft van de 956 jaarrekeningen en de helft van de 3000 jaarrekeningen. De overige jaarrekeningen vormen de testverzameling. Uit de trainingsverzameling leiden we een model af. Het is een algemeen model, aangezien het is afgeleid met jaarrekeningen uit alle drie de categorieën. Het model testen we vervolgens op de testverzameling. In de laatste kolom van tabel 6 wordt het testresultaat gegeven per leeftijdscategorie. Het zijn de resultaten bij het gebruik van LDA, maar de resultaten van nn2 wijken daar niet veel van af.¹⁵ Duidelijk is te zien dat, naarmate de leeftijd van een onderneming hoger is, het model beter kan voorspellen of de onderneming wel of niet zal failleren.¹⁶ Met name startende ondernemingen (categorie 1-4) zijn moeilijker te classificeren. Door het ontbreken van bestaansjaren kan er bij een starter geen sprake zijn van een geleidelijke ontwikkeling richting faillissement. Het lijkt aannemelijk dat bij oudere ondernemingen een dergelijke ontwikkeling vaak wel te zien is, waarbij een verliesgevende onderneming bijvoorbeeld nog een tijd kan overleven door in te teren op de opgebouwde winstreserves. Aldus zal bij jonge ondernemingen het faillissement een meer onverwacht karakter hebben en minder vaak twee jaar van tevoren te voorzien zijn. Daarnaast zijn starters meestal kleine ondernemingen. Meer dan bij grote ondernemingen speelt bij kleine ondernemingen ook andere informatie dan financiële ratio's een rol bij faillissementspredictie. Een belangrijke reden is de grotere vervlechting van zakelijke en privé-activiteiten bij kleine ondernemingen (Caouette et al., 1998).

of 86, 32 t/m 35 dan 88 of 87, 36 t/m 38 dan 89 of 88, 39 t/m 41 dan 90 of 89, 42 t/m 45 dan 91 of 90, 46 t/m 48 dan 92 of 91, 49 t/m 50 dan 93 of 92. Als het getal 11 of lager is, dan is het oprichtingsjaar 1972 of eerder. Dat dit een goede schattingsmethode is, bleek bij een grote groep van 3122 industriële ondernemingen waarvan we het werkelijke oprichtingsjaar wel wisten. Bij 99,6% van deze ondernemingen klopte het bovenstaande schattingschema.

- 15 Precies dezelfde tekst als in voetnoot 11 is hier van toepassing. Het enige verschil is dat het genoemde aantal van 450 nu 1500 moet zijn.
- 16 Het verschil tussen het testresultaat voor categorie 1-4 en het testresultaat voor categorie 14-∞ is getoetst op significantie. Er was in negen van de tien uitvoeringen van het experiment een p-waarde van kleiner dan 0,05 (eenzijdige toets).

Tabel 6.

		Lop, fai		
Leeftijd	1-4	803,342	1-4	69
	5-13	955,309	5-13	73
	14-∞	1242,305	14-∞	75
	Totaal	3000,956		

Ten slotte beschouwen we de invloed van de grootte van een onderneming. Voor de vaststelling van de grootte van een onderneming wordt over het algemeen gekeken naar de omvang van drie grootheden: het totaal vermogen, het personeelsbestand en de omzet. Gezien het feit dat zoveel jaarrekeningen de omzet niet vermelden, beperken we ons tot het totaal vermogen en het aantal personeelsleden. Er wordt gebruikgemaakt van 832 jaarrekeningen van klasse "failliet" (uit jg_2) en 2563 jaarrekeningen van klasse "lopend".¹⁷ We hanteren drie categorieën van het totaal vermogen (0 tot en met 9 miljoen BEF, 9 tot en met 30 miljoen BEF, 30 miljoen BEF en meer) en drie categorieën van de personeelsomvang (1 tot en met 5, 6 tot en met 18, 19 en hoger). Door combinatie van beide grootheden ontstaan negen nieuwe categorieën. Het linkerdeel van tabel 7 geeft aan hoe de jaarrekeningen over de negen categorieën verdeeld zijn. Vanwege een hoge correlatie tussen de twee grootheden (een correlatiecoëfficiënt van ongeveer 0,9) bevinden de meeste jaarrekeningen zich op de diagonaallijn van linksonder naar rechtsboven. De trainingsverzameling bestaat uit de helft van de 832 jaarrekeningen en de helft van de 2563 jaarrekeningen. De testverzameling wordt gevormd door de overige jaarrekeningen. Uit de trainingsverzameling leiden we een algemeen model af en dat testen we op de testverzameling. De rechterzijde van tabel 7 geeft het testresultaat voor elk van de negen categorieën. Het zijn de resultaten bij het gebruik van LDA en ook nu zijn de resultaten van nn2 niet veel anders.¹⁸ De genoemde diagonaal beschouwend, zien we dat de kleinste ondernemingen (categorie 0-9, 1-5) de

- 17 Het betreft dezelfde jaarrekeningen als gebruikt bij het onderzoek naar de invloed van de leeftijd. Wel zijn, om vervuiling van de data tegen te gaan, jaarrekeningen uitgesloten die voor de post met code 9090 (gemiddeld personeelsbestand) geen waarde ingevuld hebben staan. Nader onderzoek leerde namelijk dat in sommige gevallen het niet invullen van een waarde waarschijnlijk niet betekent dat er geen personeelsleden zijn. Reden voor dit vermoeden is de hoge waarde van het totaal vermogen in die gevallen.
- 18 Ook hier is precies dezelfde tekst als in voetnoot 11 van toepassing. Het enige verschil is dat het genoemde aantal van 450 nu 1282 moet zijn.

laagste score behalen. Om eerder genoemde redenen is dit geen onverwacht resultaat. De middelgrote ondernemingen (categorie 9-30, 6-18) zijn een stuk beter te classificeren. Heel opvallend is echter dat de score weer inzakt voor de grootste ondernemingen (30-∞, 19-∞).¹⁹ Een reden zou kunnen zijn dat bij grote ondernemingen vaker "creative accounting" plaatsvindt (wellicht mede veroorzaakt doordat grote ondernemingen meer mogelijkheden hebben om "creative accounting" toe te passen). Het mooier voorstellen van de financiële situatie bemoeilijkt vanzelfsprekend het voorspellen van een naderend faillissement.

Tabel 7.

	Lop, fai	Lop, fai	Lop, fai					
Personeel	19-∞	13,16	67,53	689,205	19-∞	68	70	73
	6-18	112,45	359,150	258,74	6-18	74	80	71
	1-5	673,179	330,92	62,18	1-5	70	75	71
		0-9	9-30	30-∞		0-9	9-30	30-∞
	Totaal vermogen				Totaal vermogen			

8. Conclusies

De drie methoden - LDA, nn1 en nn2 - hebben opvallend gelijkwaardig gepresteerd. LDA blijkt ook in dit onderzoek een robuuste methode; hoewel de gebruikte data niet of slechts in beperkte mate voldoen aan de voorwaarden die LDA stelt om een optimale scheiding voort te brengen, gaf de methode toch goede resultaten. De neurale netwerken waren niet in staat deze te verbeteren. Alleen in het geval dat er heel weinig trainingsdata waren, deden de neurale netwerken het beter dan LDA. Het is bekend dat LDA bij een erg kleine trainingsverzameling moeite heeft om een goed model voort te brengen. In dit onderzoek presteerde LDA slechter dan neurale netwerken bij een trainingshoeveelheid van ongeveer 30 jaarrekeningen of minder per klasse. De verschillen tussen de

¹⁹ Het verschil tussen het testresultaat voor categorie (0-9, 1-5) en het testresultaat voor categorie (9-30, 6-18) is getoetst op significantie. Er was in acht van de tien uitvoeringen van het experiment een p-waarde van kleiner dan 0,1 (eenzijdige toets). Ook het verschil tussen het testresultaat voor categorie (9-30, 6-18) en het testresultaat voor categorie (30-∞, 19-∞) is getoetst. Er was in acht van de tien uitvoeringen een p-waarde van kleiner dan 0,1 (eenzijdige toets).

twee versies van neurale netwerken bleven beperkt. Onder de 242 extra netwerken van nn1 bevond zich meestal geen netwerk dat aanzienlijk beter presteerde dan een van de tien netwerken van nn2. De tijdsbesparing door alleen de tweede versie te gebruiken is aanzienlijk. Vanwege het gelijkwaardig presteren van LDA en neurale netwerken hebben de modellen van LDA de voorkeur bij toepassing in de praktijk. In vergelijking met neurale netwerken zijn modellen van LDA minder complex en daardoor makkelijker interpreteerbaar. In paragraaf 6 werd getoond hoe een model eenvoudig praktisch kan worden toegepast. De classificatiepercentages bij de grenswaarden in tabel 4 moeten echter met zorg worden beschouwd. De prestaties van een model worden namelijk beïnvloed door de leeftijd en grootte van de te classificeren onderneming en door de macro-economische situatie (dat wil zeggen het kalenderjaar van de jaarrekening).

Bibliografie

- ALTMAN, E.I. (1968), "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *The Journal of Finance*, 23, blz. 589-609.
- ALTMAN, E.I., G. MARCO en F. VARETTO (1994), "Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)", *Journal of Banking and Finance*, 18, blz. 505-529.
- BARNIV, R., A. AGARWAL en R. LEACH (1997), "Predicting the outcome following bankruptcy filing: a three-state classification using neural networks", *International Journal of Intelligent Systems in Accounting, Finance and Management*, 6, nr. 3, blz. 177-194.
- BILDERBEEK, J. (1977), *Financiële ratio-analyse*, Leiden, Stenfert Kroese.
- BISHOP, C.M. (1995), *Neural networks for pattern recognition*, Oxford, Oxford University Press.
- BORITZ, J.E. en D.B. KENNEDY (1995), "Effectiveness of neural network types for prediction of business failure", *Expert Systems with Applications*, jg. 9, nr. 4, blz. 503-512.
- CAOUEITE, J.B., E.I. ALTMAN en P. NARAYANAN (1998), *Managing credit risk, the next great financial challenge*, New York, Wiley Frontiers in Finance.
- DIMITRAS, A.I., S.H. ZANAKIS en C. ZOPOUNIDIS (1996), "A survey of business failures with an emphasis on prediction methods and industrial applications", *European Journal of Operational Research*, 90, blz. 487-513.
- FLETCHER, D. en E. GOSS (1993), "Forecasting with neural networks: an application using bankruptcy data", *Information and Management*, 24, blz. 159-167.
- JAMES, M. (1985), *Classification algorithms*, New York, John Wiley & Sons.
- HAYKIN, S. (1994), *Neural networks, a comprehensive foundation*, Londen, Prentice-Hall International.

- O'LEARY, D.E. (1998), "Using neural networks to predict corporate failure", *International Journal of Intelligent Systems in Accounting, Finance and Management*, 7, nr. 3, blz. 187-197.
- OOGHE, H., P. JOOS en D. DE VOS (1993), "Risico-indicator voor een onderneming aan de hand van falingspredictiemodellen", *Accountancy en Bedrijfskunde Kwartaalschrift*, jg. 18, nr. 3, blz. 3-26.
- OOGHE, H. en C. VAN WYMEERSCH (1994), *Financiële analyse van de onderneming*, Diegem, Kluwer Editorial.
- OOGHE, H., P. JOOS en C. DE BOURDEAUDHUIJ (1995), "Financial distress models in Belgium: the results of a decade of empirical research", *The International Journal of Accounting*, jg. 30, nr. 3, blz. 245-274.
- POMPE, P.P.M. en A.J. FEELDERS (1997), "Using machine learning, neural networks, and statistics to predict corporate bankruptcy", *Microcomputers in Civil Engineering*, 12, blz. 267-276.
- SPSS for Windows (1993), *Base system user's guide*, Release 6.0.
- VENABLES, W.N. en B.D. RIPLEY (1994), *Modern applied statistics with S-Plus*, New York, Springer-Verlag.

Abstract

The Prediction of Bankruptcies: A Comparison Between Linear Discriminant Analysis and Neural Networks

In this paper we report about research comparing the performances of two methods - linear discriminant analysis (LDA) and neural networks - when used for producing a model with which bankruptcies can be predicted. A large dataset was used containing Belgian annual reports of companies that went bankrupt and companies that did not go bankrupt. A proper statistical design was used to test whether differences in performance were statistically significant. In general the two methods performed equally well. Only in the case where few training data were available neural networks defeated LDA. One model is presented in such a way that it can be applied in practical situations. The paper shows the different ways in which the model can easily be applied. It also demonstrates how model performance is influenced by the size and age of the company and by the state of the economy.

Boekbesprekingen

Algemene economie

Feminist Economics

Gillian J. Hewitson, *Feminist Economics: Interrogating the Masculinity of Rational Economic Man*
Cheltenham (VK) / Northampton (MA, VS),
Edward Elgar, 1999, vii + 277 blz. - Prijs: 55
GBP
ISBN 1-85898-946-9

Dat de economische wetenschap voornamelijk door mannen ontwikkeld werd en misschien onevenredig veel aandacht besteedt aan "mannelijke" onderwerpen, ondanks de beschikbaarheid van evenveel vrouwelijk talent (zie ook de bespreking van *Women of Value* in dit tijdschrift, jg. 1997, nr. 1), kan onmogelijk ontkend worden. De Australische Gillian Hewitson gaat in haar boek nog een stap verder door met een feministische poststructurele benadering aan te willen tonen dat de economie een masculiene wetenschap is. De redenering komt hierop neer: de klassieke economen denken dat het nutsmaximaliserende individu geslachtsloos is, hoewel het, impliciet, verschillende mannelijke trekken vertoont. Het economisch paradigma zelf is dus aan revisie toe.

Het werk bestaat uit twee delen. In deel 1 worden de feministische economie en het feministisch poststructuralisme voor het voetlicht ge-

bracht, waarna de "rational economic man", vertegenwoordigd door de dikwijls als didactisch hulpje gebruikte Robinson Crusoe, op intertextuele wijze gedeconstrueerd wordt. In de laatste twee hoofdstukken wordt een neoklassiek model van gecontracteerd draagmoederschap voorgesteld. Op het eerste gezicht hoort dat model niet in het boek thuis, maar op het tweede gezicht misschien wel, aangezien de auteur - m.i. op weinig overtuigende wijze - meent aan te kunnen tonen dat zoiets vrouwelijks als zwangerschap niet geanalyseerd kan worden binnen de "fallocentrische" (haar woorden) neoklassieke economie.

De auteur geeft zelf aan dat haar aanpak, zelfs in kringen van feministische economen, marginaal is. Maar dat is natuurlijk geen reden om er geen boek over te schrijven. Wie vlot leesbare lectuur zoekt, bijvoorbeeld om een vakantie in Benidorm te overleven, is hier wel aan het verkeerde adres. Het werk vereist een meer dan modale vrijetijdconcentratie, maar het belooft de gemotiveerde lezer door de veelheid, de rijkdom en de belezen onderbouwing van de gepresenteerde ideeën, zelfs al is die het er niet altijd mee eens.

Marc Jegers