

# **LEUGENS, GROTE LEUGENS**

## **EN KWANTITATIEVE VOORSPELLINGEN**

Rede, in verkorte vorm uitgesproken  
bij het aanvaarden van het ambt van  
hoogleraar in de kwantitatieve methoden  
in de Faculteit der Economische Wetenschappen en Econometrie  
van de Universiteit van Amsterdam  
op woensdag 20 mei 1992

door

Dr. Ir. Jan G. de Gooijer

1992

*Universiteit van Amsterdam*

*Mijnheer de Rector Magnificus,*

*Zeer gewaardeerde toehoorders,*

## Inleiding

Toen in de 19e eeuw de Engelse staatsman Benjamin Disraeli (1804-1881) publiekelijk zijn wantrouwen in officiële statistieken, als verzameling van getallen, uitte met de woorden "*There are three kinds of lies: Lies, damned lies, and statistics*" <sup>1)</sup>, kon hij niet voorzien dat daarmee tevens lange tijd een stempel werd gedrukt op de statistiek als wetenschap. Pas gedurende en na de Tweede Wereldoorlog is de betekenis en het nut van zowel de statistiek als van statistieken, op talloze wetenschappelijke en maatschappelijke terreinen ten volle duidelijk geworden. Hedenmiddag zullen enkele collegae dit nog eens met boeiende voorbeelden en resultaten trachten te schetsen. De verleiding om vanuit mijn specifiek vakgebied hieraan mee te doen was groot. Echter, zoals de titel van mijn oratie reeds aangeeft, wil ik u in dit kort tijdsbestek confronteren met een lichte variant op Disraeli's uitspraak. Niet zozeer het publiekelijk wantrouwen in statistieken, maar de validiteit van kwantitatieve voorspellingen en de integriteit van voorspellers spelen hierin de hoofdrol <sup>2)</sup>.

## Verleden en heden

Een oud gezegde luidt: Voorspellen is de kunst iets te zeggen over een gebeurtenis in de toekomst, en dan uit te leggen waarom de voorspelling niet is uitgekomen. Hoe afgezaagd deze spreuk wellicht ook klinkt het duidt op een sceptische houding, welke heden ten dage nog steeds een overweging kan zijn bij het al of niet accepteren van voorspellingen. In zoverre is er historisch bezien misschien niet zo veel veranderd. In de oudheid werd de voorspeller soms hoog geacht, zoals het orakel van Delphi met zijn dubbel- of veelzinnig interpreteerbare uitspraken. Dikwijls echter werd voorspellen beschouwd als een noodzakelijk kwaad. Soms werd het zelfs illegaal bevonden, waarbij de Romeinse keizer Constantijn in het jaar 337 na Christus ongetwijfeld het meest ver ging door een edict uit te vaardigen waarbij het een ieder werd verboden "een waarzegger, wiskundige, of een voorspeller" te consulteren. Met daaraan de toevoeging "moge nieuwsgierigheid om de toekomst te voorspellen voorgoed tot zwijgen zijn gebracht".

Tegenwoordig is voorspellen een onlosmakelijke activiteit bij het plannen, het ontwikkelen van strategieën en alle andere op de toekomst gerichte besluitvorming. Op individueel niveau rangeert het van trivialiteiten zoals het 's ochtends zo goed mogelijk afstemmen van onze kleding op de weersverwachting van die dag tot aan belangrijke beslissingen zoals het investeren van spaartegoeden. Voor de optimale besluitvorming in een onderneming zijn voorspellingen een voorname informatiebron voor ondermeer het investeren in nieuwe technologieën, het bouwen van fabrieken, en het introduceren van nieuwe producten. Op het niveau van de overheid spelen voorspellingen een cruciale rol bij het tijdig anticiperen op tal van monetaire, financiële, technologische en demografische ontwikkelingen. Dit alles brengt



met zich mee dat het gewenst is op zo nauwkeurig en wetenschappelijk mogelijke wijze te voorspellen.

Alhoewel methoden uit de wetenschap bij het voortbrengen van uitspraken over de toekomst al voor en tijdens de Tweede Wereldoorlog werden gebruikt, begint toekomstonderzoek pas vanaf de jaren zestig langzaam tot een volwassen wetenschappelijke discipline uit te groeien. Op het terrein van kwantitatieve voorspellingen, waartoe ik mij in de rest van deze oratie voornamelijk zal beperken, is deze groei mogelijk geworden dankzij grote ontwikkelingen die zich gedurende de laatste twee decennia voordeden in ondermeer de statistiek, econometrie, en tijdreeksanalyse.

De groei naar volwassenheid is mede bevorderd doordat moderne wetenschappelijke voorspellingen aan een aantal eisen moeten voldoen. Te noemen valt de eis van ondubbelzinnigheid waarmee, anders dan de profetieën uit de oudheid, bedoeld wordt dat een voorspelling duidelijk is en niet voor velerlei interpretaties vatbaar. Voorts dient een voorspelling te voldoen aan de eis van falsificeerbaarheid. Anders gezegd: de argumenten, methoden en technieken waarop een voorspelling gebaseerd is, dienen stapsgewijs verifieerbaar te zijn. Bij kwantitatieve voorspelmethoden kan dit laatste, door de gehanteerde wiskundig-statistische symboliek, nogal eens lastig zijn. Ten slotte mag men verlangen dat een wetenschappelijke voorspeller zich vooraf rekenschap geeft van de nauwkeurigheid en betrouwbaarheid van een voorspelling. Daarbij wordt met nauwkeurigheid bedoeld een bepaalde maat voor de grootte van het zogenaamde voorspellingsinterval. De betrouwbaarheid daarentegen geeft de kans aan dat een voorspelling uitkomt.

Bovenstaande vereisten zijn praktisch en doelmatig. Tezamen met het hanteren van kwantitatieve methoden hebben ze het publiekelijk wantrouwen aangaande voorspellingen en voorspellers aanzienlijk verminderd. Maar de volledige weg naar volwassenheid van de voorspellingsdiscipline is nog lang niet afgelegd. Volwassenheid kan verder worden bereikt door enige vorm van consensus tussen onderzoekers en gebruikers van voorspellingen tot stand te brengen. Hierin dient op zijn minst een erkenning te liggen van het feit dat voorspellen zowel een "kunst" als een "wetenschap" is. Het merendeel van de kunst zal komen van de gebruiker en de wetenschap van de onderzoeker. Een dergelijke consensus kan denk ik alleen worden bereikt door op nuchtere en eerlijke wijze de huidige stand van de kwantitatieve voorspelmethoden inclusief hun zwakten te schetsen.

## **Kwantitatieve voorspelmethoden**

### *Classificatie en praktijk*

Aan nagenoeg alle kwantitatieve voorspelmethoden liggen ideeën en concepten ten grondslag uit de tijdreeksanalyse, waarbij onder een tijdreeks wordt verstaan een reeks waarnemingen van een variabele gemeten in de tijd. De verkregen afzonderlijke waarnemingen kunnen worden beschouwd als een steekproef uit een onderliggend proces. Dit proces wordt verondersteld gedeeltelijk gedetermineerd te zijn of op zijn minst volledig te kunnen worden gekenschetst door een kansverdeling. In zoverre komt dit overeen met de relatie tussen steekproef en populatie in de klassieke statistiek. Het verschil met de analyse van simultane of tijdloze statistische gegevens is echter dat daarin een ideale of aangenomen onafhankelijkheid

tussen waarnemingen bestaat, terwijl dit bij tijdreeksen vrijwel nooit het geval is. Juist de vroegere waarden van een reeks zijn meestal van invloed op de latere. Het aanvankelijk doel van tijdreeksanalyse is om uit de waargenomen reeks een beschrijving of karakterisering in de vorm van een deterministisch of een stochastisch model af te leiden. In tweede instantie kunnen daaruit weer nieuwe waarden worden bepaald.

Kwantitatieve voorspelmethoden kunnen ruwweg worden geclassificeerd in enkelvoudige methoden, waarbij voorspellingen van één enkele tijdreeks plaatsvinden op grond van geconstateerde regelmatigheden en patronen in het verleden; en in meervoudige of causale methoden, waarbij voorspellingen ten minste gedeeltelijk afhangen van waarnemingen van andere tijdreeksen, ook wel de causale of verklarende variabelen genoemd. Een andere classificatie is tussen automatische en niet-automatische methoden. De eerste kunnen los van enige menselijke tussenkomst worden uitgevoerd. Daarmee wordt een in de praktijk doorgaans moeilijk en moeizaam proces van modelspecificatie vermeden. Anderzijds kunnen automatische voorspelmethoden veel eerder tot misbruik leiden door onderzoekers, die deze methoden onkritisch voorbij hun beperkingen hanteren. De methoden in de tweede groep vereisen op zijn minst enige subjectieve interpretatie van de voorspeller. Over het algemeen zijn meervoudige methoden niet-automatisch. Daarentegen kunnen de meeste enkelvoudige methoden, inclusief die van Box-Jenkins (waarover zo meteen meer), indien gewenst volledig automatisch worden gebruikt.

Een verdere onderverdeling van voorspelmethoden ontstaat door een onderscheid te maken tussen stochastische en deterministische modelcomponenten in de onderliggende tijdreeksmodellen. Afvlakmethoden zijn voornamelijk deterministisch van aard. Daarbij gaat men uit van een langzaam veranderend patroon in het verschijnsel achter de tijdreeks. Verondersteld wordt dat dit patroon zich in de reeks weerspiegelt als een trend. Talrijke varianten zijn op dit thema bedacht waaronder de in de praktijk veel gehanteerde exponentiële afvlakmethode van Holt-Winters. Binnen de klasse van stochastische methoden hebben met name (enkelvoudige en meervoudige) lineaire ARIMA modellen veel aandacht gekregen <sup>3)</sup>. Met deze modellen tracht men het toevalskarakter van een tijdreeks te karakteriseren. Het in 1970 gepubliceerde boek van de statistici Box en Jenkins bevat een coherent en pragmatisch raamwerk voor het analyseren, modelleren en voorspellen van tijdreeksen aan de hand van ARIMA modellen. Dit raamwerk staat tegenwoordig beter bekend als de Box-Jenkins methode.

Voorts is een groot aantal andere, vaak minder bekende methoden voorgesteld. Ik noem slechts: de X-11-ARIMA methode (Huot e.a., 1986), waarbij via een reeks lineaire afvlakfilters seizoen-correctie plaatsvindt; de Bayesiaanse methode (West en Harrison, 1989), die het mogelijk maakt om met gekwantificeerde subjectieve waarschijnlijkheden over de toekomst berekeningen uit te voeren; en de structurele voorspellingsmethode (Harvey, 1989), waarbij expliciet de dynamische veranderingen in trend en seizoen worden gemodelleerd.

Veel van deze methoden zijn niet of nog nauwelijks doorgedrongen tot de praktijk. Bij de overheid en het bedrijfsleven wordt, binnen de klasse van kwantitatieve methoden, vaak de voorkeur gegeven aan eenvoudige, begrijpbare methoden voor het maken van korte-termijn voorspellingen. Daarbij scoren voortschrijdende gemiddelden, exponentiële afvlakmethoden <sup>4)</sup> en lineaire regressie modellen verhoudingsgewijs hoog, terwijl statistisch meer geavanceerde methoden nog weinig worden gebruikt. Zo gaf bij een in 1983 gehouden schriftelijke enquête



onder 500 Amerikaanse bedrijven (Mentzer en Cox, 1984) ruim 70% van de respondenten, op een totale respons van 32%, te kennen met name met bovengenoemde methoden vertrouwd te zijn. Daarentegen had slechts 26% praktische ervaring met de Box-Jenkins methode, die overigens vrij veel expertise van de voorspeller vereist. Deze resultaten zijn in overeenstemming met een soortgelijk onderzoek gehouden onder Engelse bedrijven in 1984 (Sparkes en McHugh, 1984)<sup>5)</sup>. In 1991 is dit onderzoek voor Engeland nog eens herhaald (Sarantis e.a., 1991). Vergelijking van de resultaten van beide Engelse enquêtes laat zien dat de rangorde van de gehanteerde methoden in de loop der tijd niet is veranderd en dat nieuwe statistische methoden, zoals die van Box-Jenkins, slechts sporadisch hun weg naar de praktijk hebben gevonden. Enerzijds is dit vrij opmerkelijk daar deze nieuwe methoden geacht worden een hogere korte-termijn voorspelaccuratesse te hebben dan tal van andere eenvoudige kwantitatieve methoden. Anderzijds, gelet op de resultaten van de zo meteen te bespreken voorspelcompetities, is de voorkeur voor traditionele en welbeproefde methoden begrijpbaar.

### *De "beste" methode*

Hoe kwantitatieve voorspelmethode ook worden geclassificeerd en welke statistische principes aan een methode ten grondslag liggen, een vaak gestelde vraag is: welke methode kan als de "beste" worden beschouwd. Helaas valt hierop niet een eenvoudig antwoord te geven. Experts zijn thans van mening dat geen enkele methode de beste is in iedere denkbare situatie, gegeven de grote verscheidenheid aan problemen die ieder een eigen specifieke benadering vereisen (Jenkins, 1982).

De keuze voor een methode hangt af van een veelheid aan factoren, waaronder het doel waarvoor voorspellingen worden gemaakt. Met andere woorden: de besluitvorming die op de voorspelling wordt gebaseerd in een door onzekerheid en risico gekarakteriseerde omgeving. Voorts hangt de keuze af van het niveau waarop een tijdreeks is gemeten, b.v. micro of macro, en de statistische eigenschappen van een reeks zoals de aan- of afwezigheid van een trend en/of een seizoencomponent. Ook de relatieve omvang van permanent systematische componenten, tijdelijk systematische componenten en stochastische componenten is erg belangrijk. Tenslotte spelen factoren een rol die betrekking hebben op:

- de kwantiteit en kwaliteit van de beschikbare historische gegevens;
- of lange-, middellange dan wel korte-termijn voorspellingen nodig zijn;
- het aantal te voorspellen reeksen en de daaraan per reeks verbonden kosten;
- de mate van ervaring, kennis en kunde aanwezig bij een voorspeller; en
- de beschikbaarheid van computerprogramma's.

Daarnaast moet er klaarheid zijn over de aard van de gewenste voorspelling. Dient men, bijvoorbeeld, slechts één enkele periode te voorspellen, of wordt een cumulatieve voorspelling voor verschillende perioden verlangd. En tevens: in hoeverre is een puntvoorspelling voldoende, of moet een voorspellingsinterval worden afgeleid. Bij het beantwoorden van al die vragen stuit men vervolgens weer op nieuwe moeilijkheden. Zo kan de constructie van een voorspellingsinterval plaatsvinden op basis van de veronderstelling dat modelvorm en geschatte parameters in de toekomst niet veranderen. Anderzijds kan aan de hand van waargenomen éénstap-vooruit "voorspelfouten" over de "gefite" periode een voorspellingsinterval worden afleiden. Men moet zich daarbij wel realiseren dat beide benaderingen berusten op de identificatie van historische stabiliteit. Er bestaat derhalve geen enkele garantie dat toekomstige

voorspelfouten aselekt en symmetrisch verdeelde toevalsvariabelen zijn, of binnen het bereik van de variantie van voorspelfouten uit het verleden liggen. Uit empirisch onderzoek is evenwel gebleken dat voorspellingsintervallen gebaseerd op kwantitatieve methoden in het algemeen te nauw zijn, d.w.z. veel meer dan  $\alpha\%$  van de waarnemingen valt buiten de nominale  $(100-\alpha)\%$  grenzen. Oftewel, men is in de praktijk vaak te optimistisch over de nauwkeurigheid van een voorspelling.

Let wel, bovenstaande opmerkingen hebben slechts betrekking op korte-termijn voorspelmethoden. Wie geïnteresseerd is in lange-termijn-prognoses, d.w.z. minstens 10 jaar vooruit, zal weinig houvast hebben aan de huidige stand der wetenschap bij de keuze van de beste kwantitatieve voorspelmethode. Om eerlijk te zijn, ben ik uiterst pessimistisch over de mogelijkheden om m.b.v. deze methoden belangrijke ontwikkelingen in de verre toekomst te voorspellen zoals nieuwe trends, innovatieve technologieën en producten. Recent is dit nog eens bevestigd in een, ter viering van het 20-jarig bestaan van het tijdschrift *Technological Forecasting and Social Change*, speciaal uitgebracht nummer. Daarin klaagt de ene gerenommeerde auteur na de andere over de beperktheid van kwantitatieve methoden voor de vervaardiging van lange-termijn voorspellingen. Dit alles doet de hoofdredacteur (Linstone, 1989) herinneren aan het mannetje dat op het dak van zijn huisje klom om de maan beter te kunnen zien. Anders gezegd, deze methoden hebben het terrein van lange-termijnprognoses niet ver gebracht, en het ziet er nochtans niet naar uit dat ze dat ooit zullen doen. Dat enkel kwalitatieve (lees: informele) toekomstbenaderingen voor dit doel beter geschikt zouden zijn, is tot op heden ongegrond gebleken (zie b.v. Lawrence e.a., 1985).

### *Voorspelcompetities*

Teneinde de keuze tussen enkelvoudige voorspelmethoden te vergemakkelijken zijn er verschillende vergelijkende "competities" geweest, waarbij een keur aan tijdreeksen door diverse experts, ieder met een andere methode, werden voorspeld. De M-competitie uit 1982, genoemd naar de initiatiefnemer Makridakis (Makridakis e.a., 1982), is zonder twijfel de meest omvangrijke en grondige studie op dit gebied. Daarbij werden 24 methoden met elkaar vergeleken aan de hand van 1001 reeksen, afkomstig uit een breed scala aan informatiebronnen. Hiervoor werden de reeksen in twee delen gesplitst, het eerste deel diende voor de modelkeuze, terwijl op basis van het tweede deel de voorspelnauwkeurigheid in een aantal beschrijvende statistische maten werd uitgedrukt. Met uitzondering van de Bayesiaanse methode en die van Box-Jenkins werden alle methoden automatisch gehanteerd <sup>6)</sup>.

Waarschijnlijk onvermijdelijk met dit soort competities zijn er naar mijn mening net zoveel vragen opgeroepen als er zijn beantwoord. Ten dele is dit een gevolg van de slechte presentatie van de onderzoekresultaten in allerlei maatstaven en tabellen. Maar ook het feit dat er geen duidelijke conclusies werden getrokken door de grote groep participerende experts, vormt een ernstige handicap bij de interpretatie en evaluatie van de resultaten. Desondanks zien sommigen de M-competitie als een baken voor verdere ontwikkelingen in zowel de theorie als de praktijk van het voorspellen (McLaughlin, 1983). Anderen daarentegen beschouwen vergelijkende studies als volslagen belachelijk (Jenkins en McLeod, 1982, voorwoord).

De geuite kritiek richt zich ondermeer op het punt dat in dergelijke studies de gespecificeerde modellen niet worden gepubliceerd. Dit was zeker het geval bij eerdere vergelijkende studies. Maar bij de M-competitie is volledige repliceerbaarheid van alle stappen in het voorspelproces



door een ieder mogelijk. Voorts wordt als bezwaar aangevoerd dat in de praktijk onvoldoende of totaal geen kennis aanwezig is om naar behoren met een bepaalde methode om te gaan. Hiertegen kan men inbrengen dat de expertise van de bij de meeste competities betrokken deskundigen ver boven het gemiddelde ligt. Dit is naar mijn mening een veel realistischer afspiegeling van de praktijk dan uit te gaan van de stelling dat de voorspeller in de praktijk totaal geen kennis over een methode heeft. Een derde, en meer wezenlijk punt van kritiek heeft betrekking op het vergelijken van de voorspelnauwkeurigheid van methoden uitsluitend aan de hand van de beschrijvende statistische maten. Ook de verwachte kosten en het gebruiks- en interpretatiegemak zijn in de praktijk net zo belangrijke aspecten als nauwkeurigheid. Waarschijnlijk is echter het grootste bezwaar tegen competities dat deze vrijwel altijd uitgaan van een volledig automatische analyse van grote aantallen tijdreeksen met behulp van "standaard-methoden". In mijn ogen is dit een te steriele benadering van de realiteit, waar via een interactie tussen voorspeller en gebruiker vaak ook "maatwerk" voor een voorspelvariabele wordt afgeleverd.

Het voert te ver om hier op alle aspecten van de M-competitie in te gaan. Voor een samenvatting van de belangrijkste resultaten verwijs ik graag naar Makridakis (1986, Exhibit 1). Een van de hoofdconclusies is dat geen enkele methode uniform het meest nauwkeurig blijkt te zijn in alle situaties. Wel kan men de nauwkeurigheid verbeteren door meerdere voorspellingen, verkregen m.b.v. verschillende modellen, tot één nieuwe voorspelling te combineren. Een mogelijke verklaring is dat géén enkele voorspelling gebaseerd is op alle mogelijk relevante informatie over een systeem. Kwantitatieve modellen zijn slechts formele benaderingen daarvan, en bovendien kan de onderliggende processtructuur in de toekomst veranderen.

### *Combineren*

In de literatuur is recentelijk veel onderzoek op het terrein van combineren verricht. Daarbij wordt soms de slogan "twee weten meer dan een" (Jain, 1988) gebruikt. Zo kunnen combinaties van voorspellingen bijvoorbeeld de vorm aannemen van een enkelvoudig gemiddelde of van een wegingsschema waarbij expliciet rekening wordt gehouden met de historische ontwikkeling van de voorspelfouten in ieder model. Alhoewel het strikt genomen juist is dat een combinatie van voorspellingen de kans verkleint op basis van de slechtste voorspelling besluitvorming te plegen, is het even zo goed mogelijk dat deze slechter is dan de beste van de afzonderlijke componenten. In feite kan een gecombineerde voorspelling niet nauwkeuriger zijn dan ieder van de afzonderlijke componenten, tenzij een aantal daarvan systematisch de toekomstige waarde van een variabele onderschat, terwijl andere deze juist overschatten. Wijken anderzijds alle voorspellingen abusievelijk af in dezelfde richting dan zal een combinatie slechter zijn dan tenminste een van de afzonderlijke componenten.

Uit empirische studies blijkt dat combinatiestrategieën, ter verhoging van voorspelnauwkeurigheid, in het algemeen zinvol zijn (zie b.v. Winkler en Makridakis, 1983). Toch valt bij deze aanpak wel een paar kanttekeningen te plaatsen. De basisgedachte achter deze strategieën is bevat in het deel van de zinsnede "voorspellingen van dezelfde grootheid". Dit roept maar al te gemakkelijk associaties op met de theorie van de inferentiële statistiek, gebaseerd op waarnemingen aan identiek verdeelde toevalsvariabelen. In de sociale en economische wetenschappen, waar door verschillende theorieën, modellen, ervaringen en oordelen de variatie in de waarnemingen sterk kan worden beïnvloed, is deze associatie echter

niet altijd evident. Iedere waarneming of voorspelling wordt voorgesteld door een eigen theorie of model en deze hoeft niet per se identiek te zijn aan andere waarnemingen of voorspellingen. Niet iedere voorspelling zal een zuivere schatter zijn van een uniek gedefinieerde toevalsvariabele, en bovendien zullen voorspellers niet altijd identieke systemen beschouwen. In een dergelijke situatie is een gecombineerde voorspelling lastig te interpreteren in verhouding tot een theorie of een model. Het is eerder een vorm van *wishfull thinking* dat het gemiddelde van een aantal voorspellingen nauwkeuriger is dan iedere afzonderlijke component. Sommige informatie over de toekomst zal onvermijdelijk verloren gaan, en aan de gecombineerde voorspelling zal doorgaans niet een zinvol kwantitatief model van het onderliggend proces ten grondslag liggen.

Hoe dient men dan wel met diverse alternatieve voorspellingen om te gaan. Wellicht kan het de geloofwaardigheid en volwassenheid van de discipline ten goede komen indien, in plaats van te streven naar één enkele onnauwkeurige puntvoorspelling, de mogelijke aanwezigheid van vele misschien min of meer gelijkwaardige toekomstbeelden expliciet wordt erkend. In dat geval vormen alle afzonderlijke voorspellingen, elk eventueel voorzien van subjectieve betrouwbaarheden, één verdelingsfunctie van de toekomstige waarde van een voorspelvariabele. Bij een dergelijke opzet kan een gebruiker kiezen voor de meest plausibele verwachte waarde of waarden, bijvoorbeeld aan de hand van de spreidingsbreedte, zonder daarbij minder plausibele voorspellingen uit het oog te verliezen <sup>7)</sup>. Voorspellers dienen dan wel duidelijker de voorwaarden aan te geven waarop hun resultaten zijn gebaseerd.

## De realiteit

Geachte toehoorders, we hebben in vogelvlucht gezien dat gedurende de laatste 20 jaar de voorspellingsdiscipline een aantal interessante ontwikkelingen heeft doorgemaakt. Dit niet in de laatste plaats door de eerder genoemde voorspellingscompetities die weer geleid hebben tot specifieke onderzoekthema's zoals dat van de combinatiestrategieën. Ook heeft de tijdreeksanalyse en talrijke daarop geënte methoden en technieken een grote invloed gehad op de huidige stand van zaken van het vakgebied. Thans verricht een groot aantal onderzoekers met veel inzet en kunde wetenschappelijk onderzoek op het terrein van voorspellingen. Desondanks is de nuchtere realiteit dat de voorspellende waarde van moderne kwantitatieve methoden over vrijwel de gehele linie nogal eens teleurstelt. Zelfs de meest verfijnde statistische methoden blijken vaak een niet of nauwelijks hogere voorspelnauwkeurigheid te hebben dan de meer naïeve of eenvoudige methoden.

De oorzaken voor deze zwakten zijn ten dele terug te voeren op de beperkingen van het materiaal waarmee wordt gewerkt. Voor het bepalen van de verwachte stand van een planeet of een ster kan men zich verheugen op lange waarnemingsreeksen. Voor de verklaring van geconstateerde regelmatigheden in deze reeksen staan goed gestaafde universele theorieën en wetmatigheden in de astronomie ter beschikking, die uitnodigen tot extrapolatie over te gaan. Zo kan het tijdstip waarop morgen of over een jaar de zon opkomt tot op de honderdste van een seconde nauwkeurig worden vastgesteld. Voor het voorspellen in de economische en sociale werkelijkheid ontbreekt het vaak aan dit alles.



Voor een ander deel zijn de oorzaken te herleiden tot de zwakten van de grondslagen van kwantitatieve voorspelmethoden. Uiteraard kan, als de relatie tussen de variabelen in een model uitsluitend van deterministische aard is, de waarde van een voorspelvariabele exact uit de andere variabelen worden voorspeld, zo deze volledig bekend zijn. De bekende variabelen bevatten dan alle informatie over de toekomst. Statistiek komt daaraan niet te pas. Bij kwantitatieve methoden wordt daarentegen uitgegaan van een statistisch verband tussen de variabelen onderling en in de tijd. Van de sterkte van dit verband hangt af hoeveel de voorspelvariabele door zichzelf of door andere variabelen wordt verklaard en welke nauwkeurigheid aan een voorspelling kan worden toegekend. Maar een kwantitatieve voorspelling kan nooit vrij van fouten zijn. Deze fouten worden veroorzaakt door de onvoorspelbaarheid van de toevalscomponenten, door inherente statistische schattingsfouten, door waarnemingsfouten in de tijdreeksen, door gebrek aan overeenkomst tussen model en werkelijkheid en door onvoorziene veranderingen in de belangrijkste aspecten van die werkelijkheid.

Tenslotte zijn de zwakten gedeeltelijk inherent aan de methoden en technieken zelf. Vaak zijn deze te technisch verfijnd, hetgeen geen garantie voor voorspelnauwkeurigheid behoeft te zijn. Ook zijn veel kwantitatieve methoden en varianten ontleend aan andere, formele en empirische disciplines en niet specifiek ontworpen om betrouwbare voorspellingen voort te brengen. Een voorbeeld vormen de grote macro-economische modellen uit de jaren zeventig. Aanvankelijk waren deze modellen bedoeld om economische theorieën en verschijnselen te toetsen alsmede te analyseren en niet zozeer om daarmee te voorspellen. Echter het meest wezenlijke probleem binnen deze categorie is dat kwantitatieve methoden geen betrouwbare indicatie voor de voorspelfout kunnen verschaffen. Ten onrechte wordt meestal uitgegaan van toevalsvariëaties van een constant systeem en wordt geen rekening gehouden met veranderingen in systemen. Bijgevolg kan het minimaliseren van de éénstaps-vooruit voorspelfouten van een model "gefit" op historische gegevens geen garantie zijn dat toekomstige voorspelfouten minimaal zullen zijn.

### **Een blik vooruit**

Toegegeven, al deze zwakten zijn bepaald niet bevorderlijk voor het imago van voorspellers. Ze zijn misschien ook nog maar het topje van de spreekwoordelijke ijsberg. Toch verwachten veel consumenten/gebruikers van voorspellingen dat deze altijd juist zijn en dat hun onzekerheid beperkt is. Het feit blijft echter dat zulke verwachtingen volslagen onrealistisch zijn. Ongetwijfeld zullen vanwege de hierboven genoemde zwakten voorspellingen soms verkeerd zijn, maar wat is het alternatief? Als zoiets het geval is, is het veel zinvoller de oorzaken daarvan op te sporen dan deze zwakten op te vatten als een vrijbrief om formele methoden terzijde te schuiven en in te wisselen voor andere meer subjectieve en intuïtieve werkwijzen. Voor een dergelijke reactie zullen tegenargumenten moeten worden aangedragen, die aantonen dat laatstgenoemde methoden in theorie en praktijk tot betere resultaten leiden.

De kloof tussen de realiteit van voorspellingen en verwachtingen, die daarover bij consumenten/gebruikers leven, kan worden verkleind door een aantal stappen ter verbetering te zetten. Die verbeteringen liggen zowel op het beroepsmatige als op het technische vlak. Allereerst dienen producenten van voorspellingen gebruikers beter te informeren dat de

onzekerheid in iedere op de toekomst gerichte activiteit groot kan zijn. Ook het wetenschappelijk gedachtenproces achter een methode moet zonder overspannen verwachtingen, met de nodige voorbehoud en met aanduiding van de beperkingen aan de gebruiker worden gepresenteerd. Vaak vormt dit proces ook een natuurlijke ingang om extra informatie over het systeem van de gebruiker los te krijgen, waarmee deze een actieve rol in het totale voorspellingsproces krijgt. Bij dit alles zijn goede communicatieve vaardigheden van de voorspeller niet te verwaarlozen.

Een andere verbetering die het beroepsaanzien ten goede kan komen heeft betrekking op de voorspeller zelf. Gegeven de complexe wereld waarin wij verkeren, hebben de meeste gebruikers begrip voor het feit dat voorspelfouten onvermijdelijk zijn. Als een voorspeller echter de oorzaak van deze fouten nauwelijks weigert te erkennen, laat staan deze wenst te analyseren dan oogst dit weinig bijval. In feite ondermijnt een dergelijke houding de geloofwaardigheid van de professie (Chase, 1992). Voorspellers hebben tot taak altijd een kritische reflectie naar een gebruiker te blijven organiseren, ondanks het besef dat er grenzen zijn aan de mogelijkheden van gehanteerde methoden en technieken in een bepaalde context.

Op het technisch vlak bestaat er volgens velen een kloof tussen wat gebruikers werkelijk als nuttig ervaren, of als een deel van hun behoeften beschouwen, en wat onderzoekers produceren (Armstrong, 1988). Ook hier dient verbetering in te komen. Talrijke praktijkproblemen en omissies die reeds geruime tijd geleden in kaart zijn gebracht wachten dringend op een oplossing (zie b.v. Tukey (1980), Ord (1988) en de Gooijer en Kumar (1992)). Hier ligt een hernieuwde taak weggelegd voor de academisch ingestelde voorspeller/tijdreeksanalyst. De sterke wisselwerking tussen deze groep en gebruikers van voorspellingen heeft in het verleden uiterst vruchtbare gewerkt. Tegenwoordig is de kruisbestuiving tussen beide groepen veel minder groot (de Gooijer, 1990). Voorspellers en tijdreeksanalysten zijn meer geïnteresseerd in het ontwikkelen, verfijnen, en aanpassen van verschillende theoretische aspecten van bestaande methoden dan dat men oog heeft voor vraagstukken van werkelijk praktische aard. Weliswaar is de dichotomie tussen academische onderzoeksbelangstelling en de praktijk niet iets unieks voor de tijdreeks/voorspellingsdiscipline. Echter juist deze discipline ontleent bij uitstek haar bestaansrecht aan het oplossen van reële problemen. Vindt dit niet plaats dan is het risico niet ondenkbeeldig dat de professie haar zichtbaarheid, invloed en vertrouwen verliest. Door juist in de geest van deze voorstellen verder te handelen kan de voorspellingsdiscipline naar buiten treden als een bonafide wetenschap die, door een sterke verweving van theorie en praktijk, ingezet wordt om onzekerheid over de toekomst te reduceren en gebruikers van waardevolle adviezen te voorzien.

## Slot

Dames en heren.

Ik heb geprobeerd aan te geven wat de huidige ontwikkelingen zijn op het terrein van kwantitatieve voorspelmethoden. Ook heb ik gepoogd de zwakten van deze methoden op nuchtere wijze te schetsen. Daarmee hopen dat de waas van wantrouwen rond voorspellingen is weggenomen of op zijn minst is verminderd. Uiteraard zult u uit mijn betoog wel begrepen



hebben dat de titel van mijn oratie slechts bedoeld was om dit onderwerp wat aan te scherpen. Tenslotte heb ik een pleidooi gehouden om de interactie tussen theorie en praktijk te verbeteren. Naar mijn overtuiging is het belang hiervan essentieel voor een verdere groei van de voorspellingsdiscipline naar volwassenheid. Maar de toekomst zal het leren.

Aan het slot van mijn betoog gekomen, wil ik allereerst degenen danken die betrokken zijn geweest bij mijn benoeming. Tevens dank ik allen die mijn verblijf aan de Universiteit van Amsterdam al jaren zo stimulerend en plezierig maken, in het bijzonder de leden van de Vakgroep Kwantitatieve Methoden. Gaarne ben ik mijn collega's in de Faculteit der Economische Wetenschappen en Econometrie erkentelijk voor de aangename contacten die ik met velen onder u reeds jaren onderhoud. Graag spreek ik de hoop uit deze contacten op dezelfde wijze te mogen voortzetten. Ten slotte spreek ik met mijn collega's en vrienden statistici aan deze universiteit graag af, dat we gezamenlijk in de komende tijd met evenveel enthousiasme de samenwerking rond de Vrije Studierichting Statistiek inhoud en vorm geven dan we tot nu toe al hebben gedaan.

Geachte toehoorders, ik dank u voor uw belangstelling.

## Noten

- 1) Zie daartoe de autobiografie van de Amerikaanse schrijver Mark Twain.
- 2) Reeds in 1952 had de redactie van het tijdschrift *Statistica* (later omgedoopt in *Statistica Neerlandica*) de vooruitziende blik een geheel nummer aan dit onderwerp te wijden, met ondermeer bijdragen van D. van Dantzig en A.D. de Groot; *Statistica*, 6, blz. 195-279.
- 3) Soms wordt de mening verkondigd dat alle exponentiële afvlakmethoden een bijzonder geval zijn van de meer algemene Box-Jenkins methode, hetgeen zou impliceren dat men in de praktijk met de laatste methode kan volstaan. Voor wat de Holt-Winters methode betreft, is men onlangs hier anders over gaan denken (Chatfield en Yar, 1988).
- 4) Een aardig voorbeeld in dit verband is het onlangs door het Ministerie van O & W ingevoerde "Ramingen Uitkeringen Model". Op verschillende aggregatieniveaus (directies, onderwijssoorten, uitkeringscategorieën en functiecategorieën) tracht men daarmee het volume aan gelden te voorspellen dat aan uitkeringen van onderwijzend personeel zal worden besteed. Het voorspellen gebeurt met behulp van de dubbele (afvlakking van niveau én trend) exponentiële afvlakmethode van Holt-Winters.
- 5) Soortgelijke gegevens voor het Nederlands bedrijfsleven zijn, voorzover mij bekend, niet voorhanden.
- 6) Vanwege hun complexiteit en toentertijd hoge rekenkosten werden deze methoden slechts op 111 reeksen toegepast. Per reeks nam een complete analyse m.b.v. de Box-Jenkins methode circa één uur van een expert in beslag, terwijl de Bayesiaanse methode vergelijkenderwijs ongeveer 5 minuten werk vergde. Momenteel zijn de rekenkosten van de Box-Jenkins methode aanzienlijk teruggebracht, maar arbeidsintensief blijft het.
- 7) Combineren kan ook worden gezien als een soort van *bootstrapping*; (zie b.v. de Gooijer, 1990).



## Referenties

- Armstrong, J.S. (1988), "Research needs in forecasting", *International Journal of Forecasting*, 4, blz. 449-365.
- Chatfield, C. en M. Yar (1988), "Holt-Winters forecasting: Theory and practice", *The Statistician*, 37, blz. 129-140.
- Chase, C.W. (1992), "Corporate America is looking for a few good forecasters", *The Forum: Newsletter of the International Association of Business Forecasting*, 5, blz. 4-5.
- Gooijer, J.G. de (1990), "The role of time series analysis in forecasting: A personal view", *International Journal of Forecasting*, 6, blz. 449-451.
- Gooijer, J.G. de en K. Kumar (1992), "Some recent developments in non-linear time series modelling, testing and forecasting", te verschijnen in *International Journal of Forecasting*, 8.
- Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Huot, G., Chiu, K. en J. Higginson (1986), "Analysis of revisions in the seasonal adjustment of data using X-11-ARIMA model-based filters", *International Journal of Forecasting*, 2, blz. 217-229.
- Jain, C.L. (1988), "Ways to combine forecasts". In C.L. Jain (ed.), *Understanding Business Forecasting*. Flushing: Graceway, blz. 151-153.
- Jenkins, G.M. (1982), "Some practical aspects of forecasting in organizations", *Journal of Forecasting*, 1, blz. 3-21.
- Jenkins, G.M. en G. McLeod (1982), *Case Studies in Time Series Analysis*, vol. 1. Lancaster: Gwilym Jenkins & Partners Ltd.
- Lawrence, M.J., Edmundson, R.H. en M.J. O'Connor (1985), "An examination of the accuracy of judgemental extrapolation of time series", *International Journal of Forecasting*, 1, blz. 25-36.
- Linstone, H. (ed.) (1989), "Twenty years of TF&SC", *Technological Forecasting and Social Change*, 36, 1-13.
- Makridakis, S., Andersen, A., Carbone, R., fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. en R. Winkler (1982), "The accuracy of extrapolation (time series) methods: results of a forecasting competition", *Journal of Forecasting*, 1, blz. 111-153.
- Makridakis, S. (1986), "The art and science of forecasting", *International Journal of Forecasting*, 2, blz. 15-39.
- Mclaughlin, R.L. (1983), "Commentary on the Makridakis time series competition (M-competition)", *Journal of Forecasting*, 2, blz. 274.
- Mentzer, J.T. en J.E. Cox, Jr. (1984), "Familiarity, application, and performance of sales forecasting techniques", *Journal of Forecasting*, 3, blz. 27-36.
- Ord, J.K. (1988), "Future developments in forecasting: The time series connexion", *International Journal of Forecasting*, 4, blz. 389-401.
- Sarantis, N., Hadjimatheou, G. en V. Daly (1991), "The use of forecasting services in British industry". School of Economics, Kingston Polytechnic, UK.
- Sparkes, J.R. en A.K. McHugh (1984), "Awareness and use of forecasting techniques in British industry", *Journal of Forecasting*, 3, blz. 37-42.

- Tukey, J.W. (1980), "Can we predict where "Time Series" should go next?" In D.R. Brillinger en G.C. Tiao (eds.), *Directions in Time Series: Proceeding of the IMS special topics meeting on time series analysis*. Ames: Iowa, 1978, blz. 1-31.
- West, M. en J. Harrison (1989), *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
- Winkler, R.L. en S. Makridakis (1983), "The combination of forecasts", *Journal of the Royal Statistical Society, Series A*, 146, blz. 150-157.