

DR. H.J. BIERENS

Specificatie-analyse in de econometrie

REDE uitgesproken bij aanvaarding van het ambt van gewoon
hoogleraar in de Econometrie/Wiskundige Economie aan de
Faculteit der Economische Wetenschappen en Econometrie van
de Vrije Universiteit te Amsterdam op 11 september 1987.



Mijnheer de Rector, Dames en Heren.

Economen bestuderen het economisch handelen van mensen, d.w.z. het gedrag ten aanzien van het aanwenden van schaarse middelen ter bevrediging van behoeften. Een belangrijk hulpmiddel daarbij is het economische model. Een economisch model beschrijft, veelal in wiskundige vorm, de relaties tussen economische variabelen die worden geïmpliceerd door elementaire gedragshypothesen en werkhypothesen. Zoals de naam aangeeft is een gedragshypothese een aanname over het economisch gedrag van mensen, zoals de aanname dat de consument streeft naar maximale behoeftebevrediging of de aanname dat de ondernemer streeft naar maximale winst. De complexiteit van de economische werkelijkheid maakt het noodzakelijk het te bestuderen economisch verschijnsel los te koppelen van het grote geheel waarvan het deel uitmaakt of vergaand te vereenvoudigen. Deze loskoppeling of vereenvoudiging gebeurt door aan de elementaire gedragshypothesen een of meer werkhypothesen toe te voegen die impliceren dat het te bestuderen economisch verschijnsel of proces niet van buiten af wordt beïnvloed of zich in een eenvoudige wereld afspeelt. De bekendste van die werkhypothesen is de zogenaamde "ceteris paribus" conditie: al het overige blijft onveranderd. Andere voorbeelden zijn de hypothese van een gesloten economie, en de hypothese van een representatieve consument of producent waardoor collectief gedrag kan worden geïnterpreteerd als individueel gedrag.

De elementaire gedragshypothesen vormen de axioma's van de economische theorie. Zij kunnen worden beschouwd als elementaire uitspraken over het te bestuderen economische verschijnsel of proces, op basis waarvan met behulp van de logica complexere beweringen worden afgeleid. Het samenstel van beweringen gevormd op basis van de elementaire gedragshypothesen, inclusief de elementaire gedragshypothesen zelf, is een economische theorie¹⁾. Een economisch model zou kunnen worden gekarakteriseerd als een economische theorie die verder is ingeperkt en vereenvoudigd door werkhypothesen. Dus een economisch model is een stelsel van beweringen op basis van elementaire gedragshypothesen én werkhypothesen. Indien deze beweringen

in wiskundige vorm zijn gegoten is er sprake van een wiskundig-economisch model. In mijn betoog zal ik verder onder economische modellen alleen wiskundig-economische modellen verstaan.

Model en werkelijkheid verschillen in twee opzichten van elkaar. In de eerste plaats zijn de elementaire gedragshypothesen doorgaans niet volledig in overeenstemming met het werkelijke gedrag, maar schetsen een ideaalbeeld waaraan slechts bij benadering is voldaan. In de tweede plaats zijn de werkhypothesen die het te bestuderen economische verschijnsel afzonderen van zijn omgeving per definitie onwaar. Immers, als de werkhypothesen waar zouden zijn, zouden we ze niet nodig hebben. Economische modellen zijn derhalve in wezen onwaar. Dit doet op zich niets af aan hun bruikbaarheid, mits we de beperkingen goed in het oog houden bij het trekken van conclusies op basis van een model.

Economische modellen laten alleen kwalitatieve uitspraken toe over economische verbanden. Economische modellen kunnen de richting van het effect van een verandering van een economische variabele op andere economische variabelen aangeven, gegeven de onderliggende gedrags- en werkhypothesen, maar niet hoe groot dat effect is. Daartoe zijn econometrische modellen nodig, d.w.z. modellen die kenmerken van specifieke data-generende processen beschrijven.

De constructie van een econometrisch model op basis van een economisch model gebeurt doorgaans in een drietal stappen. In de eerste plaats moet de functionele vorm van de wiskundige vergelijkingen waaruit een economisch model bestaat verder worden gespecificeerd als een bekende functies van modelvariabelen en onbekende parameters. Het begrip functionele vorm moet hierbij in ruime zin worden opgevat: het omvat ook de dynamische structuur van het model, voor zover het een tijdreeksmodel betreft. Vervolgens worden storingstermen aan de modelvergelijkingen toegevoegd en wordt de kansverdeling van deze storingstermen gespecificeerd. De laatste stap in de econometrische modelbouw is het schatten van de onbekende parameters op basis van waarnemingen van de verklarende en te ver-

klaren variabelen met behulp van econometrisch-statistische technieken.

De specificatie van de functionele vorm is soms gebaseerd op economisch-theoretische overwegingen, maar vaak is het achterliggende economisch model te weinig expliciet om van veel nut te zijn. In het laatste geval wordt meestal de gemakkelijkste weg bewandeld, namelijk een lineaire specificatie. Lineaire modellen genieten een grote populariteit in de econometrie omdat ze doorgaans gemakkelijk te schatten zijn en de effecten van veranderingen in de verklarende variabelen op de te verklaren variabelen direct kunnen worden bepaald aan de hand van de geschatte parameters. In het geval dat de specificatie van de functionele vorm wel is gebaseerd op economische overwegingen, is de relatie tussen economisch model en de functionele vorm van het econometrische model doorgaans indirect. Zo wordt bij de specificatie van vraagfuncties eerst de nutsfunctie gespecificeerd en worden op basis van de nutsmaximalisatie hypothese en de budget-restrictie de vraagvergelijkingen afgeleid. Ook hier wordt echter bijna altijd de weg van de minste weerstand gekozen: de nutsfunctie wordt zodanig gespecificeerd dat de resulterende vraagvergelijkingen een handzame vorm krijgen. Op deze wijze is bijvoorbeeld het bekende "Linear Expenditure System" tot stand gekomen. Dus de specificatie van de functionele vorm van het econometrisch model is in belangrijke mate direct of indirect gebaseerd op pragmatische overwegingen. In feite vormt de functionele specificatie een additionele werkhypothese met dezelfde strekking als de werkhypothesen die mede ten grondslag liggen aan het economische model, namelijk een hypothese die de analyse vergemakkelijkt.

Het aldus gespecificeerde model is deterministisch in de zin dat gegeven numerieke waarden voor de parameters er sprake is van een exact verband tussen de te verklaren variabelen en de verklarende variabelen. Dit deterministische model zal echter nooit geheel passend gemaakt kunnen worden op de economische werkelijkheid die het tracht te beschrijven, hoe flexibel de functionele vorm van het model ook is gespecificeerd. Als tweede stap wordt daarom het gat tussen model en werkelijkheid opgevuld door aan de modelvergelijkingen storingstermen

toe te voegen. Deze storingstermen zijn kansvariabelen die bepaalde eigenschappen moeten bezitten opdat het model een adequate afbeelding is van de werkelijkheid, met name de eigenschap dat de storingstermen zelf niet worden beïnvloed door de verklarende variabelen in het model. Immers, als dat wel het geval zou zijn, dan zou het in principe mogelijk zijn de storingstermen zelf voor een deel te verklaren uit de verklarende variabelen met behulp van een model, en dus was dan het eerdere model incompleet. In het geval van een regressiemodel, waar een te verklaren variabele wordt verklaard uit een functie van een aantal verklarende variabelen en onbekende parameters, plus een storingsterm, leidt deze eis er toe dat de conditionele verwachting van de storingsterm relatief ten opzichte van de verklarende variabelen met kans 1 gelijk aan nul moet zijn.

Bij de traditionele econometrische modelbouw worden dus de hypothesen die aan het economische model ten grondslag liggen aangevuld met verdere werkhypothesen omtrent de functionele vorm van het model en de eigenschappen van de storingstermen. Omdat het economische model al onwaar was en het traditionele econometrische model wordt gebouwd op de fundamenteën van het economische model, gemetseld met verdere werkhypothesen, ligt de conclusie nu voor de hand dat het econometrische model nog meer onwaar is dan het economische model. Deze gedachte is gemeen goed in de econometrische professie. Algemeen wordt aangenomen dat econometrische modellen per definitie onwaar zijn en dat het derhalve zinloos is te zoeken naar ware modellen. Daarom spreekt men van "admissible models", d.w.z. toegelaten modellen in plaats van ware modellen, waarbij de toelaatbaarheid van een model afhangt van het doel waarvoor het wordt gebruikt. Zo heeft het Centraal Plan Bureau in de loop van haar bestaan diverse econometrische modellen van de Nederlandse economie geconstrueerd zonder daaraan de conclusie te verbinden dat alle voorgaande modellen en de daarop gebaseerde prognoses en beleidsaanbevelingen fout waren. Ook is het een gangbare opvatting onder econometristen dat analysemodellen, die tot doel hebben economische hypothesen te toetsen, niet noodzakelijkerwijs goede voorspelmodellen zijn, en omgekeerd.

De opvatting dat econometrische modellen in principe onwaar zijn, is in zoverre correct dat de zojuist geschetste traditionele manier van econometrische modelspecificatie bijna zeker tot misspecificatie leidt. Dit betekent echter niet dat een waar econometrisch model niet bestaat. Sterker nog, econometrische modellen zijn in principe te classificeren in twee categorieën, namelijk tautologieën en logisch inconsistente (en dus onware) modellen. Een model is logisch inconsistent als de modelveronderstellingen strijdig zijn, d.w.z. dat twee of meer modelveronderstellingen niet tegelijkertijd waar kunnen zijn. Is een econometrisch model wel logisch consistent, dan is er in feite sprake van een tautologie, dat is een gedefinieerde structuur, zodanig dat alle modelveronderstellingen als stukjes van een legpuzzel in elkaar passen. Om de verwarring nog groter te maken wil ik hierbij alvast opmerken dat er meerdere ware econometrische modellen bestaan die hetzelfde economische verschijnsel verklaren, zelfs onder de veronderstellingen van het klassieke lineaire regressiemodel. Tenslotte zal ik in deze oratie de stelling verdedigen dat het probleem hoe een econometrisch model te specificeren geen economisch probleem is, maar een mathematisch-statistisch probleem. Het ware econometrische model is door de economische data volledig bepaald. De restricties die door de economische theorie worden opgelegd aan de econometrische modelspecificatie vormen eerder belemmering dan een hulp bij het juist specificeren van het econometrische model. Ik pleit dus voor het terugdringen van de economische theorie bij de specificatie van econometrische modellen. De economische theorie is dus niet de kip die het econometrische model als ei legt, maar omgekeerd: het econometrische model is de kip, en de economische theorie is het ei.

De bewering dat ware econometrische modellen bestaan en volledig bepaald zijn door de data is niet gebaseerd op filosofische overwegingen, maar op een stelling uit de waarschijnlijkheidstheorie²⁾. Deze stelling zegt dat gegeven een kansvariabele y met eindig eerste moment³⁾ en een vector x van kansvariabelen, de conditionele verwachting van y relatief ten opzichte van de vector x een Borel-meetbare functie is op de ruimte waarop de vector x is gedefinieerd. Deze Borel-meetbare functie, zeg de functie g , is uniek bepaald met kans 1, d.w.z.

dat elke andere Borel-meetbare functie met deze eigenschap met kans 1 gelijk is aan deze functie g op de door de vector x gedefinieerde deelruimte. Het begrip Borel-meetbare functie zal ik verder onbesproken laten⁴⁾. Alleen wil ik opmerken dat Borel-meetbare functies een zeer grillig verloop kunnen hebben, maar zich toch zodanig fatsoenlijk gedragen dat als je er een kansvariabele of kansvector instopt er een kansvariabele uitkomt. Vatten we de vector x op als vector van verklarende variabelen en y als de te verklaren variabele, dan bepaalt de hiervoor genoemde functie g volledig de functionele vorm van het regressiemodel. De storingsterm van dit regressiemodel is simpelweg het verschil tussen y en de conditionele verwachting van y relatief ten opzichte van de vector x , en voldoet derhalve automatisch aan de eis dat de conditionele verwachting van de storingsterm relatief ten opzichte van de vector x van verklarende variabelen nul is met kans 1. Het aldus gedefinieerde regressiemodel is een tautologie: het is uniek bepaald door de definitie van conditionele verwachting. Elke specificatie van een regressiemodel dat y uit deze vector x verklaart, waarbij de storingsterm conditionele verwachting nul heeft, moet noodzakelijkerwijs gelijk zijn aan dit tautologische regressiemodel. Zo niet, dan is de gespecificeerde functionele vorm strijdig met de betreffende veronderstelling over de storingsterm. Dus een regressiemodel is ofwel een tautologie, ofwel logisch inconsistent.

De uniekheid van het tautologische regressiemodel geldt alleen voor de klasse van regressiemodellen die dezelfde te verklaren variabele y uit de zelfde vector x van verklarende variabelen verklaren. Als we aan de vector x van verklarende variabelen een extra component zouden toevoegen of een component ervan zouden weglaten, dan verandert doorgaans de structuur van het regressiemodel. Het nieuwe model is echter nog steeds een tautologie en dus per definitie waar. Het zelfde geldt als we de te verklaren variabele y transformeren. Als bijvoorbeeld y altijd positief is en de logaritme van y een eindig eerste moment heeft, dan kunnen we ook een regressiemodel definiëren dat de log van y verklaart uit x . Dit model zal eveneens een andere structuur hebben als het oorspronkelijke regressiemodel, maar het is opnieuw een tautologie en dus waar. Bij "cross-section"

modellen, die een economisch verschijnsel op een bepaald vast tijdstip beschrijven, kunnen we bovendien naar believen de rol van y en x verwisselen. We kunnen elke kansvariabele op andere kansvariabelen conditioneren, mits de geconditioneerde kansvariabele een eindig eerste moment heeft. In het geval van multivariaat normaal verdeelde kansvariabelen leidt dit er toe dat de resulterende regressiemodellen lineair zijn met storings-termen die zelf normaal verdeeld zijn met verwachting nul en eindige variantie en die bovendien onafhankelijk zijn van de verklarende variabelen⁵⁾. Bij "cross-section" modellen kan derhalve de richting van de causaliteit niet uit de data worden afgeleid. De keuze welke variabele de te verklaren variabele is en welke de verklarende variabelen zijn is geheel aan de onderzoeker. Bij elke keuze hoort echter een waar econometrisch model. Voor tijdreeksmodellen geldt in wezen hetzelfde, maar door de volgtijdelijke aard van de data ligt de richting van de causaliteit vast. Bij tijdreeksen kunnen we niet langer verklarende en te verklarende variabelen naar believen verwisselen. Het heeft immers geen zin om het verleden uit de toekomst te verklaren. Bij tijdreeksen kunnen we een uniek tautologisch regressiemodel definiëren door een variabele te conditioneren op het gehele verleden van het proces⁶⁾. Deze conditionele verwachting is een optimaal voorspelschema in de zin dat elk ander voorspelschema een voorspelfout met grotere variantie oplevert. Ook dit optimale voorspelschema is uniek, gegeven het betreffende tijdreeksproces. Als we naast het eigen verleden van de te verklaren variabele ook conditioneren op het gehele verleden van andere tijdreeksprocessen, dan verandert de structuur van het optimale voorspelschema⁷⁾. Echter, gegeven de te verklaren variabele en het vectortijdreeksproces waarop wordt geconditioneerd, is het optimale voorspelschema, en dus het regressiemodel, volledig en uniek bepaald.

Het zojuist gehouden betoog heeft in principe alleen betrekking op regressiemodellen, d.w.z. modellen die rechtstreeks te interpreteren zijn als conditionele verwachtingen. De meeste econometrische modellen, inclusief tijdreeksmodellen, zijn echter te herleiden tot regressiemodellen, waarbij de correspondentie uniek is. Dit wil zeggen dat een econometrisch model niet kan worden herleid tot twee verschillende regressiemodellen en

omgekeerd als een regressiemodel kan worden herleid uit twee verschillende econometrische modellen, dan zijn die econometrische modellen niet geïdentificeerd. In dat laatste geval zijn die econometrische modellen te ruim gespecificeerd. Door identificerende restricties op te leggen kunnen deze modellen equivalent gemaakt worden. De identificerende restricties maken dan deel uit van het uiteindelijke econometrische model en voor dit model geldt dan weer dat het in uniek verband kan worden gebracht met een regressiemodel.

De specificatie van econometrische modellen gaat soms verder dan zojuist geschetst. Niet alleen wordt dan de conditionele verwachting gemodelleerd, maar zelfs het gehele data-genererende proces. Met andere woorden, de "likelihood" functie van de dataverzameling wordt dan gespecificeerd als een bekende parametrische familie van kansdichtheden. In dat geval geldt echter hetzelfde als bij het modelleren van conditionele verwachtingen, namelijk dat het echte data-genererende proces al vast ligt in de data zelf en dat het gemodelleerde data-genererende proces ofwel het ware data-genererende proces omvat ofwel foutief gespecificeerd is.

Een en ander betekent dat ware econometrische modellen uniek bepaald zijn door de data die ze beogen te beschrijven. Zodra de modelvariabelen, onderscheiden naar verklarende en te verklaren variabelen, met eventuele transformaties zijn gekozen en de bijbehorende data zijn verzameld ligt het ware econometrische model vast. Met andere woorden, de data heeft het econometrische model al voor u gespecificeerd. De kunst is nu om uit de data de informatie te halen over hoe dit model er uitziet. Deze kunst noemen we specificatie-analyse. De economische theorie speelt bij dit alles een bescheiden rol. Alleen de keuze van de modelvariabelen en hun eventuele transformaties is een economisch probleem. Zodra deze keuze gemaakt is en de bijbehorende data zijn opgespoord is het probleem hoe het econometrisch model te specificeren in wezen een mathematisch-statistisch probleem en niet een economisch probleem. Het model is er al, alleen zit het verborgen in de data. De specificatie-analyse is een conglomeraat van econometrisch-statistische technieken om de informatie die de data bevat over dit verborgen

model te ontcijferen. Dit heeft iets weg van cryptologie, de leer van de geheimschriften en hun ontcijfering. De specificatie-analyse verschaft methoden om de in de data verborgen code over de structuur van het model te kraken.

Het specificatieprobleem kan van twee tegengestelde richtingen worden aangepakt. De eerste richting is die waarbij uitgegaan wordt van een vrij restrictieve parametrische specificatie en vervolgens deze specificatie getoetst wordt op misspecificatie. Indien dit model alle specificatietoetsen doorstaat, wordt het geaccepteerd, zo niet, dan wordt een nieuwe specificatie geprobeerd en getoetst en wordt dit specificatie- en toetsingsproces herhaald totdat een specificatie geaccepteerd wordt. Acceptatie van een specificatie hoeft echter niet te betekenen dat de betreffende specificatie juist is, maar alleen dat de informatie in de data niet toelaat het geaccepteerde model te falsificeren. De dataverzamelingen waarmee econometristen werken bevatten over het algemeen een beperkt aantal waarnemingen, vooral tijdreeksdata. Daardoor is het niet mogelijk het juiste model volledig te determineren op basis van de data-informatie. Hoe kleiner de dataverzameling, hoe groter de klasse van modellen die niet te onderscheiden zijn op basis van de data-informatie. In dit geval is zeker een rol weggelegd voor de economische theorie om uit deze klasse van modellen een keuze te maken. Deze keuze dient echter achteraf en niet vooraf te worden gemaakt. Het devies is: laat eerste de data spreken, en als er daarna nog keuzes moeten worden gemaakt, gebruik dan pas de economische theorie als leidraad.

In het afgelopen decennium heeft zich in de econometrie een duidelijke ontwikkeling afgetekend in de richting van grotere aandacht voor modelspecificatietoetsen. Deze aandacht is vooral gestimuleerd door de schoolstrijd in de economische wetenschap⁸⁾. Het Keynesiaanse economische model is in discrediet geraakt, maar de strijd om de opvolging is nog niet beslecht. Monetarisme, "supply side economics", en de nieuwe klassieke economie strijden om de suprematie, maar geen van de scholen heeft tot nu toe de overwinning behaald. De behoefte om conflicterende nieuwe theorieën empirisch te toetsen heeft met name een hausse veroorzaakt in de ontwikkeling van toetsen van

niet-geneste hypothesen⁹⁾. Daarnaast is het instrumentarium van de toetsen van geneste hypothesen, met name de Wald-, Likelihood Ratio- en Lagrange Multiplier-toetsen, verder verfijnd en theoretisch uitgediept¹⁰⁾, en zijn er nieuwe ontwikkelingen op gang gekomen op het gebied van de zuivere modelspecificatietoetsen, waarbij niet langer tegen een expliciete alternatieve hypothese wordt getoetst. De belangrijkste exponenten van de laatste categorie toetsen zijn de toetsen van Hausman (1978), White (1981), Newey (1985) en (in alle bescheidenheid) mijn eigen consistente toetsen voor modelspecificatie¹¹⁾. Deze toetsen testen de nulhypothese dat het model correct gespecificeerd is tegen de algemene alternatieve hypothese dat het model foutief gespecificeerd is, zonder de aard van de misspecificatie a priori vast te leggen. Een modelspecificatietoets is consistent, indien uiteindelijk elke misspecificatie wordt ontdekt als de steekproefomvang of de lengte van de tijdreeks aangroeit tot oneindig. De toetsen van Hausman, White en Newey zijn niet consistent¹²⁾, maar zijn in de praktijk zeker goed bruikbaar. Ikzelf heb diverse toetsen ontwikkeld en toegepast die wel consistent zijn, en naar mijn beste weten als eerste, maar de ontwikkeling ervan is te recent om nu al uitspraken te doen over hun onderscheidend vermogen in kleine steekproeven. Het onderzoek daarvan staat op mijn programma.

Een diametraal tegenovergestelde aanpak van het specificatieprobleem is die waarbij wordt uitgegaan van een alles omvattende modelspecificatie die met zekerheid het ware model omvat, of waarbij zelfs geheel wordt afgezien van een expliciete modelspecificatie. In het laatste geval hebben we te maken met niet-parametrische regressie-analyse¹³⁾; in het eerste geval met oneindig- of bijna oneindig- parametrische regressieanalyse. Hoewel op het eerste gezicht de termen "niet-parametrisch" en "oneindig-parametrisch" op een totale tegenstelling duidt, zal blijken dat een en ander nauw aan elkaar verwant is.

In de tijdreeksanalyse is "oneindige parametrisatie" geen nieuw begrip. De Wold-decompositie stelling¹⁴⁾ zegt dat stationaire tijdreeksen additief kunnen worden opgesplitst in een deterministisch gedeelte dat alleen van de tijdsindex afhangt en een oneindige gewogen som van ongecorreleerde kansvariabelen.

Deze kansvariabelen worden innovaties genoemd. Indien het deterministische gedeelte kan worden weggelaten of tijdsafhankelijk kan worden verondersteld en het tijdreeksproces normaal verdeeld is, dan geeft de Wold-decompositie aanleiding tot een lineair autoregressief model van oneindige orde. Dit betekent dat de conditionele verwachting te schrijven is als een lineaire combinatie van alle vertraagde te verklaren variabelen. Dit ware model neemt dus de vorm aan van een lineair model met oneindig veel parameters. De bekende tijdreeksmodellen, zoals de $AR(p)$, $MA(q)$ en $ARMA(p,q)$ modellen, zijn dan op te vatten als benaderingen van dit ware model. De Wold-decompositie stelling is ook van toepassing op stationaire vector-tijdreeksen. Onder de normaliteitshypothese neemt het ware model dan de vorm aan van een oneindige-orde vector-autoregressief model. Sims (1980) benadert dit model door een vector-autoregressief model van lage orde zonder restricties op te leggen aan de resterende parameter matrices, teneinde een alternatief te bieden voor de traditionele manier van macro-economische modelbouw.

Een verdere ontwikkeling op dit gebied vormen de door mij voorgestelde "ARMA memory index" modellen. Binnenkort zal over dit onderwerp een artikel met discussie verschijnen in *Econometric Theory*¹⁵⁾. Daarin heb ik langs wiskundige weg afgeleid dat het algemene tijdreeksmodel, waarbij niet langer de normaliteitsveronderstelling wordt gemaakt, kan worden geschreven als een niet-lineaire functie van een ARMA-model. Deze niet-lineaire functie kan je in principe benaderen door middel van niet-parametrische regressie¹⁶⁾ (waarover zometeen meer) en op basis daarvan parametrisch specificeren.

Ook voor "cross-section" modellen kan de algemene functionele vorm van het regressiemodel worden bepaald. Indien de verklarende variabelen kwalitatief van aard zijn en dus een eindig aantal waarden kunnen aannemen is het algemene regressiemodel te schrijven als een eindige-orde polynoom van een lineaire functie¹⁷⁾. Indien de verklarende variabelen niet kwalitatief van aard zijn maar een begreind variatiegebied hebben, biedt Gallant's (1981) "Fourier flexible functional form" een oplossing van het specificatieprobleem. Gallant laat zien dat je dan de

regressiefunctie kan schrijven als een oneindige gewogen som van sinus en cosinus functies van lineaire combinaties van de verklarende variabelen. Door nu deze oneindige parametrisatie af te knotten tot een eindige parametrisatie kan je het model langs de gebruikelijke weg schatten zonder van te voren de vorm van het model vast te leggen.

Bij niet-parametrische regressie-analyse gebeurt in feite iets soortgelijks. We kunnen een zodanig ruim niet-parametrisch model specificeren dat het ware model daarbinnen valt. In plaats van een oneindig aantal parameters wordt daarbij echter gebruik gemaakt van convoluties van kansdichtheden. De data wordt daarbij als het ware gefilterd met behulp van een van te voren gespecificeerde klasse van functies, kernel functies genaamd. De mate van filtering kan worden gecontroleerd door middel van de zogenaamde "window width" (de raambreedte), die afhangt van de steekproefomvang. Deze "window width" parameter speelt een analoge rol als het afknotten van oneindige parametrisaties. Een grote "window width" is analoog aan een sterke afknotting en een kleine "window width" is analoog aan een bescheiden afknotting (echter met verlies van vrijheidsgraden).

Zowel bij oneindig-parametrische regressie-analyse als bij niet-parametrische regressie-analyse wordt de functionele vorm van het model rechtstreeks uit de data geschat zonder die functionele vorm van te voren te beperken. Op basis van deze schattingsresultaten kunnen we dan een handzamere functionele specificatie bepalen en het model opnieuw schatten, of onze conclusies direct baseren op de eerdere schattingsresultaten. De eerlijkheid gebiedt echter te zeggen dat niet-parametrische en oneindig-parametrische regressie-analyse alleen redelijk accuraat het ware model benadert als de steekproefomvang of de lengte van de tijdreeks groot is. Op basis van zeg dertig jaarwaarnemingen kunnen we deze aanpak wel vergeten.

De stelling die ik in deze oratie heb verdedigd is dat econometrische modelbouw niet een verlengstuk dient te zijn van economische modelbouw, maar dat de specificatie van econometrische modellen zo veel mogelijk data afhankelijk moet worden

gemaakt. Ik heb ook aangegeven dat daarvoor de statistisch-econometrische middelen reeds beschikbaar zijn of in een gevorderde staat van ontwikkeling verkeren. Vanuit technisch oogpunt is het niet langer nodig het econometrische model geheel los van de data te specificeren. Terugkoppeling van de specificatie naar de data door middel van specificatietoetsen of zelfs geheel data afhankelijke econometrische analyse behoren thans tot de mogelijkheden. Econometrische modellen dienen zo goed mogelijk de economische data te beschrijven. Als blijkt dat ze niet stroken met de economische theorie dan mag niet zonder meer worden geconcludeerd dat het econometrische model fout is, maar moet ook de mogelijkheid onder ogen worden gezien dat de economische theorie niet realistisch is. Sterker nog, econometrie bedrijven langs deze lijn kan de economische theorievorming juist stimuleren. Ter afsluiting wil ik u daarvan een voorbeeld geven uit mijn eigen onderzoek.

Dit onderzoek, waarover ik onlangs gepubliceerd heb in de *Journal of Econometrics*¹⁸⁾, beoogde in eerste instantie de "Rational Expectation-Natural Rate" hypothese voor Nederland te toetsen. Ik zal u niet vermoeien met een preciese uitleg van at deze hypothese inhoudt¹⁹⁾, maar alleen vermelden wat de consequentie van deze hypothese is, namelijk dat de werkloosheid niet wordt beïnvloed door andere economische variabelen. Dit impliceert dat werkloosheid een soort natuurverschijnsel is waar we niets aan kunnen veranderen. Om deze hypothese te toetsen heb ik op basis van maandcijfers over de Nederlandse economie een model opgesteld dat de werkloosheid relateert aan vertraagde werkloosheid, rente, prijsinflatie, looninflatie, de groei van de industriële productie en de groei van de export. Dit model, dat een zogenaamde ARMAX structuur heeft, is in twee ronden geschat en getoetst, waarbij in de tweede ronde de uiteindelijke versie van het model door de specificatietoets werd geaccepteerd. De betreffende toets is een van de door mij ontwikkelde consistente modelspecificatietoetsen

De "Rational Expectation-Natural Rate" hypothese kan je nu falsificeren door te toetsen of alle variabelen behalve de vertraagde werkloosheid uit het model kunnen worden verwij-

derd. Wat bleek nu: behalve de vertraagde werkloosheid was ook de rente van significante invloed op de werkloosheid. Daarmee was de "Rational Expectation-Natural Rate" hypothese voor Nederland verworpen, hetgeen op zich een mooi resultaat is. Echter, het gevonden verband tussen werkloosheid en rente strookt niet met de gevestigde economische theorie. Moet je nu achteraf concluderen dat het model onzinnig is, of mankeert er iets aan de economische theorie? Gezien mijn betoog zal het u niet verbazen dat ik in eerste instantie van het laatste ben uitgegaan. Daarbij ben ik tot de conclusie gekomen dat het gevonden verband wel degelijk plausibel is, mits je de neo-klassieke theorie van het ondernemersgedrag over boord zet. Deze neo-klassieke theorie gaat er van uit dat ondernemers streven naar maximale winst. Deze gedragshypothese heeft als consequentie dat de vaste kosten niet van invloed zijn op de werkgelegenheid in het bedrijf, maar alleen de variabele kosten, met name de loonkosten. Dit wordt geïmpliceerd door de welbekende gelijkheid van marginale kosten en marginale opbrengsten als conditie voor winstmaximalisatie. Looninflatie blijkt echter in mijn model helemaal niet van invloed te zijn op de werkloosheid. Rente zou de werkloosheid kunnen beïnvloeden via de investeringen. Echter, in dat geval zou je een negatief effect verwachten van de groei van de industriële produktie op de werkloosheid, omdat investeringsgoederen in belangrijke mate door de industriële sector worden voortgebracht. Ook dit verband tussen de groei van de industriële produktie en de werkloosheid bleek in mijn model afwezig. Dan blijft over een mogelijke invloed van de rente via de vaste kosten van bedrijven, maar dit wordt uitgesloten door de neo-klassieke theorie.

De oplossing van dit raadsel kan worden gevonden in de theorie van Baumol (1959) over het ondernemersgedrag, aangevuld met de hypothese van een flexibele inspanningsgraad van de arbeid. Baumol stelt dat de winst-maximalisatie-hypothese alleen plausibel is voor ondernemingen die door de eigenaar zelf worden geleid. De moderne middelgrote en grote ondernemingen echter worden doorgaans geleid door beroepsmanagers die in loondienst zijn. Deze beroepsmanagers streven andere doeleinden na dan de kapitaalbezitters. De RSV-enquete heeft daarvan enkele

fraaie voorbeelden laten zien. Die doeleinden, zoals status, hoog inkomen, ruime onkostenvergoedingen, etc., kunnen meer en beter worden gerealiseerd naarmate de onderneming groter is. Aannemende dat de grootte van een onderneming wordt afgemeten aan de omzet, stelt Baumol dat deze beroepsmanagers streven naar maximale omzet in plaats van maximale winst. Deze beroepsmanagers kunnen hun gang gaan met het nastreven van hun eigen doeleinden zolang de aandeelhouders tevreden zijn, d.w.z. zolang er voldoende dividend wordt uitgekeerd. Daartoe is een minimale hoeveelheid winst nodig, de zogenaamde winstrestrictie.

In deze alternatieve theorie van het ondernemersgedrag spelen de vaste kosten wel een belangrijke rol, namelijk via deze winstrestrictie. Als de vaste kosten zodanig toenemen dat de winstrestrictie bindend wordt, dan zal de bedrijfsleiding overgaan tot reorganisatie: een deel van het werknemersbestand wordt ontslagen. Dit hoeft niet direct consequenties te hebben voor de productiecapaciteit, omdat voor de productiecapaciteit niet alleen het aantal arbeidskrachten zelf telt, maar ook de inspanning van die arbeidskrachten. Door de inefficiënte delen van het bedrijf te saneren kunnen de kosten worden verlaagd bij gelijkblijvende productiecapaciteit. Een en ander impliceert dat een rentestijging via de toename van de vaste kosten en daarmee gepaard gaande winstdaling een toename van de werkloosheid tot gevolg kan hebben. Dit effect wordt nog versterkt door het feit dat een rentestijging ook een negatief effect heeft op de omzet, deels doordat de vaste lasten van de consumenten toenemen en dus hun bestedingen afnemen, deels doordat de financiering van duurzame consumptiegoederen duurder wordt.

In principe zou langs deze weg een loonstijging via winstdaling eveneens tot werkloosheid kunnen leiden. Echter, de loonkostenstijging kan worden gecompenseerd doordat een loonstijging (met name een collectieve loonstijging) de inkomens en dus de bestedingen van de consumenten doet toenemen en daarmee de omzet van bedrijven. Dit verklaart waarom ik wel een negatief effect van de rente op de werkloosheid vond, maar geen effect van looninflatie. Tenslotte verklaart de hypothese van een flexibele inspanningsgraad van arbeid waarom een verandering

in de produktie-omvang geen waarneembaar effect hoeft te hebben op de werkloosheid.

Nu kunt u tegen dit betoog inbrengen, dat diverse econometrische onderzoeken wel een effect van loonstijging en verandering in de produktie-omvang op de werkloosheid hebben geconstateerd. Beschouwen we deze econometrische onderzoeken nader, dan zal echter blijken dat zelden of nooit een adequate toets op modelmisspecificatie is toegepast. De betreffende econometrische modellen zijn mogelijk misgespecificeerd en leiden dan tot foute conclusies.

Dit betoog illustreert tevens de plaats en het belang van de economische wetenschap in de econometrie en omgekeerd het belang van econometrie voor de economische wetenschap. De economische theorie is het referentiekader van de *resultaten* van de econometrische analyse, voor zover deze resultaten niet alleen voorspellingen betreffen. Zonder confrontatie van deze resultaten met de relevante economische theorie is de econometrische analyse zinloos. Alleen als het doel van de econometrische analyse louter het voorspellen van het toekomstige verloop van een economisch proces is, kan terugkoppeling naar de economische theorie achterwege blijven. De econometrische analyse zelf moet echter in beide gevallen zoveel mogelijk vrij gehouden worden van censurering door de economische theorie.

Zeër geachte aanwezigen. In deze oratie heb ik mijn visie geschetst op het hoe en waarom van econometrische modelbouw. Ik ben mij ervan bewust dat die visie niet door velen wordt gedeeld. Dit betekent echter geenszins dat van mij een geïsoleerde opstelling in de Faculteit der Economische Wetenschappen en Econometrie moet worden verwacht. In tegendeel, u kunt van mij, in collegiale samenwerking, een volledige inzet verwachten voor de belangen van het onderwijs en onderzoek in de econometrie en de economie aan de Vrije Universiteit.

Ik dank u voor uw aandacht.

Noten

- 1) Zie Klant (1972) voor een diepgaande beschouwing over het eigene van economische theorieën.
- 2) Chung (1974, Theorems 9.1.1, 9.1.2)
- 3) D.w.z.: $E|y| < \infty$.
- 4) Zie bijvoorbeeld Chung (1974) voor een definitie.
- 5) Dit kan worden afgeleid uit formule (2), p.28, in Anderson (1958) en het feit dat ongecorrleerde multivariaat normaal verdeelde kansvariabelen onafhankelijk zijn.
- 6) Chung (1974, Chapt. 9)
- 7) Bierens (1984, p.326)
- 8) Bell & Kristol (1981)
- 9) MacKinnon (1983)
- 10) Engle (1984)
- 11) Bierens (1982,1984,1987b,c,d,e,f) en Bierens & Hartog (1984)
- 12) Bierens (1982) en Holly (1982)
- 13) Zie Bierens (1987a) voor een literatuuroverzicht.
- 14) Harvey (1981, p.69)
- 15) Bierens (1987b,c)
- 16) Bierens (1986)
- 17) Bierens & Hartog (1984)
- 18) Bierens (1987d)
- 19) Zie Shiller (1978) voor een kritisch overzicht van de literatuur.

Literatuur

- Anderson, T.W. (1958), *Introduction to Multivariate Statistical Analysis*, New York: John Wiley
- Baumol, W.J. (1959), *Business Behaviour, Value and Growth*, New York: MacMillan
- Bell, D. & I.Kristol (eds.) (1981), *Crisis in Economic Theory*, New York: Basic Books
- Bierens, H.J. (1982), "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134
- Bierens, H.J. (1984), "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353
- Bierens, H.J. (1986), "Model-Free Asymptotically Best Forecasting of Stationary Economic Time Series", RM1986-32, Faculty of Economics, Free University
- Bierens, H.J. (1987a), "Kernel Estimators of Regression Functions", in: Truman F.Bewley (ed.), *Advances in Econometrics 1985*, New York: Cambridge University Press, 99-144
- Bierens, H.J. (1987b), "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory* (forthcoming)
- Bierens, H.J. (1987c), "Reply to Comments on ARMA Memory Index Modeling of Economic Time Series", *Econometric Theory* (forthcoming)
- Bierens, H.J. (1987d), "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics* 35, 161-190
- Bierens, H.J. (1987e), "A Consistent Hausman-Type Model Specification Test", RM1987-2, Faculty of Economics & Econometrics, Free University

Bierens, H.J. (1987f), "Tests for Model Misspecification", RM1987-22, Faculty of Economics & Econometrics, Free University

Bierens, H.J. & J. Hartog (1984), "On Specifying, Estimating and Testing Regression Models with Discrete Explanatory Variables, with an Empirical Application to the Earnings Function", RM8405, Faculty of Economics, University of Amsterdam

Chung, K.L. (1974), *A Course in Probability Theory*, New York: Academic Press

Engle, R.F. (1984), "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics", in: Z.Griliches and M.Intrilligator (eds.), *Handbook of Econometrics*, Amsterdam: North Holland

Gallant, A.R. (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Functional Form", *Journal of Econometrics* 15, 211-245

Harvey, A.C. (1981), *Time Series Models*, Oxford: Philip Alan

Hausman, J.A. (1978), "Specification Tests in Econometrics", *Econometrica* 46, 1251-1271

Holly, A. (1982), "A Remark on Hausman's Specification Test", *Econometrica* 50, 749-760

MacKinnon, J.G. (1983), "Model Specification Tests Against Non-Nested Alternatives", *Econometric Reviews* 2, 85-158

Klant, J.J. (1972), *Spelregels voor Economen*, Leiden: Stenfert Kroese

Newey, W.K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica* 53, 1047-1070

Shiller, R.J. (1978), "Rational Expectations and the Dynamic Structure of Macroeconometric Models: A Critical Review", *Journal of Monetary Economics* 4, 1-44

Sims, C.A. (1980), "Macroeconomics and Reality", *Econometrica* 48, 1-48

White, H. (1981), "Consequences and Detection of Misspecified Nonlinear Regression Models", *Journal of the American Statistical Association* 76, 419-433

Ontvangen: 9-10-1987
Geaccepteerd: 2-11-1987