FORMEEL DENKEN; DE BOUW VAN EEN EKONOMETRISCH MODEL

door Dr. M. A. Kooyman

Het oogmerk waarmee de econoom de werkelijkheid bezielt, heeft grote invloed op de uitbeelding van deze werkelijkheid in de vorm van een economisch model. Bij de start van de confrontatie van de onderzoeker met de werkelijkheid, moet eerstgenoemde over een heldere probleemstelling beschikken. Deze dient als leidraad bij het proces waarbij de voor hem relevante hoofdlijnen in de werkelijkheid worden neergeschreven in de formele vorm van een economisch model. De probleemstelling bepaalt het abstraktieniveau van het model. Hoe hoger het abstraktieniveau mag liggen, des te eenvoudiger en inzichtelijker wordt het model, doch tevens des te groter wordt de afstand tot de werkelijkheid. Een onscherp geformuleerde opdracht tot het maken van een model voor de naoorlogse economie van Nederland kan leiden tot een vrijwel onbegrensd groot model, maar ook tot een model van niet meer dan vier vergelijkingen. Het voorschrift dat het model de vier doelstellingen van de naoorlogse economische politiek (volledige werkgelegenheid, evenwichtige betalingsbalans, stabiel prijsniveau en een redelijke inkomensverdeling) moet expliciteren zou de omvang en het abstraktieniveau van het model beter vastleggen.

Een doordachte probleemstelling legt niet alleen de hoofdlijnen van het model vast, maar ook het soort model en de karakterisering van de variabelen. Naargelang de belangstelling uitgaat naar het verkrijgen van voorspellingen dan wel naar het nemen van beslissingen of naar het beschrijven c.q. „verklaren” van een economisch proces, spreekt men resp. van voorspelmodellen, decisiemodellen en beschrijvende of verklarende modellen.

De variabelen die in een model optreden, worden naar hun aard ingedeeld in twee categorieën, te weten die der endogene en die der gepredetermineerde variabelen. De gepredetermineerde variabelen worden daarbij gezien als grootheden die buiten de gezichtskring van het model tot stand komen; voor een economisch model behoren daar in elk geval toe de relevante niet-economische variabelen. De aard van de economische grootheden heeft geen invloed op de karakterisering in endogeen of gepredetermineerd, het gestelde probleem vormt hierbij de beslissende factor. Het nationale inkomen zal bijvoorbeeld in een macro-economisch model als endogeen fungeren, maar in een marktmodel voor een goed als gepredetermineerde. Vertraagde endogene grootheden worden tot de gepredetermineerden gerekend, omdat in de lopende periode geen invloed meer kan worden uitgeoefend op de waarde van grootheden in het verleden. Het model geeft de voor het gestelde doel relevante betrekkingen tussen de variabelen weer, zodanig dat de fluktuaties in de endogene variabelen worden „verklaard”, d.w.z. de fluktuaties worden opgebouwd uit de „bijdragen” van de verschillende gepredetermineerde variabelen. Een volledige „verklaring” is slechts mogelijk indien het model evenveel vergelijkingen telt als endogene variabelen, men kan in dat geval de endogene variabelen op eenduidige wijze uitdrukken in alle gepredetermineerden; laatstbedoelde vorm van het model wordt aangeduid met de term „herleide vorm”.

In een model kunnen veelal een aantal typen van vergelijkingen worden onderscheiden, zoals gedragsvergelijkingen, technische en institutionele relaties en definitievergelijkingen. De gedragsvergelijkingen vormen het hart van het model; elke gedragsvergelijking afzonderlijk beschrijft een economisch proces. Bouwt men bijvoorbeeld een model voor een volkshuishouding dan blijft de beschrijving in eerste instantie beperkt tot het opstellen van een lijst met variabelen, die voor de beschrijving van het desbetreffende proces relevant worden geacht. Voor de beschrijving van de konsumptieve uitgaven zal men bijvoorbeeld kiezen: het looninkomen, het winstinkomen, het prijsniveau, de consumptie uit de vorige periode, de likwiditeitspositie en een trend.

Onder technische relaties in de economie verstaat men verbanden die voortvloeien uit de gebruikte produktietechniek, de produktiefunctie van Cobb-Douglas is hiervan een voorbeeld. Institutionele relaties vormen de neerslag van wettelijke kaders, zoals bijvoorbeeld een belastingstelsel, waarbinnen de economische subjekten zich dienen te bewegen. Tenslotte komen in een model veelal definitievergelijkingen voor, die worden gebruikt om variabelen te definiëren of om evenwichtsvoorwaarden weer te geven; de relatie $Y = C + I$ behoort tot deze categorie.

Het model is tot dusver omschreven als een aantal impliciete funkties. Het aantal dezer funkties wordt gekozen met als leidraad de probleemstelling. De indeling der variabelen naar endogeen of gepredetermineerd wordt eveneens bepaald door de desbetreffende probleemstelling, terwijl de eis van oplosbaarheid van het model met zich brengt dat het aantal relaties gelijk dient te zijn aan het aantal endogene variabelen. Ter illustratie van een model in dit stadium van zijn ontwikkeling wordt hieronder Model I van Klein voor de Verenigde Staten in de periode 1921-1941 gegeven.

Gedragsvergelijkingen

konsumptiefunctie	$C = f_C (P, P_{-1}, W)$	(1)
investeringsfunctie	$I = f_I (P, P_{-1}, K_{-1})$	
loonsomvergelijking	$W^* = f_W (E, E_{-1}, A)$	

Definitievergelijkingen

produkt	$Y + T = C + I + G$
inkomen	$Y = P + W$
kapitaal	$K = K_{-1} + I$
loonsom	$W = W^* + W^{**}$
produkt minus overheid	$E = Y + T - W^{**}$

waarin C = consumptie, I = investeringen, W^* = loonsom bedrijven, P = winsten, Y = nationaal inkomen, K = kapitaalgoederenvoorraad aan het einde van het jaar, W = totale loonsom, E = produkt bedrijven, W^{**} = loonsom overheid, T = indirecte belastingen, G = overheidsuitgaven en A = trend. De endogene variabelen van het model zijn C, I, W^*, P, Y, K, W en E . Er komen geen prijsvariabelen in het model voor omdat alle grootheden zijn gedefleerd.

Het bovenstaande model toont, dat Klein bijvoorbeeld de variabele K niet opneemt in zijn konsumptiefunctie, hetgeen betekent dat hij de invloed van

deze variabele op de konsumptieve bestedingen verwaarloosbaar acht. A priori informatie van deze vorm wordt aangeduid met de term „nulrestriktie”. Anderssoortige restrikties aangaande de relaties komen voor, bijvoorbeeld homogeniteit van de orde nul of een indicatie van de richting der beïnvloeding der verklarende variabelen op de te verklaren variabele. Uit het bovenstaande model kan men aflezen dat a priori wordt vastgesteld dat in de konsumptiefunctie de loonsom van de overheid eenzelfde invloed heeft als de loonsom van bedrijven, er wordt immers de som van beide genomen in plaats van elk afzonderlijk. De overvloed aan economische theorieën vormt een rijk jachtgebied bij het zoeken naar a priori informatie.

Het gestelde probleem zal veelal zodanig zijn geformuleerd, dat van de onderzoeker numerieke antwoorden worden verwacht, zoals bijvoorbeeld een voorspelling van de loonkostenstijging in het komende jaar. Modellen in het stadium (1) dienen te worden gekwantificeerd. In tegenstelling tot de tot nog toe uiteengezette werkwijze zal bij een gebleken noodzaak tot kwantificering de ekonometrische benaderingswijze de boventoon gaan voeren. Daarbij is weinig plaats voor een formele benadering, maar overheerst een pragmatische aanpak die noodzakelijk is om een brug te kunnen slaan tussen de formele eisen van de economische theorie enerzijds en het feitelijke gedrag van de economische subjecten anderzijds.

De stap van het impliciete model naar het gekwantificeerde model kan slechts worden gezet met behulp van de wiskunde en de statistiek. De statistiek begint haar taak bij het verzamelen van waarnemingen van de in het model gebruikte variabelen. Men kan hierbij stuiten op in principe onmeetbare variabelen zoals bijvoorbeeld geografisch gebied, sociaal milieu, aanwezigheid van crises e.d. Daarnaast komen soms moeilijk kwantificeerbare grootheden voor zoals gewenste kapitaalgoederenvoorraad, toekomstverwachtingen e.d.; slechts benaderingsgewijs kan men hiervoor „waarnemingen” vinden. De in principe onmeetbare variabelen geeft men veelal weer met behulp van onechte (dummy) variabelen. Bijvoorbeeld de aanwezigheid van een crisis wordt aangegeven met het getal 1 en de afwezigheid met het getal 0, aldus verkrijgt men een onechte variabele die als „waarnemingen” de getallen 1 en 0 vertoont.

Beschikt men over waarnemingen in cijfermatige vorm voor alle betrokken grootheden, dan dient men de wiskundige vorm van de impliciete functies te kiezen. Er is weinig bekend over de mathematische vorm van de relaties die het gedrag van economische subjecten moeten benaderen. Twee redenen voert men veelal aan voor de keuze van lineariteit der vergelijkingen, namelijk men ontleent aan de wiskunde de stelling dat vele typen functies kunnen worden benaderd door lineaire functies indien het variatiegebied van de desbetreffende variabelen niet groot is en men hanteert het principe om te streven naar eenvoud. De gedragsvergelijkingen van model (1) luiden na linearisatie:

konsumptiefunctie	$C = a_0 + a_1 P + a_2 P_{-1} + a_3 W + e_1$	
investeringsfunctie	$I = b_0 + b_1 P + b_2 P_{-1} + b_3 K_{-1} + e_2$	(2)
loonsomvergelijking	$W^* = c_0 + c_1 E + c_2 E_{-1} + c_3 A + e_3$	

In (2) komen een twaalfstal onbekende constanten voor, die worden aangeduid met a, b of c. Zij karakteriseren het desbetreffende economische proces en worden aangeduid met „structuurparameters” of kortweg „parameters”.

De e's stellen de zogenaamde storingstermen voor. Hierin worden de „fouten” samengevoegd die men maakt bij:

- 1e het verwaarlozen van a priori niet relevant geachte grootheden (de nul-restricties),
- 2e het benaderen van moeilijk kwantificeerbare grootheden,
- 3e het meten van de overige grootheden en
- 4e het lineariseren van de relaties.

De storingsterm fungeert dus als vergaarbak van al die factoren, die bij het beschrijven van een proces in de werkelijkheid worden verwaarloosd of onjuist worden weergegeven. Veronderstelt men dat de vele invloeden die in e zijn samengevoegd elk een geringe betekenis bezitten en van elkaar onafhankelijk zijn, dan kan men een beroep doen op de centrale limietstelling uit de statistiek om een normale verdeling te mogen veronderstellen. De specificatie van de verdeling van de storingsterm is noodzakelijk voor het kunnen uitvoeren van hypothesetoetsingen en voor het aanleggen van betrouwbaarheidsintervallen ten aanzien van de parameters.

De lineariteitsrestriktie knelt minder dan men op het eerste gezicht denkt. Zo is de relatie

$$Y = d_0 + d_1 (1/X_1) + d_2 \log X_2 + d_3 X_3 + e$$

lineair in de parameters d_0 t/m d_3 , doch geenszins in de variabelen X_1 en X_2 , aan de lineariteitsrestriktie wordt voldaan omdat deze alleen betrekking heeft op de parameters.

Alvorens men met behulp van de waarnemingen der variabelen kan overgaan tot het schatten van de parameters, dient men te onderzoeken of de wiskundige vorm van het model toelaat dat er überhaupt schattingen worden gemaakt. Dit vraagstuk staat bekend als het identifikatieprobleem. Aan de hand van een eenvoudig marktmodel van een landbouwprodukt zal de problematiek worden geschetst.

$$\begin{array}{ll} \text{vraagvergelijking} & Q = a_{10} + a_{11} P \\ \text{aanbodvergelijking} & Q = a_{20} + a_{21} P + a_{22} Z \end{array} \quad (3)$$

waarin Q = gevraagde hoeveelheid, a = aangeboden hoeveelheid, P = evenwichtsprijs en Z = regenval. Storingstermen zijn niet opgenomen omdat het een wiskundig probleem betreft en niet een statistisch. De waarnemingen van de variabelen Q , P en Z moeten voldoen aan stelsel (3), maar eveneens aan de onderstaande herleide-vormvergelijkingen van het model.

$$\begin{array}{l} Q = h_{10} + h_{11} Z \\ P = h_{20} + h_{22} Z \end{array} \quad (4)$$

waarin

$$\begin{array}{l} h_{10} = (a_{11} a_{20} - a_{21} a_{10}) / (a_{11} - a_{21}), \quad h_{20} = (a_{20} - a_{10}) / (a_{11} - a_{21}), \\ h_{11} = a_{11} a_{22} / (a_{11} - a_{22}) \quad \text{en} \quad h_{22} = a_{22} / (a_{11} - a_{22}). \end{array}$$

Er bestaan derhalve twee mogelijkheden om tot bepaling van de parameters uit model (3) te geraken, te weten rechtstreeks uit model (3) of via de h 's uit model (4). Indien beide wegen tot identieke resultaten leiden, noemt men de parameters eksakt geïdentificeerd. Parameters die niet uit de herleide-vorm-koefficienten kunnen worden afgeleid noemt men niet-geïdentificeerd, deze parameters kunnen niet worden geschat. Zeer frekwent komt de situatie voor dat men voor een en dezelfde parameter verschillende uitkomsten kan verkrijgen met behulp van de herleide vorm, men spreekt dan van over-geïdentificeerde parameters. Verschillende schattingsmethoden staan voor deze situatie ter beschikking, zij leiden echter doorgaans tot verschillende schattingen voor dezelfde parameter. De noodzaak om een schattingsmethode te kiezen introduceert onvermijdelijk een element van willekeur in de schattingsresultaten.

Zoals men voor de stelsels (3) en (4) eenvoudig kan nagaan, zijn de parameters a_{20} , a_{21} en a_{22} niet geïdentificeerd en a_{10} en a_{11} eksakt geïdentificeerd, immers $a_{11} = h_{11}/h_{22}$ en $a_{10} = h_{10} - h_{11}h_{20}/h_{22}$. Introduceert men in de vraagvergelijking het inkomen als tweede verklarende variabele, dan zijn alle parameters eksakt geïdentificeerd.

Er geldt de volgende algemene regel bij het onderzoek op identificeerbaarheid. De parameters uit de i -de vergelijking ($i = 1, 2, \dots, M$) van een model met M endogene en N gepredetermineerde variabelen zijn geïdentificeerd, indien het aantal wel in het model voorkomende maar niet in de vergelijking opgenomen gepredetermineerden (N_i^{**}) in de volgende relatie staat tot het aantal wel in de vergelijking opgenomen endogene variabelen (M_i^*) van het model.

$$N_i^{**} \geq M_i^* - 1$$

Thans kan het stadium van het schatten der parameters worden betreden. De beschikbare schattingsprocedures kan men indelen naar de mate waarin de modelspecificatie wordt gebruikt. De methode der „gewone kleinste kwadraten” gebruikt alleen de nulrestricties in de desbetreffende vergelijking. „Kleinste kwadraten in twee ronden” en de methode der „maximale aannemelijkheid met beperkte informatie”, gebruiken daarenboven de niet in de beschouwde vergelijking doch wel in het model voorkomende gepredetermineerden. „Kleinste kwadraten in drie ronden” en de methode der „maximale aannemelijkheid met volledige informatie” benutten alle informatie die in de modelspecificatie ligt geborgen. De kwaliteit der schatters neemt toe naarmate men meer gebruik maakt van a priori informatie. Men dient daarbij in gedachten te houden dat onjuiste a priori informatie de kwaliteit der schatters doet dalen evenredig met het gebruik dat men van deze a priori informatie maakt. Aangezien de kwaliteit van de a priori informatie ontleend aan economisch-theoretische beschouwingen niet immer hoog geschat kan worden, heeft men aan de bovenstaande beschouwing weinig houvast bij het kiezen van een schattingsmethode.

De berekeningen, behorend bij de genoemde schattingsmethoden, nemen in dezelfde volgorde toe wat betreft de complexiteit. De afweging van kwaliteit tegen ingewikkeldheid doet niet zelden de keuze vallen op de methode

der kleinste kwadraten in twee ronden. Daar verschillende akseptabele schattingsmethoden in het algemeen tot verschillende schattingsuitkomsten leiden, vormt de steekproefvariatie niet de enige bron voor de onzekerheid omtrent de parameterwaarde; een oplossing hiervoor is nog niet gevonden.

In de schattingsfase komen veelal de a priori gekozen specificaties van de gedragsvergelijkingen uit het model weer ter discussie. Vindt men bijvoorbeeld dat de parameter, die de invloed van de prijs op de consumptie geeft, niet significant van nul verschilt, dan moet men een keuze maken uit de volgende twee mogelijkheden. Men laat de steekproefinformatie zwaarder wegen dan de a priori informatie uit de economische theorie en past de specificatie aan door de prijsvariabele weg te laten, waarna opnieuw wordt geschat, of men handhaaft de specificatie. Restrikties die volgen uit algemeen aanvaarde economische theorieën zullen sterker staan bij een botsing met de realiteit dan minder goed gefundeerde restrikties. Een nauwkeurig omschreven strategie bestaat niet voor dit vraagstuk.

Het opstellen van de uiteindelijke specificatie van het model met behulp van de steekproefgegevens leidt tot het gebruik van informatie, die ligt opgeborgen in de steekproef. Methodologisch gezien mag men slechts een gedeelte van de waarnemingen bij het specificatieonderzoek gebruiken, de rest van de steekproef dient men te reserveren voor het schatten en toetsen van de parameters uit het model in zijn definitieve vorm.

Aandacht besteed aan de specificatie van het model loont zonder enige twijfel, omdat specificatiefouten de kwaliteit van de geschatte parameters uit de desbetreffende vergelijking aantasten. Voor de methoden die alle modelrestrikties gebruiken bij elke parameterschatting, heeft een specificatiefout reperkussies op alle schattingen.

In het voorgaande is getracht te schetsen dat een ekonometrisch model in zijn uiteindelijke gedaante wel de sporen van het formele denken vertoont, maar er niet in belangrijke mate door wordt bepaald. Veel en soms misschien te veel knopen moeten bij de opbouw van een model worden doorgehakt; de gegeven opsomming is nog geenszins uitputtend. Ter verdediging van de pragmatische werkwijze van de ekonometrist moge dienen dat zijn opdrachtgever niet om schoonheid uit economisch-theoretisch gezichtspunt vraagt, maar om resultaten die van praktisch nut zijn.