



centraal bureau voor de statistiek

Hoofdafdeling Statistische Methoden

Postbus 959, 2270 AZ VOORBURG

BOX-JENKINS MODELLEN EN COHORT-VRUCHTBAARHEID

Joop de Beer ^{*})

Samenvatting

De constatering dat het aantal echtparen dat kinderloos blijft gedurende de afgelopen tien jaar is toegenomen, lijkt weinig verrassend. Minder triviaal is de vraag naar de omvang van dit verschijnsel. Het is een onmiskenbaar feit dat het aantal eerste geboorten bij recente huwelijken is gedaald. Tegelijkertijd is echter de gemiddelde huwelijksduur tot de eerste geboorte toegenomen. Het gevolg is dat de daling van het aantal eerste geboorten niet tot een overeenkomstige toename van de kinderloosheid heeft geleid. In dit rapport worden enkele tijdreeksmodellen ontwikkeld met behulp waarvan een beeld kan worden verkregen van de recente ontwikkeling van de kinderloosheid.

^{*}) De in dit rapport weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek of met de opvattingen van andere stafmedewerkers van het CBS. De auteur dankt C.A.W. Citteur, A.J. Hundepool, W.J. Keller, B.T. Rijken van Olst en T.J. Wansbeek voor het kritisch lezen van een eerdere versie van dit rapport.

1. Inleiding

Bij een analyse van veranderingen in het vruchtbaarheidsgedrag dient onderscheid te worden gemaakt tussen een longitudinale en een transversale benadering. Een longitudinale analyse neemt de vruchtbaarheid van opeenvolgende generaties of cohorten tot uitgangspunt van onderzoek. Een cohort bestaat uit een groep mensen die in eenzelfde periode eenzelfde gebeurtenis hebben meegemaakt. In dit rapport wordt uitgegaan van huwelijkscohorten bestaande uit vrouwen die in hetzelfde jaar zijn gehuwd. Bij een transversale analyse gaat men uit van de vruchtbaarheid die in een bepaalde periode gerealiseerd wordt. Deze periode-waarnemingen hebben betrekking op een doorsnede van de cohorten die in een bepaalde periode (bijvoorbeeld een kalenderjaar) aan de vruchtbaarheid bijdragen. Bij longitudinale waarnemingen van de vruchtbaarheid kan men onderscheid maken tussen het uiteindelijk bereikte niveau van de vruchtbaarheid (intensiteit) en het patroon in de tijd volgens welke deze gerealiseerd wordt (lokalisatie). Dit onderscheid kan ook worden gemaakt voor transversale waarnemingen. De huwelijks-duurspecifieke vruchtbaarheidscijfers die in eenzelfde kalenderjaar worden waargenomen kunnen namelijk worden toegerekend aan een 'fictieve' cohort, waarvan de intensiteit en lokalisatie kunnen worden bepaald.

En probleem bij een longitudinale analyse van vruchtbaarheidsgedrag is dat voor recente cohorten de intensiteit onbekend is. Weliswaar is bekend hoeveel geboorten tot op het tijdstip van waarneming bij recente cohorten hebben plaatsgevonden, maar het is niet mogelijk hierbij onderscheid te maken tussen veranderingen in de intensiteit en in de lokalisatie, zodat niet duidelijk is of een afname in het aantal geboorten kan worden toegeschreven aan een daling van het vruchtbaarheidsniveau. Aangezien van recente 'fictieve' cohorten wel intensiteit en lokalisatie bekend zijn, ligt het voor de hand bij een analyse van recente veranderingen in de vruchtbaarheid gebruik te maken van transversale (periode-) waarnemingen. Een analyse op grond van periode-waarnemingen kan evenwel een vertekend beeld geven van veranderingen in het vruchtbaarheidsgedrag van opeenvolgende cohorten. Veranderingen van het in achtereenvolgende jaren waargenomen vruchtbaarheidsniveau (transversale intensiteit) worden namelijk niet alleen veroorzaakt door veranderingen van het vruchtbaarheidsniveau van cohorten

(longitudinale intensiteit), maar ook door veranderingen van het patroon volgens welke die vruchtbaarheid in de tijd wordt gerealiseerd (longitudinale lokalisatie).

De vraag naar de omvang van de kinderloosheid gedurende de afgelopen tien jaar kan derhalve niet zonder meer worden beantwoord door een verwijzing naar periode-waarnemingen van aantallen eerste geboorten. Maar ook cohort-waarnemingen kunnen hierover geen uitsluitsel geven. Om die reden lijkt het zinvol na te gaan of een combinatie van beide soorten waarnemingen een beter beeld van recente ontwikkelingen kan verschaffen. Een mogelijkheid hiertoe biedt translatie-analyse, waarbij getracht wordt een model te formuleren dat de relatie tussen periode- en cohort-waarnemingen kan beschrijven. Ryder (1964) geeft een theoretische beschrijving van het translatie-model. Frinking en De Roo (1979) en De Beer (1981) ontwikkelen uitgaande van het translatie-principe modellen voor het voorspellen van cohort- vruchtbaarheid.

In dit rapport wordt nagegaan in hoeverre tijdreeksanalyse een bijdrage kan leveren aan het ontwikkelen van dergelijke modellen. Uitgangspunt vormt daarbij de benadering van Box en Jenkins (1970). Met behulp van zogenaamde transfer-modellen kan de relatie tussen cohort- en periode-waarnemingen worden vastgesteld. Als inleiding op de behandeling van transfermodellen zullen eerst univariate ARIMA-modellen van de cohort- vruchtbaarheid worden beschreven. Daar voor een analyse van de ontwikkeling van de kinderloosheid slechts de eerste geboorten van belang zijn, hebben in dit rapport de begrippen 'vruchtbaarheid' en 'intensiteit' betrekking op het relatieve aantal gehuwde vrouwen dat minstens één kind voortbrengt, ongeacht het totaal aantal kinderen dat uiteindelijk gemiddeld per vrouw wordt geboren. De gegevens zijn voor de periode 1937-1971 verkregen uit CBS (1974) en voor de daaropvolgende jaren uit de jaarlijkse supplementen van de Maandstatistiek van Bevolking en Volksgezondheid. De data voor de cohorten 1937-1978 zijn weer-gegeven in de bijlage.

2. Het univariate ARIMA-model

De toepassing van univariate tijdreeksanalyse heeft in de afgelopen tien jaar een sterke stimulans ontvangen door het verschijnen (in 1970) van het in korte tijd tot standaardwerk uitgegroeide boek van Box en Jenkins, "Time series analysis: forecasting and control". In dit boek beschrijven beide auteurs een algemeen - stochastisch - model waarvan veel methoden van tijdreeksanalyse als bijzonder geval kunnen worden afgeleid. Het univariate model voor een tijdreeks Y_t wordt aangeduid als ARIMA(p,d,q) waarbij p=aantal autoregressieve (AR) parameters, d=aantal differenties van Y_t en q=aantal 'moving average' (MA) parameters (i.e. de coëfficiënten van vertraagde storingstermen):

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} \quad (1)$$

waarbij $\nabla Y_t = \nabla^d Y_t$ is de d^e differentie van de te voorspellen variabele Y_t , zodat

$$\nabla Y_t = Y_t - Y_{t-1}$$

en

$$\nabla^2 Y_t = \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2};$$

u_t is een normaal verdeelde storingsterm met $E(u_t) = 0$, $E(u_t^2) = \sigma_u^2$ en $E(u_t u_{t+k}) = 0$, $k \neq 0$; ϕ_i en θ_i zijn de te schatten coëfficiënten. Als we nu een autoregressieve en een moving average operator definiëren als respectievelijk

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

en

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

waarbij $B^i Y_t = Y_{t-i}$, kan model (1) worden herschreven als

$$\phi(B) \nabla^d Y_t = \theta(B) u_t \quad (3)$$

Box en Jenkins leggen de nadruk op toepassing van een iteratieve gefaseerde methode die uit de volgende stappen bestaat: identificeren (vaststellen van de waarden van p, d en q), schatten (bepaling van de waarde van de coëfficiënten die een optimale 'fit' oplevert), 'diagnostic checking' (analyse van de residuen) met terugkoppeling naar de identificatie en tenslotte - wanneer het 'optimale' model is verkregen - voorspellen.

De identificatie geschiedt aan de hand van de autocorrelatiecoëfficiënten (ac's) en partiële autocorrelatiecoëfficiënten (pac's) van de te analyseren reeks als functie van de vertraging ('lag') k . De autocorrelatiecoëfficiënt r_k geeft de correlatie tussen Y_t en Y_{t+k} aan, terwijl de partiële autocorrelatiecoëfficiënt ϕ_{kk} aangeeft in hoeverre Y_{t-k} een bijdrage kan leveren aan de verklaring van Y_t wanneer rekening is gehouden met de invloed van $Y_{t-1}, \dots, Y_{t-k+1}$. Door vergelijking van de geschatte ac's en pac's met het theoretische verloop zoals dat uit de verschillende specificaties van het ARIMA-model kan worden afgeleid, kan een indicatie worden verkregen van de te kiezen waarden voor p , d en q . Allereerst wordt bepaald hoeveel differenties moeten worden genomen tot de trend uit de reeks is verwijderd. Een trend manifesteert zich in een geleidelijk dalende ac-functie. Uit fig. 1a blijkt duidelijk de invloed van de trend op de ac's in de door ons onderzochte reeks van fracties van gehuwde vrouwen (van de cohorten 1937-1969) die binnen tien jaar na huwelijksluiting minstens één kind hebben voortgebracht. Uit fig. 1b kan worden geconcludeerd dat het berekenen van eerste verschillen voldoende is om de effecten van de trend te elimineren.

Op grond van de ac's en de pac's van de na differentie verkregen reeks wordt de keuze van p en q bepaald. Bij een MA-proces (d.w.z. $p=0$) zijn de ac's voor $k > q$ gelijk aan nul (hetgeen volgt uit de veronderstelling dat $\text{cov}(u_t, u_{t+k}) = 0$ voor $k \neq 0$), terwijl bij een AR-proces (d.w.z. $q=0$) de pac's voor $k > p$ gelijk zijn aan nul. De pac-functie bij een MA-proces en de ac-functie bij een AR-proces vertonen een verloop dat wordt gekenmerkt door een exponentieel dalende functie of een gedempte sinus. Bij een gemengd proces (d.w.z. $p > 0$ en $q > 0$) wordt een dergelijk verloop vertoond door de ac als $q > p$ en door de pac als $p > q$. De in de figuren 1b en 1c weergegeven ac's en pac's laten niet zonder meer een éénduidige conclusie toe over het te kiezen model. Het verloop van de ac's wijst in de richting van een ARIMA-(0,1,4) model, terwijl de pac's een keuze voor het ARIMA(3,1,0) model lijken te rechtvaardigen.

De parameters van deze specificaties van het ARIMA model zijn geschat met een niet-lineaire kleinste-kwadratenmethode die een benadering van de maximale-aannemelijkheidsschatting oplevert (Box en Jenkins (1970), p. 213). Daarbij is gebruik gemaakt van de Marquardt-algoritme. De schattingsresultaten zijn vermeld in tabel 1. De beoordeling van de kwaliteit van een geschat model (door Box en Jenkins aangeduid als 'diagnostic checking') kan plaatsvinden aan de hand van een analyse van de residuen. Bovendien is het mogelijk door 'overfitting' (toevoeging van extra parameters) een indruk te

Figuur 1. Autocorrelatiecoëfficiënten (ac's) en partiële autocorrelatiecoëfficiënten (pac's)

a. Ac's van de oorspronkelijke reeks Y_t



b. Ac's van de eerste verschillen ∇Y_t



c. Pac's van de eerste verschillen ∇Y_t



verkrijgen of het model als adequaat kan worden beschouwd, dan wel of een uitbreiding van het model gewenst is. Voor wat betreft een toetsing van het model op grond van de residuen, kan in eerste instantie worden gekeken naar de grootte van die residuen zoals die zich bijvoorbeeld manifesteert in de 'mean square error' (MSE). Verder is het van belang het verloop van de residuen in de beschouwing te betrekken. Wanneer daarin systematische elementen aanwezig blijken te zijn, kan worden geconcludeerd dat blijkbaar niet het optimale model is geselecteerd. Een dergelijke analyse kan plaatsvinden aan de hand van de ac's van de residuen.

Tabel 1. Schattingsresultaten ARIMA-modellen (schattingsperiode 1937-1969)

model	geschatte parameters ^{b)}									MSE ^{a)} ×10 ⁻⁵
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1	θ_2	θ_3	θ_4	θ_5	
(3,1,0)	0,39	-0,40	0,38							9,96
	(0,16)	(0,16)	(0,16)							
(4,1,0)	0,47	-0,45	0,42	-0,08						9,87
	(0,19)	(0,18)	(0,18)	(0,18)						
(3,1,1)	0,12	-0,33	0,30		-0,37					9,90
	(0,28)	(0,17)	(0,20)		(0,30)					
(2,1,0)	0,27	-0,32								11,91
	(0,17)	(0,17)								
(0,1,4)					-0,56	0,35	-0,23	-0,56		8,48
					(0,15)	(0,18)	(0,18)	(0,15)		
(0,1,5)					-0,66	0,47	-0,24	-0,63	0,18	7,17
					(0,17)	(0,19)	(0,17)	(0,18)	(0,19)	
(1,1,4)	-0,33				-0,98	0,17	-0,17	-0,68		6,95
	(0,33)				(0,38)	(0,28)	(0,23)	(0,22)		
(0,1,3)					-0,44	0,25	-0,02			10,29
					(0,18)	(0,19)	(0,18)			

a) 1942 - 1969.

b) Tussen haakjes staan de geschatte standaardfouten vermeld.

Voor de residuen die resulteren na schatting van model (3,1,0) is uit het verloop van de ac's geen systematisch patroon af te leiden. Als toetsingsgrootheid is door Box en Jenkins voorgesteld de 'portmanteau test' te berekenen $Q = n \sum_{k=1}^K r_k^2(u)$ waarbij n=aantal waarnemingen na differenties, $r_k^2(u)$ = de

correlatie tussen u_t en u_{t-k} en K = het aantal ac's dat in de beschouwing wordt betrokken. Onder de veronderstelling dat het geschatte model correct is, is Q bij benadering χ^2 verdeeld met $K-p-q$ vrijheidsgraden. Ljung en Box (1978) geven een modificatie die voor een beperkt aantal waarnemingen dichter bij de asymptotische verdeling ligt.

$$Q' = n(n+2) \sum_{k=1}^K (n-k)^{-1} r_k^2(u). \quad (4)$$

Voor het (3,1,0) model is $Q' = 14,6$ voor $K = 15$ (12 vrijheidsgraden), terwijl de tabelwaarde van de χ^2 -verdeling bij 5% 21,0 bedraagt, zodat volgens dit criterium geen indicatie wordt verkregen dat het model dient te worden verworpen. Overigens dient bij korte tijdreeksen de nodige voorzichtigheid te worden betracht bij de interpretatie van de 'portmanteau test' (zie bijvoorbeeld Newbold (1980)).

Vervolgens is het model getest door 'overfitting' toe te passen. Daarbij wordt gekeken of bij uitbreiding van het model de schattingen van de additionele parameters significant van nul verschillende waarden opleveren. Uit de tweede en derde regel van tabel 1 blijkt dat noch toevoeging van een AR-parameter, noch toevoeging van een MA-parameter tot verbetering leidt. Ten slotte is gekeken of een kleiner model tot even goede resultaten zou kunnen leiden. Dit blijkt niet het geval te zijn: model (2,1,0) voldoet duidelijk minder goed dan model (3,1,0). In dit verband kan nog worden opgemerkt dat de correlatie tussen de parameterschattingen van model (3,1,0) niet groot is: de correlatiecoëfficiënten liggen tussen -0,26 en +0,33.

De wortels van de karakteristieke vergelijking $\lambda^3 - 0,39\lambda^2 + 0,40\lambda - 0,38 = 0$ geven een indicatie van het gedrag in de tijd van de onderzochte reeks volgens het hier geschatte model (3,1,0). De reële wortel 0,65 geeft aan dat hier sprake is van een stabiel proces, waarbij volgens de complexe wortels $0,76$ (cos 1,74 \pm i sin 1,74) een gedempte fluctuatie met een periode van $\frac{2\pi}{1,74} = 3,6$ jaar plaatsvindt.

Door volgens de ac's van de reeks van eerste verschillen ook model (0,1,4) een geschikte specificatie zou kunnen zijn hebben we ook dat model geschat. Volgens tabel 1 voldoet het model redelijk. De correlatiecoëfficiënten van de parameterschattingen liggen tussen -0,21 en +0,52. De 'portmanteau test' leidt tot de conclusie dat het model niet verworpen kan worden: $Q' = 3,6$ tegen een tabelwaarde van 19,5 bij 5% (11 vrijheidsgraden). Om te kunnen beoordelen

in hoeverre dit model verschilt van de eerder geschatte specificatie (3,1,0) hebben we dit laatste model herschreven in de 'random shock' vorm (waarbij de autotegressieve parameters worden gesubstitueerd voor moving average-termen):

$$VY_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3} u_t = \Psi(B)u_t = \left(1 - \sum_{j=1}^{\infty} \psi_j B^j\right) u_t \quad (5)$$

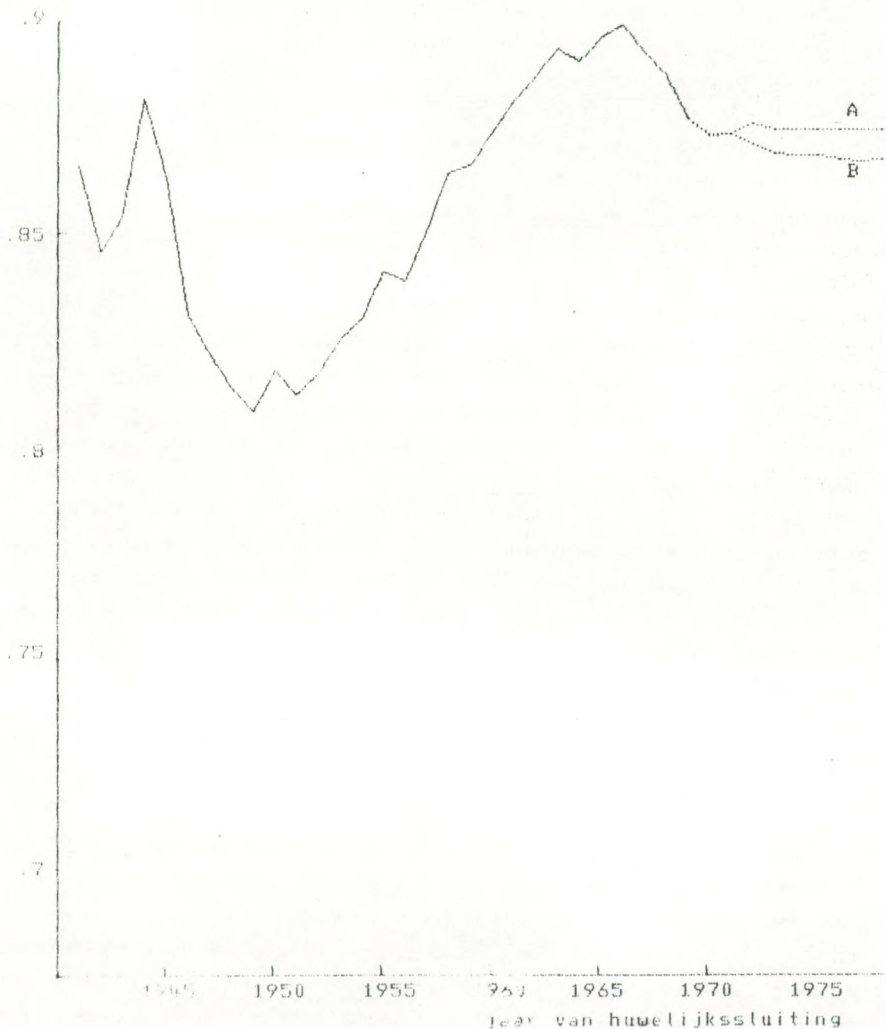
Dit levert voor de eerste 4 parameters ψ_j respectievelijk de waarden -0,39, 0,26, -0,12 en -0,30 op. Beperking tot deze vier parameters leidt tot een residuele kwadratensom van $3,10 \times 10^{-3}$ (tegen $2,85 \times 10^{-3}$ voor model (3,1,0)). Het patroon van deze schattingen vertoont overeenkomst met dat van de geschatte parameters van model (0,1,4). De modellen (3,1,0) en (0,1,4) verschillen dus minder sterk van elkaar dan op grond van de duidelijk uiteenlopende specificaties verwacht zou kunnen worden.

In tegenstelling tot hetgeen het geval is voor model (3,1,0) leidt uitbreiding van model (0,1,4) wel tot een betere 'fit'. De waarden van de additionele coëfficiënten zijn evenwel klein ten opzichte van de geschatte standaardfouten. Bovendien is bij het gemengde model (1,1,4) sprake van een sterke correlatie tussen de parameterschattingen. De conclusie kan derhalve luiden dat de twee specificaties van het ARIMA model die op grond van de identificatie aan de hand van de ac's en pac's werden gekozen, namelijk (3,1,0) en (0,1,4), inderdaad het beste blijken te voldoen. Uit figuur 2 blijkt dat de ex ante voorspellingen van beide methoden niet veel van elkaar verschillen. Het feit dat de identificatie niet tot een eenduidige conclusie leidt, heeft derhalve geen ernstige gevolgen voor de keuze van het te hanteren univariate voorspellingsmodel.

3. Transfermodellen:

Bij de schatting van de univariate ARIMA-modellen is alleen gebruik gemaakt van informatie over vrouwen die op het voorspeltijdstip minstens tien jaar zijn getrouwd. Informatie over meer recente cohorten is evenwel aanwezig op kalenderjaarbasis. Deze informatie kan benut worden door periode-waarnemingen als verklarende variabelen op te nemen in een transfermodel. 'Verklarend' moet hier overigens niet worden geïnterpreteerd als 'causaal'. Er wordt niet

Figuur 7. Waarnemingen en voorspellingen van de fractie gehuwde vrouwen die binnen tien jaar na huwelijkssluiting minstens één kind voortbrengt. De voorspellingen zijn afgeleid met univariate ARIMA-modellen: A = (0,1,4) en B = (3,1,0).



beweerd dat een bepaalde ontwikkeling van de longitudinale intensiteit wordt 'veroorzaakt' door het waargenomen verloop van de transversale variabelen. Er wordt slechts verondersteld dat de ontwikkeling van de laatste variabelen voorspellende waarde kan hebben voor de ontwikkeling van de longitudinale intensiteit.

De huwelijksduurspecifieke vruchtbaarheidscijfers die in eenzelfde kalenderjaar worden waargenomen en die betrekking hebben op verschillende cohorten (in de tabel in de bijlage: de diagonalen van links onder naar rechts boven) kunnen worden samengevoegd alsof ze tot eenzelfde cohort behoren, een zogenaamde 'fictieve' of 'synthetische' cohort. Van een dergelijke cohort kunnen op overeenkomstige wijze als voor reële cohorten intensiteit en lokalisatie van de vruchtbaarheid worden bepaald. Dit zijn de transversale (periode-) waarnemingen. Het voordeel van deze periode-waarnemingen is dat ze tot op het jaar van waarneming bekend zijn, zodat recentere gegevens bekend zijn dan voor reële cohorten het geval is. Om die reden worden in de praktijk veelvuldig periode-waarnemingen gehanteerd voor een analyse van recente vruchtbaarheidsontwikkelingen. Een probleem is daarbij evenwel dat systematische afwijkingen kunnen optreden tussen periode- en cohort-waarnemingen. Deze kunnen ertoe leiden dat bepaalde periode-waarnemingen verkeerd worden geïnterpreteerd. Zo zal uitstel van geboorten bij recente cohorten leiden tot een daling van de transversale intensiteit (d.w.z. de intensiteit van fictieve cohorten), X_t , terwijl de longitudinale intensiteit, Y_t , geen of in ieder geval een minder sterke wijziging ondergaat. Een indicator van de mate waarin uitstel van geboorte plaatsvindt is de gemiddelde huwelijksduur bij de geboorte van het eerste kind van fictieve cohorten, Z_t . Wanneer bij jongere reële cohorten relatief minder eerste geboorten in de eerste huwelijksjaren plaatsvinden dan bij oudere cohorten, zal dit zich manifesteren in een toename van Z_t . Bij een schatting van het niveau van de cohort- vruchtbaarheid op grond van recente periode-waarnemingen dient derhalve deze indicator in de beschouwing te worden betrokken.

Bij het opstellen van een model waarin een relatie wordt gelegd tussen cohort- en periode-waarnemingen doet zich de moeilijkheid voor dat niet zonder meer duidelijk is hoe de waarnemingen dienen te worden gedateerd. Een cohortwaarneming heeft namelijk niet betrekking op één specifiek kalenderjaar, maar op een reeks van jaren volgend op het jaar van huwelijksluiting. Hiermee kan rekening worden gehouden door het opstellen van een transfermodel. Daarbij wordt in de identificatiefase bepaald welke vertragingstructuur het best de relatie tussen cohort- en periode-waarnemingen

kan beschrijven. Als uitgangspunt zal een transfermodel worden geschat met één verklarende variabele, namelijk de transversale intensiteit x_t . Onder de veronderstelling dat de veranderingen in het spreidingspatroon van de eerste geboorten over de huwelijksduur in de voorspelperiode hetzelfde effect hebben op het verschil tussen de cohort- en de periode-intensiteit als in de schattingsperiode, kunnen met behulp van dit model voorspellingen worden afgeleid voor de intensiteit van recente cohorten op grond van waarnemingen van de periode-intensiteit. Wanneer evenwel deze veronderstelling niet opgaat, bijvoorbeeld tengevolge van een versnelling van het uitstel van de eerste geboorte, zullen de voorspellingen op een te laag niveau uitkomen (of op een te hoog niveau wanneer de eerste geboorte sneller na het tijdstip van huwelijksluiting zou plaatsvinden). In dat geval zullen de voorspellingen die worden afgeleid met een transfermodel waarin ook de gemiddelde huwelijksduur bij de eerste geboorte Z_t voorkomt als verklarende variabele, afwijken van die van het model met slechts de transversale intensiteit.

In het geval van één verklarende variabele luidt het transfermodel

$$y_t = v(B)x_t + n_t \quad (6)$$

$$v(B) = \sum_{i=0}^{\infty} v_i B^i \quad (7)$$

waarbij $y_t = \sqrt{d} y_t$; $x_t = \sqrt{d} x_t$ en n_t is een storingsterm die niet gecorreleerd is met x_t . Uitgaande van het streven naar 'parsimony' (i.e. beperking tot het kleinste mogelijk aantal parameters), kan het nuttig zijn te zoeken naar een specificatie van de 'transfer function' $v(B)$ als quotiënt van twee polynomen:

$$v(B) = \frac{\omega(B)}{\delta(B)} B^b \quad (8)$$

$$\omega(B) = \sum_{i=0}^s \omega_i B^i \quad (9)$$

$$\delta(B) = 1 - \sum_{i=1}^r \delta_i B^i \quad (10)$$

De storingsterm n_t kan worden toegeschreven aan de effecten van niet in het model opgenomen variabelen voorzover die niet gecorreleerd zijn met x_t en niet kunnen worden verklaard door de autoregressieve operator $\delta(B)$. Onder de veronderstelling dat het verloop van de storingsterm n_t kan worden beschreven door een ARIMA(p,d,q) model

$$\psi(B)n_t = \theta(B)u_t \quad (11)$$

resulteert het 'transfer function-noise model' (hierna kortweg aangeduid als 'transfermodel')

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} u_t, \quad (12)$$

ofwel

$$y_t = v(B)x_t + \psi(B)u_t, \quad (13)$$

waarbij u_t witte ruis is en $\psi(B) = \zeta^{-1}(B)\theta(B)$.

Een probleem bij het onderzoeken van relaties tussen twee of meer tijdreeksen is dat wanneer bij de afzonderlijke reeksen autocorrelatie optreedt, daarvoor correlatie tussen de reeksen kan optreden, ook als de reeksen geen enkel verband met elkaar hebben ('spurious correlation')(Jenkins, 1979, p. 16). Om die reden is het zinvol eerst de autocorrelatie uit de verklarende variabele te verwijderen door vorming van een univariaat ARIMA-model, waarmee vervolgens de te voorspellen variabele kan worden gefilterd ('prewhitening'). In het geval dat een 'feedback' optreedt (y_t kan wel door x_t worden verklaard, maar x_t niet door y_t), kan aan de hand van de cross-correlatiefunctie tussen de aldus getransformeerde reeksen identificatie van het model plaatsvinden. Op grond van de veronderstelling dat het verloop van de verklarende variabele x_t kan worden beschreven door het univariate ARIMA-model

$$\phi_x(B)x_t = \theta_x(B)\alpha_t, \quad (14)$$

waarbij α_t een witte-ruis-proces vormt, resulteert na 'prewhitening' het transfermodel

$$y_t = v(B)\alpha_t + \varepsilon_t, \quad (15)$$

waarbij $\beta_t = \phi_x(B) \sigma_x^{-1}(B) y_t$ en $\epsilon_t =$ storingsterm die onafhankelijk is van α_t . Hieruit volgt dat de k^e coëfficiënt van de polynoom $v(B)$ gelijk is aan

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2} \quad (16)$$

waarbij $\gamma_{\alpha\beta}(k) = E(\alpha_{t-k} \beta_t)$ en $\sigma_\alpha^2 = E(\alpha_t^2)$, zodat uitgedrukt in cross-correlatiecoëfficiënten $\rho_{\alpha\beta}(k)$

$$v_k = \rho_{\alpha\beta}(k) \cdot \frac{\sigma_\beta}{\sigma_\alpha} \quad (17)$$

Mit de geschatte cross-correlatiecoëfficiënten kan dus een indruk worden verkregen van het patroon van de transferfunctie $v(B)$. Dit patroon kan vervolgens worden vergeleken met het theoretische verloop zoals dat uit verschillende specificaties van de polynomen $\delta(B)$ en $\omega(B)$ kan worden afgeleid (zie Box en Jenkins (1970), p. 350, voor een overzicht van de combinaties voor $r=0,1,2$ en $s=0,1,2$).

Het verloop van de transversale intensiteit x_t in de periode 1946-1978 is te beschrijven met het ARIMA(1,1,0) model

$$(1 - 0,48B) \nabla x_t = \alpha_t \quad (18)$$

waarbij α_t een witte ruis proces vormt. De variabele y_t kan dan getransformeerd worden tot

$$\beta_t = (1 - 0,48B) \nabla y_t \quad (19)$$

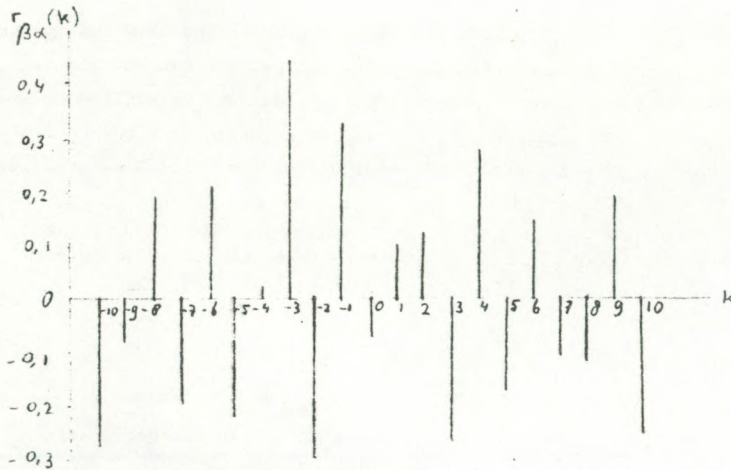
Aan de hand van de cross-correlatiecoëfficiënten van α_{t-k} en β_t voor verschillende waarden van k kan vervolgens getracht worden het 'juiste' transfermodel te selecteren (d.w.z. de waarden van r , s en b te bepalen). De correlatiecoëfficiënten zijn weergegeven in figuur 3. Opvallend hierbij is vooral dat enkele relatief hoge waarden van de correlatiecoëfficiënten voorkomen voor negatieve waarden van k . Dit betekent dat de 'verklarende' variabele een vertraging vertoont ten opzichte van de te voorspellen variabele, een op het eerste gezicht wellicht wat merkwaardige constatering. De oorzaak moet gezocht worden in het feit dat de longitudinale intensiteit van ieder cohort toegerekend is aan het

jaar waarin de vrouwen, die deel uitmaken van dat cohort, zijn gehuwd, terwijl deze vruchtbaarheid pas gedurende een reeks van jaren volgend op het jaar van de huwelijksluiting wordt gerealiseerd. Het model dat de relatie tussen α_t en β_t het beste blijkt te kunnen beschrijven is

$$(1 + 0,34B - 0,12B^2 - 0,49B^3)\beta_t = (0,19 + 0,16B)\alpha_{t+2} + \varepsilon_t \quad (20)$$

$$(0,16) \quad (0,17) \quad (0,17) \quad (0,07) \quad (0,07)$$

Figuur 3. Cross-correlatiecoëfficiënten tussen α_{t-k} en β_t .



Dit model dient als uitgangspunt voor het schatten van het transfermodel van de oorspronkelijke waarnemingen, hetgeen tot het volgende resultaat leidt:

$$(1 + 0,13B - 0,15B^2 - 0,39B^3)VY_t = (0,19 + 0,19B)VX_{t+2} + n_t \quad (21)$$

$$(0,19) \quad (0,18) \quad (0,18) \quad (0,07) \quad (0,08)$$

De storingsterm u_t kan worden beschreven door het univariate ARIMA(0,0,3) model:

$$u_t = (1 - 0,12B - 0,40B^2 + 0,65B^3)u_t. \quad (22)$$

(0,11) (0,08) (0,12)

De autocorrelatiecoëfficiënten van u_t laten geen systematisch patroon zien. De 'portmanteau test' levert een waarde op van $Q' = 11,5$ (tabelwaarde bij 12 vrijheidsgraden voor 5% is 21,0). Het lijkt derhalve gerechtvaardigd er vanuit te gaan dat u_t een witte ruis proces vormt, zodat het model kan worden geaccepteerd. Simultane schatting van alle parameters leidt uiteindelijk tot het volgende resultaat

$$(1 + 0,59B - 0,73B^2 - 0,49B^3)Y_t = (0,21 + 0,22B) \nabla X_{t+2} + (0,20) (0,10) (0,21) (0,03) (0,04) + (1 - 0,32B - 1,31B^2 + 1,79B^3)u_t. \quad (23)$$

(0,49) (0,68) (0,53)

Naast een analyse van de ac's van de residuen kan ter beoordeling van het geschatte model de cross-correlatie tussen de verklarende variabele X_t en de storingsterm u_t worden bepaald. Uit (13) en (24)

$$Y_t = \hat{v}(B)x_t + \hat{\psi}(B)u_t, \quad (24)$$

waarbij $\hat{v}(B)$, $\hat{\psi}(B)$ en \hat{u}_t schattingen weergeven, kan worden afgeleid dat

$$\hat{u}_t = \hat{\psi}^{-1}(B) \{ \hat{v}(B) - \hat{v}(B) \} x_t + \hat{\psi}^{-1}(B) \psi(B) u_t, \quad (25)$$

zodat bij een onjuist model cross-correlatie tussen de verklarende variabele en de geschatte storingsterm zal optreden. Voor het hier geschatte model blijkt dat de cross-correlatiecoëfficiënten klein zijn ($r^2 < 0,1$ voor verschillende vertragingen). Dit levert derhalve geen indicatie op dat het model onjuist zou zijn.

Als maatstaf voor de 'fit' wordt weer de gemiddelde kwadratische fout berekend: $MSE = 0,39 \times 10^{-5}$. Deze waarde kan niet zonder meer worden vergeleken met die van de univariate modellen uit tabel 1, daar die betrekking hebben op een andere - langere - schattingsperiode. Ter vergelijking is voor het model (3,1,0) de MSE bepaald voor dezelfde periode als waar de MSE van het transfermodel betrekking op heeft (1950-1969): $MSE = 6,64 \times 10^{-5}$. Daarbij dient wel

te worden bedacht, dat deze periode voor het univariate model niet overeenkomt met de schattingsperiode, waardoor bij de schatting de kwadratensom van de residuen niet over die periode is geminimaliseerd. Het verschil met het transfermodel is evenwel duidelijk. Opvallend is dat de waarden van de geschatte geschatte standaardfouten van de parameters die betrekking hebben op de 'verklarende' variabele relatief klein zijn. In samenhang met de aanzienlijk betere 'fit' leidt dit tot de conclusie dat het transfermodel is te verkiezen boven een univariaat ARIMA-model. De voorspellingen die met het transfermodel zijn berekend, zijn weergegeven in figuur 4 (curve A). Duidelijk manifesteert zich hier de invloed van informatie over het verloop van de periode- vruchtbaarheid in de jaren zeventig op de voorspellingen. Deze informatie die niet in het univariate ARIMA-model wordt verwerkt geeft aanleiding tot het in twijfel trekken van de plausibiliteit van de voorspellingen zoals die uit de univariate modellen volgen.

Het is in principe mogelijk meer dan één verklarende variabele in het transfermodel op te nemen. Bij twee variabelen luidt het algemeen model (Jenkins (1979), p. 104):

$$Y_t = \frac{\omega_1(B)}{\delta_1(B)} x_t + \frac{\omega_2(B)}{\delta_2(B)} z_t + \frac{\theta(B)}{\phi(B)} u_t. \quad (26)$$

Daar, zoals reeds eerder opgemerkt, het verschil tussen de longitudinale en de transversale intensiteit mede wordt beïnvloed door het verloop van de gemiddelde huwelijksduur bij de geboorte van het eerste kind, is deze variabele in het transfermodel opgenomen als additionele verklarende variabele

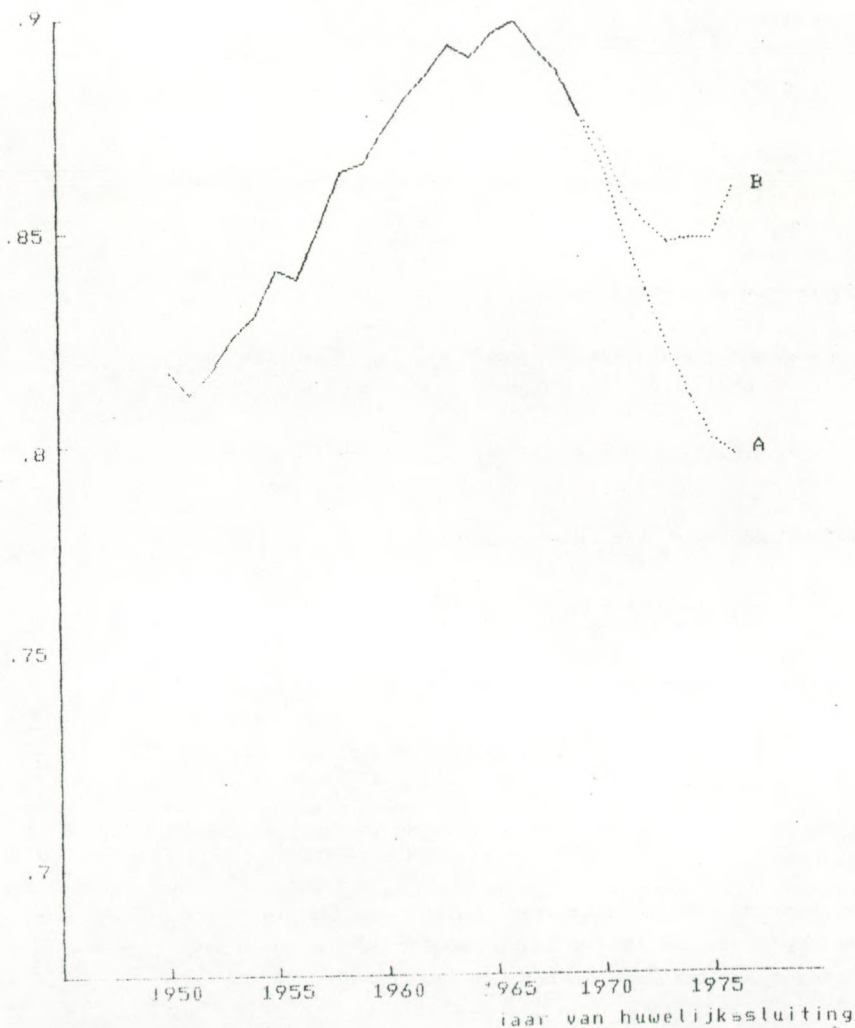
Het verloop van de transversaal waargenomen gemiddelde huwelijksduur bij de geboorte van het eerste kind Z_t kan worden beschreven met het (1,1,0) model

$$(1 - 0,78B) \nabla Z_t = \alpha'_t. \quad (27)$$

Met behulp van dit model kan de te voorspellen variabele Y_t getransformeerd worden tot de reeks β'_t . De relatie tussen α'_t en β'_t is te beschrijven met

$$(1 + 0,50B + 0,05B^2 - 0,45B^3) \beta'_t = \begin{matrix} 0,08\alpha'_{t+1} \\ (0,07) \end{matrix} + \varepsilon'_t. \quad (28)$$

Figuur 4. Voorspellingen van de fractie gehuwde vrouwen die binnen tien jaar na huwelijkssluiting minstens één kind voortbrengt. De voorspellingen zijn afgeleid met behulp van transfermodellen met als verklarende variabelen resp. (A) de transversale intensiteit en (B) de transversale intensiteit en de gemiddelde huwelijksduur bij de geboorte van het eerste kind.



Dit betekent dat de autoregressieve operator $\delta_1(B)$ van de transfer functie van Y_t en Z_t dezelfde vorm heeft als $\delta_1(B)$. Daar X_t en Z_t niet sterk gecorreleerd zijn ($r^2 = 0,31$) is het volgende transfermodel geschat met X_t en Z_t als verklarende variabelen:

$$\delta(B) \nabla Y_t = \omega_1(B) \nabla X_t + \omega_2(B) \nabla Z_t + n_t \quad (29)$$

waarbij

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \delta_3 B^3 \quad (30)$$

$$\omega_1(B) = (\omega_{10} + \omega_{11} B) B^{-2} \quad (31)$$

$$\omega_2(B) = \omega_{20} B^{-1}. \quad (32)$$

Dit leverde de schattingen op:

$$\begin{aligned} (1 + 0,16B - 0,12B^2 - 0,40B^3) \nabla Y_t &= (0,22 + 0,20B) \nabla X_{t+2} + \\ (0,18) \quad (0,17) \quad (0,17) &\quad (0,07) \quad (0,07) \\ &+ 0,07 \nabla Z_{t+1} + n_t. \end{aligned} \quad (33)$$

Het verloop van n_t kan worden beschreven met het (0,0,3) model

$$n_t = (1 - 0,02B + 0,32B^2 - 0,34B^3) u_t. \quad (34)$$

(0,22) (0,22) (0,24)

Simultane schatting van alle parameters leidt tot

$$\begin{aligned} (1 + 0,42B - 0,53B^2 - 0,38B^3) \nabla Y_t &= (0,23 + 0,20B) \nabla X_{t+2} + \\ (0,15) \quad (0,17) \quad (0,19) &\quad (0,05) \quad (0,04) \\ &+ 0,05 \nabla Z_{t+1} + (1 + 0,55B + 1,90B^2 - 4,91B^3) u_t \end{aligned} \quad (35)$$

(0,02) (1,34) (1,41) (1,50)

De geschatte correlaties tussen de residuen en de beide verklarende variabelen zijn klein. De MSE bedraagt $0,06 \times 10^{-5}$. Dit levert geen sterke verbetering op ten opzichte van het model met slechts de

transversale intensiteit als verklarende variabele. Bij de interpretatie van de resultaten moet rekening worden gehouden met het feit dat het aantal geschatte parameters groot is ten opzichte van het aantal waarnemingen. Overigens kan worden opgemerkt dat geen sterke correlatie optreedt tussen de geschatte parameters, met uitzondering van de storingstermen en de vertraagde afhankelijke variabelen (als verklarende variabelen), voorzover die op de zelfde periode betrekking hebben.

De voorspellingen die met dit model zijn berekend, zijn weergegeven in figuur 4 (curve B). Opvallend is vooral het duidelijke verschil met de voorspellingen van het transfermodel waarin Z niet voorkomt als verklarende variabele. De oorzaak hiervan is gelegen in het feit dat de transversaal waargenomen gemiddelde huwelijksduur bij de geboorte van het eerste kind juist in de jaren zeventig (d.w.z. buiten de schattingsperiode) sterk is gestegen. Deze nam in een periode van tien jaar (1968-1978) met één jaar toe. Volgens de hierboven weergegeven schattingen van het transfermodel zouden bij een gelijkblijvend niveau van de transversale intensiteit de voorspellingen van de longitudinale intensiteit met 98 geboorten per 10 000 vrouwen per jaar toenemen als de gemiddelde huwelijksduur elk jaar met 0,1 jaar toeneemt. Immers, de 'gain' die de uiteindelijke verandering in de afhankelijke variabele als gevolg van een verandering in een verklarende variabele aangeeft, is gelijk aan

$$\frac{0,05}{1 + 0,42 - 0,53 - 0,38} = 0,098$$

Hierbij dient wel te worden bedacht dat deze schatting betrekking heeft op een periode waarin de gemiddelde huwelijksduur bij de geboorte van het eerste kind niet sterk veranderde. Het feit dat het verloop in de voorspellingsperiode hiervan sterk afwijkt is een reden om de voorspellingen met het nodige voorbehoud te hanteren.

4. Conclusies

Toepassing van de Box-Jenkins methode van tijdreeksanalyse levert een meer verfijnde analyse op dan de gebruikelijke regressie van verschillende tijdreeksen tegen elkaar (zie bijvoorbeeld Jenkins (1979), p.88 e.v.). Ter

voorkoming van 'spurious correlation' krijgt bij de Box Jenkins methode vooral de identificatie van het model aandacht. Daarbij is een belangrijke rol weggelegd voor een analyse van de residuen. Het gaat niet alleen om de omvang van de residuen, maar evenzo om het patroon dat daarin kan worden onderkend.

Hoewel de reeks van cijfers betreffende de longitudinale intensiteit van echte eerste geboorten kort is, blijkt het toch mogelijk om zowel univariate ARIMA- als transfermodellen te vormen die het statistisch gezien redelijk goed doen. Ook na eliminatie van trendmatige ontwikkelingen, ter vermindering van trendcorrelatie, kunnen nog significante verbanden worden aangetroffen. Dit neemt uiteraard niet weg dat hantering van de met deze modellen berekende voorspellingen met de nodige voorzichtigheid dient te geschieden. Uiteindelijk is de informatie die in de beschouwing is betrokken bij de formulering van die modellen zeer beperkt, zowel wat de lengte van de periode betreft als wat het aantal en de aard van de verklarende variabelen betreft. Slechts binnen dit beperkte kader worden 'optimale' voorspellingen (dat wil zeggen 'minimum mean square error' voorspellingen) verkregen, uitgaande van de veronderstelling dat het model ook buiten de schattingsperiode geldt.

Met betrekking tot de resultaten die worden verkregen met behulp van univariate ARIMA-modellen valt vooral op, dat slechts geringe wijziging in de longitudinale intensiteit wordt voorspeld. Voor wat het (0,1,4) model betreft, is dit geenszins verrassend. Hierbij wordt immers verondersteld, dat de omvang van de eerste veranderingen in de longitudinale intensiteit wordt bepaald door een reeks van vier storingstermen met verwachting nul, zodat voor tijdstippen van vier of meer jaar vooruit geen veranderingen worden verwacht. Dat ook het (3,1,0) model geen grote veranderingen voorspelt, kan worden toegeschreven aan het feit dat in de schattingsperiode slechts vrij korte perioden van ononderbroken stijging of daling voorkomen. In de schattingsperiode vindt twee keer een omslag in de richting van de reeks plaats, terwijl bovendien de stijging in de periode 1949-1966 herhaaldelijk wordt onderbroken, zodat een model dat grote veranderingen genereert deze reeks niet goed had kunnen voorspellen, respectievelijk niet met grote aannemelijkheid aan deze reeks ten grondslag had kunnen liggen.

De cruciale vraag ten aanzien van het afleiden van voorspellingen is uiteraard in hoeverre de veronderstelling dat dit model ook buiten de schattingsperiode geldt, gerechtvaardigd is. Hierbij kan niet worden volstaan met

een vergelijking van de 'fit' van de verschillende modellen in de schattingsperiode. In dit verband rijst de vraag of de - impliciete - veronderstelling van constante coëfficiënten terecht is. Wanneer ervan wordt uitgegaan dat de coëfficiënten een stochastisch karakter hebben als gevolg van het feit dat niet in het model opgenomen factoren een versturende invloed uitoefenen, kan worden gedacht aan toepassing van Kalman filtering. Dit kan ertoe leiden dat bij het afleiden van voorspellingen meer gewicht wordt gehecht aan de daling van de longitudinale intensiteit na 1966, zodat een lager niveau voor de cohorten van de jaren zeventig wordt voorspeld dan het niveau dat in figuur 2 is weergegeven.

Door het opnemen van transversale grootheden als verklarende variabelen in een transfermodel kan informatie over recente vruchtbaarheidsontwikkelingen zoals die zich manifesteren in periode-waarnemingen, worden betrokken bij het afleiden van voorspellingen van longitudinale intensiteit. De transversale - periode - waarnemingen geven aan dat in de eerste helft van de jaren zeventig een sterke daling van de vruchtbaarheid optrad die evenwel gepaard ging met een sterke toename van de gemiddelde huwelijksduur bij de geboorte van het eerste kind. Schatting van een transfermodel met alleen de transversale intensiteit als verklarende variabele levert een beeld op van het patroon dat de ontwikkelingen van de longitudinale en de transversale intensiteit ten opzichte van elkaar vertonen. Volgens dit model wordt weliswaar een daling voor de cohorten in de eerste helft van de jaren zeventig voorspeld, maar deze daling is beduidend minder sterk dan de daling die de - transversale - periode-waarnemingen in de jaren zeventig te zien geven.

Daar de samenhang tussen de longitudinale en de transversale intensiteit mede wordt beïnvloed door de ontwikkeling van de gemiddelde huwelijksduur bij de geboorte van het eerste kind, ligt opnemen van deze variabele in het transfermodel voor de hand. In een periode van uitstel van geboorten, dat wil zeggen wanneer jongere cohorten hun eerste kind in een latere fase van hun huwelijk krijgen dan oudere cohorten, zal de transversale intensiteit lager zijn dan de longitudinale intensiteit. Dit betekent dat gezien de sterke toename van de gemiddelde huwelijksduur zoals die uit de periode-waarnemingen voor de

jaren zeventig naar voren komt, verwacht kan worden dat de longitudinale intensiteit beduidend minder sterk is gedaald dan de transversale intensiteit. Het feit dat de voorspellingen van het tweede transfermodel (curve B in figuur 4) duidelijk afwijken van die van het eerste (curve A van figuur 4), kan worden toegeschreven aan de omstandigheid dat het proces van uitstel van de geboorte van het eerste kind zich in de schattingsperiode in veel minder sterke mate heeft voorgedaan dan in de voorspelperiode. Met andere woorden, de impliciete - ceteris-paribus-veronderstelling gaat voor het eerste transfermodel niet op: de relatie tussen de longitudinale en de transversale intensiteit ondergaat wijziging onder invloed van een niet in het model opgenomen variabele.

Een en ander lijkt tot de conclusie te leiden dat de voorspellingen van het tweede transfermodel (curve B) zijn te verkiezen boven de andere voorspellingen. Hierbij is evenwel een kanttekening op zijn plaats. Het verloop van de verklarende variabelen wijkt - zoals reeds eerder opgemerkt - in de voorspelperiode vrij sterk af van dat in de steekproefperiode. Dit geldt met name voor de gemiddelde huwelijksduur. In een dergelijk geval past een vraagteken bij de veronderstelling dat de waarden van de coëfficiënten die in de schattingsperiode zijn vastgesteld ook in de voorspelperiode gelden. Ook al is de specificatie van het tweede transfermodel te verkiezen boven die van de andere modellen, toch dient de nodige voorzichtigheid te worden betracht bij de interpretatie van de voorspellingen. In dit verband is van belang op te merken dat de geschatte coëfficiënt van de gemiddelde huwelijksduur vrij klein is ten opzichte van de geschatte standaardfout. Het lijkt derhalve gerechtvaardigd zich af te vragen of het niveau van de voorspellingen volgens curve B niet te hoog is. Loslaten van de veronderstelling van constante coëfficiënten kan in dit verband zinvol zijn. Toepassing van Kalman filtering onder de veronderstelling dat de coëfficiënt van de gemiddelde huwelijksduur een autoregressief proces van de eerste orde vormt, blijkt tot voorspellingen te leiden die tussen de curves A en B in figuur 4 inliggen (De Beer, 1982).

Literatuur

- Beer, J. de, 1981, Translatie en extrapolatie van vruchtbaarheidscijfers. Intern CBS-rapport (Hoofdafdeling Statistische Methoden, Centraal Bureau voor de Statistiek, Voorburg).
- Beer, J. de, 1982, Voorspelmodellen met variërende coëfficiënten: een toepassing van het Kalman-filter. Intern CBS-rapport (Hoofdafdeling Statistische Methoden, Centraal Bureau voor de Statistiek, Voorburg).
- Box, G.E.P. en G.M. Jenkins, 1970, Time series analysis. Forecasting and control (Holden-Day, San Francisco).
- C.B.S., 1974, Huwelijksvruchtbaarheid, een cohortanalyse 1937-1971 (Staatsuitgeverij, Den Haag).
- Frinking, G. en Y. de Roo, 1979, The translation process as forecast instrument (N.I.D.I., Voorburg).
- Granger, C.W.J. and P. Newbold, 1977, Forecasting economic time series (Academic Press, New York).
- Jenkins, G.M., 1979, Practical experiences with modelling and forecasting time series (Gwilym Jenkins & Partners, Jersey).
- Ljung, G.M. and G.E.P. Box, 1978, On a measure of lack of fit in time series models. *Biometrika* 65, pp. 297-303.
- Newbold, P., 1981, Some recent developments in time series analysis. *International Statistical Review* 49, pp. 53-66.
- Ryder, N.B., 1964, The process of demographic translation. *Demography* 1, pp. 74-82.