

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEHYGIENE
BILTHOVEN

Rapport nr. 958606 005

Demografische modellen voor scenario-onderzoek.

R.T.Hoogenveen

november 1991

Dit onderzoek werd verricht in opdracht van DGVGZ
en ten laste van de Directie RIVM.

Verzendlijst

- 1-6 Mr. R.F. Schreuder, Stafbureau Beleidsontwikkeling van het
Ministerie van WVC
7 Directeur-Generaal van Volksgezondheid
8 Plaatsvervangend Directeur-Generaal van Volksgezondheid,
tevens Hoofddirecteur Financiering en Planning
9 Drs. H.B. Haveman (Ministerie van W.V.C.)
10 Drs. S.E. Kooiker (NIVEL)
11 Prof.dr. P.J. van der Maas (EUR)
12 Dr. G.J. Truin (KUN)
13 Prof.dr.ir.drs. O.J. Vrieze (RL)
14 Depot Nederlandse Publicatie en Nederlandse Bibliografie
15 Directie Rijksinstituut voor Volksgezondheid en
Milieuhygiene
16 Hoofd Bureau Voorlichting & Public Relations
17 Drs. J. van den Berg Jeths
18 Drs. A. van der Giessen
19 Ir. M.L.L. van Genugten
20 Ir. B.J. de Haan
21 Drs. S.H. Heisterkamp
22 Drs. O.G. van der Heijden
23 Dr. J.C. Jager
24 Dr. P.G.N. Kramers
25 Ir. F. Langeweg
26 Drs. P. van der Mheen
27 Drs. M.J.J.C. Poos
28 Drs. M.J. Postma
29 Drs. D. Ruwaard
30 Dr. E.J.M. Veling
31 Dr. H. Verkleij
32 Auteur
33 Bureau Projecten- en Rapportenregistratie
34-35 Bibliotheek RIVM
36-55 Reserve-exemplaren

Voorwoord

In het kader van het project 'Methodologie-ontwikkeling ten behoeve van scenarioprojecten' is een demografisch rekenmodel ontwikkeld en geïmplementeerd. Dit model is toegepast voor vooruitberekeningen van het vóórkomen van o.a. ischaemische hartziekten en diabetes mellitus. Een verantwoording van dit model ontbrak echter vooralsnog. Naar aanleiding van het verschijnen van het scenario-rapport Chronische Ziekten Deel 1 (Casparie en Verkleij, 1990) is besloten om een RIVM-rapport uit te brengen, waarin het aan de berekeningen ten grondslag liggende demografische model beschreven wordt. Als uitgangspunt voor de modelstructuur is gekozen voor het wiskundige concept van de zogeheten Markov-keten. Bij de gebruikte gegevens is uitgegaan van gepubliceerde CBS-cijfers.

Tevens wordt dit model in een ruimer methodologisch kader geplaatst. Dit kader omvat een inleiding in de theorie van Markov-ketens, van overlevingsduuranalyse, van maten van demografische gebeurtenissen en van meerdimensionale demografische modellen. Na deze wiskundig getinte beschouwingen wordt in het tweede deel van het rapport op modeltoepassingen en verschillende modelversies ingegaan.

Inhoudsopgave

Verzendlijst	ii
Voorwoord	iii
Inhoudsopgave	iv
Summary	vii
Samenvatting	viii
Inleiding	1
1 Markov-modellen van toestandsovergangen	
1.1 Inleiding	3
1.2 De formele definitie	3
1.3 Stochastische matrices	4
1.4 De Chapman-Kolmogorov vergelijkingen	5
1.5 De toestandsovergangsimpulsintensiteit	5
1.6 De Kolmogorov differentiaalvergelijkingen	7
1.7 De kansovergangsmatrix geschreven als exponentiële functie	7
2 Een eenvoudig mortaliteitsmodel	
2.1 Inleiding	9
2.2 De overlevingscurve en -functie	9
2.3 Twee voorbeelden van aannames over de sterfte	11
2.4 Verschillende sterfthematen	12
3 Meerdimensionale demografische modellen	
3.1 Inleiding	13
3.2 De formele beschrijving van het demografisch proces	13
3.3 Constante overgangsimpulsintensiteiten	14
3.4 Lineaire verblijfskansen	15
3.5 De Leslie-matrix structuur	16
4 Maten van demografische gebeurtenissen	
4.1 Inleiding	19
4.2 Herhaalbare en versturende gebeurtenissen	19
4.3 Het Lexis-diagram	20
4.4 Demografische maten	21
4.5 De centrale maat	22
4.6 De onafhankelijke maat	23
4.7 De afhankelijke maat	24
4.8 De semi-afhankelijke maat	24

4.9	De meerdimensionale demografische modellen met gebruik van de demografische maten	25
5	Het geïmplementeerde uitgebreide demografische model	
5.1	Inleiding	27
5.2	Algemene kenmerken	27
5.3	Het demografische model	28
5.4	De beschreven demografische veranderingen	29
5.5	De structuur van het programma	31
5.6	De onderdelen van het hoofdprogramma toegelicht	32
6	De resultaten van het demografische model voor het verleden	
6.1	Inleiding	33
6.2	De basisversie	35
6.3	Enkele alternatieve versies	39
6.4	Enkele gevoeligheidsanalyses	40
6.5	Conclusies	46
7	De resultaten van het demografische model voor de toekomst	
7.1	Inleiding	49
7.2	De mogelijke aannames	50
7.3	De vooronderstellingen aan de CBS varianten	51
7.4	De CBS bevolkingsprognoses en modeluitkomsten vergeleken	55
7.5	De middenvariant	56
7.6	De lage variant	59
7.7	De hoge variant	62
7.8	De geboorte- en sterfte-aantallen	65
7.9	De totaalaantallen	68
7.10	Conclusies	70
8	Spaarzame demografische modellen.	
8.1	Inleiding	73
8.2	Het algemene demografische model	74
8.3	De basisversie	75
8.4	De fertiliteitsversie	82
8.5	De 'lineaire uitstroom' versie	84
8.6	De heterogene versie met expliciete geboorte-aantallen	87
8.7	De heterogene versie met fertiliteitscijfers	91
8.8	Conclusies	93
	Literatuurlijst	95

Bijlagen

A De gebruikte CBS-cijfers

A.1	Sterftecijfers	97
A.2	Migratiecijfers	103
A.3	Geboortecijfers	106
A.4	Administratieve correcties	108

B De belangrijkste programmateksten van het uitgebreide demografische model

B.1	Inleiding	109
B.2	De gebruikte symbolen	109
B.3	De indeling in leeftijdsklassen	110
B.4	De bewerking van de mortaliteitscijfers	111
B.5	De bewerking van de migratiecijfers	111
B.6	De bewerking van de geboortecijfers	112
B.7	De bewerking van de administratieve correcties	113
B.8	De berekening van de nieuwe bevolkingsaantallen	113
B.9	De verschillende invoeropties	115

C De belangrijkste programmateksten van het spaarzame demografische model

C.1	Inleiding	117
C.2	De basisversie	117
C.3	De heterogene versie	118
C.4	Het aantal nieuwgeborenen	120
C.5	De uitstroomfractie	121

Summary

A demographic model is described that has been used in scenario-research for calculating future population numbers. The demographic model developed is a discrete Markov-model. It calculates the changes in Netherlands population numbers for each year successively. The model is specially developed to analyse the consequences of alternative assumptions with respect to future death, birth, and migration developments. If necessary the model also makes use of administrative corrections.

The results of the demographic model are compared with the known population numbers for the past. For the future the results are compared with the projections published by the Netherlands Central Bureau of Statistics. The results of this extensive demographic model are also compared with those of some sparse models. These sparse models have a fewer number of transition probabilities between age classes to be estimated.

It is concluded that the more extensive demographic models generally give better results than the sparse models. These sparse models however give rather good results for general approximations when the heterogeneity within the age classes is modelled explicitly. The future population numbers for the oldest age group(s) are generally overestimated, presumably due to the effect of using the registration numbers in the models. For the CBS medium variant the calculated future population numbers differ less from the CBS population forecasts than for the low and high variant.

Samenvatting

Het rapport bevat een beschrijving van een demografisch model, dat toegepast is voor vooruitberekeningen bij scenario-onderzoek. Het ontwikkelde demografisch model vormt een discrete Markov-keten. Voor ieder opeenvolgend jaar worden de veranderingen in de Nederlandse bevolking berekend. Het model is vooral ontworpen om de consequenties van alternatieve aannames voor sterfte-, geboorte- en migratie-ontwikkelingen te analyseren. Behalve met genoemde processen houdt het model ook rekening met eventuele administratieve correcties, die op de CBS gegevens moeten worden toegepast.

De resultaten van het model zijn voor het verleden vergeleken met de geregistreerde aantallen en voor de toekomst met de CBS bevolkingsprognoses. Tevens zijn de resultaten van het model vergeleken met de resultaten van "minder verfijnde" demografische modellen, dat wil zeggen modellen met een kleiner aantal te schatten overgangsfracties tussen de verschillende onderscheiden leeftijdsgroepen.

Geconcludeerd wordt, dat uitgebreide demografische modellen over het algemeen (duidelijk) betere resultaten geven dan de minder verfijnde - spaarzame - modellen. Deze spaarzame modellen voldoen overigens voor globale benaderingen redelijk, indien de heterogeniteit binnen de diverse leeftijdsgroepen expliciet in rekening wordt gebracht. Prognoses voor de omvang van de oudste leeftijdsgroep(en) blijken vrijwel steeds te hoog uit te vallen, vermoedelijk door het effect, dat het gebruik van geregistreerde sterftcijfers op de modelresultaten heeft. Bij scenario-berekeningen blijkt, dat de modelvoorspellingen meer van de CBS prognoses verschillen voor de lage en hoge variant dan voor de middenvariant.

Inleiding.

In dit rapport wordt een verantwoording gegeven van het demografisch model, dat ontwikkeld is ten behoeve van scenario-onderzoek in het kader van de Stuurgroep Toekomstscenario's Gezondheidszorg. Dit model beschrijft de (veranderingen in de) bevolkingsaantallen ten gevolge van geboorte, sterfte, migratie en eventueel administratieve correcties. Deze karakterisering geeft aan, dat toepassing van dit model in plaats van de resultaten van de bevolkingsprognoses van het Centraal Bureau voor de Statistiek (CBS) eerst dan zinvol is bij scenario-onderzoek, indien ontwikkelde scenario's aannames bevatten, die afwijken van de CBS kernhypothesen omtrent veranderingen voor de genoemde demografische grootheden. Het door het CWM ontwikkelde model is derhalve bedoeld om scenario's door te rekenen voor andere dan de door het CBS gehanteerde aannames omtrent geboorte, migratie en/of sterfte.

Hiermee is reeds één belangrijk referentiepunt, anders gezegd: mogelijkheid van validering, genoemd, namelijk het demografisch model van het CBS. Vooruitberekeningen, die met het door het CWM ontwikkelde model gedaan zijn, moeten getoetst (gevalideerd) worden aan de CBS prognoses. Het andere referentiepunt vormen de waargenomen bevolkingsaantallen uit het verleden. Het ontwikkelde model moet vanuit een beginsituatie, in het onderhavige geval het jaar 1955, de bevolkingsaantallen in de daaropvolgende jaren kunnen reproduceren. Beide validaties zijn in het rapport uitvoerig beschreven.

Het onderhavige rapport bevat overigens meer dan slechts een beschrijving van het ontwikkelde demografische model. Het plaatst dit model in een ruimer theoretisch kader. Daartoe wordt aandacht besteed aan o.a. Markov-ketens, maten van demografische gebeurtenissen, mortaliteitsmodellen en mogelijke modelaannames (constante intensiteit en de lineaire hypothese). Daarnaast worden ter vergelijking enkele zogenaamde spaarzame demografische modellen gepresenteerd. Deze spaarzame modellen worden gekenmerkt door een grovere indeling van de bevolking in leeftijdsgroepen en een bijbehorend kleiner aantal mogelijke overgangen tussen de groepen.

In de hoofdstukken 1 t/m 4 wordt vooral op wiskundige aspecten van de modellering ingegaan. De meer in modeltoepassingen geïnteresseerde lezer komt vanaf hoofdstuk 5 aan zijn/haar trekken en kan zich vooral op dat deel van het rapport richten.

Hoofdstuk 1 van het rapport bevat een inleiding in de theorie van Markov-ketens of modellen van toestandsovergangen. Markov-modellen worden gekenmerkt door een vorm van geheugenloosheid: gegeven het heden is de toekomst onafhankelijk van het verleden. Dergelijke modellen zijn geschikt voor het beschrijven van tijdsafhankelijke systemen als de demografie. Hoofdstuk 2 beschrijft een tweede meer statistisch geïoriënteerde ingang naar demografische modellen,

namelijk overlevingsfuncties en sterftetabellen. Deze geven een beschrijving van de leeftijdsafhankelijke sterfte onder een geboortecohort c.q. een bevolkingsgroep. Hoofdstuk 3 combineert de inhoud van de hoofdstukken 1 en 2 tot een beschrijving van zogenaamde meerdimensionale demografische modellen, waarin meer toestandsovergangen zijn opgenomen. Twee aannames, die in hoofdstuk 2 voor mortaliteitsmodellen toegepast zijn, worden ook op meerdimensionale demografische modellen toegepast. Hoofdstuk 4 beschrijft een derde meer empirisch geöriënteerde ingang naar demografische modellen, namelijk demografische maten. Deze maten kunnen empirisch bepaald worden en vervolgens gebruikt om de in hoofdstuk 3 beschreven meerdimensionale demografische modellen 'in te vullen'.

Hoofdstuk 5 bevat een samenvattende beschrijving van (de computer implementatie van) het door het CWM ontwikkelde demografisch model. Aandacht wordt geschonken aan zowel modelmatige als implementatieaspecten. In de hoofdstukken 6 en 7 wordt het model gevalideerd, ofwel getoetst aan de hand van beschikbare gegevens. In hoofdstuk 6 gebeurt deze validatie voor het verleden aan de hand van de door het CBS geregistreerde bevolkingsaantallen, in hoofdstuk 7 voor de toekomst aan de hand van de CBS bevolkingsprognoses. In het eerste geval wordt het model 'opgestart' in het jaar 1955 en berekent het vervolgens de bevolkingsaantallen tot het jaar 1989. In het tweede geval wordt het model opgestart in het jaar 1989 en berekent het voor de lage, midden- en hoge scenario-variant - aannames met betrekking tot sterfte, geboorte en migratie - de bevolkingsaantallen tot het jaar 2010.

In hoofdstuk 8 worden enkele zogenaamde spaarzame demografische modellen gepresenteerd. Deze spaarzame modellen worden gekenmerkt door een indeling van de totale bevolking in een beperkt aantal leeftijdsgroepen in plaats van naar afzonderlijke leeftijdsjaren. De bijbehorende overgangsfracties worden geschat aan de hand van de CBS bevolkingsprognoses. De verschillende versies onderscheiden zich naar de berekeningswijze van de geboorte-aantallen, de directe of vertraagde realisatie van de te schatten overgangsfracties, en de wijze, waarop de afzonderlijke leeftijdsgroepen worden samengesteld.

In de bijlagen worden de demografische modellen expliciet beschreven. In bijlage A worden de invoercijfers gepresenteerd, zoals die door het CBS gepubliceerd en in het model verwerkt zijn. Het gaat hierbij om sterftefracties, migratie-aantallen, geboorte-aantallen en fertiliteitscijfers, en tenslotte administratieve correcties. In bijlage B worden de belangrijkste wiskundige vergelijkingen van het uitgebreide demografische model gegeven. Dit betreffen met name de voorbewerkingen van de invoercijfers, en de voor ieder jaar uitgevoerde berekeningen van de nieuwe bevolkingsaantallen. In bijlage C worden de belangrijkste vergelijkingen van de spaarzame demografische modellen beschreven, voorzover deze afwijken van die van het uitgebreide model.

1 Markov-modellen van toestandsovergangen.

1.1 Inleiding.

Een **stochastisch proces** $\{ \mathbf{x}_t, t \in T \}$ is een verzameling van stochastische variabelen, die gedefinieerd zijn op dezelfde kansruimte. Bijvoorbeeld: de burgerlijke staat, waarin een gegeven individu zich op een reeks achtereenvolgende tijdstippen bevindt, vormt een stochastisch proces. De indexverzameling T wordt opgevat als een verzameling tijdstippen; t wordt de tijdparameter genoemd. In het onderhavige geval van demografische modellen wordt uitgegaan van een continue of discrete tijdparameter t . De toestandsruimte, bestaande uit de mogelijke uitkomsten van de stochastische variabelen, wordt eindelijk verondersteld.

In de volgende paragrafen van dit hoofdstuk wordt eerst een formele definitie van een Markov-keten (met eindige toestandsruimte) gegeven. Hierbij worden het geval van een continue en een discrete tijdparameter parallel behandeld. Vervolgens wordt aandacht besteed aan drie concepten, die in het kader van Markov-modellen een belangrijke rol spelen: stochastische matrices, meerstaps overgangskansen, en intensiteitsmatrices.

1.2 De formele definitie.

De formele definitie van een **Markov-keten** wordt gegeven, parallel voor een discrete en een continue tijdparameter. Een Markov-keten is een stochastisch proces, dat aan de volgende eigenschap voldoet (notatie: I is de toestandsruimte):

Discreet	Continu
$\Pr\{ \mathbf{x}_{n+1}=j \mid \mathbf{x}_n=i, \mathbf{x}_{n-1}=i', \dots \}$ $= \Pr\{ \mathbf{x}_{n+1}=j \mid \mathbf{x}_n=i \}$ $= p_{ij}(n)$	$\Pr\{ \mathbf{x}_t=j \mid \mathbf{x}_s=i, \mathbf{x}_{s'}=i', \dots \}$ $= \Pr\{ \mathbf{x}_t=j \mid \mathbf{x}_s=i \}$ $= p_{ij}(s, t)$

voor alle tijdparameters n voor alle tijdparameters $t > s > s'$

voor alle $j, i, i' \in I$

Anders gezegd: van een Markov-keten $\{ \mathbf{x}_n, n=0, 1, \dots \}$ resp. $\{ \mathbf{x}_t, t \geq 0 \}$ zijn, gegeven de waarden op tijdstip n resp. s , de waarden op tijdstip $n+1$ resp. $t > s$ onafhankelijk van de waarden op de tijdstippen $n-1, n-2, \dots$ resp. $\dots < s' < s$ (geheugenloosheid). Ofwel: in een Markov-keten is, gegeven het heden, de toekomst onafhankelijk van het verleden. Indien bovendien wordt verondersteld, dat de overgangskansen onafhankelijk zijn van n resp. de absolute waarden van s en t , spreken we van een (stabiele of homogene) Markov-keten met stationaire overgangskansen. In dat geval geldt:

Discreet

$$p_{ij}(n) = p_{ij} \quad (= p_{ij}^1)$$

Continu

$$p_{ij}(s, t) = p_{ij}(t-s)$$

In het discrete geval worden de kansen p_{ij} de éénstaps-overgangskansen genoemd (vergelijk de notatie tussen haakjes). In het continue geval is de overgangskans slechts afhankelijk van het verschil van de beschouwde tijdstippen en niet van de tijdstippen zelf.

1.3 Stochastische matrices.

De beschreven (éénstaps-)overgangskansen kunnen in het geval van een eindige toestandsruimte in matrixvorm genoteerd worden. De elementen van de matrix P bevatten de overgangskansen p_{ij} :

$$P = \begin{matrix} p_{11} & p_{12} & p_{13} & p_{14} & \dots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \dots & & \dots \\ p_{31} & \dots & & & & \\ \dots & & & & & \dots \\ p_{n1} & \dots & & \dots & & p_{nn} \end{matrix}$$

Een vierkante matrix P wordt een **stochastische matrix** genoemd, indien:

- $(A:i, j \in I): p_{ij} \in \mathbb{R}, p_{ij} \geq 0$
- $(A:i \in I) \sum (j \in I) p_{ij} = 1$

Duidelijk is, dat de matrix behorende bij de toestandsovergangskansen van een Markov-keten een stochastische matrix is. Omgekeerd kan bij iedere stochastische matrix een Markov-keten geconstrueerd worden.

1.4 De Chapman-Kolmogorov vergelijkingen.

De vergelijkingen van Chapman-Kolmogorov ontlede een overgangskans over verscheidene tijdstappen - een zogenaamde meerstaps-overgangskans - in overgangskansen over kortere tijdsperiodes. Daartoe wordt - in het geval van een discrete tijdparameter - van de volgende notatie gebruik gemaakt:

$$p_{ij}^m = \Pr \{ \mathbf{x}_{n+m} = j \mid \mathbf{x}_n = i \}$$

De vergelijkingen van Chapman-Kolmogorov luiden:

Discreet

Continu

$$p_{ij}^m = \sum_{k \in I} p_{ik}^r * p_{kj}^{m-r}$$

$$p_{ij}(t+s) = \sum_{k \in I} p_{ik}(t) * p_{kj}(s)$$

voor alle $0 \leq r \leq m$ resp. $s, t \geq 0$. Aangenomen wordt, dat voor de nulstapsovergangskansen geldt:

$$p_{ij}^0 = p_{ij}(0) = \delta_{ij} = \begin{cases} 1 & \text{i.g.v. } i=j \\ 0 & \text{i} \neq j \end{cases}$$

De grootte δ wordt de Kronecker deltafunctie genoemd. In de matrixnotatie kunnen deze vergelijkingen geschreven worden als:

Discreet

Continu

$$P^m = P^r P^{m-r}$$

$$P(t+s) = P(t) P(s)$$

met: $P^0 = P(0) = I$, de eenheidsmatrix.

1.5 De toestandsvergangintensiteit.

In het geval van een continue tijdparameter kan het begrip toestandsvergangintensiteit geïntroduceerd worden. De intensiteit van een toestandsvergang is gelijk aan de verandering van de toestandsvergangskans:

$$q_{ij} = \lim_{h \downarrow 0} \frac{p_{ij}(h) - p_{ij}(0)}{h} = \frac{d}{dt} p_{ij}(t)_{t=0} = p'_{ij}(0)$$

Stilzwijgend is er hierbij van uitgegaan, dat de Markov-keten stationair is, ofwel de toestandsvergangskansen stationair zijn. De intensiteiten q_{ij} kunnen, analoog aan de kansen p_{ij} , in matrixvorm genoteerd worden. De elementen van de matrix Q bevatten de intensiteiten q_{ij} (notatie: $q_i = -q_{ii}$):

$$Q = \begin{matrix} -q_1 & q_{12} & q_{13} & q_{14} & \cdots & q_{1n} \\ q_{21} & -q_2 & q_{23} & \cdots & & \cdots \\ q_{31} & \cdots & & & & \\ \cdots & & & & & \cdots \\ q_{n1} & \cdots & & \cdots & & -q_n \end{matrix}$$

Een vierkante matrix Q wordt (in dit verband) een **intensiteitsmatrix** genoemd, indien:

- $(A:i, j \in I): q_{ij} \in \mathbb{R}$
- $(A:i, j \in I): q_{ij} \geq 0 \ (i \neq j) \ , \ q_{ii} \ (= -q_i) \leq 0$
- $(A:i \in I) \ \Sigma(j \in I) \ q_{ij} = 0$

Voor een eindige toestandsruimte I vormt de matrix van de toestandsovergangintensiteiten van een Markov-keten een intensiteitsmatrix. Omgekeerd kan bij iedere intensiteitsmatrix een continue Markov-keten geconstrueerd worden.

Als q_i begrensd is wordt toestand i stabiel genoemd, als q_i onbegrensd is ogenblikkelijk, ofwel: met verwachte verblijfstijd nul. Als $q_i = 0$ wordt toestand i absorberend genoemd, ofwel: met verwachte verblijfstijd oneindig. In geval van een eindige Markov-keten geldt tussen de intensiteiten van de overgangen vanuit een toestand i het volgende verband:

$$q_i = \Sigma(j \neq i) \ q_{ij}$$

De interpretatie hiervan in demografische termen luidt: er verdwijnen geen individuen uit de modelbeschrijving. In de mathematische fysica worden vergelijkbare matrices toegepast, bijvoorbeeld:

Metzler-matrices, gedefinieerd op \mathbb{R} , met elementen buiten de hoofddiagonaal niet-negatief.

Compartimentsmatrices, gedefinieerd op \mathbb{R} , met niet-diagonaal elementen niet-negatief, diagonaal elementen niet-positief en kolomsommen niet-negatief.

Compartimentsmatrices vormen de getransponeerde van intensiteitsmatrices in de ruime zin (met rijssom niet-positief in plaats van nul). De in deze paragraaf gedefinieerde intensiteitsmatrices (met rijssom nul) worden in de literatuur conservatief genoemd. In het vervolg van dit hoofdstuk wordt - vanwege de eindigheid van de toestandsruimte I zonder beperking van de algemeenheid - uitgegaan van conservatieve intensiteitsmatrices.

1.6 De Kolmogorov differentiaalvergelijkingen.

Onder zekere voorwaarden met betrekking tot de toestandsovergangskansen $p_{ij}(t)$ en -intensiteiten q_{ij} kunnen de achterwaartse en voorwaartse differentiaalvergelijkingen van Kolmogorov afgeleid worden (in som- resp. matrixnotatie):

Achterwaartse differentiaalvergelijkingen:

$$p'_{ij}(t) = \sum(j) q_{ik} * p_{kj}(t) \quad P'(t) = Q P(t)$$

Voorwaartse differentiaalvergelijkingen:

$$p'_{ij}(t) = \sum(j) p_{ik}(t) * q_{kj} \quad P'(t) = P(t) Q$$

De Kolmogorov differentiaalvergelijkingen bieden de mogelijkheid om de toestandsovergangskansen van een continue Markov-keten (na transponeren) te schrijven in de vorm van (de fundamenteelvorm van) een beginwaardeprobleem (notatie: I is de eenheidsmatrix):

$$P'(t) = P(t) Q, \quad P(0) = I$$

De matrix $P(t)$ vormt de fundamentealmatrix bij de volgende matrix-differentiaalvergelijking ($x(t)$ rijvector, 1 is de kolomvector met als elementen de getallen 1 resp. de scalair 1):

$$x'(t) = x(t) Q, \quad x(0) 1 = 1$$

Deze herformulering in de vorm van een matrix-differentiaalvergelijking wordt in de volgende paragraaf verder uitgewerkt. Tevens laat deze matrixvorm zien, dat het voor de hand ligt om vectoren van verblijfskansen in rijvorm te noteren.

1.7 De kansovergangsmatrix geschreven als exponentiële functie.

Indien Q een (conservatieve) intensiteitsmatrix is, dan bezit de fundamenteelvorm van de matrix-differentiaalvergelijking :

$$P'(t) = P(t) Q, \quad P(0) = I$$

de volgende (unieke) **oplossing**:

$$P(t) = \exp(Q t) = \sum(i \geq 0) Q^i t^i / i!$$

De oplossing $P(t)$ vormt voor ieder tijdstip t een stochastische matrix. Omgekeerd is het onder zekere voorwaarden mogelijk om van een stochastische matrix de (een) bijbehorende intensiteitsmatrix te bepalen. Het probleem van het omzetten van een stochastische

matrix, die de overgangskansen over een discreet tijdsinterval beschrijft, in een intensiteitsmatrix, die de instantane veranderingen van de overgangskansen beschrijft, wordt aangeduid als het inbeddingsprobleem - Bestaat een dergelijke intensiteitsmatrix? - resp. het identificatieprobleem - Is een gevonden intensiteitsmatrix uniek? - (zie Singer, Spilerman, 1976).

2 Een eenvoudig mortaliteitsmodel.

2.1 Inleiding.

In dit hoofdstuk wordt een eenvoudig mortaliteitsmodel geïntroduceerd. Dit model wordt gebruikt als een voorbereiding voor de in hoofdstuk 3 plaats vindende beschrijving van meerdimensionale demografische modellen. Uitgangspunt vormt de cumulatieve kansverdelingsfunctie van het overlijdenstijdstip van een willekeurig individu uit een bevolkingsgroep. Deze functie beschrijft voor ieder tijdstip de kans, dat het individu niet ouder wordt. De overlevingsfunctie kan geschat worden uit de overlevingscurve van de bevolkingsgroep bestaande uit individuen met identieke en onderling onafhankelijke sterftekansen. Aan de laatste voorwaarde wordt voldaan, indien de groep voldoende homogeen is.

2.2 De overlevingscurve en -functie.

Een (geboorte)cohort, bestaande uit gelijkvormige individuen, wordt in de tijd gevolgd. Voor ieder tijdstip x en daarmee ook de leeftijd wordt het aantal nog levende individuen $n(x)$ genoteerd:

$n(x)$ het aantal nog levende individuen op tijdstip x

De grootte $n(x)$ wordt de **overlevingscurve** genoemd. Uittrede uit het cohort is slechts mogelijk in de vorm van sterfte. Censurering, beëindiging van de waarneming met onbekende reden, wordt onmogelijk verondersteld. Dit aantal $n(x)$ kan eventueel zo geschaald worden, dat het beginaantal $n(0)$ gelijk aan 1 is. In het laatste geval beschrijft de curve $n(x)$ voor iedere leeftijd x de overlevingsfractie.

Naast de gemeten overlevingscurve $n(x)$ wordt een hypothetische overlevingsfunctie $l(x)$ onderscheiden. De overlevingsfunctie vormt het complement van de kansverdelingsfunctie van het overlijdenstijdstip van een willekeurig individu uit het cohort:

$$l(x) = \Pr(I(x)=1)$$

met:

$$I(x) = 1 \quad \text{indien het individu op (leef)tijd } x \text{ nog in leven is}$$

$$0 \quad \text{anders}$$

ofwel:

$$l(x) = \Pr\{ \mathbf{x} > x \}$$

met:

\mathbf{x} het stochastische overlijdenstijdstip van het individu

De overlevingsfunctie vormt een niet-negatieve, niet-stijgende functie van de (leef)tijd x met $l(0)=1$. De indicatorfuncties $I(x)$ voor opeenvolgende tijdstippen x vormen een stochastisch proces (zie paragraaf 1.1). Onder de aanname van onderlinge onafhankelijkheid en identiek gedrag van de individuen is de (geschaalde) overlevingscurve in verwachting gelijk aan de overlevingsfunctie van (willekeurig) elk van de individuen.

In het volgende schema worden de belangrijkste grootheden van de bovengenoemde overlevingscurve en -functie genoemd en zo nodig formeel gedefinieerd:

symbool	omschrijving
x	tijd, of leeftijd of (leef)tijdsinterval $[x, x+1)$
$[x, x+h)$	(leef)tijdsinterval; $[x, x+1)$ wordt ook genoteerd als x
\mathbf{x}	het stochastische overlijdenstijdstip van een willekeurig individu (de vette notatie geeft aan, dat het symbool een stochastische grootheid representeert)
$n(x)$	het gemeten aantal individuen (met) op (leef)tijd x
$I(x)$	de indicatorfunctie (voor het nog in leven zijn) van een willekeurig individu, $\Pr\{ I(x)=1 \} = \Pr\{ \mathbf{x} > x \}$
$l(x)$	de overlevingsfunctie, $l(x) = \Pr(I(x)=1)$
$L(x, x+h)$	de geïntegreerde overlevingsfunctie c.q. het aantal persoonsjaren tussen x en $x+h$, $L(x, x+h) = \int_0^h l(x+s) ds$ notatie: $L(x) = L(x, x+1)$

$d(x)$ afname van de overlevingsfunctie c.q. -curve,

$$d(x) = l(x) - l(x+1)$$

$p(x, x+h)$ de overlevingskans (een voorbeeld van een toestandsovergangskans),

$$\begin{aligned} p(x, x+h) &= \Pr\{ I(x+h)=1 \mid I(x)=1 \} \\ &= \Pr\{ \mathbf{x} > x+h \mid \mathbf{x} > x \} \end{aligned}$$

$h(x)$ de mortaliteitsdruk,

$$h(x) = - l'(x) / l(x) = - \frac{d}{dx} \ln l(x)$$

De overlevingsfunctie $l(x)$ kan geschreven worden als functie van de mortaliteitsdruk $h(x)$, zodat ook de laatstgenoemde grootheid gekozen kan worden als basisgrootheid:

$$l(x) = \Pr(\mathbf{x} > x) = \exp - \int_0^x h(s) ds$$

2.3 Twee voorbeelden van aannames over de sterfte.

In deze paragraaf worden twee voorbeelden van aannames over de optredende sterfte gedurende een tijdsinterval $[x, x+1)$ gegeven en de gevolgen onderzocht: constante mortaliteitsdruk resp. uniforme verdeling van de sterfte. Voor beide aannames kan de overlevingskans geschreven worden als functie van een nog onbekende parameter. Deze parameter kan met behulp van de volgende aannemelijkheidsfunctie geschat worden:

$$L \approx (1-p(x, x+h))^{d(x, x+h)} p(x, x+h)^{n(x) - d(x, x+h)}$$

met:

$p(x, x+h)$ de onbekende overlevingskans als functie van de te schatten parameter

$d(x, x+h)$ de waargenomen sterfte tussen (leef)tijd x en $x+h$

$n(x)$ de omvang van de groep op (leef)tijd x

In het vervolg van deze paragraaf worden de grootheden $p(x, x+h)$, $d(x, x+h)$ en $n(x)$ geschreven als p , d resp. n .

Constante mortaliteitsdruk.

In geval van een constante mortaliteitsdruk kan de overlevingskans p geschreven worden als:

$$p = \exp(-ch)$$

Maximalisering van de aannemelijkheidsfunctie leidt tot de volgende maximum likelihood (ML) schatter van c :

$$\hat{c} = - \ln(1 - d/n) / h$$

Uniforme verdeling van sterfte.

Onder de lineaire hypothese (ofwel: uniforme verdeling van de sterfte, lineaire vorm van de overlevingsfunctie) kan de overlevingskans p geschreven worden als:

$$p = 1 - bh$$

Maximalisering van de aannemelijkheidsfunctie leidt tot de volgende ML-schatter van b :

$$\hat{b} = d / nh$$

2.4 Verschillende sterftematen.

Verschillende grootheden kunnen onderscheiden worden, die een beschrijving geven van de voorwaardelijke relatieve sterfte, d.w.z. van de relatieve verandering van de overlevingsfunctie:

$$q(x) = d(x) / l(x) \quad \text{de sterftekans}$$

$$m(x) = d(x) / L(x) \quad \text{de centrale sterftegraad (centrale maat)}$$

$$h(x) = - l'(x) / l(x) \quad \text{de mortaliteitsdruk}$$

De grootte $m(x)$ wordt ook wel de occurrence-exposure rate genoemd, en is vergelijkbaar met het concept dosis-effect relatie. Onder aanname van een constante mortaliteitsdruk (zie paragraaf 2.3) is de centrale sterftegraad gelijk aan de mortaliteitsdruk:

$$m(x) = c$$

Onder aanname van een lineaire overlevingsfunctie - de lineaire hypothese - is de centrale sterftegraad gelijk aan:

$$m(x) = - b / (1 + bh/2)$$

De schatters van de sterftegraad onder de twee verschillende aannames verschillen slechts in de derde orde.

3 Meerdere dimensionale demografische modellen.

3.1 Inleiding.

In hoofdstuk 2 is een eenvoudig mortaliteitsmodel beschreven, waarin slechts één toestandsovergang mogelijk was, namelijk sterfte. Dit mortaliteitsmodel werd beschreven in de vorm van een geparаметriseerde overlevingsfunctie. Twee voorbeelden van overlevingsfuncties zijn gegeven, namelijk onder de aanname van een constante mortaliteitsdruk resp. een lineaire overlevingsfunctie. In dit hoofdstuk wordt het eenvoudige mortaliteitsmodel uitgebreid tot een meerdimensionaal demografisch model. In plaats van slechts de toestandsovergang sterfte zijn in een dergelijk model ook overgangen in de vorm van geboorte, migratie en veroudering mogelijk. Een dergelijk meerdimensionaal demografisch model wordt ook wel een increment-decrement model of multitoestandsmodel genoemd. De twee genoemde aannames voor het eenvoudige model worden eveneens toegepast op het multidimensionale model.

3.2 De formele beschrijving van het demografisch proces.

Het demografisch proces van het verblijf in en de overgangen tussen toestanden wordt beschreven met behulp van de grootheden $l_i(x)$ en $p_{ij}(x,y)$:

- $l_i(x)$ de kans op verblijf in toestand i op (leef)tijd x
- $p_{ij}(x,y)$ de kans op verblijf in toestand j op (leef)tijd y ,
gegeven verblijf in toestand i op (leef)tijd $x \leq y$
- $q_{ij}(x)$ de overgangsintensiteit tussen de toestanden i en j
op (leef)tijd x

In geval van een aftelbare toestandruimte I kunnen de overgangskansen en -intensiteiten in matrixvorm genoteerd worden en de verblijfskansen in rijvector-vorm (zie paragraaf 1.7).

Onder de Markov-aanname van geheugenloosheid (zie paragraaf 1.2) kunnen de toestandsovergangskansen volgens de **voorwaartse Kolmogorov differentiaalvergelijkingen** geschreven worden als (zie ook paragraaf 1.7):

$$\frac{\delta}{\delta y} p_{ij}(x,y) = \sum(k) p_{ik}(x,y) q_{kj}(y)$$

ofwel in matrixvorm:

$$P_y(x,y) = P(x,y) Q(y)$$

Vanwege de veronderstelde eindigheid van de toestandsruimte I vormt Q een conservatieve intensiteitsmatrix:

$$\sum(j) q_{ij}(x) = 0 \text{ (voor alle toestanden } i \text{ en (leef)tijden } x)$$

Daarnaast kan een **boekhoudkundige gelijkheid** (Engels: accounting equation) opgesteld worden voor de verblijfskansen op twee opeenvolgende tijdstippen en de verschillende optredende toestandsovergangen:

$$l_j(x+h) - l_j(x) = \sum(k \neq j) \{ d_{kj}(x, x+h) - d_{jk}(x, x+h) \}$$

ofwel in matrixvorm:

$$l(x+h) - l(x) = l D(x, x+h)$$

met:

1 de één(rij)vector met als elementen de getallen 1

D(x, x+h) de (verwachte) toestandsovergangen gedurende tijdsinterval [x, x+h),

$$d_{ii}(x, x+h) = - \sum(j \neq i) d_{ij}(x, x+h)$$

De zo geconstrueerde matrix D vormt een intensiteitsmatrix (zie paragraaf 1.5). Dit algemene demografische model wordt onder de twee bekende aannames - constante intensiteiten resp. lineaire verblijfskansen - in de volgende paragraaf uitgewerkt.

3.3 Constante overgangsintensiteiten.

Onder de aanname van constante overgangsintensiteiten worden de voorwaartse Kolmogorov differentiaalvergelijkingen (zie paragraaf 3.1):

$$P'(t) = P(t) Q, \quad P(0) = I$$

waarbij de tijdparameter t staat voor de tijdsduur vanaf een gegeven tijdstip 0, en Q voor de tijdsonafhankelijke intensiteitsmatrix. De oplossing van deze matrix-differentiaalvergelijking luidt:

$$P(t) = \exp^{Qt} = \sum(k \geq 0) Q^k t^k / k!$$

Voor een gegeven tijdsduur h en bekende overgangsaantallen n_{ij} kunnen de bijbehorende overgangskansen $p_{ij}(h)$ als volgt geschat worden:

$$\hat{p}_{ij}(h) = n_{ij} / \sum(j) n_{ij}$$

Het is echter minder eenvoudig om de overgangssintensiteiten te schatten. Hierbij spelen de problemen van inbedding (Engels: embeddability) - Bestaat een intensiteitsmatrix Q, waarvan de e-macht de stochastische matrix P oplevert? - en identificatie - Is een dergelijke intensiteitsmatrix uniek? - (zie paragraaf 1.8) een rol. In de praktijk van de modellering van de demografie blijken de laatstgenoemde problemen overigens meestal slechts van theoretische aard te zijn (Keilman, Gill, 1987).

3.4 Lineaire verblijfskansen.

Onder de lineaire hypothese - verblijfskansen lineair in de tijd - kunnen de centrale maten voor de verschillende toestandsovergangskansen geschreven worden als functie van de verblijfskansen aan het begin en het einde van het betreffende interval:

$$\begin{aligned} m_{ij}(x, x+h) &= d_{ij}(x, x+h) / L_i(x, x+h) \\ &= 2 * d_{ij}(x, x+h) / (l_i(x) + l_i(x+h)) \end{aligned}$$

Gecombineerd met de boekhoudkundige gelijkheid voor de verblijfskansen en de veranderingen daarin (zie paragraaf 3.1) kan de volgende modelvergelijking afgeleid worden:

$$l(x+h) = l(x) A(x, x+h)$$

met:

$$A(x, x+h) = \{ I + h*M(x, x+h)/2 \} \{ I - h*M(x, x+h)/2 \}^{-1}$$

Bij deze modelvergelijking kunnen de volgende kanttekeningen geplaatst worden:

- De matrix A(x, x+h) vormt in geval van relatief veel toestandsveranderingen geen stochastische matrix, namelijk: indien een i bestaat waarvoor $\sum(j \neq i) m_{ij}(x, x+h) > 2/h$.
- De matrix A(x, x+h) vormt onder zekere voorwaarden een zuivere schatter van de kansovergangsmatrix P(x, x+h) (Keilman, Gill, 1987)

De overgangsmatrix A(x, x+h) kan als volgt in de eerste orde benaderd worden (zie ook paragraaf 4.9):

$$A(x, x+h) \approx I + h M(x, x+h)$$

3.5 De Leslie-matrix structuur.

In deze paragraaf wordt de specifieke structuur beschreven van de intensiteitsmatrix Q (zie de paragrafen 1.5 en 3.2) en de stochastische overgangsmatrix A (zie de paragrafen 1.3 en 3.3) voor een eenvoudig demografische model. Dit model beschrijft de veranderingen van een cohort in termen van veroudering en sterfte. Als mogelijke toestanden worden gekozen het verblijf in een reeks opeenvolgende leeftijdsklassen en de toestand 'overleden'. De volgende (éénstaps) overgangen tussen toestanden worden onderscheiden:

Veroudering

De overgang van een leeftijdsklasse naar de eerstvolgende. De bijbehorende overgangsfractie is de verouderingsfractie: het deel van de leeftijdsgroep, dat overgaat.

Sterfte

De overgang van een leeftijdsklasse naar de toestand 'overleden' met als overgangsfractie de mortaliteitsfractie: het deel van de leeftijdsgroep, dat overlijdt.

In het geval van een toestandsovergangsmatrix (over een discreet tijdsinterval) wordt de tijdsstap zo klein gekozen, dat veroudering slechts optreedt in de vorm van de overgang naar de eerstvolgende klasse en niet naar een hogere. De structuur van de intensiteitsmatrix en de toestandsovergangsmatrix is weergegeven in figuur 1, uitgaande van (bijvoorbeeld) vier leeftijdsklassen.

Figuur 1: een toestandsovergangs(intensiteits)matrix.

		leeftijdsklassen				overleden
		1	2	3	4	
leeftijds- klassen	1	x	x			x
	2		x	x		x
	3			x	x	x
	4				x	x
overleden						x

De vakken met symbool 'x' beschrijven mogelijke toestandsovergangen, de overige zijn onmogelijk. De mogelijke overgangen betreffen het blijven in een leeftijdsklasse (hoofddiagonaal), het overgaan naar de eerstvolgende klasse ten gevolge van veroudering (nevendiagonaal) en het overgaan naar de klasse 'overleden' ten gevolge van sterfte (laatste kolom). Een dergelijke matrix wordt aangeduid met de term **Leslie-matrix**, waarbij in het onderhavige geval afgezien is van geboorte (Flipse, Veling, 1984). De toestand 'overleden' vormt een zogenaamde absorberende toestand. Hieruit zijn geen overgangen naar andere toestanden mogelijk.

In het geval van een intensiteitsmatrix bevatten de vakken overgangssintensiteiten. Op de hoofddiagonaal zijn de waarden van de intensiteiten negatief, voor de overige mogelijke overgangen - aangekruiste vakken - positief, en voor de onmogelijke overgangen - niet-aangekruiste vakken - gelijk aan 0. Voor iedere rij sommeren de intensiteiten tot de waarde 0. In het geval van een (stochastische) overgangsmatrix zijn alle overgangsfracties niet-negatief. Voor de onmogelijke overgangen zijn de fracties gelijk aan 0. Voor iedere rij sommeren de fracties tot de waarde 1.

Bij het bovenstaande moet een aantal **kanttekeningen** geplaatst worden. **Ten eerste:** zowel de stochastische matrix P onder de aanname van constante overgangssintensiteiten (paragraaf 3.2) als de overgangsmatrix A onder de lineaire hypothese (paragraaf 3.3) voldoen slechts bij benadering aan de Leslie-structuur. De overgangsfracties voor de onmogelijke overgangen zijn namelijk van de orde 'tijdsstap kwadraat' of hoger. Bij een voldoende kleine tijdsstap zijn deze fracties verwaarloosbaar klein in vergelijking met de fracties voor de 'mogelijke toestandsovergangen'.

Ten tweede: onderscheid moet gemaakt worden naar (theoretische) overgangskansen en (gemeten) overgangsfracties. Aan de hand van gemeten overgangsfracties en voor zekere modelaanname worden de theoretische overgangskansen bepaald. Mogelijke modelaanname zijn: in de (continue) tijd constante overgangssintensiteiten (zie paragraaf 3.2), lineaire verblijfskansen (zie paragraaf 3.3), en een discreet Markov-model (zie hoofdstuk 1). In het geval van constante overgangssintensiteiten zijn voor kleine tijdstappen ook de kansen voor de onmogelijke overgangen naar hogere klassen dan de eerstvolgende positief, in tegenstelling tot de gemeten fracties. In het geval van een discreet model zijn de éénstaps overgangskansen gelijk aan de gemeten éénstaps overgangsfracties. Indien éénstaps overgangskansen geschat moeten worden met behulp van gemeten meerstaps overgangsfracties moet uit theoretisch oogpunt rekening gehouden worden met de problemen van inbedding en identificatie (zie paragraaf 1.8). In de praktijk van demografische modellen zijn de éénstaps overgangskansen echter eenduidig.

Ten derde: in de literatuur wordt de Leslie-matrix meestal gebruikt om de veranderingen in een gesloten populatie te beschrijven, waarbij ook de geboorte van nieuwe individuen mogelijk is. Deze geboorte wordt beschreven in de vorm van niet-negatieve getallen in de eerste rij van de matrix. Deze getallen worden geïnterpreteerd

als fertiliteitscijfers: het aantal nieuwgeborenen als fractie van het aantal individuen in de betreffende leeftijdsklasse. Deze uitgebreide Leslie-matrix vormt echter geen intensiteits- of stochastische matrix meer. Immers, de som van de cijfers in de eerste rij hoeft niet gelijk aan 0 resp. 1 te zijn. De stabiele leeftijdsverdeling van de betreffende populatie kan bepaald worden met behulp van een eigenwaarden-analyse van de betreffende Leslie-matrix (Flipse, Veling, 1984).

4 Maten van demografische gebeurtenissen.

4.1 Inleiding.

In de voorgaande hoofdstukken zijn verschillende maten voor demografische gebeurtenissen geïntroduceerd: de intensiteit, de kans, de occurrence-exposure rate, en de centrale maat. Daarnaast zijn verschillende demografische modellen beschreven, die geformuleerd zijn in termen van overgangssintensiteiten en verblijfskansen. Onder de aanname van constante overgangssintensiteiten worden de overgangskansen beschreven als functie van deze intensiteiten (zie paragraaf 3.2). Onder de lineaire hypothese worden deze kansen geschreven als functie van de centrale maten (zie paragraaf 3.3).

In dit hoofdstuk wordt eerst een beschrijving gegeven van verschillende mogelijke maten van demografische gebeurtenissen, vergelijkbaar met de centrale maat. Vervolgens wordt aangegeven, hoe de wiskundige vergelijkingen van de meerdimensionale demografische modellen met gebruik van genoemde demografische maten geschreven worden.

4.2 Herhaalbare en verstorende gebeurtenissen.

In de demografie wordt onderscheid gemaakt tussen herhaalbare en niet-herhaalbare gebeurtenissen. Bij een herhaalbare gebeurtenis blijft het individu tot de risicogroep (bijvoorbeeld een leeftijds-klasse) behoren. Een voorbeeld van een herhaalbare gebeurtenis is voor een vrouw het krijgen van een kind, indien vrouwen niet onderverdeeld worden naar het aantal kinderen, dat zij voortgebracht hebben. Is dat wel het geval, dan vormt het krijgen van een kind een niet-herhaalbare gebeurtenis. Een niet-herhaalbare gebeurtenis heeft tot gevolg, dat een individu de risicogroep verlaat. Een ander voorbeeld van een niet-herhaalbare gebeurtenis is sterfte. Herhaalbare gebeurtenissen worden ook wel aangeduid als niet-autopreventief, niet-herhaalbare als autopreventief.

Verstorende gebeurtenissen, zijn gebeurtenissen, anders dan de beschouwde, die van invloed zijn op het aantal individuen. In de overlevingsduuranalyse worden veelal (mogelijke) verstorende gebeurtenissen, die leiden tot onttrekking aan de onderzoeksgroep, samengevoegd onder de term censurering. Indien voor vrouwen in de vruchtbare leeftijd het krijgen van een kind als te onderzoeken gebeurtenis gekozen wordt, vormen (bijvoorbeeld) sterfte en migratie verstorende gebeurtenissen. Deze verstorende gebeurtenissen zijn van invloed op de omvang van het risico in termen van aantal personen of aantal persoonsjaren.

4.3 Het Lexis-diagram.

Bij de beschrijving van het verloop van een cohort is ervan uitgegaan, dat de waarnemingstijd en de leeftijd parallel lopen, anders gezegd: dat de geboortetijdstippen constant gehouden worden. Dit vormt een voorbeeld van de volgende gelijkheid tussen de grootheden geboortetijdstip, leeftijd en waarnemingstijdstip:

$$\text{geboortetijdstip} + \text{leeftijd} = \text{waarnemingstijdstip}$$

Een Lexis-diagram (**figuur 1**) kan worden gebruikt om de bevolking te beschrijven in termen van de drie genoemde grootheden (zie ook Hoogenveen en Brouwer, 1989).

Een **verticale lijn** in het diagram geeft een dwarsdoorsnede (beschrijving van de momentane samenstelling) van de bevolking voor een gegeven waarnemingstijdstip. Een **horizontale lijn** in het diagram geeft een vergelijking van individuen weer met gelijke leeftijd, maar met verschillend geboortetijdstip. Bijvoorbeeld: de opmerking 'de sterfte is gedurende de laatste decennia afgenomen' betreft een horizontale vergelijking, uitgevoerd voor één of meer leeftijdsklassen. Een **diagonale lijn** (van linksonder naar rechtsboven) geeft het verloop van een groep personen weer met constant geboortetijdstip, maar met eventueel instroom in de vorm van immigratie, gedurende een zekere waarnemingsperiode. Een variatie op de diagonale lijn is het concept **cohort**: een vaste groep personen, zonder instroom in de vorm van immigratie, maar wel met een mogelijke uitstroom in de vorm van sterfte en censurering, wordt gedurende een zekere waarnemingsperiode gevolgd.

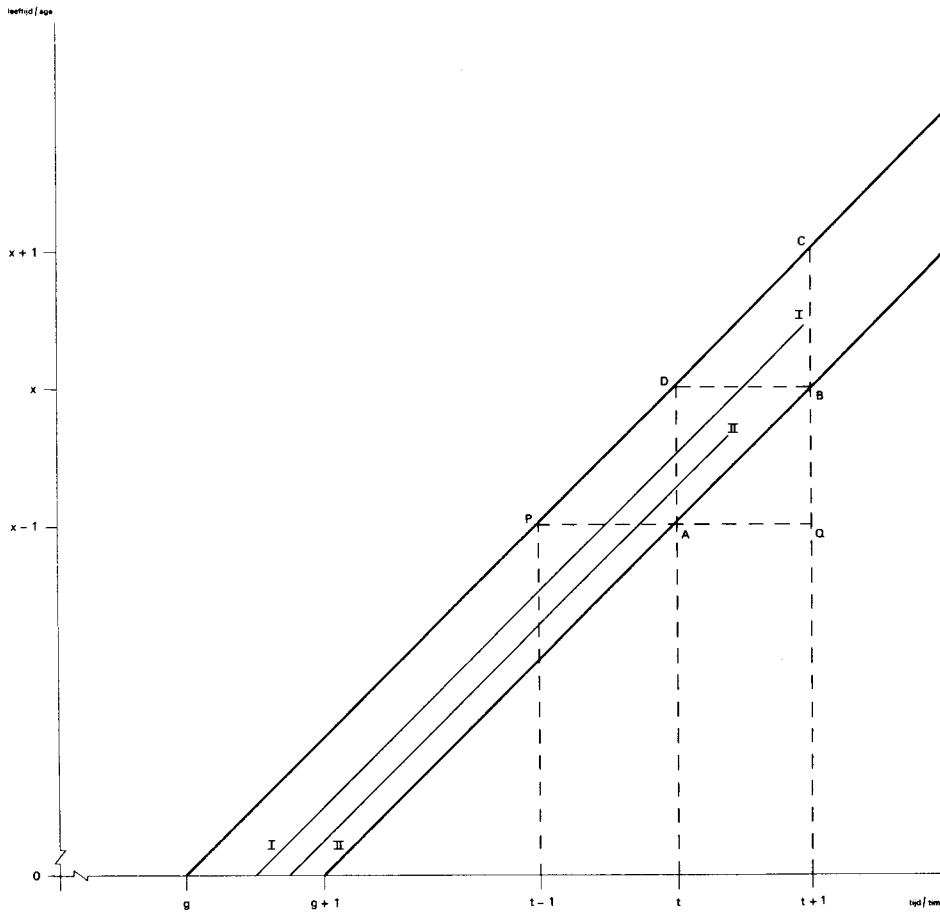
Indien een waarnemingstijd t samenvalt met het begin van een kalenderjaar, dan wordt de periode $[t, t+1)$ meestal aangeduid als het kalenderjaar t . Analoog wordt gesproken van een geboortejahr g in plaats van het interval $[g, g+1)$ en van een leeftijd x in plaats van het leeftijdsinterval $[x, x+1)$. De twee bovengenoemde concepten, een lijn in het Lexis diagram als beschrijving van een vorm van waarneming en het identificeren van intervallen met puntelementen, komen samen in de volgende drie mogelijkheden van waarneming:

Periode waarneming

Hierbij worden de demografische gebeurtenissen vastgelegd voor gegeven kalenderjaar en leeftijd. In het diagram komt dit overeen met het vierkant AQBD.

Cohort waarneming

Hierbij worden van individuen met gelijk geboortejahr en voor gegeven leeftijdsjaar de gebeurtenissen vastgelegd. De waarnemingsperiode betreft in dit geval twee kalenderjaren. In het diagram komt dit overeen met parallellogram ABDP.



Bron: CBS, 1984

Figuur 1: Het Lexis-diagram.

Periode-cohort waarneming

Hierbij worden van individuen met gelijk geboortjaar gedurende een kalenderjaar de gebeurtenissen vastgelegd. De leeftijden van de individuen beslaan in dit geval twee leeftijdsjaren. In het diagram komt dit overeen met parallellogram ABCD.

4.4 Demografische maten.

In de volgende paragrafen wordt een aantal grootheden beschreven, die een maat vormen voor het optreden van demografische gebeurtenissen. Een dergelijke maat kan geïnterpreteerd worden als een risicomat en wordt geschreven als een breuk. De teller van de breuk wordt gevormd door het aantal optredende gebeurtenissen; de noemer door de omvang van het **risico** op het optreden ervan. Het risico kan bestaan uit het aantal individuen, dat risico loopt, of het aantal persoonsjaren, gedurende welke individuen risico lopen.

Om de verschillende maten formeel te kunnen definiëren worden de twee volgende grootheden geïntroduceerd (vergelijk het Lexis-diagram, figuur 1):

$l(x,t)$ het aantal individuen met (continue) leeftijd x op (continue) waarnemingstijdstip t

$g(x,t)$ het aantal gebeurtenissen, dat optreedt, voor individuen met leeftijd x op tijdstip t

4.5 De centrale maat.

De centrale maat beschrijft het optreden van een gebeurtenis in relatie tot een risico in termen van aantallen persoonsjaren. De centrale maat kan gedefinieerd worden voor elk van de drie genoemde typen waarneming. Ofwel: met betrekking tot (zie Lexis diagram):

periode waarneming $c = \int_{ABQD} g(x,t) dxdt / \int l(x,t) dxdt$

cohort waarneming $\int_{ABDP} g(x,t) dxdt / \int l(x,t) dxdt$

periode-cohort waarneming $\int_{ABCD} g(x,t) dxdt / \int l(x,t) dxdt$

In het vervolg wordt uitgegaan van periode-cohort waarnemingen. De eenheid van de centrale maat is: het aantal gebeurtenissen per persoonsjaar. Onder aanname van een lineair verloop van het aantal individuen - de zogenaamde **lineaire hypothese** - kan de noemer van de centrale maat, het aantal persoonsjaren onder risico, geschreven worden als:

$$\int l(x,t) dxdt = \frac{l_0 + l_1}{2} h$$

met:

l_0 het aantal individuen aan het begin van de waarnemingsperiode

$$= \int_x l(x, t_0) dx$$

l_1 het aantal individuen aan het einde van de waarnemingsperiode

$$= \int_x l(x, t_1) dx$$

h de waarnemingsduur

$$= t_1 - t_0$$

De centrale maat beschrijft het optreden van een gebeurtenis voor een gegeven risico. Deze maat corrigeert echter niet voor het optreden van versturende gebeurtenissen. Een voorbeeld hiervan kan gegeven worden voor concurrerende doodsoorzaken. De centrale maat voor de sterfte aan hart- en vaatziekten houdt geen rekening met de sterfte aan andere doodsoorzaken (zie Hoogenveen, Jager, 1990). Daarom wordt de zogenaamde onafhankelijke maat geïntroduceerd (zie de volgende paragraaf).

4.6 De onafhankelijke maat.

De onafhankelijke maat kan uit de centrale maat afgeleid worden onder de lineaire hypothese (zie ook paragraaf 3.3): het verloop van de bevolking is lineair. Het aantal optredende gebeurtenissen wordt aangeduid met $n(G)$. De onafhankelijke maat wordt gedefinieerd als:

i.g.v. herhaalbare gebeurtenissen

$$a = \frac{n(G)}{(\bar{l}_0 + \bar{l}_1)/2}$$

i.g.v. niet-herhaalbare gebeurtenissen

$$a = \frac{n(G)}{(\bar{l}_0 + \bar{l}_1 + n(G))/2}$$

De eenheid van de onafhankelijke maat is: het aantal gebeurtenissen per persoon. Het volgende verband tussen de centrale maat en de onafhankelijke maat kan afgeleid worden:

i.g.v. herhaalbare gebeurtenissen

$$a = h * c$$

i.g.v. niet-herhaalbare gebeurtenissen

$$a = \frac{2 * h * c}{2 + h * c}$$

4.7 De afhankelijke maat.

De afhankelijke maat wordt gedefinieerd als:

$$a^* = \frac{n(G)}{l_0}$$

De eenheid van de afhankelijke maat is identiek aan die van de onafhankelijke maat: het aantal gebeurtenissen per persoon. De afhankelijke maat is identiek aan de onafhankelijke maat, indien aan de volgende voorwaarden voldaan wordt:

i.g.v. herhaalbare gebeurtenissen

$$l_1 = l_0$$

i.g.v. niet-herhaalbare gebeurtenissen

$$l_1 = l_0 - n(G)$$

Deze voorwaarden kunnen geïnterpreteerd worden als: er vinden (netto) geen verstorende gebeurtenissen plaats.

4.8 De semi-afhankelijke maat.

Bij de semi-afhankelijke maat, zoals die in de demografie wordt toegepast, wordt onderscheid gemaakt tussen de verschillende verstorende gebeurtenissen. Eén groep van verstorende gebeurtenissen wordt wél meegenomen bij de berekening van de demografische maat; de overige gebeurtenissen niet. De semi-afhankelijke maat wordt gedefinieerd als:

$$a^+ = \frac{n(G)}{l_0 + \frac{n(G')}{2}}$$

met:

$n(G')$ het (netto) saldo van de wél meegenomen verstorende gebeurtenissen

Bijvoorbeeld: indien het risico van echtscheiding onder gehuwde mannen gemeten wordt, kan gekozen worden voor wél een correctie voor migratie als verstorende factor, maar niet voor huwelijks-sluiting, sterfte en verweduwing.

4.9 De meerdimensionale demografische modellen met gebruik van de demografische maten.

In paragraaf 3.3 is het meerdimensionale stochastische demografische model beschreven onder de lineaire hypothese: de verblijfskansen worden lineair in de tijd verondersteld. Deze aanname komt overeen met die, welke aan de definiëring van de belangrijkste demografische maten ten grondslag ligt: het aantal individuen wordt lineair verondersteld. De hierna te beschrijven meerdimensionale demografische modellen gaan uit van de **lineaire hypothese**.

In het algemeen kan de (verandering van de) bevolkingsomvang geschreven worden in de vorm van de volgende vectoriële vergelijking:

continue tijdparameter

$$D l(x) = l(x) M(x) + i'(x)$$

discrete tijdparameter

$$l(x+h) = l(x) A(x, x+h) + i(x, x+h) B(x, x+h)$$

met:

- $l(x)$ de bevolkingsomvang op tijdstip x (een rijvector)
- $M(x)$ de overgangsintensiteitsmatrix op tijdstip x
- $A(x, x+h)$ de toestandsovergangsmatrix voor het tijdsinterval $[x, x+h)$
- $i'(x)$ de migratie-intensiteit op tijdstip x
- $i(x, x+h)$ het migratie-saldo over het tijdsinterval $[x, x+h)$
- $B(x, x+h)$ de zogenaamde ingangsmatrix

Deze demografische vergelijkingen lijken sterk op de vergelijkingen uit de wiskundige **systemtheorie**, die de (veranderingen van de) waarden van de toestandsvariabelen van een systeem beschrijft in de loop van de tijd in de aanwezigheid van invoer.

De overgangsmatrix A en ingangsmatrix B kunnen als volgt als functie van de gedefinieerde demografische maten geschreven worden:

overgangsmatrix A	ingangsmatrix B
centrale maat	
$(I + M/2) (I - M/2)^{-1}$	$(A + I)/2$
onafhankelijke maat	
$(I + \bar{M}/2) (I - \bar{M}/2)^{-1}$	$(A + I)/2$
afhankelijke maat	
$I + M$	0
semi-afhankelijke maat	
$I + M$	$(A + I)/2$

met:

$M = M(x, x+1)$ de intensiteitsmatrix, die de betreffende (niet-getransformeerde) demografische maten bevat

\bar{M} de intensiteitsmatrix met als elementen de onafhankelijke maten getransformeerd tot centrale maten

5 Het geïmplementeerde uitgebreide demografische model.

5.1 Inleiding.

Op basis van de theorie van hoofdstuk 4 is een demografisch model ontwikkeld en geïmplementeerd. Dit door het CWM ontwikkelde model is bedoeld om demografische scenario's door te kunnen rekenen, waarbij aan de aannames, die in de vorm van de kernhypothesen aan de CBS bevolkingsprognoses ten grondslag liggen, wordt gesleuteld (zie ook de Inleiding). De genoemde aannames betreffen met name in scenario's veronderstelde veranderingen in de sterfte, die van invloed zijn op de toekomstige bevolkingsomvang.

Het ontwikkelde model is eenvoudiger van opzet dan het CBS model. De burgerlijke staat vormt er geen onderdeel van. Het model maakt gebruik van de beschikbare CBS cijfers betreffende de sterfte, migratie en geboorte, en eventueel de administratieve correcties. De verschillende overgangsfracties - geboorte, sterfte en migratie - en de beginsamenstelling van de bevolking vormen een gegeven en worden niet geschat (zie echter ook hoofdstuk 7). In de volgende paragrafen wordt het model meer uitvoerig beschreven. In hoofdstuk 8 wordt dit uitgebreide demografische model vergeleken met modellen, waarin de bevolking in een kleiner aantal leeftijdsgroepen verdeeld is.

5.2 Algemene kenmerken.

Het uitgebreide demografische model bezit de volgende kenmerken:

De modeltijd

Het model rekent met discrete tijdstappen van 1 jaar. De bevolkingsaantallen worden telkens gegeven per 1 januari van ieder jaar. Het model kan de berekeningen in principe vanaf elk jaar tussen 1955 en 1988 starten en, onder zekere aannames over de verschillende modelfracties na 1988, in ieder toekomstig jaar beëindigen. De betrouwbaarheid van de uitkomsten neemt echter af met de tijd.

Deterministisch

Gegeven de beginsamenstelling van de bevolking en de waarden van de overgangsfracties ligt de samenstelling in de daarop volgende jaren vast.

Verdeling naar geslacht en in leeftijdsjaren

Bij de berekeningen van de demografische veranderingen wordt uitgegaan van de bevolking verdeeld naar geslacht en in afzonderlijke leeftijdsjaren.

Gebruik van fracties

De jaarlijkse sterfte, migratie en geboorte worden berekend met gebruik van CBS sterftefracties, migratie-aantallen en fertilitateitscijfers of geboorte-aantallen (zie bijlage A). De verschillende cijfers zijn voorzover mogelijk gespecificeerd naar geslacht en in vijfjaars leeftijdsklassen.

Aannames over trends

Vanaf het laatste jaar, waarvoor CBS cijfers beschikbaar zijn, in het actuele geval 1988, moeten voor de genoemde overgangsfracties vanaf 1989 veronderstellingen gedaan worden. Deze veronderstellingen kunnen bestaan uit trends op basis van de laatst beschikbare cijfers of uit door de gebruiker op te geven waarden. Het is bijvoorbeeld mogelijk om de overgangsfracties constant te houden of een trend te dempen c.q. bij te stellen.

Hard- en software

Het model is geïmplementeerd in de vorm van een computerprogramma, geschreven in de programmeertaal Turbo Pascal (Borland), en kan gedraaid worden op een IBM-compatible PC. Het model is modulair opgebouwd met behulp van zogenaamde including files.

5.3 Het demografische model.

Het demografische model bestaat uit een beschrijving van de jaarlijkse veranderingen van de samenstelling van de Nederlandse bevolking in de loop van de tijd. Uitgegaan wordt van de samenstelling in een zeker beginjaar. Voor elk van de daarop volgende jaren worden de jaarlijkse veranderingen in de samenstelling bepaald, waaruit de nieuwe samenstelling volgt. Veranderingen in de samenstelling vinden plaats ten gevolge van veroudering, sterfte, migratie en geboorte. Daarnaast kunnen in navolging van het CBS administratieve correcties op de berekende bevolkingsaantallen uitgevoerd worden.

De samenstelling van de bevolking wordt gegeven in de vorm van aantallen individuen verdeeld naar geslacht en naar afzonderlijke leeftijdsjaar, per 1 januari van ieder jaar. De veranderingen in de samenstelling in de loop van het jaar worden geacht instantaan plaats te vinden. Met behulp van gegeven éénjaarlijkse fracties wordt de samenstelling van de bevolking namelijk aan het einde van

het jaar en daarmee aan het begin van het volgende jaar bepaald. De toegepaste vergelijkingen gaan uit van afhankelijke maten (zie de paragrafen 4.7, 4.9 en bijlage B).

Op de volgende manieren verandert de samenstelling van de bevolking:

- Veroudering** De overgang van een leeftijdsjaar naar het eerstvolgende.
- Sterfte** Uittrede uit een bevolkingsgroep.
- Migratie** Intrede in (immigratie) en uittrede uit (emigratie) een bevolkingsgroep.
- Geboorte** Intrede in de groep met leeftijdsjaar 0.

Administratieve correctie (eventueel)

Een correctie van de door het CBS berekende bevolkingsaantallen aan de hand van de door de gemeenten opgegeven aantallen.

Er wordt geen expliciet verband geformuleerd tussen de sterfte-, migratie- en geboorte-fracties. In de volgende paragraaf worden de genoemde demografische veranderingen toegelicht. Beschreven wordt, hoe de CBS cijfers in de modelvergelijkingen ingepast zijn.

5.4 De beschreven demografische veranderingen.

In deze paragraaf worden de in paragraaf 5.3 genoemde demografische veranderingen meer uitvoerig beschreven.

Veroudering.

Veroudering houdt voor het demografisch model in, dat individuen gedurende een jaar van een leeftijdsjaar overgaan naar het volgende. In het geval van leeftijdsklassen, die verscheidene leeftijdsjaren omvatten gaat bij veroudering ieder jaar slechts het oudste jaar van een leeftijdsgroep over naar de volgende klasse. De in hoofdstuk 8 beschreven spaarzame demografische modellen maken gebruik van leeftijdsklassen.

Een probleem bij demografische modellen, die gebruik maken van meerjaars leeftijdsklassen, is het bepalen van het deel van iedere klasse, dat ieder jaar ten gevolge van veroudering overgaat naar de volgende klasse. De eenvoudigste oplossing is aan te nemen, dat de klasse homogeen is. Dit wil zeggen, dat de verdeling van de individuen van de klasse over de (continue) leeftijd homogeen is. In dat geval is de overgangsfractie omgekeerd evenredig aan de breedte van

de klasse - het aantal jaren, dat de klasse omvat. Deze homogeniteitsaannname is toegepast bij een aantal demografische modellen, beschreven in hoofdstuk 8. Daarnaast zijn modellen gepresenteerd, waarbij een (lineaire) vorm van heterogeniteit verondersteld is.

De genoemde aanname van homogeniteit van de klassen is echter discutabel voor de Nederlandse situatie, waarin bijvoorbeeld sprake is van in de loop van de tijd veranderende geboorte-aantallen. Hierbij moet gedacht worden aan de geboortegolf van direct na de tweede wereldoorlog en de daling van het aantal geboorten vanaf de jaren zeventig tot halverwege de jaren tachtig. Het onterecht toepassen van de aanname van homogeniteit leidt ertoe, dat deze 'hobbel' in de bevolkingsfrequentie voor een leeftijdsklasse in de loop van de simulatietijd versneld wordt uitgesmeerd over de hogere klassen. Hoe groter de klassebreedte des te erger het euvel kan worden.

Sterfte.

Sterfte leidt tot een uittrede van individuen uit de verschillende bevolkingsgroepen. Bij de toegepaste sterftefracties wordt gebruik gemaakt van CBS cijfers (zie bijlage A.1). Het CBS publiceert (in dit verband) verschillende soorten sterftecijfers: absolute aantallen en (relatieve) fracties, gespecificeerd naar geslacht en vijfjaars leeftijdsklassen. De absolute aantallen zijn gelijk aan het aantal individuen, dat in een jaar sterft op een leeftijd behorend bij de betreffende klasse. De sterftefracties zijn gelijk aan het quotiënt van de absolute aantallen en de omvang van de klasse aan het begin c.q. gedurende het jaar (zie ook hoofdstuk 6).

Het geïmplementeerde demografische model maakt gebruik van sterftefracties, en past deze toe op de bevolkingsaantallen aan het begin van ieder jaar. Deze fracties zijn tot het jaar 1980 gelijk aan de door het CBS gepresenteerde cijfers. Vanaf 1980 worden de fracties berekend uit de gepresenteerde absolute sterfte- en de bevolkingsaantallen.

Migratie.

Migratie leidt tot een intrede in (immigratie) en uittrede uit (emigratie) de verschillende bevolkingsgroepen. Bij de berekening van de migratiefracties wordt gebruik gemaakt van CBS cijfers (zie bijlage A.2). Het CBS publiceert absolute aantallen, die in het onderhavige demografische model omgezet worden in fracties. Deze fracties worden bepaald door (voor de verschillende leeftijdsklassen) de absolute migratie-aantallen te delen door de berekende bevolkingsaantallen.

Geboorte.

Geboorte leidt tot een jaarlijkse intrede in de eerste (jongste) leeftijdsklasse. Hierbij wordt weer gebruik gemaakt CBS cijfers (zie bijlage A.3). Bij de berekening van het aantal nieuwgeborenen kan gebruik worden gemaakt van twee soorten cijfers: absolute geboorte-aantallen en fertiliteitscijfers. In het tweede geval

wordt het aantal nieuwgeborenen berekend als (de som over de verschillende leeftijdsklassen van) het product van het aantal vrouwen met de fertiliteit (het aantal nieuwgeborenen per 1000 vrouwen). Het model maakt gebruik van beide cijfers. Tot 1967 worden de direct gegeven geboorte-aantallen gebruikt, vanaf 1967 - het eerste jaar, waarvoor fertiliteitscijfers door het CBS gepubliceerd zijn - fertiliteitscijfers.

Administratieve correcties.

Het CBS registreert de jaarlijkse sterfte-, migratie- en geboorte-aantallen, die leiden tot veranderingen in de bevolkingsaantallen (zie bijlage A.4). Gegeven de samenstelling van de bevolking aan het begin van het jaar worden met behulp van deze veranderingen de bevolkingsaantallen aan het einde van het jaar en daarmee aan het begin van het nieuwe jaar bepaald. Vervolgens worden de laatstgenoemde aantallen vergeleken met de door de gemeentes opgegeven aantallen. Tussen deze aantallen kunnen verschillen bestaan, aan de hand waarvan het CBS zijn cijfers corrigeert.

Over het algemeen zijn deze verschillen klein, maar stijgen vanaf de jaren zeventig en met name vanaf halverwege de jaren tachtig. Het CBS publiceert de totale omvang van de correcties voor ieder jaar, maar slechts vanaf 1985 verdeeld naar leeftijd en geslacht. Deze verdeling is tot 1988 vrijwel constant. In de basisversie van het geïmplementeerde demografische model zijn de correcties wegge laten. Daarnaast is een variant doorgerekend, waarin deze correcties wél toegepast worden.

5.5 De structuur van het programma.

In deze paragraaf wordt de structuur van het hoofdprogramma beschreven, dat de implementatie vormt van het bovenbeschreven demografisch model. Het hoofdprogramma bestaat uit de volgende onderdelen:

Declaraties

De declaratie van constanten, typen, variabelen, procedures en functies.

Initialisaties

Het inlezen en zo nodig bewerken van de invoercijfers.

De berekeningen voor ieder simulatiejaar

Op basis van de bevolkingssamenstelling aan het begin van ieder jaar en de verschillende overgangsfracties wordt de samenstelling aan het begin van het volgend jaar berekend.

De genoemde onderdelen van het hoofdprogramma worden in de volgende paragraaf verder uitgewerkt.

5.6 De onderdelen van het hoofdprogramma toegelicht.

In deze paragraaf worden de in paragraaf 5.5 genoemde onderdelen van het hoofdprogramma toegelicht.

De declaraties.

In het declaratie gedeelte van het programma worden de toegepaste constanten, variabele-typen, variabelen (Engels: instances), procedures en functies gedeclareerd. Deze declaraties kunnen onderscheiden worden in specifieke en algemene. De specifieke declaraties hebben met name betrekking op de variabelen, die een specifieke simulatie van het demografisch verloop van de Nederlandse bevolking karakteriseren. Tot de specifieke variabelen behoren het beginjaar van de simulatie en de simulatie-periode. De overige declaraties zijn algemeen van aard.

De initialisaties.

De volgende initialisaties kunnen onderscheiden worden. Ten eerste: het construeren van de verschillende toegepaste indelingen in leeftijdsklassen. Deze indelingen worden vastgelegd door voor iedere indeling voor iedere klasse het eerste jaar, dat tot deze klasse behoort, op te geven.

Ten tweede: het inlezen en eventueel bewerken van de CBS cijfers, die de demografische veranderingen beschrijven. Deze cijfers zijn opgeslagen in verschillende tekstfiles en worden na het inlezen tijdelijk opgeslagen in een kleiner aantal gestructureerde (geformatteerde) files.

Ten derde: het inlezen uit twee tekstfiles van de samenstelling van de bevolking voor mannen resp. vrouwen in het beginjaar van de simulatie.

De berekeningen voor ieder simulatiejaar.

De berekeningen voor ieder simulatiejaar kunnen in twee onderdelen verdeeld worden. Ten eerste: het inlezen van de (bewerkte) gegevens uit de invoerfiles, die bij de initialisaties geconstrueerd zijn. Ten tweede: op basis van de bevolkingsaantallen aan het begin van het jaar en de jaarlijkse overgangsfractionen worden de aantallen berekend aan het einde van het jaar en daarmee aan het begin van het nieuwe jaar. Deze berekende aantallen worden zo nodig weggeschreven naar een uitvoermedium.

6 De resultaten van het demografische model voor het verleden.

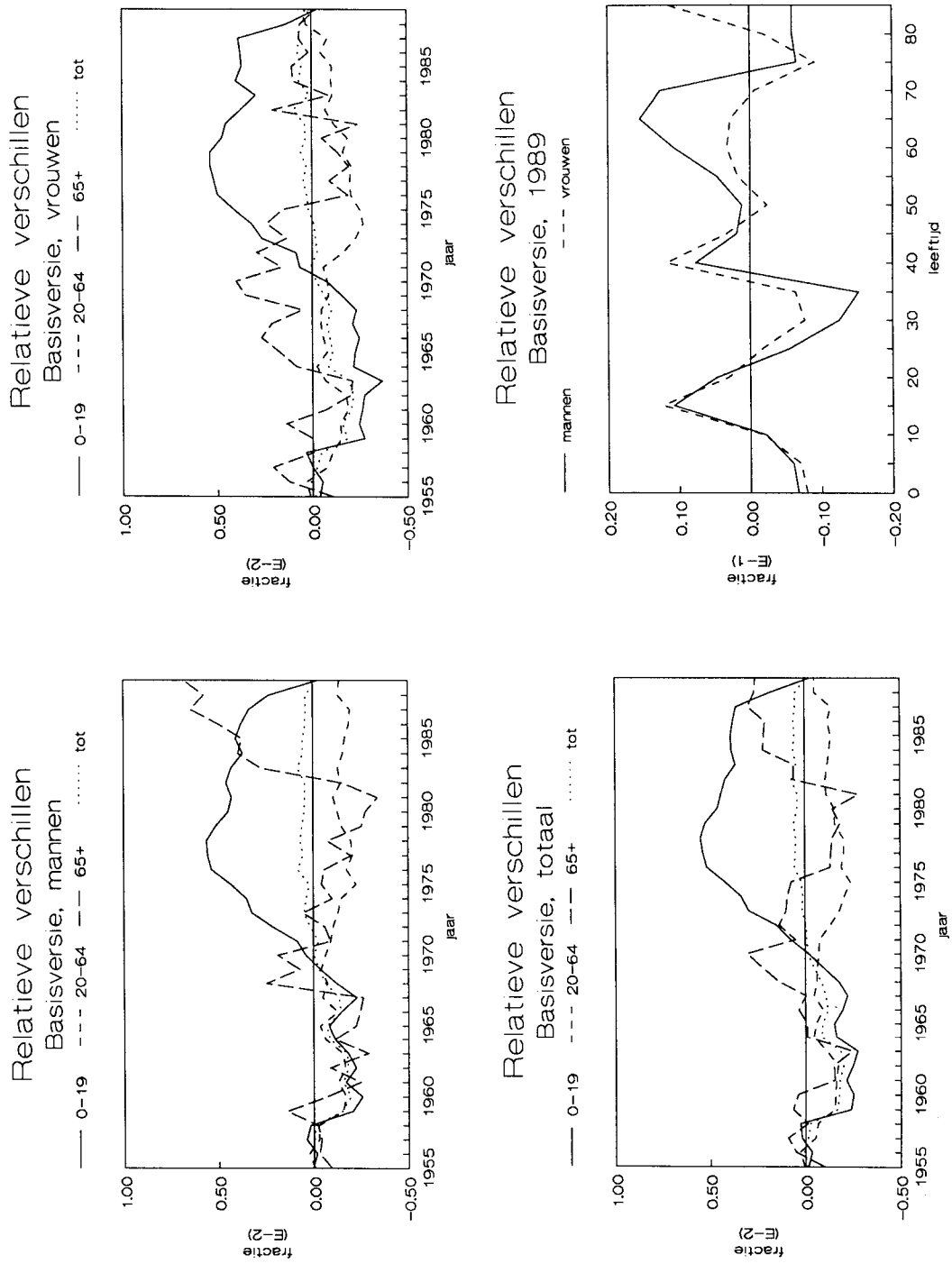
6.1 Inleiding.

In dit hoofdstuk wordt aandacht besteed aan de bruikbaarheid van het ontwikkelde demografische model voor een beschrijving van het verleden. Hiermee wordt bedoeld de mate van overeenkomst tussen de berekende bevolkingsaantallen en de door het CBS waargenomen aantallen. Hiertoe wordt het model 'opgestart' in 1955, waarna achtereenvolgens de aantallen in de jaren daarop berekend worden. Het laatste jaar, waarvoor bevolkingsaantallen berekend worden, is 1989.

Bij de berekeningen wordt uitgegaan van een basisversie. Hierbij wordt uitgegaan van basisbewerkingen op de invoercijfers, waardoor deze betrokken worden op de beginpopulatie voor ieder jaar in plaats van de gemiddelde populatie (zie bijlage B). Bovendien worden administratieve correcties toegepast. Vervolgens worden verschillende alternatieve versies doorgerekend, waarbij uitgegaan wordt van deels andere voorbewerkingen op de invoercijfers. De resultaten van deze versies worden vergeleken met die van de basisversie. Deze vergelijking met verschillende varianten kan beschouwd worden als een vorm van structurele gevoeligheidsanalyse. Daarnaast worden ook enkele meer klassieke vormen van gevoeligheidsanalyse toegepast middels het variëren van parameterwaarden voor de basisversie.

Van de berekeningen volgens de basisversie en de verschillende andere versies wordt de volgende **grafische uitvoer** gegeven. **Ten eerste:** de relatieve verschillen tussen de waargenomen en berekende bevolkingsaantallen vanaf 1955 tot aan 1989. Deze verschillen worden gegeven enerzijds voor mannen en vrouwen afzonderlijk en tezamen, anderzijds verdeeld in drie leeftijdscategorieën en tezamen. De relatieve verschillen worden gedefinieerd als het verschil tussen de berekende en waargenomen aantallen gedeeld door de waargenomen aantallen.

Ten tweede: de relatieve verschillen tussen de berekende en waargenomen bevolkingsaantallen in het jaar 1989, het laatste jaar van de simulatie voor mannen en vrouwen afzonderlijk over de leeftijd.



Grafiek 1: De uitkomsten van de basisversie van het demografische model.

6.2 De basisversie.

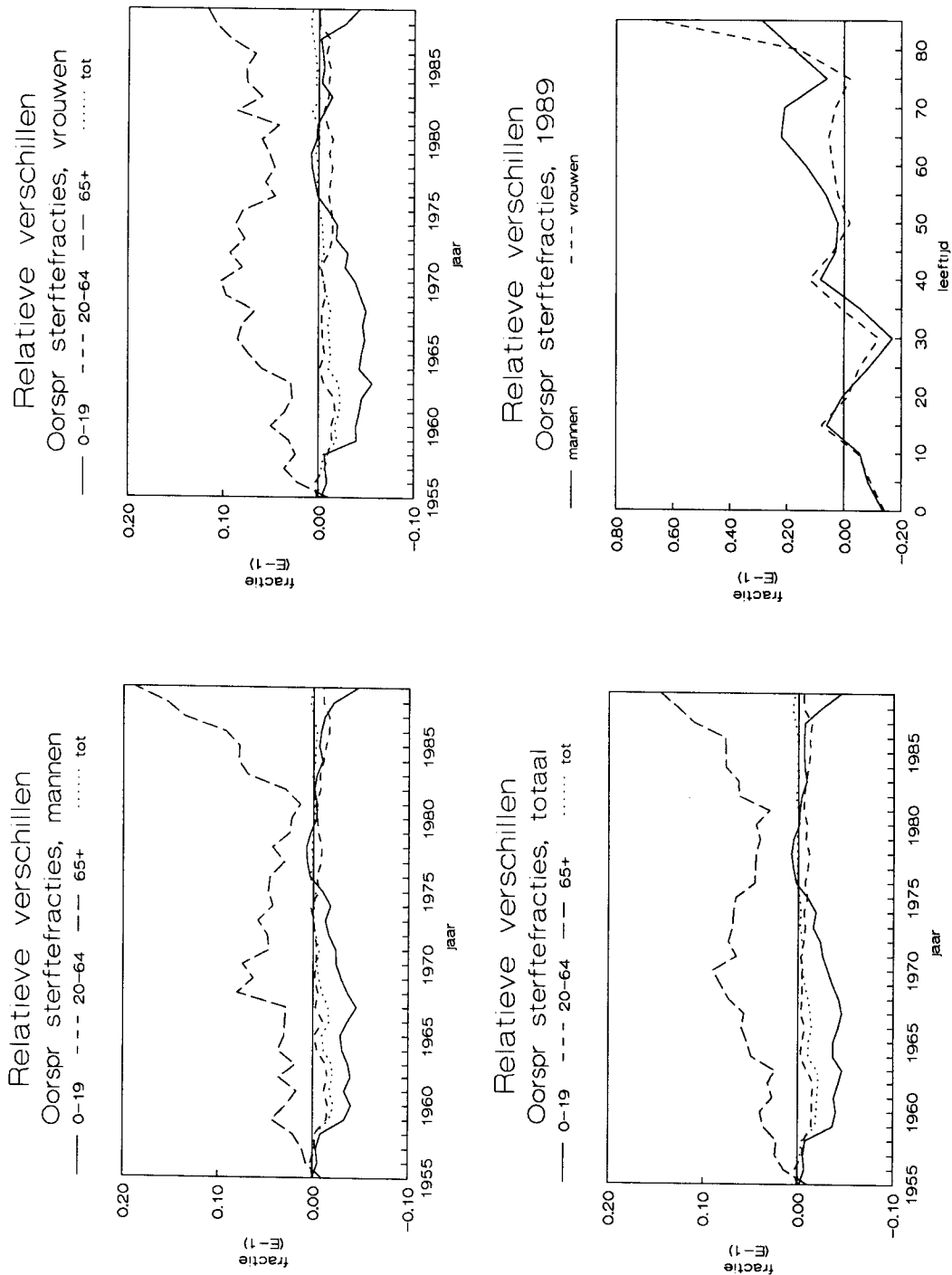
De basisversie en de verschillende alternatieve versies gaan uit van verschillende voorbewerkingen van de invoercijfers voor sterfte, migratie en geboorte (zie de paragrafen B.3, B.4 en B.5). Bij de voorbewerkingen van de invoercijfers worden de sterftefracties toegepast als afhankelijke maten. Daarbij worden de fracties, die gegeven zijn voor een reeks leeftijdsklassen, omgezet in fracties voor de afzonderlijke leeftijdsjaren. Deze omzetting vindt plaats onder de aanname, dat de sterftefracties stuksgewijs lineair van de leeftijd afhangen. De absolute migratiecijfers worden omgezet in afhankelijke maten op basis van berekende aantallen. De geboorteaantallen vormen tot 1967 een direct gegeven. Vanaf 1967 worden deze berekend als het product van fertiliteitscijfers, onafhankelijke maten, en het gemiddelde over een jaar van de berekende aantallen vrouwen. Bovendien worden administratieve correcties toegepast. De resultaten in de vorm van relatieve verschillen tussen de waargenomen en berekende bevolkingsaantallen zijn weergegeven in **grafiek 1**.

Op basis van de resultaten van de basisberekeningen kunnen de volgende conclusies getrokken worden:

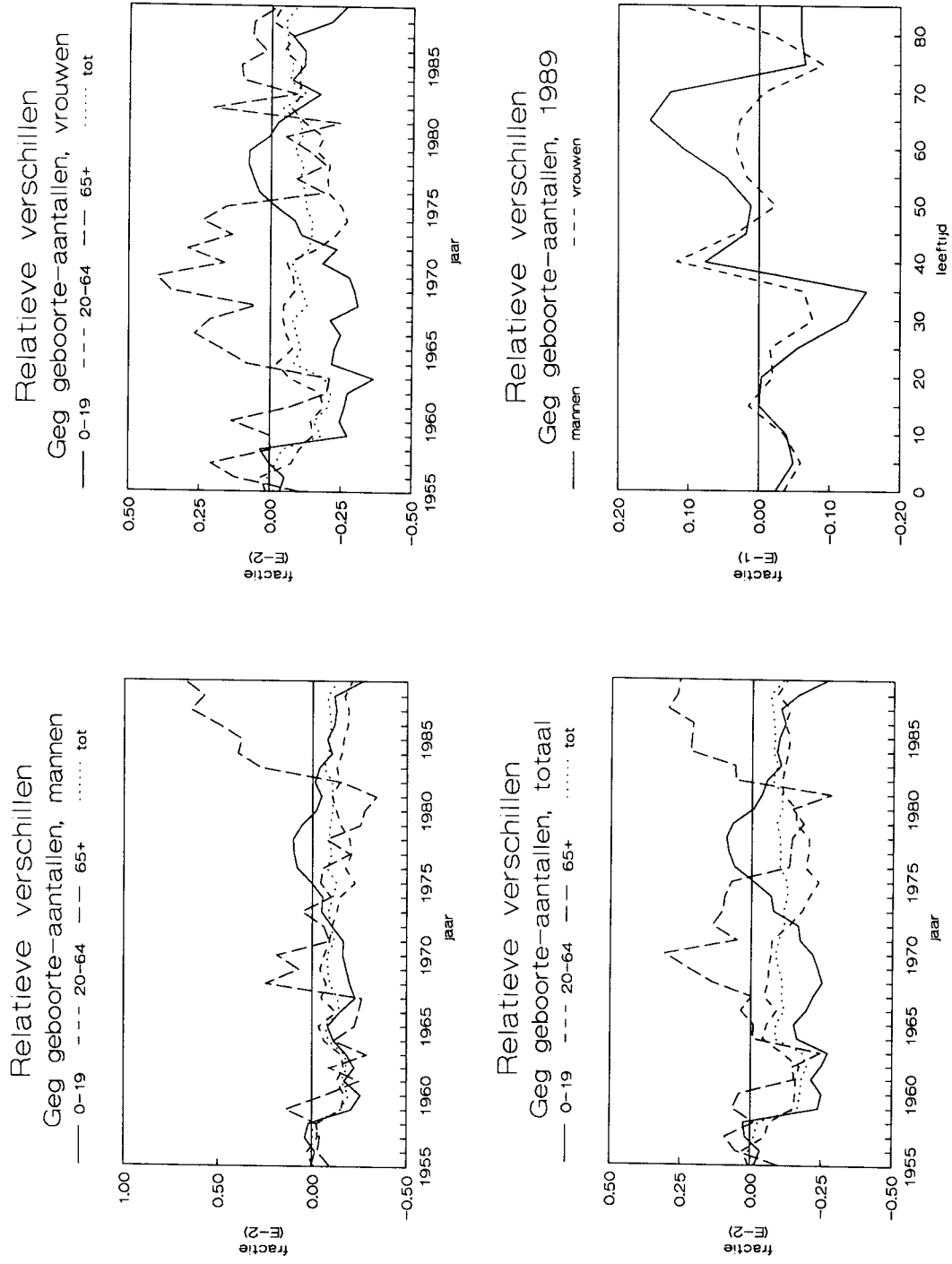
- De relatieve verschillen over de tijd tussen de (geaggregeerde) berekende en waargenomen bevolkingsaantallen zijn zeer klein (tot ca 0.5%). Voor mannen en vrouwen nemen de aantallen voor de jongste leeftijdsgroep vanaf ca 1967 toe om vanaf ca 1978 weer te dalen. Voor mannen nemen de aantallen voor de oudste leeftijdsgroep vanaf ca 1982 toe.
- De meeste curven verlopen redelijk vlak, behalve die voor de oudste leeftijdsgroepen.
- De relatieve verschillen in het jaar 1989 zijn duidelijk groter; let op de verschillende schalen in de grafieken! De curves voor mannen en vrouwen zijn vrijwel identiek tot de leeftijd van 80 jaar - ook even grillig. Voor de oudste leeftijdsgroep is het verschil voor mannen en vrouwen groter.

Enige opvallende 'knikken' in de grafieken zijn:

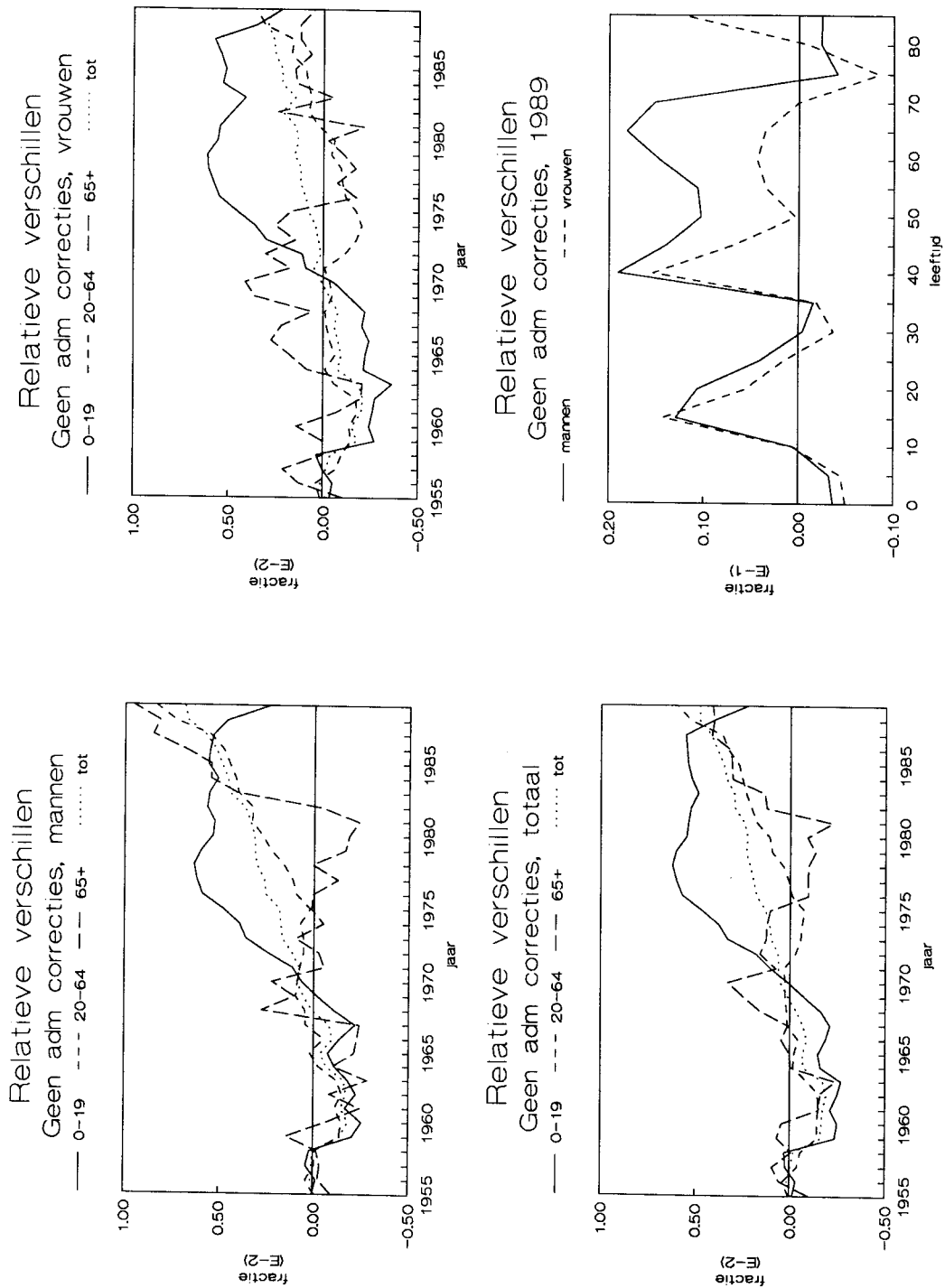
- De overgang rond het jaar 1967 van een licht toenemende onderschatting van de aantallen voor de jongste leeftijdsgroep naar een licht toenemende overschatting. Deze overgang valt samen met de overgang van gegeven geboorte-aantallen naar met behulp van fertiliteitscijfers berekende. Kennelijk wordt de sterfte onder de nieuwgeborenen (vanaf 1967) in het model overschat.
- De toename van de overschatting van de oudste leeftijdsgroep voor mannen vanaf het jaar 1980. Deze overgang valt samen met het gebruik van andere sterftecijfers (zie paragraaf 5.4).
- De overgang rond het jaar 1963 van een licht toenemende onderschatting van de totaalaantallen naar een juiste schatting in het jaar 1970, waarna het verschil tussen de geschatte en werkelijke aantallen zeer klein blijft.



Grafiek 2: De uitkomsten van de alternatieve sterfteversie van het demografische model.



Grafiek 3: De uitkomsten van de alternatieve geboorteverisie van het demografische model.



Grafiek 4: De uitkomsten van de oorspronkelijke versie van het demografische model zonder administratieve correcties.

- De relatieve verschillen in het jaar 1989 zijn tot aan de leeftijdsgroep 55-59 jaar voor mannen en vrouwen vrijwel gelijk. Daarna lopen de curven meer uiteen.

De beschreven basisversie van het demografisch model wordt in de volgende paragrafen vergeleken met enkele alternatieve versies.

6.3 Enkele alternatieve versies.

In de basisversie worden tot het jaar 1980 door het CBS gegeven sterftefracties toegepast, zoals gepubliceerd in 'Maandstatistiek van de bevolking' (zie paragraaf A.1). Deze fracties zijn gebaseerd op de gemiddelde omvang van de betreffende bevolkingsgroepen over ieder jaar en vormen onafhankelijke maten. Een probleem is echter, dat deze cijfers soms achteraf gewijzigd worden. Bijvoorbeeld: voor de sterfte onder 0- en 1-jarigen in de jaren '80 bestaan verschillende cijfers. De oudste cijfers beschrijven een hogere mortaliteit voor de 0-jarigen en een lagere voor de 1-jarigen dan de nieuwste. Dit kan wijzen op een verandering van de definitie van de gemiddelde bevolking over een jaar.

Daarnaast publiceert het CBS ook absolute sterfte-aantallen en (relatieve) -fracties in de 'Mortaliteitsstatistieken'. Deze sterftefracties zijn eveneens gebaseerd op de gemiddelde omvang van de betreffende bevolkingsgroepen, maar zijn over het algemeen hoger dan de eerstgenoemde. Dit geldt met name voor de oudere bevolkingsgroepen.

Een mogelijke onderschatting van de sterfte door de eerstgenoemde gepresenteerde sterftefracties blijkt uit de resultaten van de **alternatieve sterfteversie**, waarbij gebruik wordt gemaakt van de oorspronkelijke sterftefracties betrokken op de gemiddelde omvang van de leeftijdsgroepen (**grafiek 2**). Inderdaad treden voor de oudste leeftijdsgroepen de belangrijkste verschillen op. De alternatieve versie geeft een duidelijk (grotere) overschatting in vergelijking met de oorspronkelijke. Voor de overige leeftijdsgroepen zijn de verschillen gering en is de alternatieve versie tot ca 1985 zelfs beter. Voor het jaar 1989 geeft de alternatieve versie lagere schattingen van de omvang van de jongere leeftijdsgroepen dan de oorspronkelijke.

Daarnaast is een alternatieve geboorteverisie beschouwd (**grafiek 3**). Deze alternatieve versie gaat ook vanaf 1967 uit van gegeven geboorte-aantallen in plaats van fertiliteitscijfers. Deze versie leidt tot iets lagere bevolkingsaantallen voor de jongere leeftijdsgroepen dan de oorspronkelijke.

De laatste versie komt overeen met de oorspronkelijke, maar zonder toepassing van administratieve correcties (**grafiek 4**). Voor alle leeftijdsgroepen, maar met name de midden-groepen, geeft deze

alternatieve versie hogere bevolkingsaantallen. De verschillen zijn het duidelijkst te zien in het eindjaar 1989 van de simulatie.

6.4 Enkele gevoeligheidsanalyses.

Van het demografische model zijn enkele gevoeligheidsanalyses uitgevoerd. Bij deze analyses wordt geanalyseerd hoe de modeluitkomsten afhangen van variaties in parameterwaarden. Deze modeluitkomsten betreffen de berekende bevolkingsaantallen. In de vorige paragraaf is een vorm van structurele gevoeligheidsanalyse toegepast middels het variëren van de voorbewerkingen op de invoercijfers. In deze paragraaf wordt een meer klassieke vorm van gevoeligheidsanalyse toegepast middels het variëren van parameterwaarden.

Achtereenvolgens zijn de jaarlijkse mortaliteitsfracties, migratie-aantallen en geboorte-aantallen met 5% verhoogd om de gevolgen te bepalen van een constante verandering van één van de parameters. Vervolgens zijn de geboorte-aantallen in 1955 met 50% verhoogd om de gevolgen te bepalen van een éénmalige verandering van een parameter. De resultaten van deze berekeningen in de vorm van relatieve verschillen met de waargenomen aantallen zijn gepresenteerd (voor de mortaliteit) in **grafiek 5** resp. (voor de migratie) in **grafiek 6** resp. (voor de jaarlijkse 5% toename van de geboorte) in **grafiek 7** resp. (voor de 50% hogere geboorte-aantallen in 1955) **grafiek 8**.

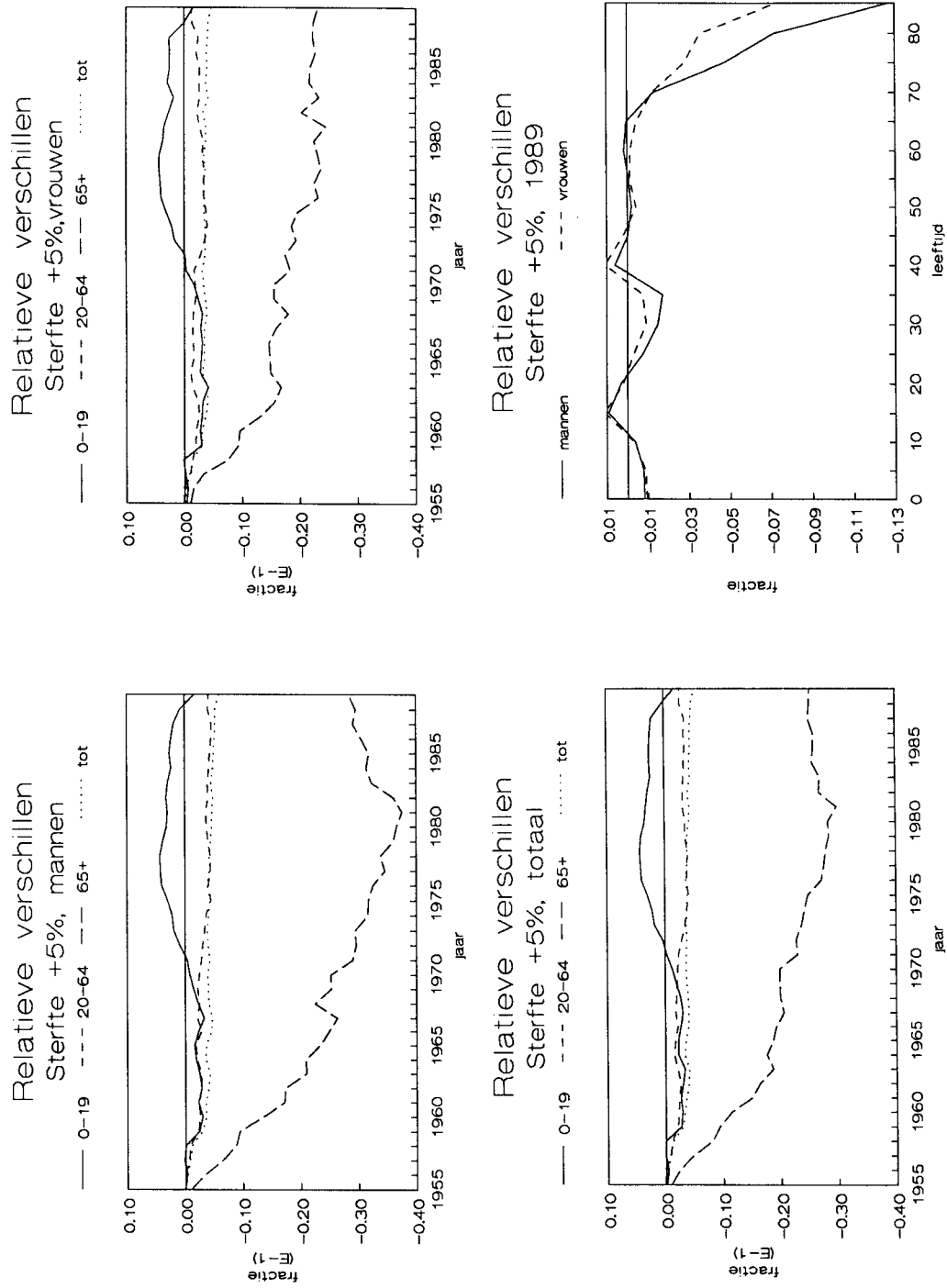
Op basis van de uitkomsten van deze gevoeligheidsanalyses kunnen de volgende uitspraken gedaan worden:

Met betrekking tot de gevoeligheid voor de sterftefracties:

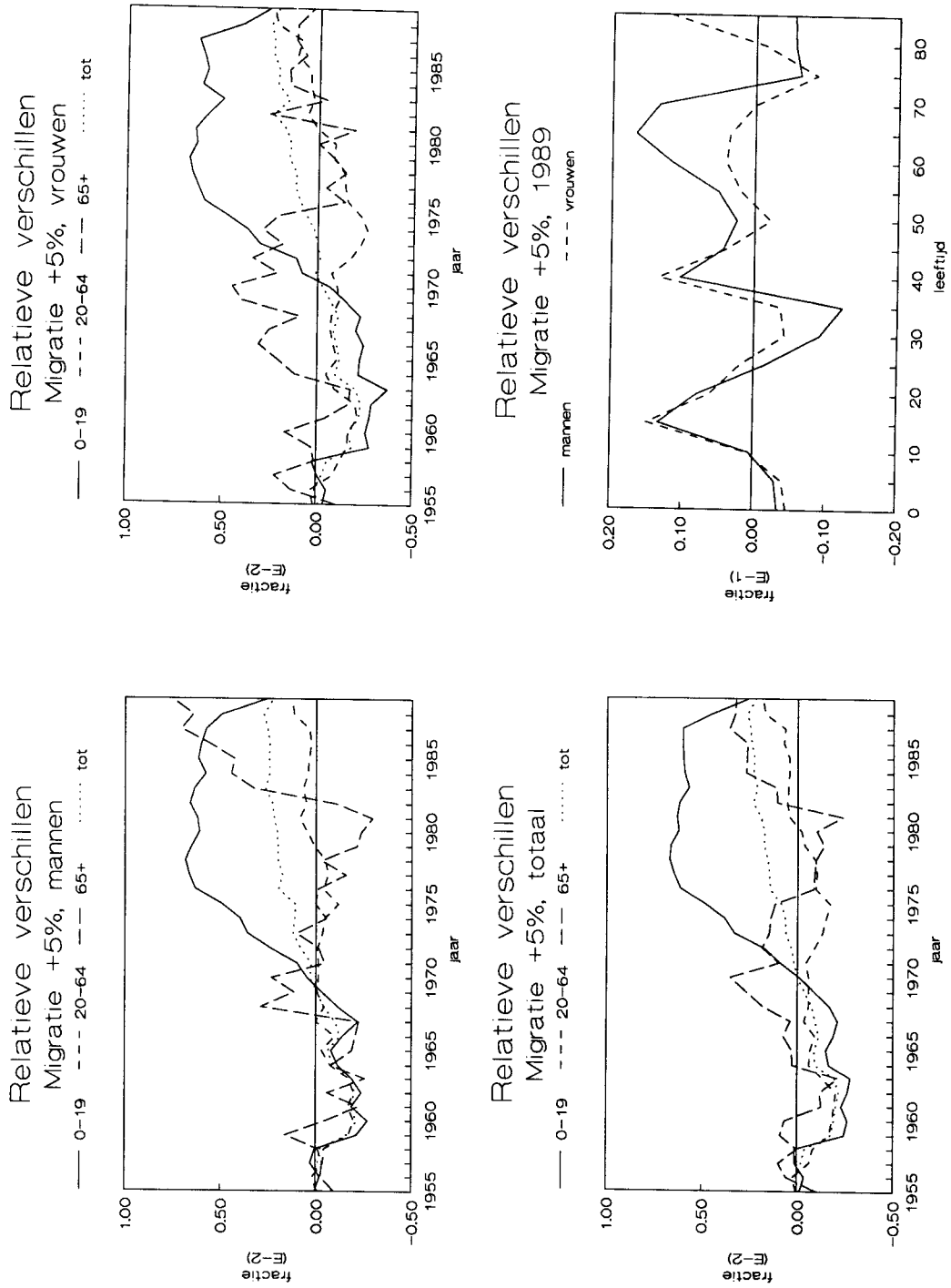
- Overeenkomstig de verwachting is de hogere sterfte met name van invloed op de berekende aantallen voor de oudste leeftijdsgroepen. Bij een 5% hogere sterftefractie worden deze aantallen ca 3% te laag geschat. Deze beperkte onderschatting is mede een gevolg van de overschatting voor de basisversie.
- Als gevolg van de onderschatting van de oudste leeftijdsgroepen en de goede schatting van de overige groepen worden ook de totaalaantallen onderschat.
- De midden-leeftijdsgroepen worden bij een 5% hogere sterfte goed geschat, de oudste leeftijdsgroepen zwaar onderschat. Dit is een gevolg van de cumulatieve werking van de hogere sterfte voor de laatstgenoemde groepen.

Met betrekking tot de gevoeligheid voor de migratie-aantallen:

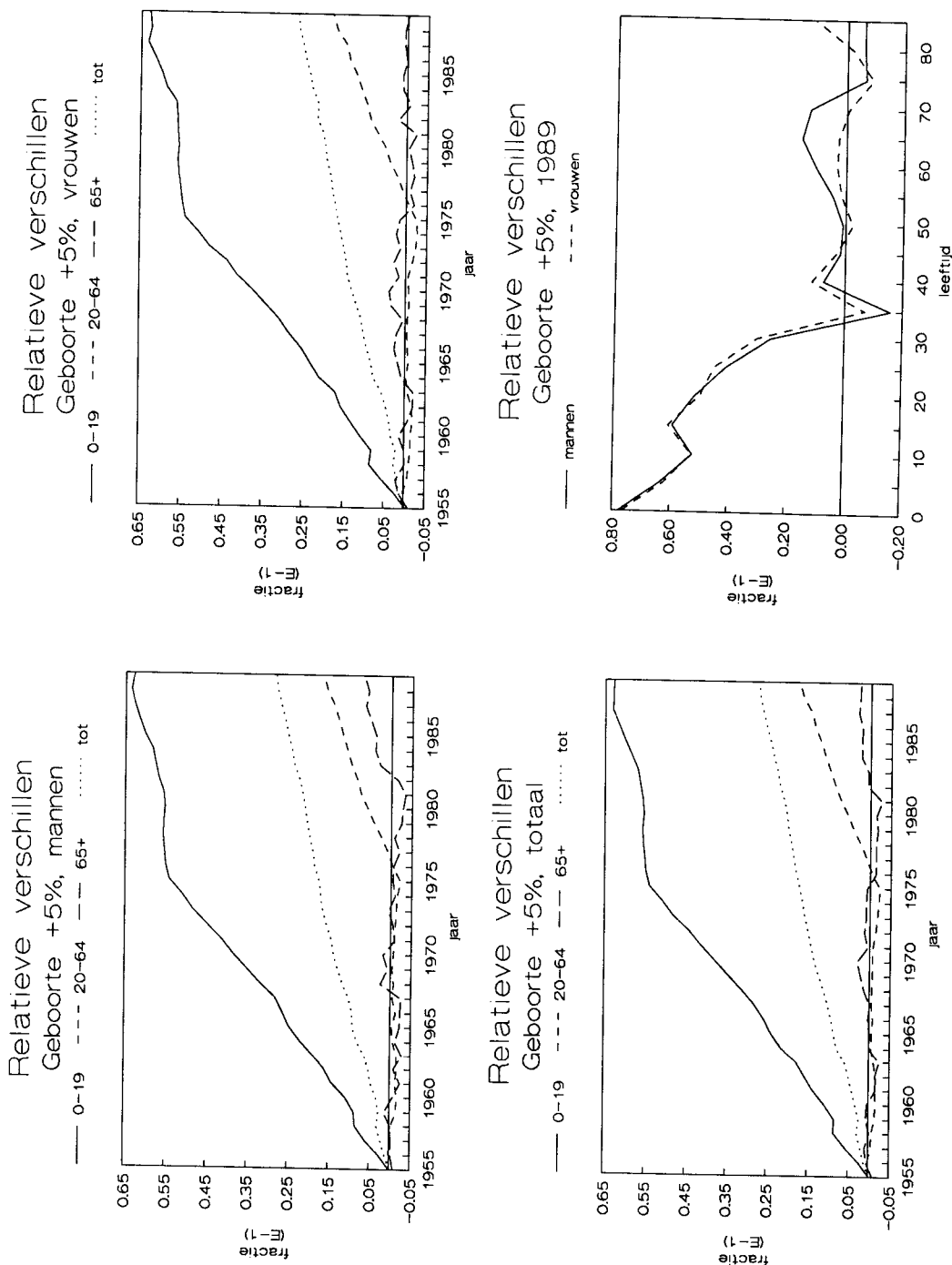
- De verschillen tussen de uitkomsten van de basisversie en van de versie met verhoogde migratie zijn beperkt. Ze zijn slechts merkbaar voor de jongste en middelste leeftijdsgroepen, waarvoor de migratie-aantallen dan ook het grootst zijn.



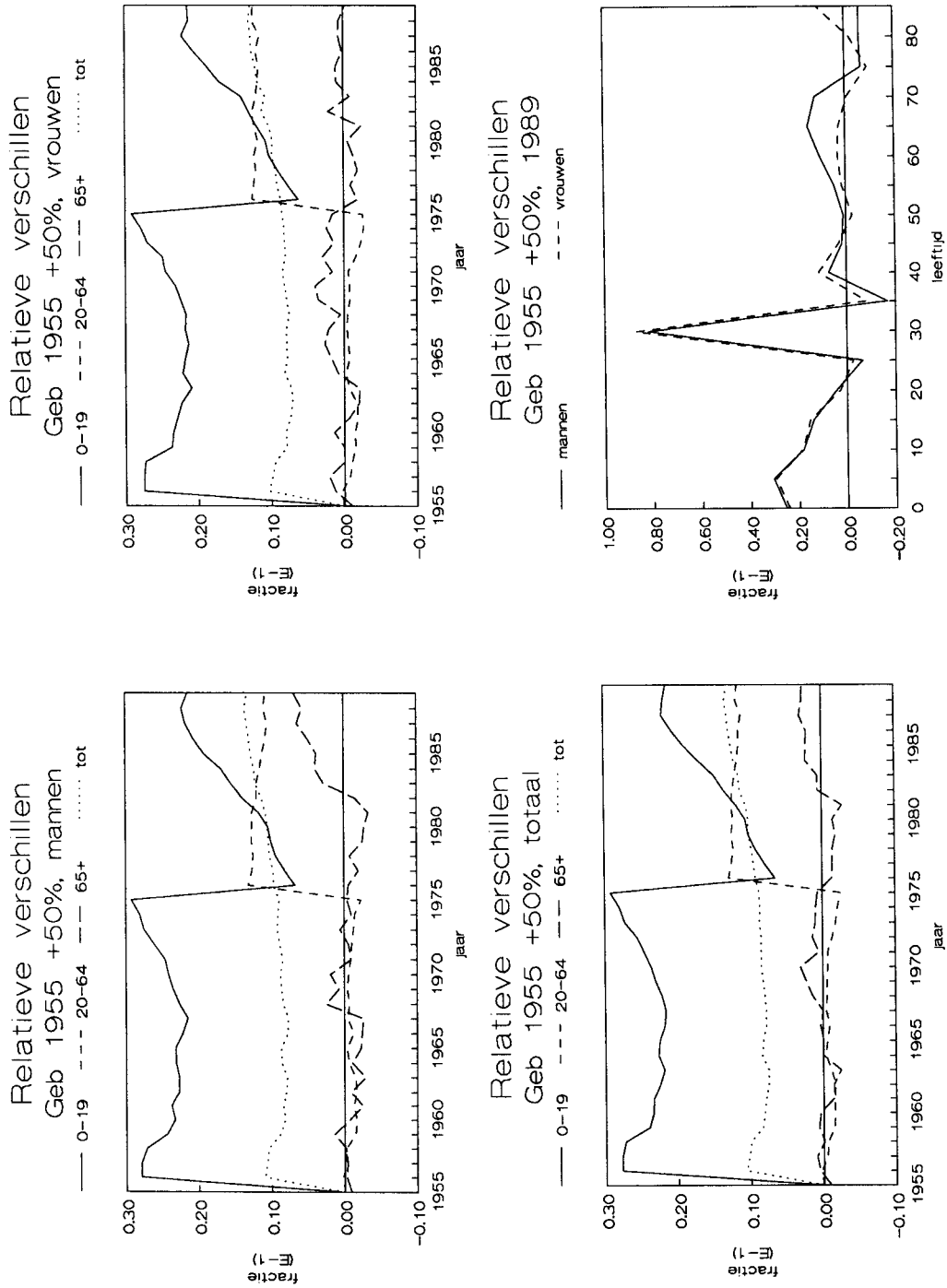
Grafiek 5: De uitkomsten van de basisversie van het demografische model met gebruik van 5% hogere sterftecijfers.



Grafiek 6: De uitkomsten van de basisversie van het demografische model met gebruik van 5% hogere migratiecijfers.



Grafiek 7: De uitkomsten van de basisversie van het demografische model met gebruik van 5% hogere geboortecijfers.



Grafiek 8: De uitkomsten van de basisversie van het demografische model met gebruik van 50% hogere geboortecijfers in 1955.

- De verschillen in het jaar 1989 voor de vijfjaars leeftijds-groepen afzonderlijk zijn zeer gering.

Overigens is in het model de fertiliteit onafhankelijk van de migratie-aantallen verondersteld. Dit kan leiden tot een onder-schatting van de geboorte-aantallen bij toenemende migratie.

Met betrekking tot de gevoeligheid voor de geboorte-aantallen:

- De invloed van de 50% hogere geboorte-aantallen in 1955 is duidelijk merkbaar. Tot het jaar 1975 lopen de curves van de relatieve verschillen voor de jongste leeftijdsgroep voor de basisversie en de versie met verhoogde geboorte-aantallen vrij-wel parallel. Het verschil bedraagt ca 2.5%. Dit verschil is gelijk aan het quotiënt van de toename voor het ene leeftijds-jaar (50%) en de klassebreedte (20 jaar). In 1976 zijn de relatieve verschillen weer vrijwel gelijk geworden. Vanaf 1975 treedt hetzelfde effect op voor de middelste leeftijdsgroep, met een vrijwel constant verschil van ruim 1% (50% gedeeld door 45 jaar).
- Tegelijkertijd neemt vanaf 1975 het relatieve verschil voor de jongste leeftijdsgroep weer toe. Dit is een gevolg van de extra geboorten vanwege de toename van het aantal vrouwen - in de vruchtbare leeftijd - met geboortjaar 1955. Dit leidt tot een nieuwe, geleidelijke stijging van de met deze modelversie bere-kende geboorte-aantallen.
- Dit effect is ook te zien in het jaar 1989. Vanwege de vertraagde toename van de geboorte-aantallen, omdat de extra vrouwelijke nieuwgeborenen van 1955 op hun beurt weer kinderen krijgen, zijn de nieuwe geschatte aantallen tot de leeftijds-groep van 25-29 jaar hoger dan voor de basisversie. De piek voor de groep 30-35 jaar is een rechtstreeks gevolg van de veronderstelde geboorte-explosie in 1955.
- De gevolgen van een 5% toename van de aantallen nieuwgeborenen vanaf 1955 zijn aanvankelijk minder spectaculair dan die van de eenmalige geboorte-explosie in 1955. Na verloop van tijd zijn de gevolgen echter eveneens fors om de volgend redenen. Ten eerste: na ca tien jaar leidt een jaarlijkse 5% toename even-eens tot een 50% toename in vergelijking met de oorspronkelijk-ke situatie. Ten tweede: een cumulatief effect, omdat hogere aantallen vrouwelijke nieuwgeborenen na verloop van tijd leiden tot weer nieuwe hogere aantallen geboorten.

Op basis van de resultaten van de gevoeligheidsanalyses kunnen de volgende conclusies getrokken worden. De gevoeligheid van het demografisch model is een te algemeen begrip. Zo blijken de totaal-aantallen even gevoelig te zijn voor een verandering in de sterfte als in de migratie. Voor de bevolkingsaantallen voor de verschil-lende leeftijdsgroepen zijn de gevolgen echter zeer verschillend. In het geval van sterfte worden de veranderingen geconcentreerd in de oudste leeftijdsgroepen. Anders gezegd: deze aantallen zijn zeer

gevoelig voor veranderingen in de sterfte. In het geval van migratie worden de veranderingen meer uitgesmeerd over de leeftijdsgroepen met een lichte nadruk op de midden-leeftijdsgroepen. Met betrekking tot de gevoeligheid van het model voor veranderingen in de geboorte-aantallen kunnen de volgende drie opmerkingen geplaatst worden. (1) De gevolgen hebben met name betrekking op (aanvankelijk) de jongste leeftijdsgroepen. (2) Het effect is cumulatief, omdat hogere geboorte-aantallen na verloop van tijd weer tot hogere geboorte-aantallen leiden. (3) De fertiliteit is in het model onafhankelijk van de migratie verondersteld.

6.5 Conclusies.

Op basis van een vergelijking van de basisversie met de beschreven alternatieve versies en de gevoeligheidsanalyses kunnen de volgende conclusies getrokken worden:

- De belangrijkste verschillen tussen de waargenomen en berekende bevolkingsaantallen treden op voor de groepen tot 20 jaar en vanaf 65 jaar en van de laatstgenoemde groep met name de groep 85+.
- De aantallen voor de oudste leeftijdsgroepen zijn het meest gevoelig voor veranderingen in de sterfte, die voor de jongste voor veranderingen in de geboorte-aantallen. Het model is minder gevoelig voor veranderingen in de migratie (zie echter opmerking in paragraaf 6.4).
- De invloed van de administratieve correcties op de bevolkingsaantallen is (in het verleden) waarneembaar, maar gering.

Bij de uiteindelijke specificatie van het demografische model spelen verschillende argumenten een rol, namelijk:

De mate, waarin de berekende en werkelijke bevolkingsaantallen overeenkomen (zie voorgaande paragrafen).

Uit de voorgaande paragrafen komt naar voren, dat de resultaten van de basisversie redelijk overeenkomen met de waargenomen bevolkingsaantallen. De eerste sterfteversie (voor de oudste mannengroep) en de versie met administratieve correcties (voor de middelste leeftijdsgroepen) leiden echter tot betere resultaten.

De mate, waarin de modelvergelijkingen overeenkomen met het type van de toegepaste CBS-cijfers (zie hoofdstuk 6).

De gegeven sterftefracties en fertiliteitscijfers betreffen onafhankelijke maten. De modelvergelijkingen gaan echter uit van afhankelijke maten. Daarom dienen uit theoretisch oogpunt de invoercijfers met gebruik van (werkelijke of berekende) absolute bevolkingsaantallen bewerkt te worden tot afhankelijke maten. De sterfte- en fertiliteitscijfers zijn relatieve cij-

fers en moeten middels vermenigvuldiging met absolute aantallen omgezet worden in absolute cijfers. De migratie-cijfers betreffen reeds absolute aantallen.

Deze absolute aantallen moeten middels deling door de absolute bevolkingsomvang aan het begin van het jaar omgezet in afhankelijke maten. Een keuze is mogelijk van de bevolkingsaantallen, waardoor de absolute aantallen veranderingen gedeeld worden om te komen tot afhankelijke maten: berekende aantallen of werkelijke aantallen. Dit argument leidt tot de keuze van de varianten, waarbij de invoercijfers omgezet worden in afhankelijke maten.

De robuustheid van het model.

De achterliggende vraag is: Bestaat een zodanige terugkoppeling, dat te hoge of te lage berekende bevolkingsaantallen gecorrigeerd worden in de richting van de werkelijke aantallen (negatieve terugkoppeling) of versterken afwijkingen zichzelf (positieve terugkoppeling)? Bijvoorbeeld: het gebruik van absolute sterfte-aantallen in plaats van sterftefracties leidt ertoe, dat te lage berekende bevolkingsaantallen te laag blijven. Om redenen van robuustheid dient het toepassen van fracties de voorkeur.

Eenvoud van de berekeningen.

Het verdient de voorkeur om de voorbewerkingen van de gebruikte gegevens zo beperkt mogelijk te houden. Dit argument pleit voor de tweede sterfteversie, waarin de klasse-specifieke sterftefracties niet verder gespecificeerd worden naar leeftijdsjaar.

Uit een vergelijking van de basisversie met de verschillende andere versies blijkt bij hantering van de bovengenoemde criteria de basisversie met gebruik van administratieve correcties goed te voldoen. Deze basisversie wordt daarom gebruikt bij de vooruitberekeningen van de Nederlandse bevolking (zie hoofdstuk 7). Hierbij dienen de volgende **kanttekeningen** geplaatst te worden.

Ten eerste: de CBS bevolkingsprognoses zijn gebaseerd op kernhypothese. De kernhypothese omtrent het migratiesaldo of vestigingsoverschot zijn inclusief administratieve correcties. Dit heeft tot gevolg, dat bij vooruitberekeningen met behulp van het ontwikkelde model, waarbij gebruik wordt gemaakt van de door het CBS opgegeven migratiesaldo's, de administratieve correcties buiten beschouwing kunnen blijven.

Ten tweede: in theoretisch opzicht 'wringen' de modelvergelijkingen in de basisversie enigszins. Deze vergelijkingen gaan uit van afhankelijke maten, de toegepaste sterfte- en fertiliteitscijfers betreffen onafhankelijke maten.

Ten derde: hoewel de uitkomsten (de gepresenteerde relatieve verschillen) van de basisversie binnen redelijke marges rond het

nulpunt blijven zijn toch zekere trends waarneembaar. Bijvoorbeeld: de relatieve verschillen voor de middelste leeftijdsgroep 20-64 jaar nemen vanaf ca 1962 voortdurend toe (van een ondertschatting naar een overschatting) evenals de totaalaantallen. Voor mannen nemen vanaf ca 1980 de relatieve verschillen voor de oudste leeftijdsgroep 65+ spectaculair toe.

7 De resultaten van het demografische model voor de toekomst.

7.1 Inleiding.

In dit hoofdstuk wordt aandacht besteed aan de voorspellende resultaten van het ontwikkelde demografische model. Hiermee wordt bedoeld de overeenkomsten van de door het model berekende toekomstige bevolkingsaantallen vanaf het jaar 1990 met de CBS prognoses (CBS, 1990). Hiertoe wordt het model 'opgestart' in 1989, waarna achtereenvolgens de bevolkingsaantallen in de jaren daarop berekend worden. Het laatste jaar, waarvoor bevolkingsaantallen vooruitberekend worden, is 2020. Vergelijk hiermee hoofdstuk 6, waarin de resultaten van het model voor het verleden gegeven zijn.

Bij het onderzoek naar de voorspellende resultaten van het demografische model moeten verschillende kanttekeningen geplaatst worden. Ten eerste: het CBS geeft drie verschillende prognoses van de bevolking, namelijk een lage, midden- en hoge variant. Deze varianten zijn gebaseerd op verschillende aannames met betrekking tot (de in dit verband relevante grootheden) sterfte, migratie en geboorte. Het ontwikkelde demografische model moet de drie varianten kunnen reproduceren.

Ten tweede: de aannames, die aan de CBS prognoses ten grondslag liggen, zijn niet direct toepasbaar in het ontwikkelde demografische model. Het gevolg is, dat de CBS aannames vertaald moeten worden in termen van de in het onderhavige model toegepaste invoergrootheden (zie paragraaf 7.3).

Ten derde: de CBS aannames in de vorm van veronderstelde toekomstige levensverwachting, gemiddeld kindertal en migratiesaldo worden niet direct in 1989 gerealiseerd, maar pas vanaf een zeker 'realisatiejaar'. Dit realisatiejaar is, in overeenstemming met het CBS, het jaar 2010.

Ten vierde: in hoofdstuk 6 zijn verschillende versies van het demografische model onderzocht. Deze versies onderscheiden zich van elkaar in de voorbewerkingen op de gebruikte CBS invoercijfers. Deze bewerkingen betreffen met name de omzetting van onafhankelijke maten en absolute aantallen in afhankelijke maten. Hierbij werd in sommige gevallen gebruik gemaakt van gegeven (waargenomen) bevolkingsaantallen. Omdat het demografische model in dit hoofdstuk gebruikt wordt voor vooruitberekeningen kan deze werkwijze niet meer gevolgd worden (zie ook paragraaf 6.7).

7.2 De mogelijke aannames.

Het demografische model maakt gebruik van specifieke sterftefracties, absolute migratie-aantallen en fertiliteitscijfers. De modelaannames hebben op deze cijfers betrekking (zie ook paragraaf 7.1) De volgende typen aannames zijn mogelijk.

Het gemiddelde.

Het gemiddelde wordt genomen over de (maximaal) laatste vijf jaren, waarvoor leeftijds- en geslachtsspecifieke CBS cijfers beschikbaar zijn. De huidige versie maakt gebruik van cijfers tot en met 1988. Deze gemiddelden worden vervolgens gebruikt om de bevolkingsaantallen vanaf het jaar 1989 te berekenen, waarbij de cijfers over de jaren constant worden verondersteld.

Deze gemiddelden kunnen met een correctiefactor vermenigvuldigd worden. Deze correctiefactor wordt leeftijds-, geslachts- en tijdsonafhankelijk verondersteld. Daarnaast wordt verondersteld, dat deze correctiefactor pas in 2000 volledig gerealiseerd wordt. Tussen 1988 en 2000 verandert deze factor volgens een rekenkundige reeks tussen de waarden 1 en de opgegeven waarde. Bijvoorbeeld: voor de berekening van de verschillende varianten van het demografische model zijn verschillende correctiefactoren voor de sterftefracties toegepast. Een grotere correctiefactor vormt de vertaling van de CBS aanname van een verkorte levensverwachting (in vergelijking met de middenvariant).

Een trend.

Door de cijfers voor de maximaal laatste vijf jaren wordt de (lineaire) regressielijn bepaald volgens de kleinste kwadraten methode. Deze regressielijn kan beschouwd worden als de lineaire trend, die de betreffende cijfers het beste beschrijft. Deze trend wordt vervolgens doorgetrokken om de toekomstige bevolkingsaantallen te berekenen. De doorgetrokken trend kan op twee manieren gestuurd worden. Een vermenigvuldigingsfactor bepaalt de richting van de trend. In geval van een factor 1 wordt de trend doorgetrokken, bij een factor -1 wordt de trend gespiegeld. Een dempingsfactor wordt gebruikt om de trend exponentieel te laten afvlakken. Een factor 0 betekent geen demping. Hoe groter de factor, hoe groter de demping. Een trend met zeer grote dempingsfactor komt overeen met een constante waarde.

Het laatst beschikbare cijfer.

Het laatst beschikbare cijfer (in de huidige versie van het model voor 1988) wordt gebruikt als invoer bij de berekeningen. Dit cijfer is gelijk aan het gemiddelde van de laatst beschikbare cijfers over precies één jaar. Het cijfer kan (vergelijk het gemiddelde) zo nodig met een correctiefactor vermenigvuldigd worden. Deze factor verandert van de waarde 1 tot de opgegeven waarde tussen 1988 en 2010.

Expliciet geformuleerde aannames.

Tenslotte is het mogelijk om expliciet invoercijfers voor de toekomst op te geven. Dit is bijvoorbeeld zinvol in het geval van migratie. Het toekomstig (totaal) migratiesaldo wordt door het CBS tijdsafhankelijk verondersteld. Dit overschot varieert tussen 1989 en 2000, waarna het constant blijft.

Bij de in de volgende paragrafen gepresenteerde resultaten van het demografische model wordt gebruik gemaakt van de tweede optie - de trend - en de vierde - expliciete cijfers. In het eerste geval wordt gebruik gemaakt van correctiefactoren, die verschillen voor de drie CBS varianten. De tweede optie wordt toegepast voor de sterfte en de geboorte, de vierde voor de migratie.

7.3 De vooronderstellingen aan de CBS varianten.

De volgende zogenaamde kernhypothesen liggen aan de lage, midden- en hoge variant van de CBS bevolkingsprognose ten grondslag. Als eindwaarden, te realiseren na een zekere overgangstijd, worden opgegeven:

		laag	midden	hoog
sterfte				
levensverwachting bij geboorte (in jaren)	mannen	74	75	76
	vrouwen	80.5	81.5	82.5
migratie				
migratiesaldo (*1000)	mannen + vrouwen	10	20	30
geboorte				
gemiddeld kindertal per vrouw		1.45	1.65	1.85
gemiddelde leeftijd bij het krijgen van een kind		30.1	id.	id.

In het vervolg van deze paragraaf wordt meer in detail ingegaan op deze kernhypothesen en het gebruik ervan in het ontwikkelde demografische model. Bij deze kernhypothesen dienen twee **kanttekeningen** geplaatst te worden. De **eerste kanttekening** is, dat de kernhypothesen waarden betreffen, die aangenomen worden vanaf ongeveer het jaar 2010. Tot die tijd vindt een geleidelijke overgang plaats vanaf de (waargenomen) waarde voor het jaar 1988.

Deze overgangstijd is als volgt in het ontwikkelde demografische model geïmplementeerd. In het geval van sterfte en geboorte wordt voor iedere leeftijdsklasse de trend bepaald voor de jaren 1984 tot 1989. Onderzocht wordt hoever deze trend doorgetrokken moet worden om een doelwaarde - de levensverwachting bij geboorte resp. gemiddeld kindertal per vrouw in 2010 - te bereiken. Op deze wijze zijn ook de waarden voor de leeftijdsklassen vastgesteld. Vervolgens wordt verondersteld, dat voor iedere leeftijdsklasse de waarde lineair verandert tussen 1989 en 2010. In het geval van migratie zijn de expliciet door het CBS opgegeven jaarlijkse migratiesaldo's toegepast. De leeftijds- en geslachtsverdeling hiervan wordt gelijk verondersteld aan de gemiddelde verdeling over de jaren 1984 tot 1989.

De **tweede kanttekening** - die reeds bij de vorige kanttekening ter sprake is gekomen - is, dat deze kernhypothese niet direct toe te passen zijn op het ontwikkelde demografische model, zoals in paragraaf 7.1 reeds aangegeven is. De CBS aannames betreffen de levensverwachting bij geboorte resp. het gemiddeld kindertal per vrouw, terwijl het model gebruikt maakt van sterftefracties resp. fertilitateitscijfers. Voor de drie verschillende invoergrootheden worden de modelaannames uit de kernhypothese afgeleid.

Sterfte.

De CBS aannames met betrekking tot de sterfte betreffen de levensverwachting bij geboorte, het ontwikkelde demografische model maakt gebruik van sterftefracties. Bepaald is hoever de trend in de sterftefracties voor alle leeftijdsklassen voor de jaren 1984 tot 1989 doorgetrokken moet worden om de genoemde levensverwachting te realiseren. In de volgende tabel is voor de drie varianten aangegeven hoeveel jaren deze trend doorgetrokken is. De resulterende levensverwachting is vergeleken met de CBS opgave.

	Levensverwachting (in jaren)	
	Mannen	Vrouwen
Lage variant	74	80.5
Aantal jaren, waarover trend doorgetrokken is: 4	74.1	80.5
Middenvariant	75	81.5
9.5	75.1	81.4
Hoge variant	76	82.5
15	76.2	82.3

De door het CBS toegepaste procedure om de toekomstige sterftefracties te bepalen kan als volgt beschreven worden:

- De beste lineaire benadering van de levensverwachting bij geboorte voor de jaren 1984 tot 1988 wordt vijf jaar doorgetrokken; vanaf het jaar 2010 wordt de gestelde levensverwachting constant verondersteld.
- Voor de tussenliggende jaren wordt de levensverwachting bij geboorte beschreven volgens de functie (Engels: spline function), die leidt tot een glatte overgang voor de jaren 1993 en 2010.
- Verondersteld wordt, dat voor de onderscheiden leeftijdsklassen de sterftefracties tussen 1988 en 2010 in dezelfde onderlinge verhouding veranderen als tijdens de vijf voorafgaande jaren. De absolute grootte van deze veranderingen wordt zo gekozen, dat de bijbehorende levensverwachting overeenkomt met de eerder berekende levensverwachting.

De procedure, die toegepast is in het ontwikkelde demografische model en die hiervoor reeds beschreven is, verschilt hier enigszins van. Deze procedure leidt tot lineair veranderende sterftefracties, terwijl de CBS procedure leidt tot fracties, die 'concaaf' in de tijd zijn, dat wil zeggen: als functie van de tijd grafisch weergegeven 'bol van boven' zijn.

Migratie.

De CBS aannames met betrekking tot migratie betreffen totale absolute aantallen inclusief administratieve correcties. Het ontwikkelde demografisch model maakt gebruik van leeftijds- en geslachts-specifieke absolute aantallen. Het totale migratiesaldo voor 1988 bedraagt 35.5 (*1000), exclusief administratieve correcties, resp. 27.3 (*1000), inclusief correcties. De door het CBS opgegeven waarden (inclusief administratieve correcties) zijn, te realiseren vanaf 2000 (1995 in geval van de middenvariant):

	CBS aanname (*1000)
Lage variant	10
Middenvariant	20
Hoge variant	30

De waarden voor de jaren tot 2010 zijn gegeven in bijlage A.2. De totaalaantallen worden in het demografische model voor de jaren vanaf 1989 expliciet opgegeven. De leeftijdsverdeling ervan wordt gelijk verondersteld aan het gemiddelde over de periode 1984 tot 1988.

Geboorte.

De CBS aannames met betrekking tot de geboorte betreffen het gemiddelde kindertal per vrouw en de gemiddelde leeftijd van de vrouw bij het krijgen van een kind voor het cohort met geboortjaar 1985. Het gemiddeld kindertal en gemiddelde leeftijd worden berekend uit fertiliteitscijfers en bevolkingsaantallen. Voor de bevolkingsaantallen worden de overlevingsaantallen van het genoemde fictieve cohort gekozen.

Uitgegaan wordt van een geboortecohort van vrouwen, waarbij slechts uitstroom mogelijk is in de vorm van sterfte. Voor de optredende leeftijdsspecifieke sterftefracties worden de waarden voor het jaar 2010 gekozen, zoals die berekend zijn op basis van de opgegeven levensverwachting in 2010 en de trend in de fracties tussen 1984 en 1989 (zie boven). Het gemiddeld kindertal kan geïnterpreteerd worden als (circa) tweemaal de vervangingsfactor. Deze vermenigvuldigingsfactor 2 heeft betrekking op de verhouding tussen het totaal aantal nieuwgeborenen en het aantal vrouwen. Daarnaast is de gemiddelde leeftijd van de moeder bij de geboorte berekend. De uitkomsten zijn:

	kindertal	leeftijd bij geboorte
geboortecohort 1988	1.53	30.0
1984-1988	1.21	29.4
CBS aanname lage variant	1.45	29.6
CBS aanname middenvariant	1.65	29.6
CBS aanname hoge variant	1.85	29.6

Het gemiddeld kindertal is berekend als de som over de verschillende leeftijdscategorieën van het product van het aantal nog in leven zijnde vrouwen en het fertiliteitscijfer. De gemiddelde leeftijd bij de geboorte als die van het product van het aantal vrouwen, het fertiliteitscijfer en de betreffende leeftijd, gedeeld door het berekende aantal geboorten.

Opgemerkt moet worden, dat het gemiddeld kindertal per vrouw op basis van een geboortecohort - cohort waarneming - verschilt van die op basis van de geboorten in een jaar - periode waarneming. Vanwege de niet-stabiele leeftijdsopbouw is het aantal vrouwen behorende tot de klasse met de hoogste vruchtbaarheid oververtegenwoordigd. Dit leidt ertoe, dat het gerealiseerde gemiddelde kindertal (voorlopig) iets hoger ligt dan het fictieve.

Het berekende gemiddelde kindertal voor het geboortecohort van vrouwen is kleiner dan 2. Dit wijst erop, dat de bevolkingsomvang - vanwege de instabiele opbouw met een tijdsvertraging - afneemt.

7.4 De CBS bevolkingsprognoses en modeluitkomsten vergeleken.

In deze en de volgende paragrafen worden de uitkomsten van de drie CBS prognoses in de vorm van toekomstige bevolkingsaantallen vergeleken met de resultaten van het ontwikkelde demografische model.

De volgende twee uitvoercijfers worden **grafisch weergegeven** (zie paragraaf 6.1). **Ten eerste:** de relatieve verschillen tussen de CBS prognoses en die van het ontwikkelde demografische model vanaf 1989 tot 2020. Deze verschillen worden gegeven enerzijds voor mannen en vrouwen afzonderlijk en totaal, anderzijds verdeeld in drie leeftijdscategorieën en totaal. De relatieve verschillen worden gedefinieerd als het verschil tussen de modelvoorspelling en de CBS prognose gedeeld door de CBS prognose. De CBS prognoses zijn gegeven voor iedere vijf jaar vanaf 1990. Om de bevolkingsaantallen voor de tussenliggende jaren te bepalen zijn deze prognoses stuksgewijs lineair verondersteld. Dit verklaart de aanwezigheid van knikpunten voor sommige 'overgangsjaren' in de grafieken en van concave of convexe lijnstukken tussen deze 'knikpunten'.

Ten tweede: de relatieve verschillen tussen de CBS prognose en de berekende aantallen in het laatste jaar van de simulatie (2010), voor mannen en vrouwen afzonderlijk verdeeld over de leeftijd.

Iedere CBS variant is vergeleken met twee modelversies. Bij de eerste modelversie is gekozen voor de modelaanname, die volgen uit de omzetting van de CBS kernhypothese. Bij de tweede modelversie is gekozen voor andere aanname met betrekking tot de sterfte en (eventueel) geboorte, die leiden tot minder grote verschillen met de CBS uitkomsten. De aanname voor de tweede modelversie zijn via trial and error gekozen.

7.5 De middenvariant

De middenvariant beschrijft de toekomstige bevolkingsaantallen onder de meest waarschijnlijke aannames met betrekking tot de sterfte - een toename van de levensverwachting - , de migratie - vergelijkbaar met het huidige niveau - en geboorte - een lichte toename van het gemiddeld kindertal per vrouw.

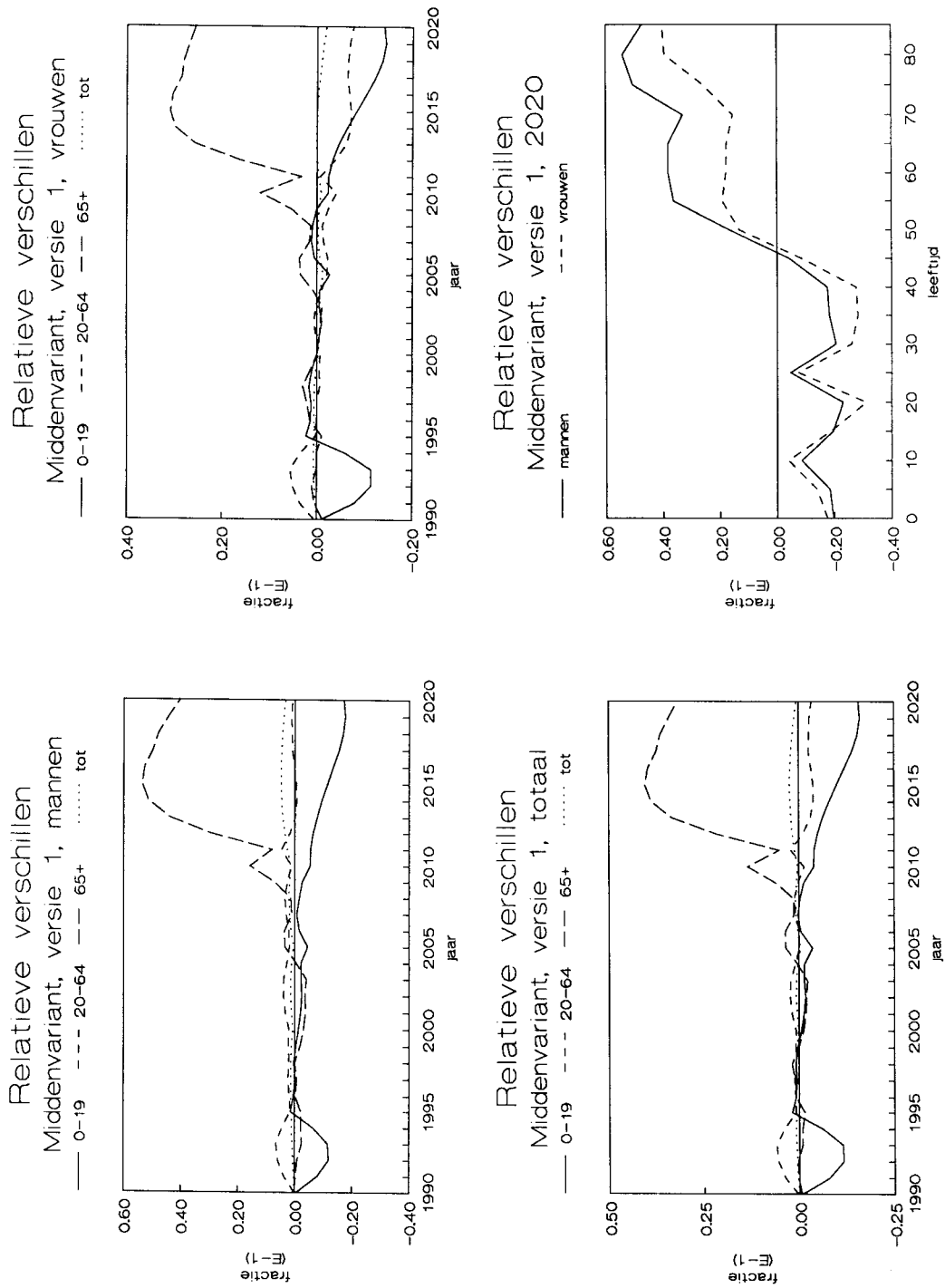
Twee versies zijn doorgerekend. Beide versies gaan uit van de tweede optie - doortrekken van de trend tussen 1984 en 1989 - , en in geval van het migratiesaldo de vierde optie - expliciet opgegeven cijfers. De twee versies verschillen in hoever de trends doorgetrokken verondersteld worden in het jaar 2010. De levensverwachtingen (in jaren) en gemiddeld kindertal per vrouw zijn:

	versie 1	versie 2
mannen	75.1	74.6
vrouwen	81.4	81.0
gemiddeld kindertal	1.65	1.68

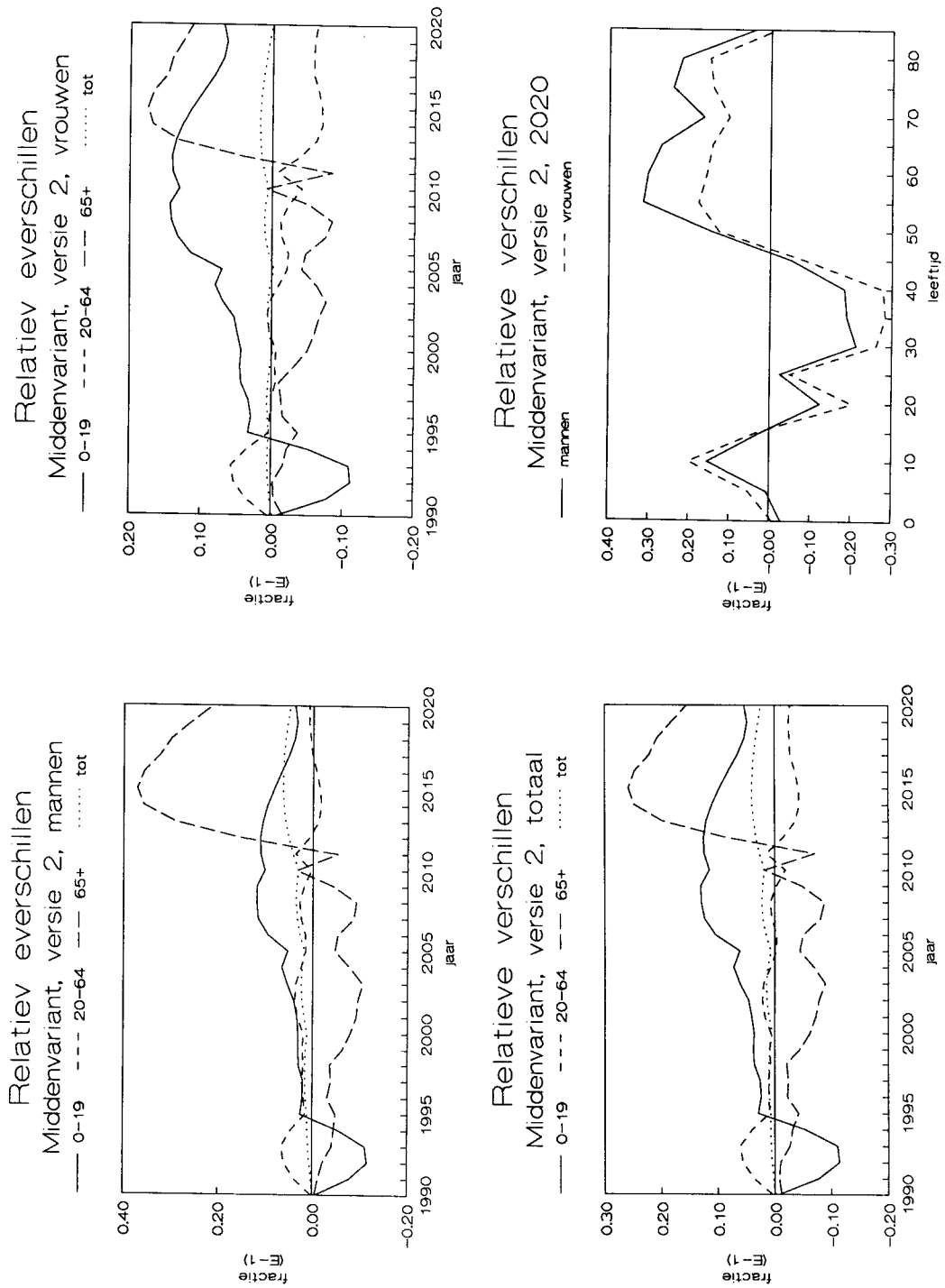
De uitkomsten van de eerste versie van de middenvariant van het demografische model zijn weergegeven in **grafiek 9**, die van de tweede in **grafiek 10**. De resultaten van beide versies verschillen aanzienlijk. Nb: let op de verschillende schalen in de twee grafieken.

- De eerste versie geeft goede schattingen van de aantallen tot ca 60 jaar en de totaalaantallen. Ze geeft overschattingen van de aantallen vanaf 60 jaar, met name van de 85+ groep.
- De tweede versie geeft betere schattingen voor de oudere leeftijdsgroepen, maar geeft een lichte overschatting van met name de totaalaantallen.
- Opvallend is, dat de eerste versie voor alle leeftijdsgroepen goede schattingen geeft tot 2010, maar daarna mindere voor met name de oudere leeftijdsgroepen.

De resultaten worden enigszins vertekend door de linearisatie van de CBS prognoses tussen met name de jaren 2010 en 2020. Worden slechts de aantallen voor deze jaren vergeleken, dan 'ogen' de resultaten van met name de eerste modelversie veel beter. De resultaten van de berekeningen wijzen erop, dat de vertaling van CBS kernhypothese naar modelveronderstellingen niet volledig in overeenstemming is met de modelvergelijkingen, waardoor na verloop van tijd de modeluitkomsten niet meer 'sporen' met de CBS prognoses.



Grafiek 9: De uitkomsten van de eerste middenvariant van het ontwikkelde demografische model.



Grafiek 10: De uitkomsten van de tweede middenvariant van het ontwikkelde demografische model.

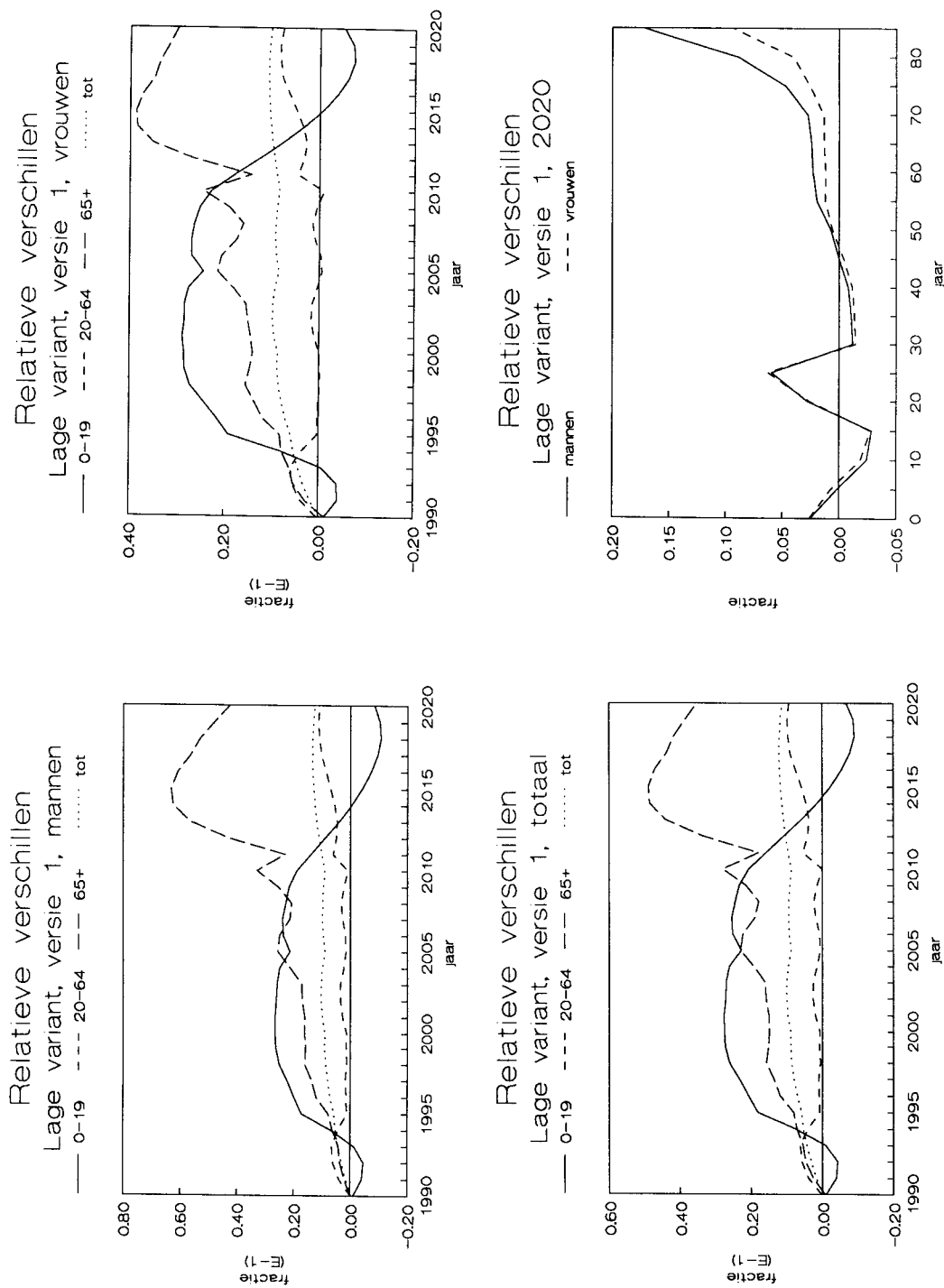
7.6 De lage variant.

De lage variant beschrijft een lage raming van de toekomstige bevolkingsaantallen. Deze lage raming is het gevolg van door het CBS in vergelijking met de middenvariant veronderstelde lagere levensverwachting, lager migratiesaldo en een lager gemiddeld kindertal per vrouw. Deze 'pessimistische' veronderstellingen zijn vertaald in hogere sterftefracties, lagere migratie-aantallen en fertiliteitscijfers.

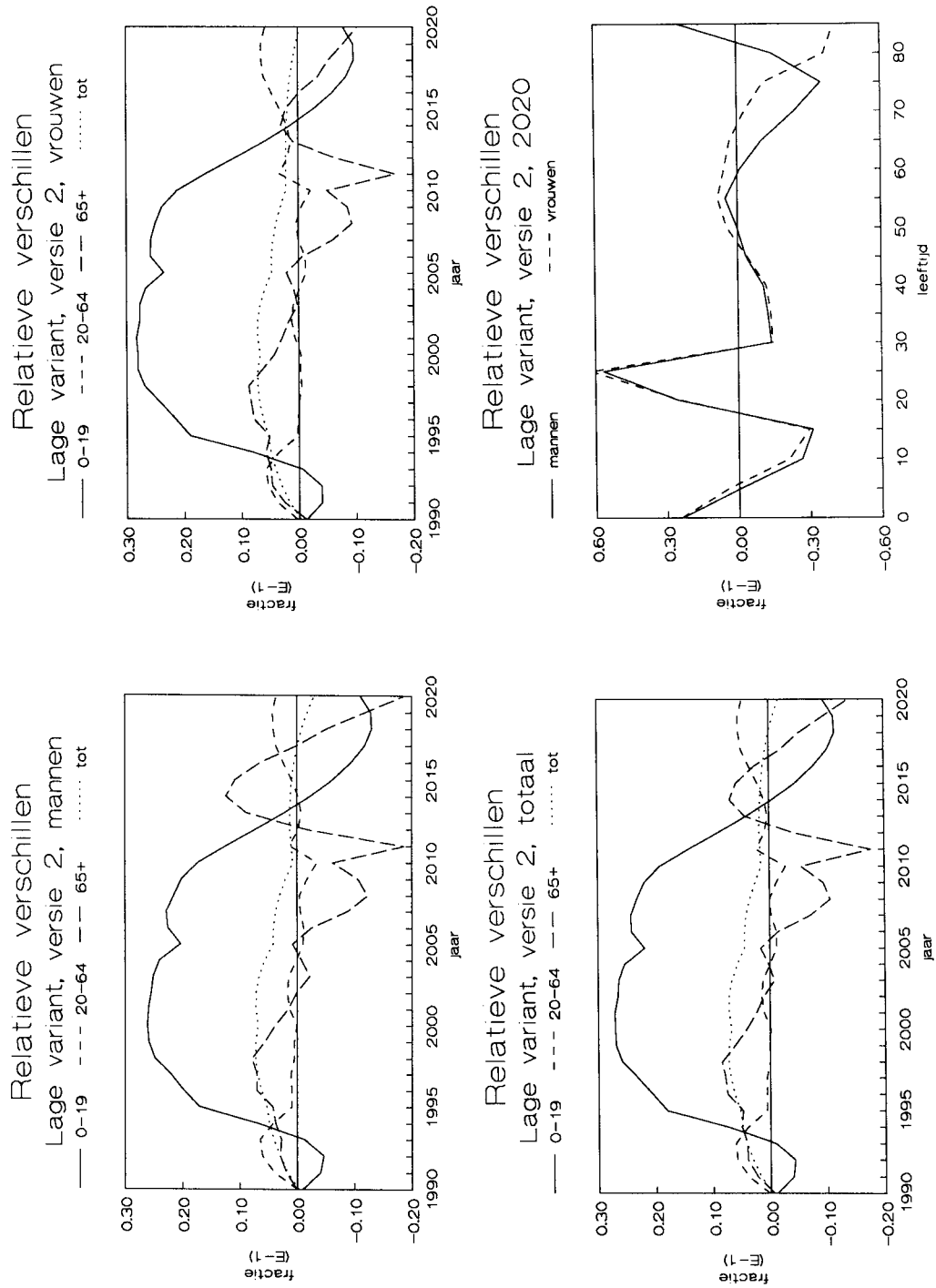
Twee versies zijn weer doorgerekend. Beide versies gaan met betrekking tot de invoer weer uit van de tweede optie voor de sterfte en geboorte - doortrekken van de trend - en de vierde optie voor de migratie - gegeven aantallen. Het gemiddeld kindertal per vrouw is voor beide versies gelijk. De modelaannames leiden tot de volgende levensverwachtingen:

	versie 1	versie 2
mannen	74.1	72.6
vrouwen	80.5	79.3

Zowel de uitkomsten van de eerste versie van de lage variant van het demografische model (zie **grafiek 11**) en die van de tweede versie (zie **grafiek 12**) als de verschillen ertussen vertonen analogieën met die van de middenvariant (zie vorige paragraaf). Zo geeft de eerste versie redelijke schattingen voor de leeftijdsgroepen tot ca 75 jaar, maar mindere voor oudere groepen. Daarentegen overschat de eerste versie de totaalaantallen. Beide versies geven forse overschattingen voor de oudste leeftijdsgroepen, behalve de tweede voor vrouwen. Opvallend is, dat de uitkomsten van de tweede modelversie voor de geaggregeerde leeftijdsgroepen in het jaar 2020 wel goed zijn; afwijkingen tot ruim 1%.



Grafiek 11: De uitkomsten van de eerste lage variant van het ontwikkelde demografische model.



Grafiek 12: De uitkomsten van de tweede lage variant van het ontwikkelde demografische model.

7.7 De hoge variant.

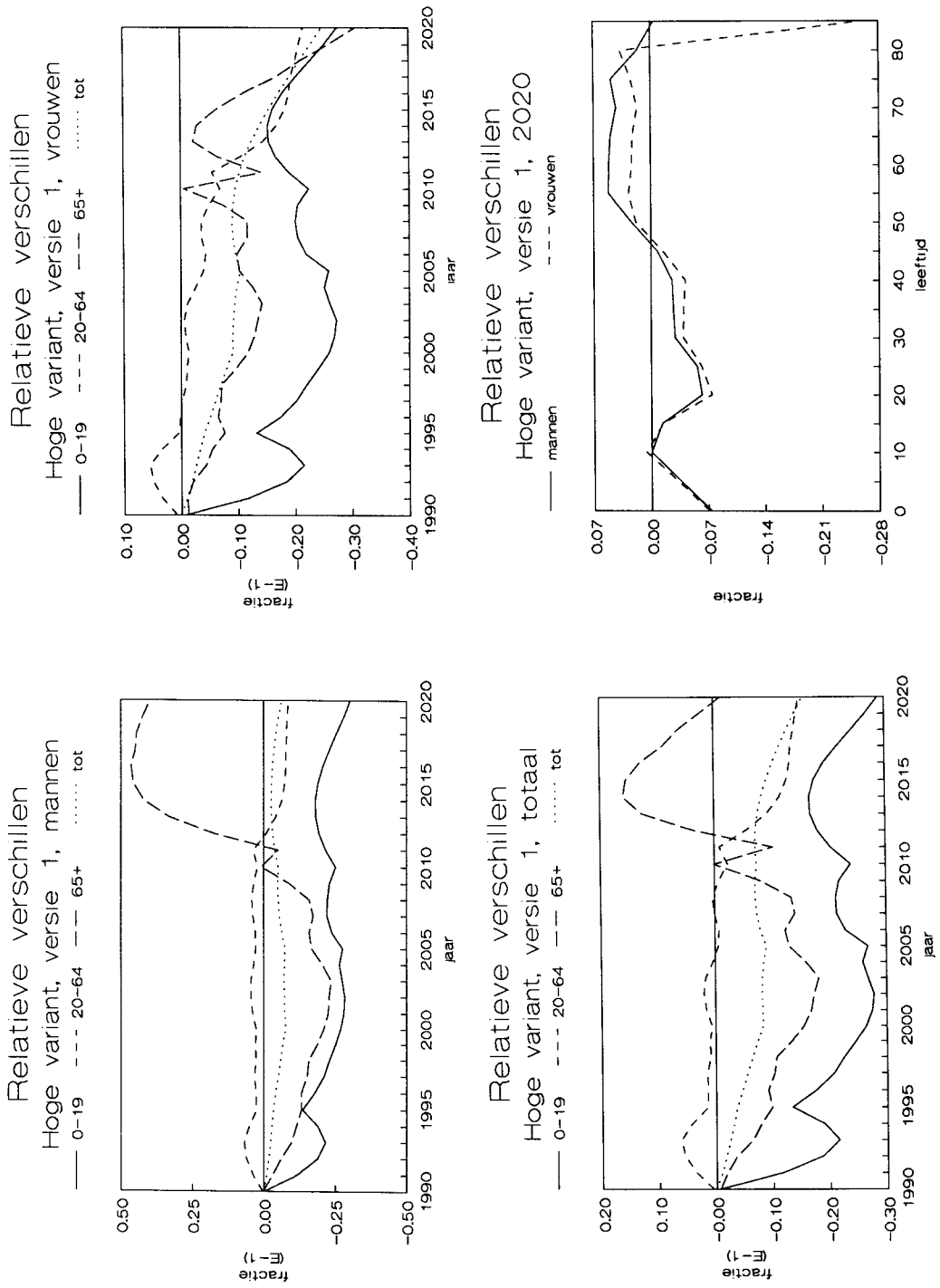
De hoge variant beschrijft een hoge raming van de toekomstige bevolkingsaantallen. Deze hoge raming is het gevolg van door het CBS in vergelijking met de middenvariant veronderstelde hogere levensverwachting, hoger migratiesaldo en een hoger gemiddeld kindertal per vrouw. Deze 'optimistische' veronderstellingen zijn vertaald in lagere sterftefracties, hogere migratie-aantallen en fertiliteitsfracties.

Twee versies (zie **grafiek 13** resp. **grafiek 14**) zijn weer doorgerekend, voor migratie de door het CBS opgegeven migratiesaldo's, voor sterfte en migratie de fracties gebaseerd op de voor het jaar 2010 gestelde levensverwachting resp. gemiddeld kindertal per vrouw en de trends tussen de jaren 1984 en 1989. De versies 1 en 2 van de hoge variant van het demografische model verschillen in de resulterende levensverwachting vanaf het jaar 2010 en het gemiddeld kindertal per vrouw:

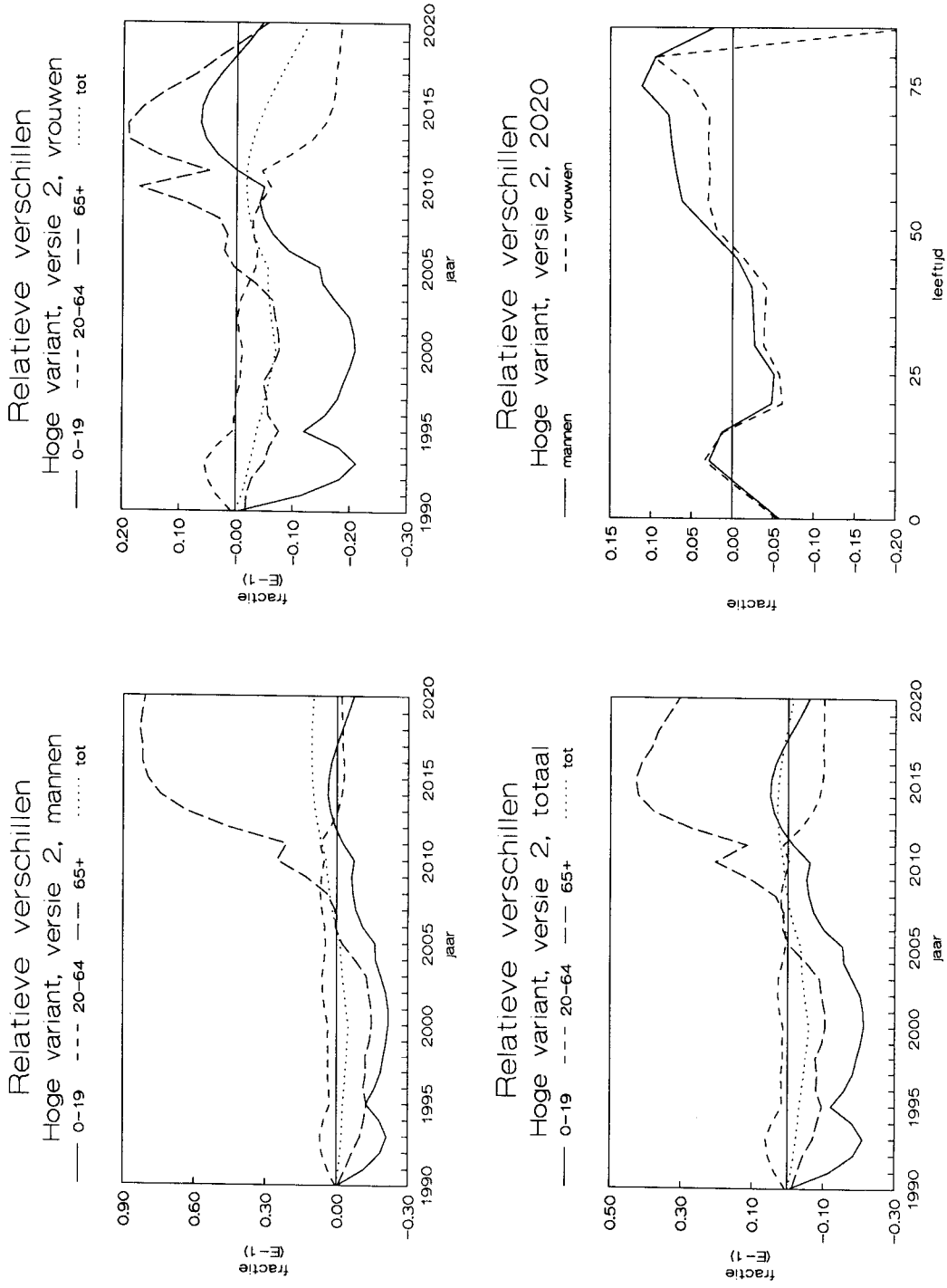
	versie 1	versie 2
mannen	76.2	77.3
vrouwen	82.3	83.2
gemiddeld kindertal	1.85	1.89

De resultaten van de eerste versie van de hoge variant van het demografische model zijn weergegeven in **grafiek 12**, van de tweede versie in **grafiek 13**. De versies vertonen duidelijke verschillen voor mannen en vrouwen.

- De eerste versie geeft voor mannen redelijke tot goede resultaten voor mannen tot 2010. Vanaf 2010 zijn de schattingen voor de oudere en jongere leeftijdsgroepen minder goed.
- De eerste versie geeft voor vrouwen in de tijd toenemende onderschattingen voor alle leeftijdsgroepen.
- De tweede versie geeft voor mannen voor alle jaren en voor alle leeftijdsgroepen goede schattingen, behalve voor de oudere vanaf 2010.
- De tweede versie geeft voor vrouwen goede schattingen tot en met name in het jaar 2010, maar vanaf 2010 toenemende onderschattingen.
- De tweede versie geeft voor alle leeftijdsgroepen en voor alle jaren goede schattingen, hoewel vanaf 2010 de aantallen in (licht) toenemende mate onderschat worden.
- In het jaar 2020 zijn de schattingen voor de eerste en tweede versie vergelijkbaar. De eerste versie geeft voor mannen voor de oudere leeftijdsgroepen betere resultaten, de tweede voor vrouwen voor de 85+ groep een duidelijk minder grote onderschatting.



Grafiek 13: De uitkomsten van de eerste hoge variant van het ontwikkelde demografische model.



Grafiek 14: De uitkomsten van de tweede hoge variant van het ontwikkelde demografische model.

7.8 De geboorte en sterfte-aantallen.

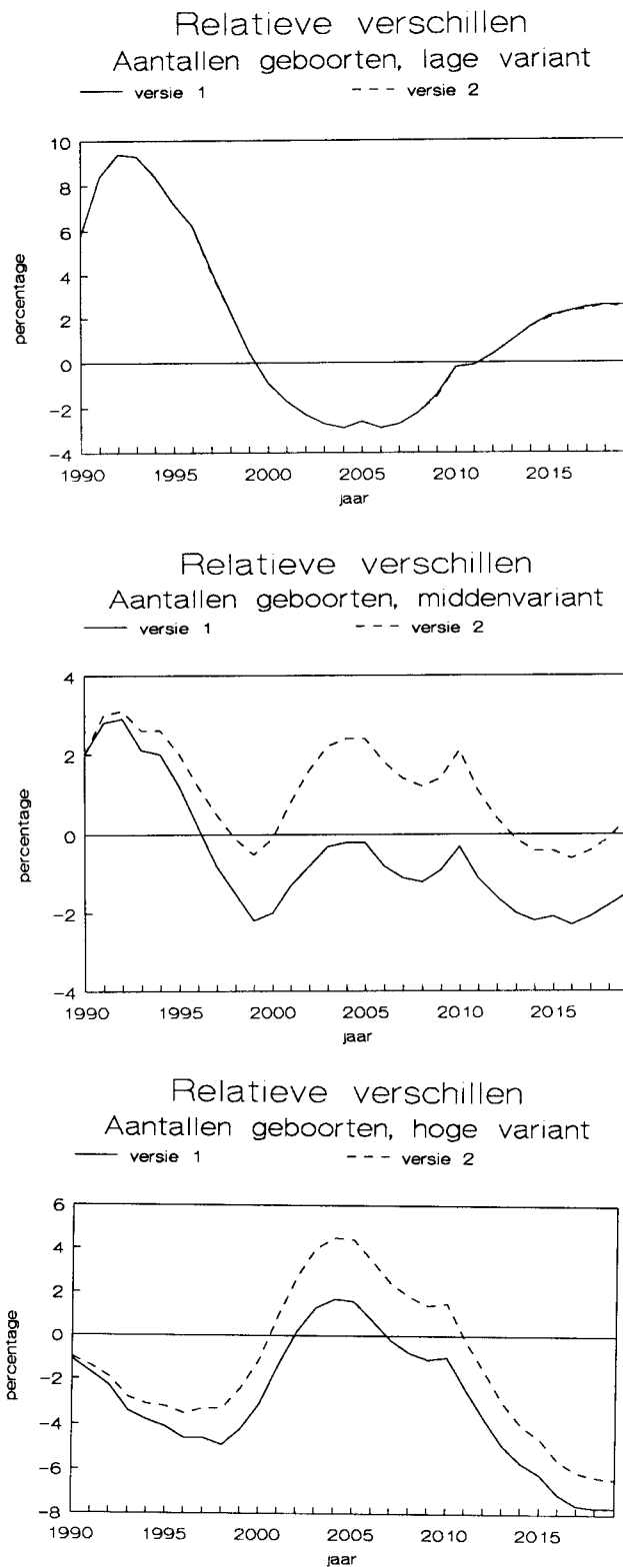
In deze paragraaf worden twee geaggregeerde resultaten van het ontwikkelde demografische model vergeleken met de betreffende CBS uitkomsten, te weten de totale geboorte-aantallen en sterfte-aantallen. Voor elk van de drie varianten zijn de relatieve verschillen tussen de met het ontwikkelde model berekende geboorte- resp. sterfte-aantallen en die van het CBS grafisch gegeven. De relatieve verschillen tussen de geboorte-aantallen zijn weergegeven in **grafiek 15**, die tussen de sterfte-aantallen in **grafiek 16**.

De volgende opmerkingen kunnen op basis van grafiek 15 gemaakt worden met betrekking tot de relatieve verschillen tussen de berekende geboorte-aantallen:

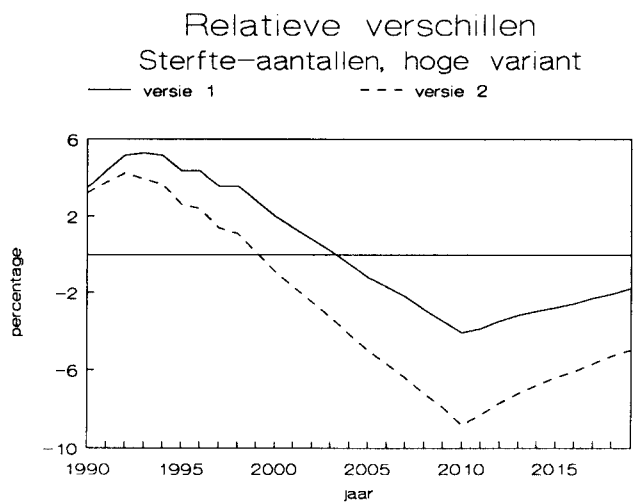
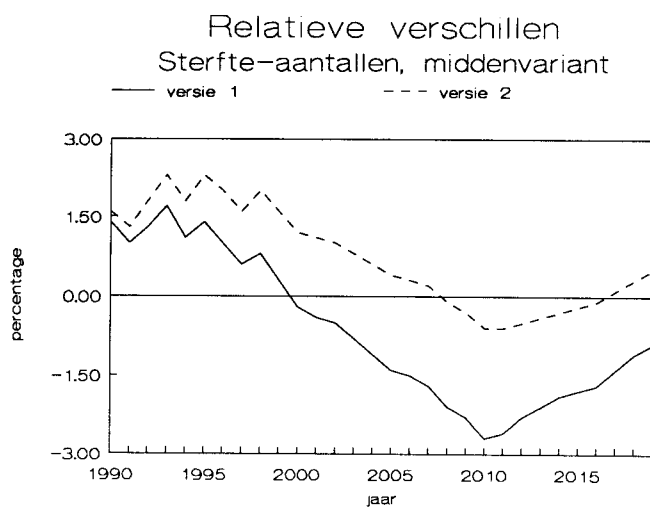
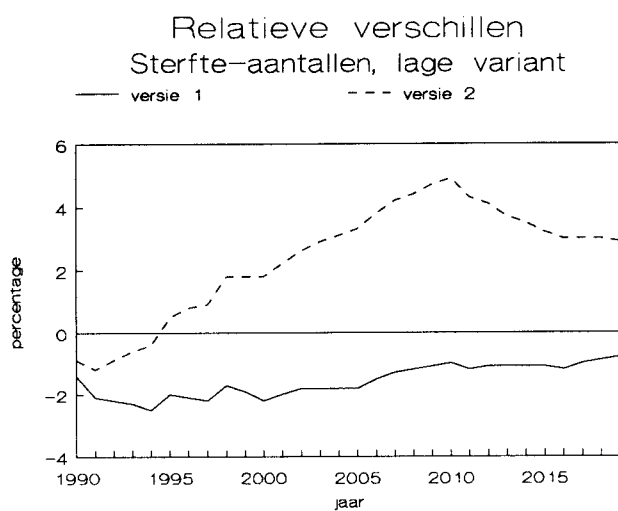
- Voor de lage variant verschillen de twee modelversies slechts in de mortaliteit (zie grafiek 16), niet in de fertiliteit. Voor de midden- en hoge variant verschillen de versies in beide factoren. Dit leidt ertoe, dat de geboorte-aantallen voor de lage variant vrijwel identiek zijn, maar dat voor de overige varianten deze verschillen. Opvallend is, dat de verschillen na een zekere tijd vrijwel constant blijven.
- Voor de lage variant worden de aantallen aanvankelijk enorm overschat (tot bijna 10%), maar de schattingen stabiliseren zich na verloop van tijd. De verschillen zijn uiteindelijk minder dan 2%.
- Voor de middenvariant zijn de schattingen het best. De verschillen zijn aanvankelijk ca 2%. Na het jaar 2010 zijn de verschillen voor de tweede modelversie miniem.
- Voor de hoge variant worden de geboorte-aantallen over het algemeen onderschat, met een lichte overschatting rond het jaar 2005.

De verschillen tussen de berekende geboorte-aantallen en de CBS prognoses zijn (deels) terug te vinden in de eerder gepresenteerde resultaten in de vorm van verschillen met betrekking tot de bevolkingsaantallen voor de leeftijdsklasse 0-19. Deze bevolkingsaantallen worden - voor een belangrijk deel - bepaald door de geboorte-aantallen in de voorafgaande jaren. De verschillen tussen deze aantallen en de CBS prognoses lopen daarom achter bij de verschillen voor de geboorte-aantallen.

De verschillen tussen de berekende aantallen en de CBS prognoses zijn overigens (onverwacht) groot, met name voor de lage variant. De verschillen nemen toe tot het jaar 2010. Dit wijst erop, dat de overgangen van de huidige parameterwaarden naar de nieuwe verschillend gemodelleerd zijn.



Grafiek 15: De geboorte-aantallen voor de drie varianten en de twee modelversies met de CBS cijfers vergeleken.



Grafiek 16: De sterfte-aantallen voor de drie varianten en de twee modelversies met de CBS cijfers vergeleken.

Enkele **opmerkingen** op basis van grafiek 16 met daarin de relatieve verschillen tussen de berekende sterfte-aantallen en de CBS prognoses:

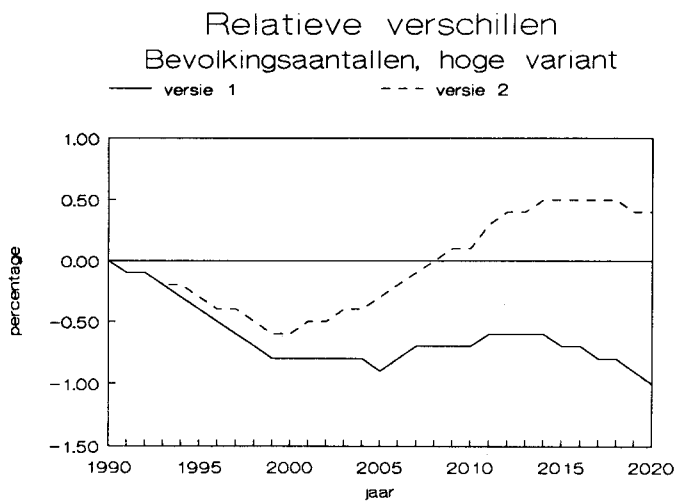
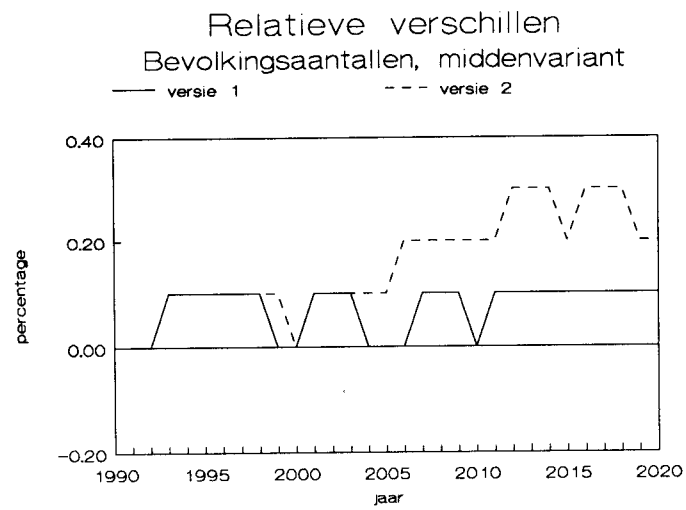
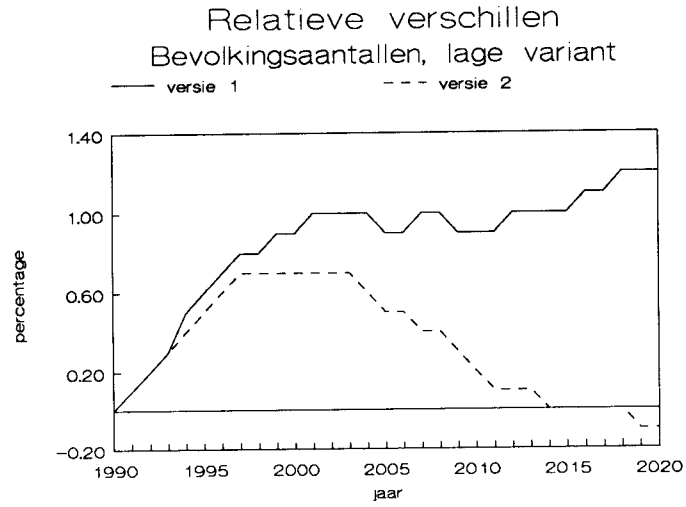
- De sterfte-aantallen worden over het algemeen slecht geschat, met uitzondering van - evenals voor de geboorte-aantallen - de middenvariant.
- Opvallend is, dat de tweede modelversies, die over het algemeen leiden tot betere schattingen van de bevolkingsaantallen voor verschillende leeftijdsgroepen, niet leiden tot betere schattingen van de sterfte-aantallen, met uitzondering van (weer) de middenvariant.
- Het jaar 2010 vormt voor alle varianten en modelversies - met uitzondering van de eerste modelversie van de lage variant een 'knikpunt'.

7.9 De totaalaantallen.

In de voorgaande paragrafen zijn de resultaten van het ontwikkelde demografische model in de vorm van leeftijds- en geslachts-specifieke bevolkingsaantallen tot het jaar 2020 vergeleken met de CBS prognoses. In deze paragraaf worden de door het ontwikkelde model vooruitberekende totaalaantallen vergeleken met de CBS prognoses. Dit gebeurt weer voor de drie varianten voor twee versies van het model (zie paragraaf 7.4). De aannames van de eerste modelversie weerspiegelen telkens de CBS kernhypothesen. De aannames van de tweede modelversie zijn via trial and error zo gekozen, dat de de CBS uitkomsten in het algemeen beter benaderd worden voor de verschillende leeftijdsgroepen. In deze paragraaf wordt onderzocht, hoe goed de benadering is voor de totaalaantallen tot het jaar 2020.

De resultaten van deze vergelijking zijn grafisch weergegeven in **grafiek 17**. Het getrapte verloop van de grafieken wordt veroorzaakt door de discrete verticale stappen van 0.1%. Voor de midden-, lage en hoge variant zijn de relatieve verschillen tussen de modeluitkomsten en de CBS prognoses gegeven, namelijk: $(\text{modeluitkomst} - \text{CBS prognose}) / \text{CBS prognose}$. Bij deze resultaten kunnen de volgende opmerkingen geplaatst worden:

- De totaalaantallen worden over het algemeen beter geschat dan die voor de afzonderlijke leeftijdsgroepen. Dit wijst erop, dat de aantallen voor de afzonderlijke leeftijdsgroepen niet systematisch onder- of overschat worden. De onder- en overschattingen compenseren elkaar, zodat de totaalaantallen relatief goed geschat worden.
- Voor de lage variant schat de tweede modelversie de totaalaantallen vooral op de langere termijn beter dan de eerste. Voor de midden- en hoge variant ontlopen de beide schattingen elkaar weinig.



Grafiek 17: De totaalaantallen van de twee modelversies en de CBS prognoses voor de drie varianten vergeleken.

- De schattingen voor de middenvariant (beide modelversies) zijn duidelijk beter dan die voor de lage en hoge variant.

Concluderend kan gesteld worden, dat het ontwikkelde demografische model voor de middenvariant de geboorte-, sterfte- en totaal-aantallen beter schat dan voor de lage en hoge variant. De mindere schattingen voor de lage en hoge variant spelen zowel tijdens de overgangperiode tussen 1990 en 2010, waarin de fertiliteits- en mortaliteitscijfers veranderen, als vanaf 2010, waarin deze cijfers hun definitieve waarde houden.

7.10 Discussie en conclusies.

Het is moeilijk om op basis van deze beperkte analyse van het ontwikkelde demografische model reeds definitieve conclusies te trekken met betrekking tot het ontwikkelde demografische model, met name in vergelijking met het demografisch model van het CBS. In plaats daarvan worden enkele kanttekeningen (opnieuw) geplaatst bij het model.

Verschillende invoergrootheden.

De CBS modelaannames voor de sterfte zijn geformuleerd in termen van gemiddelde levensverwachting, die van het ontwikkelde model in termen van leeftijdsafhankelijke sterftefracties. In het CBS model worden de sterftefracties berekend op basis van een gefitte curve van de levensverwachting bij geboorte. De huidige trends in de sterftefracties worden zover doorgetrokken, dat de bijbehorende levensverwachting overeenkomt met de genoemde gefitte curve. In het ontwikkelde model worden eveneens de trends in de sterftefracties zover doorgetrokken, dat de gestelde levensverwachting bereikt wordt. Vervolgens worden voor de tussenliggende jaren de sterftefracties berekend als een (gewogen) gemiddelde van die aan het begin en het einde van de overgangperiode. De verschillen tussen beide benade-ringswijzen zijn overigens klein.

Met betrekking tot de geboorte gelden analoge opmerkingen. Hierbij speelt bovendien het probleem, dat het gemiddeld kindertal niet slechts bepaald wordt door de fertiliteitscijfers, maar ook door de leeftijdsopbouw van de vrouwelijke bevolking. Dit houdt in, dat bij een veranderende leeftijdsopbouw (een niet-stabiele situatie) het gemiddeld kindertal en de fertiliteitscijfers niet beide constant verondersteld kunnen worden. Het CBS veronderstelt een constant gemiddeld kindertal (vanaf 2010), het ontwikkelde model daarentegen een constante fertiliteit.

Overschatting van de 85+ groep.

Bij alle vooruitberekeningen wordt de omvang van de oudste leeftijdsgroepen overschat, met name voor mannen en voor de eerste modelversies. De aan de eerste modelversies ten grondslag liggende

aannames zijn afgeleid van de CBS kernhypothesen. Maar zelfs de tweede modelversies, waarbij de gemiddelde levensverwachting geforceerd verlaagd wordt, leiden tot (kleinere) overschattingen.

Hierbij speelt het volgende waarschijnlijk een rol. Als invoer gebruikt het model gegeven jaarlijkse sterftefracties. Deze hebben betrekking op de gemiddelde omvang van de bevolkingsgroepen over het betreffende jaar (onafhankelijke maat), maar worden toegepast op de omvang aan het begin van het jaar (afhankelijke maat). Vanwege de optredende vergrijzing nemen met name de oudste leeftijdsgroepen in de loop van een jaar in aantal toe. Dit leidt tot een onderschatting van de sterfte en daarmee een overschatting van de omvang aan het einde van het jaar.

8 Spaarzame demografische modellen.

8.1 Inleiding.

In de voorafgaande hoofdstukken 5 tot en met 7 is een demografisch model gepresenteerd, dat voor zijn invoer gebruik maakt van door het CBS gepubliceerde cijfers. Dit geldt met name voor de simulaties van het model, zoals beschreven in hoofdstuk 6. Hierbij is het 'beschrijvende vermogen' van het model getoetst aan de hand van een vergelijking voor het verleden van de modeluitkomsten met de waargenomen bevolkingsaantallen.

In zekere zin geldt dit ook voor de simulaties van het model, zoals beschreven in hoofdstuk 7. Hierbij is het 'voorspellende vermogen' van het model getoetst aan de hand van een vergelijking voor de toekomst van de uitkomsten van het model met de door het CBS vooruitberekende aantallen. Dit is gedaan voor de lage, midden- en hoge variant. Voor elke variant zijn telkens twee modelversies beschouwd. Bij de eerste versies werden als invoercijfers de bewerkte CBS kernhypothesen gekozen. Deze bewerkingen betroffen met name het omzetten van levensverwachtingscijfers in sterftefracties en het omzetten van het gemiddelde kindertal en de gemiddelde leeftijd van de moeder bij geboorte in fertiliteitscijfers. Bij de tweede versies werden de invoercijfers zo gekozen, dat het model een in vergelijking met de eerste versies betere fit met de CBS vooruitberekende aantallen te zien gaf.

De laatstgenoemde tweede versies van het model vormen een overgang naar de modellen, zoals beschreven in het onderhavige hoofdstuk. In dit hoofdstuk worden enkele demografische modellen gepresenteerd, waarvan de onbekende parameters geschat worden aan de hand van gegeven bevolkingsaantallen. Deze modellen worden gebruikt om de door het CBS vooruitberekende bevolkingsaantallen (met name voor de middenvariant) te reproduceren.

De beschouwde modellen zijn spaarzaam in die zin, dat het aantal te schatten parameters klein is in vergelijking met het aantal invoergegevens van het oorspronkelijke demografische model (zie hoofdstuk 5). Het oorspronkelijke demografische model maakt gebruik van sterftefracties, migratie-aantallen, en fertiliteitscijfers, voorzover mogelijk verdeeld naar geslacht en in vijfjaars leeftijdsklassen. De in dit hoofdstuk te beschrijven modellen maken gebruik van slechts één of twee soorten fracties, verdeeld naar geslacht en in bredere leeftijdsklassen. Deze fracties (parameters) kunnen geïnterpreteerd worden als netto-uitstroomfracties resp. fertiliteitscijfers.

De parameters zijn geschat middels het minimaliseren van een doelfunctie, die de fouten in de berekende aantallen aggregereert. De doelfunctie bestaat uit de som over de beide geslachten en de onderscheiden leeftijdscategorieën van de gewogen kwadraatverschillen tussen de met het model berekende aantallen en de CBS prognoses.

8.2 Het algemene demografische model.

De verschillende demografische modellen bezitten de volgende kenmerken (zie ook hoofdstuk 5). In het spaarzame model wordt de Nederlandse bevolking verdeeld naar geslacht en in verschillende leeftijds-categorieën. In principe worden de volgende (vijftienjaars) leeftijdscategorieën onderscheiden: 0-14, 15-29, 30-44, 45-59, 60-74, 75-84 en 85+.

Het model beschrijft de jaarlijkse veranderingen in de aantallen voor de genoemde bevolkingsgroepen. Deze veranderingen betreffen een uitstroom uit elk van de groepen, een doorstroom van elk van de groepen naar de eerstvolgende - behalve voor de laatste restgroep - en een instroom in de eerste groep. De genoemde uitstroom kan geïnterpreteerd worden als de som van sterfte en netto-migratie. De doorstroom betreft veroudering, de instroom betreft geboorte. De wiskundige modelvergelijkingen zijn beschreven in bijlage C.

De parameters betreffen de fracties behorende bij de uitstroom en in sommige versies ook de fertiliteit. De fracties worden in eerste instantie vanaf het jaar 1989 constant over de tijd gekozen. Daarnaast zijn varianten beschouwd, waarbij tot het jaar 2005 de fracties lineair veranderen. De geboorte-aantallen worden in eerste instantie als gegeven beschouwd, namelijk de aantallen zoals gepubliceerd door het CBS. Daarnaast zijn varianten beschouwd, waarbij van te schatten fertiliteitscijfers gebruik wordt gemaakt.

In de basisversie wordt uitgegaan van homogene bevolkingsgroepen. Dit wil zeggen, dat voor elk van de onderscheiden bevolkingsgroepen de verdeling over de afzonderlijke leeftijdsjaren constant is. De jaarlijkse uitstroom uit elk van de groepen is omgekeerd evenredig met de breedte van het leeftijdsinterval van elk van de groepen. Daarnaast zijn varianten beschouwd, waarbij een zekere heterogeniteit van elk van de onderscheiden groepen verondersteld wordt. Deze heterogeniteit is lineair verondersteld; het aantal individuen verandert lineair (volgens een rechte lijn) met de leeftijd.

De toegepaste doelfunctie beschrijft de som over mannen en vrouwen en over de onderscheiden bevolkingsgroepen van de gewogen kwadraatverschillen tussen de berekende aantallen en de CBS uitkomsten. De wegingscoëfficiënten zijn de omgekeerden van de betreffende aantallen. De kwadraatsommen zijn berekend voor een reeks van jaren en vervolgens gesommeerd. Deze jaren zijn: 1990, 1995, 2000, 2005, 2010 en 2020.

Elk van de beschouwde demografische modellen is 'doorgerekend' voor de drie CBS varianten. Voor de basisversie worden voor alle drie varianten de resultaten gepubliceerd. Omdat de verschillende modelversies voor de drie CBS varianten tot vergelijkbare resultaten leiden, zijn van de overige modelversies slechts de uitkomsten voor de middenvariant gepubliceerd.

Overeenkomstig de uitvoer van de verschillende versies van het oorspronkelijke uitgebreide demografische model worden de volgende uitvoercijfers **grafisch weergegeven** (zie paragraaf 6.1). **Ten eerste:** de relatieve verschillen tussen de door het model berekende aantallen en de CBS uitkomsten, de relatieve fouten. Deze worden gegeven enerzijds voor mannen en vrouwen afzonderlijk en tezamen, anderzijds verdeeld in drie leeftijdscategorieën en tezamen. De relatieve verschillen worden gedefinieerd als het verschil tussen de model- en CBS-uitkomsten gedeeld door de CBS-uitkomsten.

Ten tweede: de relatieve verschillen tussen de berekende aantallen en de CBS uitkomsten in het jaar 2020, het laatste jaar van de simulatie, voor mannen en vrouwen afzonderlijk en verdeeld naar leeftijd.

Van een aantal versies van het demografische model wordt tevens een gevoeligheidsanalyse uitgevoerd. Deze gevoeligheidsanalyse is bedoeld om de invloed van de afzonderlijke parameters op de modelresultaten te bepalen. Tot de relevante modelresultaten worden gerekend de aantallen in het jaar 2020 van de simulatie en de waarde van de doelfunctie. Onderzocht wordt telkens hoe de genoemde modelresultaten veranderen onder een - overigens willekeurig gekozen - 1 procentse toename van elk van de parameters. De toename wordt gerekend vanaf de berekende optimale parameterwaarden voor de middenvariant.

8.3 De basisversie.

De basisversie bezit de volgende kenmerken (zie paragraaf 8.2):

- de (te schatten) overgangsfracties worden vanaf 1989 constant over de tijd verondersteld,
- voor de geboorte-aantallen worden de gegeven CBS cijfers gebruikt.

De uitkomsten van de gevoeligheidsanalyse zijn weergegeven in **tabel 1**. Omdat de modellen voor mannen en vrouwen niet gekoppeld zijn, zijn de modeluitkomsten voor mannen resp. vrouwen niet gevoelig voor de parameterwaarden voor vrouwen resp. mannen.

Tabel 1: De uitkomsten van de gevoeligheidsanalyse voor de basisversie van de middenvariant.

relatieve fout in bevolkingsaantallen in 2020								
	0-14	15-29	30-44	45-59	60-74	75-84	85+	doelfunctie
parameter								
mannen								
std	3.8	-7.0	10.6	1.6	-8.1	0.1	2.7	275.251
0-14								275.250
15-29								275.252
30-44			10.7					275.254
45-59				1.7				275.272
60-74					-8.3	-0.2	2.4	275.257
75-84						-0.5	2.0	275.308
85+							1.7	275.269
vrouwen								
std	5.8	-6.4	4.5	-6.4	0.3	6.6	4.8	275.251
0-14								275.263
15-29								275.252
30-44								275.252
45-59				-6.3				275.253
60-74								275.250
75-84						6.2	4.3	275.313
85+							3.8	275.354

Noot: relatieve fout = $100 * (\text{berekende waarde} / \text{CBS-waarde} - 1)$; doelfunctie = som van gewogen kwadraatverschillen, std = relatieve fouten voor optimale parameterwaarden (overgangsfracties), overige regels: relatieve fouten na 1% verandering van waarde van parameter voor leeftijdsklasse, zoals aangegeven in eerste kolom. De relatieve fouten in de lege cellen in rechterbovenhelft van tabel zijn verwaarloosbaar klein, in de linkeronderhelft per definitie gelijk aan 0.

Bij de uitkomsten van de gevoeligheidsanalyse kunnen de volgende **kanttekeningen** geplaatst worden:

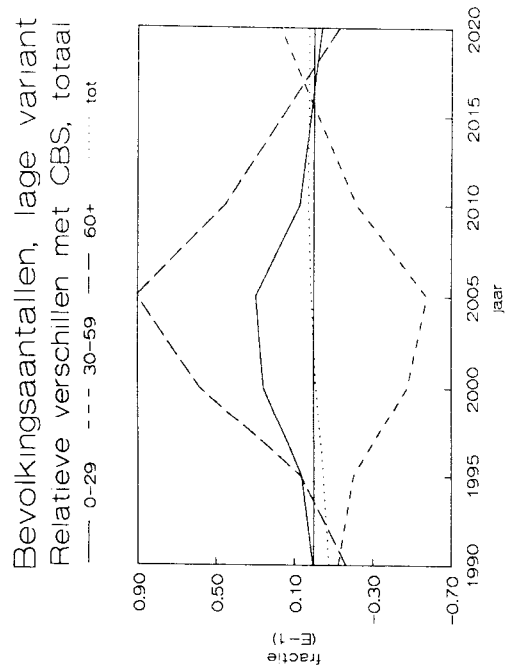
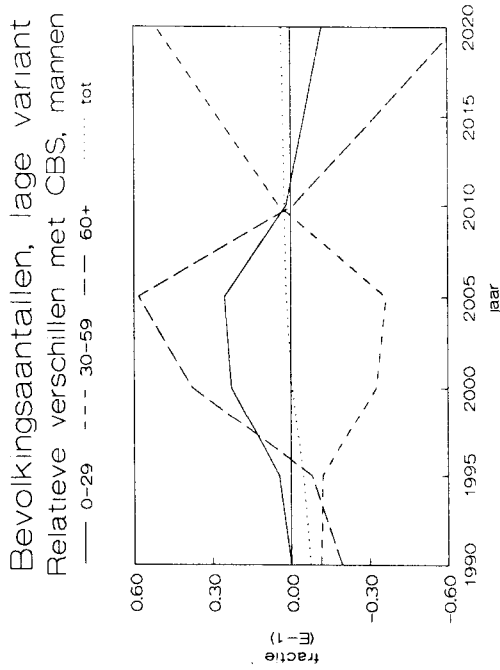
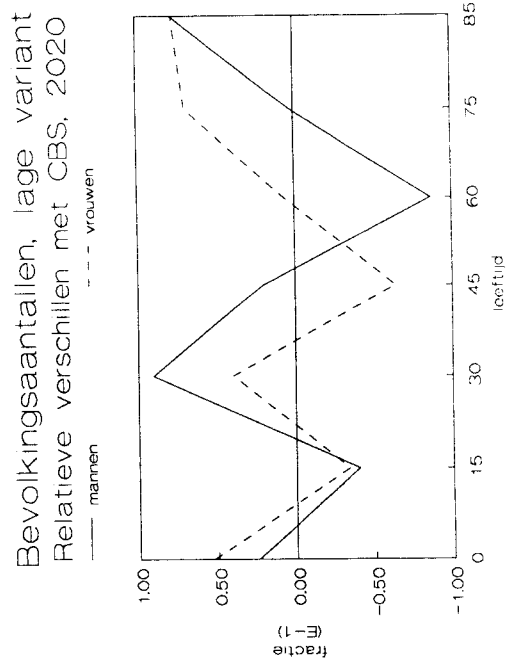
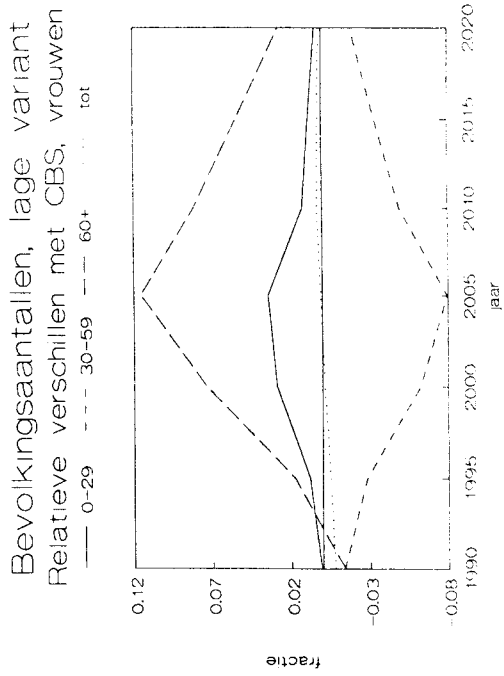
- De demografische modellen voor mannen en vrouwen zijn niet slechts ontkoppeld, maar bovendien hiërarchisch in de leeftijdscategorieën. Een verandering van een parameterwaarde heeft slechts invloed op de betreffende en oudere leeftijdsgroepen, maar niet op de jongere. Dit uit zich in een driehoek-vorm van de tabel.
- Het 'theoretische uitsmeereffect' als gevolg van de gekozen modelstructuur is nauwelijks waarneembaar. Een verandering voor de leeftijdsgroep 45-59 wordt - theoretisch - in de loop van de tijd versneld uitgesmeerd over de oudere leeftijdsgroepen. De modelresultaten geven aan, dat de grootte van het 'uitsmeereffect' beperkt is.
- De veronderstelde andere parameterwaarden leiden tot grotere veranderingen van de relatieve fout in de bevolkingsaantallen dan van de doelfunctie. Meetbare veranderingen voor de bevolkingsaantallen treden slechts op voor de hogere leeftijdsgroepen.

De conclusie uit de gevoeligheidsanalyse luidt daarom, dat met name de aantallen voor de oudere leeftijdsklassen gevoelig zijn voor de gekozen modelstructuur en de parameterwaarden.

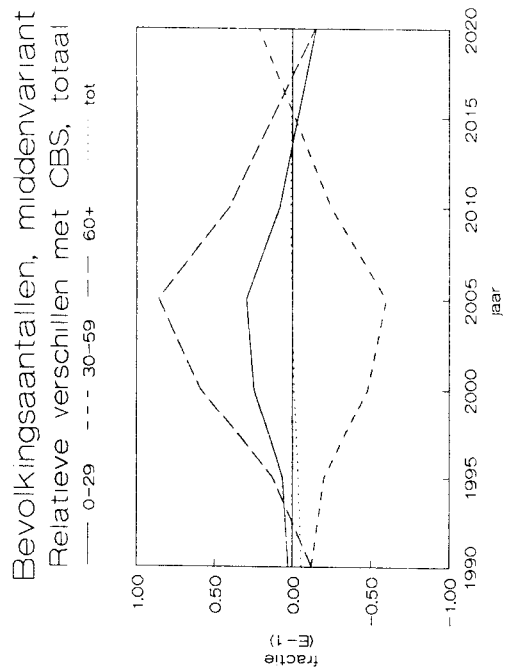
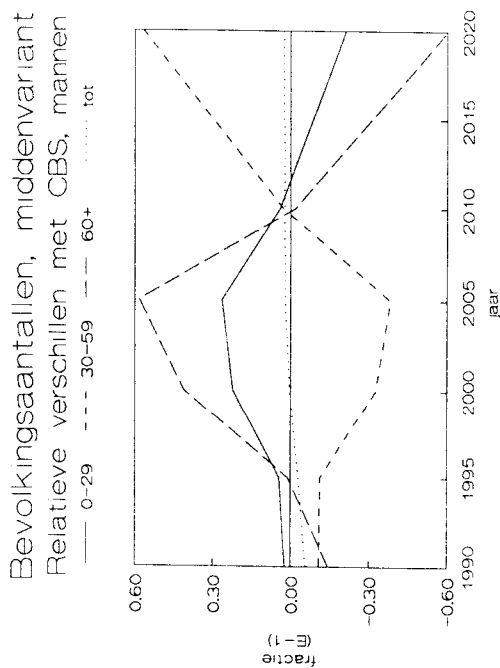
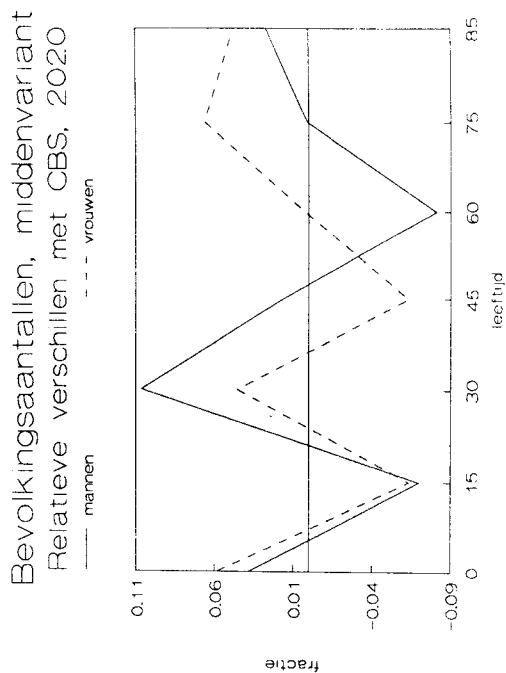
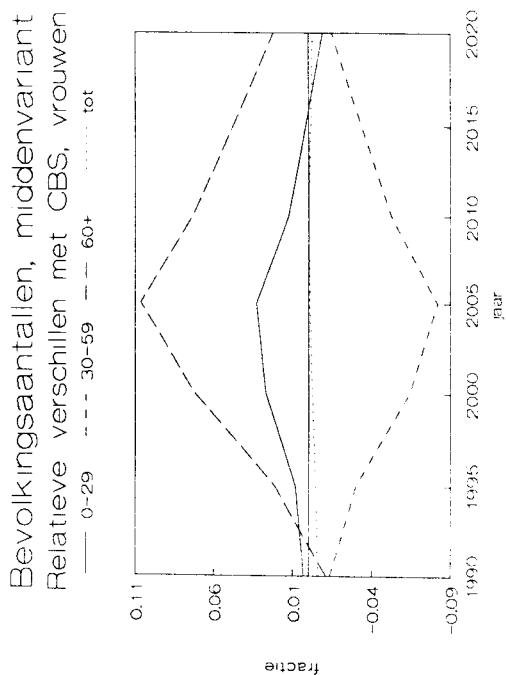
De uitkomsten van de basisversie van het spaarzame demografische model zijn voor de drie CBS varianten in de volgende grafieken weergegeven: voor de lage variant in **grafiek 18**, voor de middenvariant in **grafiek 19**, en voor de hoge variant in **grafiek 20**.

Bij de resultaten van het model voor de drie varianten kunnen de volgende **kanttekeningen** geplaatst worden:

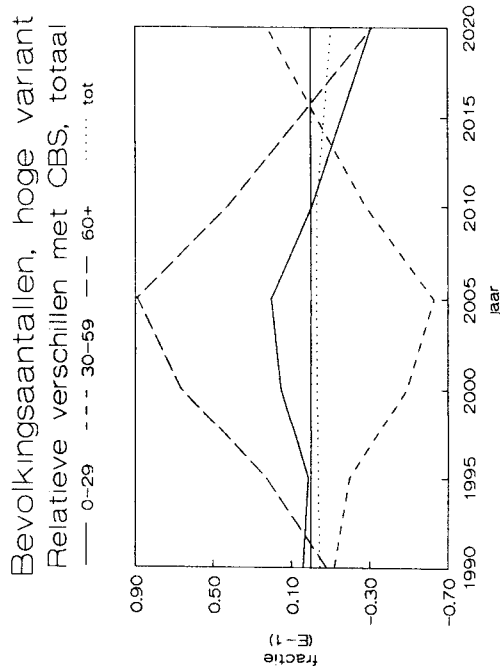
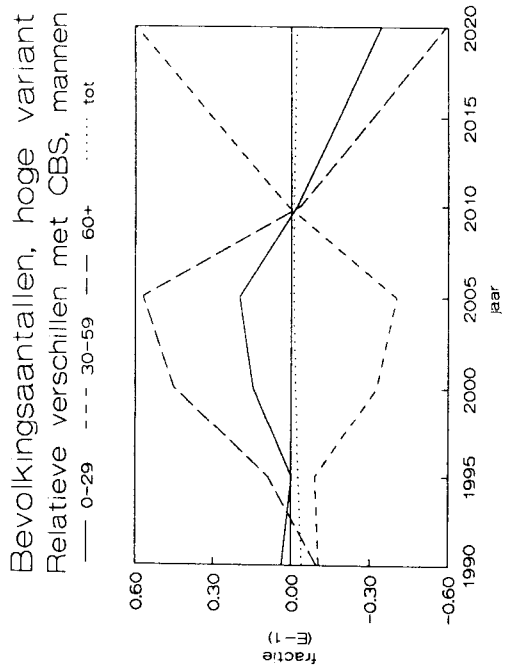
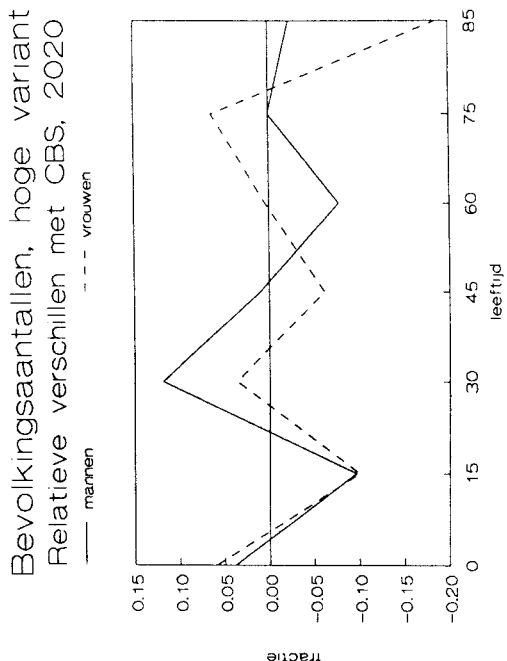
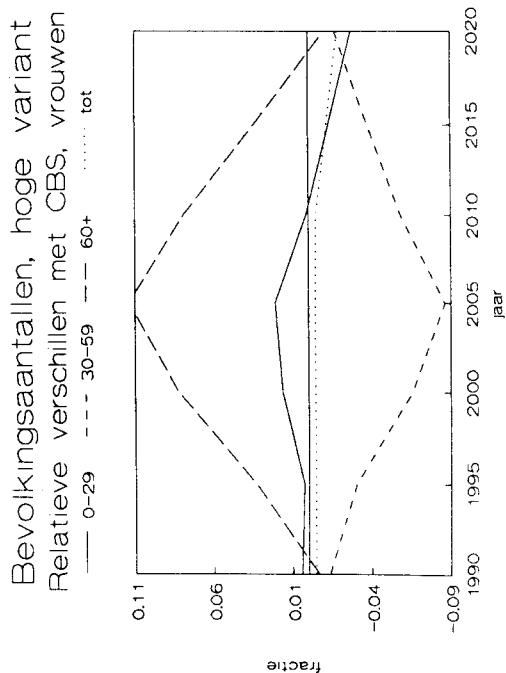
- De modelresultaten zijn voor mannen en vrouwen zeer verschillend. Tot het jaar 2005 verlopen de curves gelijkvormig, maar na het 'knikpunt' voor het jaar 2005 worden voor vrouwen de over- en onderschattingen slechts langzaam 'gecorrigeerd', terwijl voor mannen deze doorschieten naar onder- resp. overschattingen.
- Voor elk van de leeftijdscategorieën voor mannen en vrouwen afzonderlijk zijn de relatieve fouten voor de drie varianten vrijwel gelijkvormig over de tijd. Dit geldt ook voor de relatieve fouten in het jaar 2020. Een uitzondring hierop vormt de hoge variant voor de oudste leeftijdsgroepen, die in vergelijking met de overige twee varianten een grote onderschatting voor met name vrouwen geeft.
- De relatieve fouten voor de oudste en jongste leeftijdsgroep vormen ongeveer het spiegelbeeld van die voor de middelste. De relatieve fouten vertonen enkele duidelijke 'knikpunten'. De relatieve fouten voor de totale bevolking zijn veel kleiner en verlopen veel 'gladder'.



Grafiek 18: De uitkomsten van de lage variant van de basisversie van het spaarzaam demografische model.



Grafiek 19: De uitkomsten van de middenvariant van de basisversie van het spaarzame demografische model.



Grafiek 20: De uitkomsten van de hoge variant van de basisversie van het spaarzame demografische model.

- Het jaar 2005 vormt een 'knikpunt'. Voor de oudste en jongste groep gaan de relatieve fouten over van over- naar onderschattingen. Voor de middelste groep (vrijwel het spiegelbeeld) gaan deze over van onder- naar overschattingen.
- De relatieve fouten in het jaar 2020 vertonen een 'grillig' verloop over de leeftijd. Een onderschatting voor een leeftijdsgroep wordt gevolgd door een overschatting van de eerstvolgende, gevolgd door een onderschatting van de daaropvolgende enzovoort.

Onderzocht is in hoeverre de keuze van de doelfunctie en het eindjaar 2020 van de simulatie van invloed zijn op de modelresultaten. Een versie is beschouwd, waarvan de doelfunctie ongewogen in plaats van gewogen kwadraatverschillen bevat. De modelresultaten van deze versie verschillen slechts zeer weinig van de oorspronkelijke versie. Tevens is een versie beschouwd, waarvan de doelfunctie eveneens absolute kwadraatverschillen bevat, waarbij als eindjaar van de simulatie 2010 gekozen is. De relatieve fouten in de berekende bevolkingsaantallen blijken duidelijk kleiner te zijn, met name voor vrouwen over de tijd. Voor mannen en in het eindjaar zijn de verschillen minder groot.

De verschillende uitkomsten van het model, zoals in de drie grafieken weergegeven, geven duidelijk enkele zwakke punten van het model aan. Deze zijn:

- De overgangsfracties worden over de tijd constant verondersteld, terwijl het CBS uitgaat van geleidelijke veranderingen van de invoergegevens vanaf 1989 tot het jaar 2010, waarna deze constant zijn.
- De onderscheiden bevolkingsgroepen worden homogeen verondersteld, terwijl met name op hogere leeftijd de aantallen afnemen met de leeftijd. Deze onterechte aanname van homogeniteit leidt voor iedere leeftijdsgroep tot overschattingen van de doorstroom en daarmee tot overschattingen van de uitstroomfractie voor de eerstvolgende leeftijdsgroep.
- Het model veronderstelt, dat alle nieuwgeborenen aan de jongste leeftijdsgroep toegevoegd worden, terwijl de sterfte onder de nieuwgeborenen toch relatief hoog is.

In de volgende paragrafen worden daarom enkele modelversies gepresenteerd, die aan één of meer van de genoemde zwakke punten tegemoet komen.

8.4 De fertiliteitsversie.

De tweede versie verschilt op het volgende punt van de eerste versie (zie paragraaf 8.3):

- De geboorte-aantallen worden bepaald met behulp van te schatten fertiliteitscijfers. Deze zijn om evidente redenen slechts gedefinieerd voor de tweede (15-29) en derde (30-44) leeftijdsgroep van vrouwen.

Ook voor dit model is een gevoeligheidsanalyse uitgevoerd. De uitkomsten van deze gevoeligheidsanalyse zijn weergegeven in **tabel 2**. De demografische modellen voor mannen en vrouwen zijn in deze versie wel gekoppeld via de geboorte, in één richting, namelijk van vrouwen naar mannen.

Tabel 2: De uitkomsten van de gevoeligheidsanalyse voor de fertiliteitsversie van de middenvariant.

	relatief verschil tussen bevolkingsaantallen in 2020							
	0-14	15-29	30-44	45-59	60-74	75-84	85+	doelfunctie
parameter								
mannen								
std	4.9	-6.7	10.6	1.5	-8.2	0.2	2.3	281.554
overgangsfracties								
0-14	5.0							281.564
15-29								281.554
30-44			10.7					281.557
45-59				1.6		0.3	2.4	281.566
60-74					-8.4	0.0	2.1	281.574
75-84						-0.3	1.7	281.596
85+							1.3	281.593
fertiliteitsfracties								
15-29	5.5	-6.4	10.8					281.844
30-44	5.5	-6.4	10.8					281.843

Tabel 2 (vervolg)

vrouwen

std	6.8	-7.0	3.7	-6.5	0.3	6.8	4.9	281.554
-----	-----	------	-----	------	-----	-----	-----	---------

overgangsfracties

0-14	6.9	-6.9						281.557
------	-----	------	--	--	--	--	--	---------

15-29			3.6					281.566
-------	--	--	-----	--	--	--	--	---------

30-44								281.555
-------	--	--	--	--	--	--	--	---------

45-59								281.559
-------	--	--	--	--	--	--	--	---------

60-74								281.554
-------	--	--	--	--	--	--	--	---------

75-84					6.4	4.5		281.586
-------	--	--	--	--	-----	-----	--	---------

85+						3.9		281.637
-----	--	--	--	--	--	-----	--	---------

fertiliteitsfracties

15-29	7.4	-6.7	3.8					281.844
-------	-----	------	-----	--	--	--	--	---------

30-44	7.4	-6.7	3.8					281.843
-------	-----	------	-----	--	--	--	--	---------

Noot: zie tabel 1.

Bij de uitkomsten van de gevoeligheidsanalyse kunnen de volgende **kanttekeningen** geplaatst worden (vergelijk ook die bij de gevoeligheidsanalyse van de basisversie, paragraaf 8.3):

- De waarde van de doelfunctie is iets groter, maar nog wel vergelijkbaar met die voor de basisversie. Beide modellen geven een overeenkomstige algemene 'fit'. Ook de relatieve fouten in de bevolkingsaantallen voor het jaar 2020 zijn vrijwel gelijk.
- De gevoeligheid van de modeluitkomsten - de relatieve fout in de bevolkingsaantallen en de doelfunctie - is voor mannen en vrouwen vergelijkbaar.
- Het model blijkt gevoelig te zijn voor de fertiliteitscijfers, duidelijk meer dan voor de overgangsfracties. Deze gevoeligheid is zelfs waar te nemen bij de leeftijdsgroep 30- 44.

De conclusies uit de gevoeligheidsanalyse luiden daarom, dat met name (1) de aantallen voor de oudere leeftijdsklassen voor mannen en (2) de aantallen voor de jonge leeftijdsklassen gevoelig zijn voor de gekozen modelstructuur en de parameterwaarden.

De resultaten van de fertiliteitsversie van het demografische model verschillen weinig van die van de basisversie. Naar verwachting treden verschillen op voor de jongste leeftijdsgroep. Daarnaast

zijn er ook verschillen voor de oudste leeftijdsgroep(en). De 'structurele' fout in de eerste modelversie, dat onder de nieuwgeborenen geen sterfte plaatsvindt, wordt gecorrigeerd door onder-schattingen van de werkelijke fertiliteitscijfers. De betere schattingen van de aantallen voor de jongste groep over de tijd leiden overigens niet tot betere schattingen voor de oudere groepen over de tijd noch tot betere schattingen van de aantallen voor zowel jonge als oude leeftijdsgroepen in het jaar 2020. Daarnaast is de fertiliteits-versie voor mannen voor sommige overgangsfracties gevoeliger dan de basisversie.

8.5 De 'lineaire uitstroom' versie.

De volgende versie van het demografische model gaat uit van tot 2005 lineair veranderende uitstroomfracties. In het beginjaar van elk van de simulaties worden de uitstroomfracties gelijk gesteld aan de mortaliteitsfracties, zoals die gelden voor 1988. Tussen 1988 en 2005 veranderen deze uitstroomfracties lineair (volgens een rechte curve) tot het jaar 2005. Vanaf het laatstgenoemde jaar worden de fracties constant verondersteld. Deze constante waarden vormen de onbekende te schatten parameters. De geboorte-aantallen worden direct gegeven.

Analoog aan de basisversie en de fertiliteitsversie wordt eerst een gevoeligheidsanalyse van de modelversie uitgevoerd. De resultaten ervan zijn weergegeven in **tabel 3**. De volgende **kanttekeningen** kunnen bij de resultaten van de gevoeligheidsanalyse geplaatst worden:

- De waarde van de doelfunctie is nu duidelijk groter dan voor de basisversie, wat wijst op een slechtere algemene fit. De relatieve fouten in de berekende aantallen voor mannen in het jaar 2020 zijn over het algemeen groter, voor vrouwen over het algemeen echter van dezelfde orde.
- De gevoeligheid van het model voor de verschillende parameterwaarden is kleiner dan voor de basisversie. Met name geldt dit voor de doelfunctie. Dit is in overeenstemming met de verwachting. Tot het jaar 2005 veranderen de overgangsfracties vanaf de gegeven beginwaarde geleidelijk tot de opgegeven waarde.

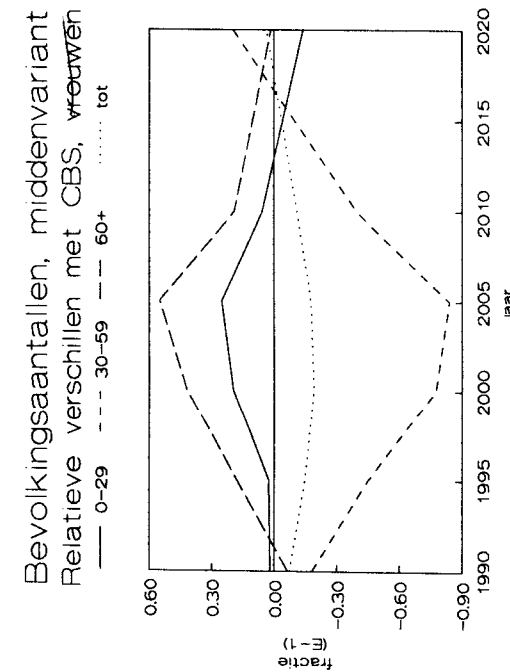
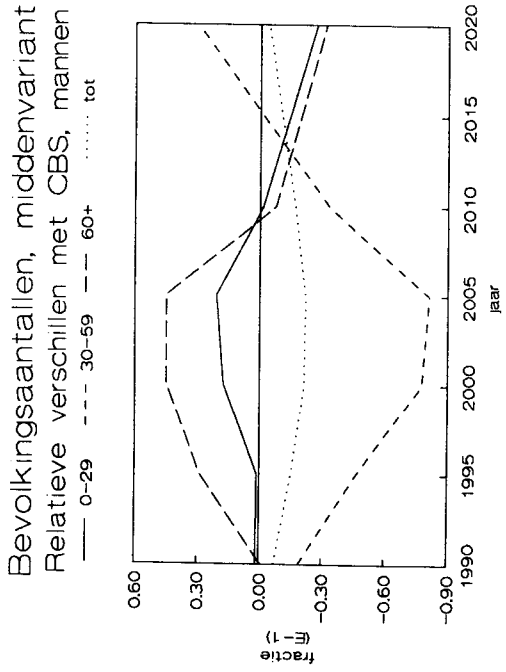
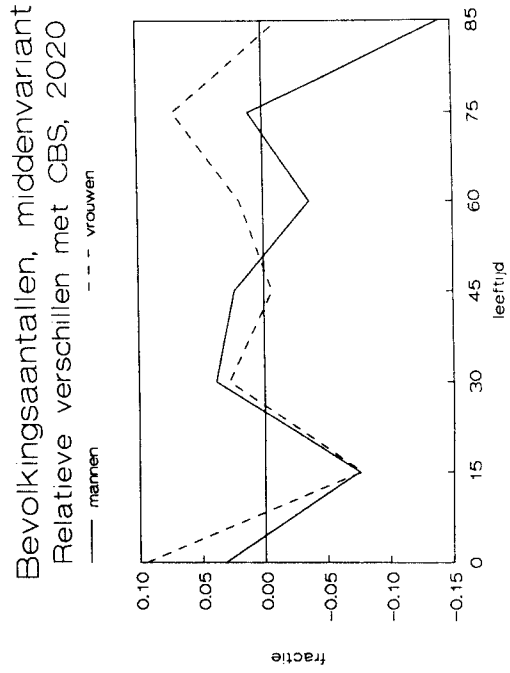
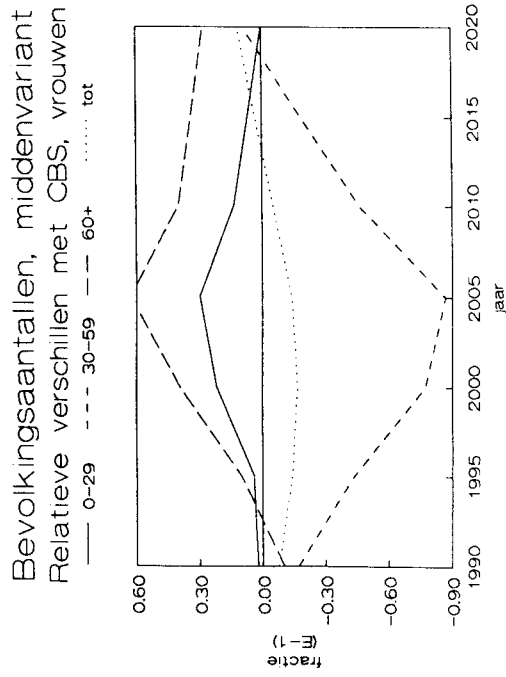
De modeluitkomsten zijn voor de middenvariant grafisch weergegeven in de **grafiek 21**. De volgende **opmerkingen** kunnen hierbij geplaatst worden:

- Voor mannen geeft de basisversie betere resultaten, voor vrouwen de 'lineaire uitstroom' versie. Over het algemeen geeft de laatstgenoemde versie slechte resultaten voor de middelste leeftijdsgroepen.
- De 'lineaire uitstroom' versie geeft een onderschatting van de omvang van de totale bevolking, terwijl de basisversie deze wel juist schat.

Tabel 3: De resultaten van de gevoeligheidsanalyse van de 'lineaire uitstroom' versie van de middenvariant.

relatief verschil tussen bevolkingsaantallen in 2020								
	0-14	15-29	30-44	45-59	60-74	75-84	85+	doelfunctie
parameter								
mannen								
std	7.2	-8.7	11.1	5.0	-2.8	1.6	-14.4	358.707
overgangsfracties								
0-14	7.3							358.709
15-29								358.707
30-44								358.718
45-59				5.2	-2.7	1.7		358.719
60-74					-2.9	1.5		358.713
75-84						1.1	-15.0	358.733
85+							-15.2	358.730
vrouwen								
std	8.8	-7.6	-0.1	-0.1	2.0	7.1	-1.4	358.707
overgangsfracties								
0-14								358.709
15-29								358.702
30-44								358.706
45-59				0.1	2.1			358.715
60-74								358.708
75-84						6.7	-1.8	358.728
85+							-2.3	358.745

Noot: zie tabel 1.



Grafiek 21: De uitkomsten van de 'lineaire uitstroom' versie van het spaarzame demografische model.

- De relatieve fouten in de berekende aantallen voor het jaar 2020 verschillen duidelijk van die voor de basisversie. Met name worden de omvang van de jongste leeftijdsgroep voor vrouwen en die van de oudste groep voor mannen duidelijk verkeerd geschat. Voor de middelste leeftijdsgroepen zijn de schattingen veel beter. Voor de voorafgaande jaren worden deze groepen juist sterk onderschat.

Concluderend kan gesteld worden, dat de 'lineaire uitstroom' versie slechter voldoet dan de basisversie.

8.6 De heterogene versies met expliciete geboorte-aantallen.

De volgende twee versies van het spaarzame demografische model vertonen een modelstructuur, die duidelijk verschilt van die van de voorgaande. De hiervoor beschreven versies veronderstelden homogene bevolkingsgroepen. Onder deze aanname is de jaarlijkse uitstroom uit iedere leeftijdsgroep omgekeerd evenredig met de breedte van de groep. Deze breedte is gedefinieerd als het aantal leeftijdsjaren, dat tot deze groep behoort.

In werkelijkheid wordt aan deze homogeniteitsvoorwaarde niet voldaan. Voor lagere leeftijden nemen de aantallen met de leeftijd toe, vanwege de afnemende geboorte-aantallen van de laatste decennia en vanwege het migratiesaldo. Voor middelbare en met name voor hogere leeftijden nemen de aantallen juist af op grond van de met de leeftijd exponentieel toenemende mortaliteit. Daarom is een modelversie ontwikkeld, waarin de heterogeniteit expliciet beschreven wordt.

Aangenomen wordt, dat de verdeling van de aantallen over de leeftijd binnen elke groep lineair is, ofwel: een rechte lijn volgt. Deze verdeling wordt vastgelegd door twee grootheden:

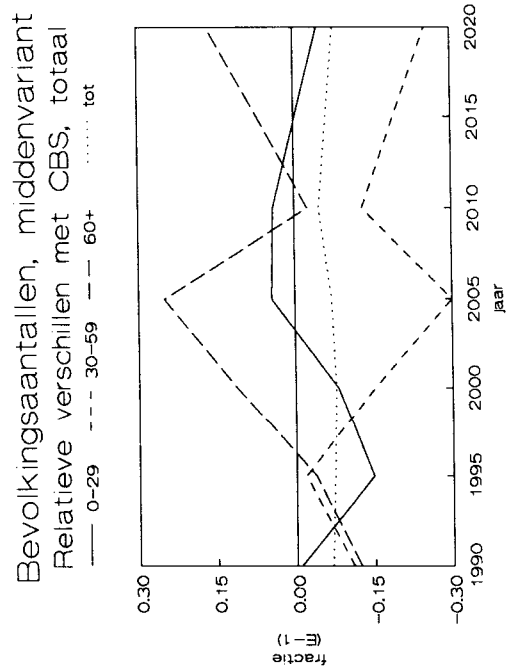
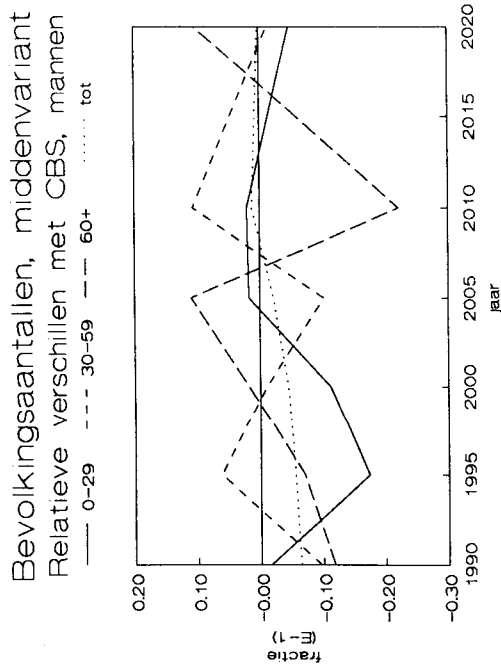
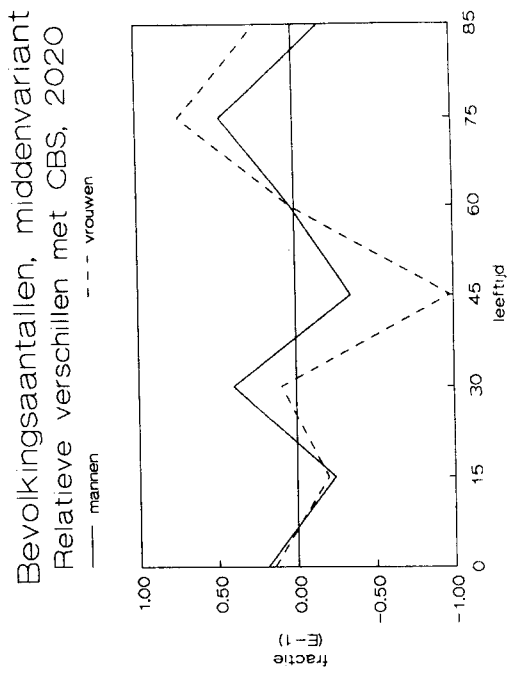
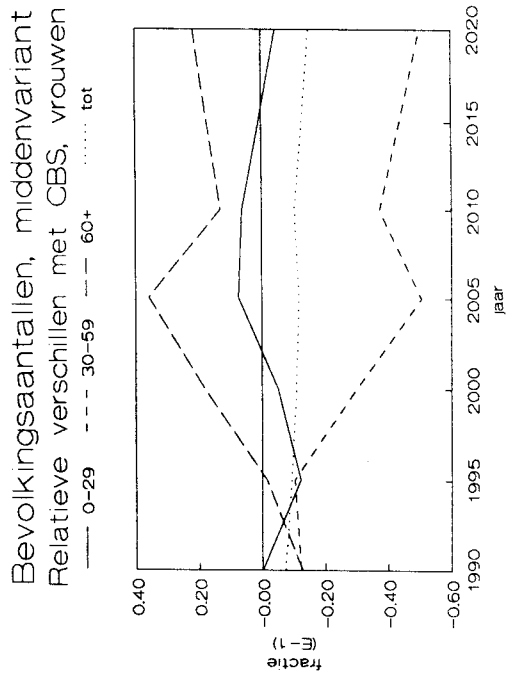
- de reeds bekende omvang van de betreffende bevolkingsgroep,
- de omvang van de groep met als leeftijd het eerste jaar van de betreffende leeftijdsklasse.

Gegeven de klassebreedte - het aantal leeftijdsjaren behorende tot de klasse - is de omvang voor ieder leeftijdsjaar van de klasse bepaald. Ieder jaar gaan voor iedere bevolkingsgroep de individuen met als leeftijd het laatste jaar behorende tot de betreffende klasse over naar de eerstvolgende bevolkingsgroep. Hierdoor verandert niet alleen de omvang van beide groepen, maar ook de samenstelling ofwel de heterogeniteit.

Tabel 4: De resultaten van de gevoeligheidsanalyse van de heterogene versie van de middenvariant met expliciete geboorte-aantallen.

relatief verschil tussen bevolkingsaantallen in 2020								
	0-14	15-29	30-44	45-59	60-74	75-84	85+	doelfunctie
parameter								
mannen								
std	2.7	-3.8	3.9	-4.0	0.6	4.2	-1.7	82.795
overgangsfracties								
0-14								82.796
15-29								82.798
30-44			4.0	-3.9	0.7			82.792
45-59				-4.1	0.4	4.0	-1.8	82.812
60-74					0.5	3.9	-2.0	82.814
75-84						3.7	-3.2	82.869
85+							-2.6	82.800
vrouwen								
std	3.5	-3.0	0.4	-10.1	0.6	8.0	1.9	82.795
overgangsfracties								
0-14		-2.9	0.5					82.807
15-29								82.798
30-44								82.796
45-59					0.5			80.792
60-74								82.794
75-84						7.6	1.0	82.868
85+							1.0	82.851

Noot: zie tabel 1.



Grafiek 22: De uitkomsten van de heterogene versie met geboorteaantallen van het spaarzaam demografische model.

De twee heterogene modelversies verschillen in de modellering van de geboorte. In de eerste versie worden de aantallen nieuwgeborenen gegeven, in de tweede versie worden deze met behulp van fertiliteitscijfers berekend. De eerste versie komt wat betreft de geboorte overeen met de basisversie (zie paragraaf 8.3), de tweede met de fertiliteitsversie (zie paragraaf 8.4).

De resultaten van de gevoeligheidsanalyse van de eerste versie met expliciete geboorte-aantallen van het heterogene model zijn weergegeven in **tabel 4**. Enkele **conclusies** op basis van de resultaten van de gevoeligheidsanalyse zijn (zie ook de kanttekeningen bij die voor de basisversie):

- De relatieve fouten in de berekende bevolkingsaantallen zijn over het algemeen (veel) minder groot dan bij de voorgaande versies. Zie de veel lagere waarden van de doelfunctie. Voor de leeftijdsgroepen 45-59 voor vrouwen en 75-84 voor mannen zijn de relatieve fouten echter groter.
- De gevoeligheid van de heterogeniteitsversie van het model voor de verschillende parameters is minder groot dan die onder de homogeniteitsaannname.

De toegepaste wiskundige modelvergelijkingen zijn beschreven in bijlage C. De resultaten van de heterogene versie met expliciete geboorte-aantallen zijn voor de middenvariant weergegeven in **grafiek 22**. De belangrijkste **kanttekeningen** c.q. verschillen met de resultaten van de (homogene) basisversie zijn:

- De relatieve verschillen voor de geaggregeerde aantallen over de tijd zijn voor de heterogene versie duidelijk beter. In de meeste gevallen zijn deze verschillen niet groter dan 1% in het jaar 2010, terwijl voor de homogene versie deze verschillen tot 3% en 4% oplopen. Een uitzondering hierop vormt de jongste leeftijdsgroep, waarvan de omvang tot 2005 duidelijk onderschat wordt.
- Voor zowel de homogene (basis)versie als de heterogene versie veranderen de schattingen tussen 2005 en 2010. Overschattingen gaan over in onderschattingen, onder- in overschattingen. De uitzondering hierop vormt de oudste bevolkingsgroep vrouwen bij de heterogeniteitsversie. Opvallend is, dat de heterogeniteitsversie vervolgens wisselende resultaten geeft voor het jaar 2020. Voor mannen treden weer overgangen op tussen onder- en overschattingen. Voor vrouwen worden de over- en onderschattingen groter, behalve voor de jongste leeftijdsgroep.
- Opvallend is het verschil tussen de modelresultaten voor mannen en vrouwen. Niet alleen zijn de relatieve fouten voor mannen veel kleiner dan voor vrouwen, ook de tijdsafhankelijkheid van de fouten is anders. Voor mannen wisselen over- en onderschattingen elkaar af. Voor vrouwen wordt de nullijn - de juiste schatting - niet overschreden, behalve voor de jongste leeftijdsgroep.
- Voor het jaar 2020 zijn de relatieve fouten voor de mannen redelijk, maar niet beter dan bij de basisversie. Voor de vrouwen zijn de relatieve fouten groot.

8.7 De heterogene versie met fertiliteitscijfers.

De tweede heterogene versie van het spaarzame demografische model verschilt van de eerste in de berekening van de geboorte-aantallen. In de tweede versie worden deze berekend als functie van de aantallen vrouwen in de vruchtbare leeftijd en de fertiliteitscijfers. Van deze versie is eerst een gevoeligheidsanalyse uitgevoerd rond de optimale waarden van de parameters. De resultaten van deze analyse zijn weergegeven in **tabel 5**.

Tabel 5: De resultaten van de gevoeligheidsanalyse van de heterogene versie van de middenvariant met gebruik van fertiliteitscijfers.

	relatief verschil tussen bevolkingsaantallen in 2020							
	0-14	15-29	30-44	45-59	60-74	75-84	85+	doelfunctie
parameter								
mannen								
std	3.2	-3.9	5.2	-3.9	0.6	4.1	-0.9	84.408
overgangsfracties								
0-14	3.3							84.419
15-29								84.408
30-44				-3.8	0.7			84.411
45-59				-4.0	0.5	4.0		84.417
60-74					0.5	3.8	-1.1	84.404
75-84						3.7	-2.5	84.477
85+							-1.8	84.375
fertiliteitsfracties								
15-29	3.7	-3.5						84.751
30-44	4.0	-3.4						84.838

Tabel 5 (vervolg)

vrouwen

std	3.6	-3.3	1.9	-9.8	1.6	8.1	1.9	84.408
------------	-----	------	-----	------	-----	-----	-----	--------

overgangsfracties

0-14	3.7	-3.1						84.462
-------------	-----	------	--	--	--	--	--	--------

15-29								84.408
--------------	--	--	--	--	--	--	--	--------

30-44								84.408
--------------	--	--	--	--	--	--	--	--------

45-59					0.1			84.429
--------------	--	--	--	--	-----	--	--	--------

60-74								84.406
--------------	--	--	--	--	--	--	--	--------

75-84						7.8	1.0	84.442
--------------	--	--	--	--	--	-----	-----	--------

85+							1.0	84.468
------------	--	--	--	--	--	--	-----	--------

fertiliteitsfracties

15-29	4.2	-2.9						84.751
--------------	-----	------	--	--	--	--	--	--------

30-44	4.4	-2.8						84.838
--------------	-----	------	--	--	--	--	--	--------

Noot: zie tabel 1.

Enkele **kanttekeningen** bij de resultaten van de gevoeligheidsanalyse van de heterogene versie met fertiliteitscijfers in vergelijking met die van de versie met expliciete geboorte-aantallen:

- De relatieve fouten in de bevolkingsaantallen in het jaar 2020 zijn vrijwel gelijk. De algemene fit is, hoewel iets minder, vergelijkbaar.
- Beide modellen zijn vrijwel even gevoelig voor variaties in de parameterwaarden.
- Het tweede model, met fertiliteitscijfers, is redelijk gevoelig voor variaties in de fertiliteit.

De toegepaste wiskundige modelvergelijkingen zijn beschreven in bijlage C. De heterogene versie met fertiliteitscijfers is doorge-rekend voor de middenvariant. De verschillen met de heterogene versie met expliciete geboorte-aantallen blijken minimaal te zijn, ook voor de jongste leeftijdsgroepen.

8.8 Conclusies.

De conclusies, die uit de voorgaande paragrafen getrokken kunnen worden, hebben enerzijds betrekking op een vergelijking van de spaarzame demografische modellen onderling, anderzijds met (de basisversie van) het uitgebreide demografische model.

Een vergelijking van de verschillende versies met de basisversie van het spaarzame model geeft duidelijk aan, waar eventuele verbeteringen in de nauwkeurigheid van de uitkomsten te vinden valt. Het vervangen van expliciet gegeven aantallen door fertilitateitscijfers, het onderscheiden van een tijdsperiode, waarin de parameters hun definitieve (te schatten) waarde bereiken, of het veranderen van de doelfunctie hebben elk weinig invloed op de resultaten van het model.

Wel een sterke verbetering van de nauwkeurigheid van de uitkomsten van het model is te verkrijgen door het heterogeen veronderstellen van elk van de leeftijdsgroepen. De 'lineair' veronderstelde heterogeniteit vormt nog steeds een simplificatie van de werkelijkheid, maar de resultaten blijken al veel beter te zijn dan van de basisversie. Met name de kwaliteit van de simulatie over de tijd is verbeterd. Voor het eindjaar 2020 van de simulatie is de verbetering overigens marginaal.

Opvallend is het verschil in nauwkeurigheid van het heterogene model voor mannen in vergelijking met die voor vrouwen. Zo vormen de onderschatting met 10% van de omvang van de groep 45-59 en de overschatting met 8% van die van de groep 75-84 een dissonant vergeleken met de overige schattingen. De bij de calibratieprocedure gevonden oplossing voor de instelling van de parameters in het model is waarschijnlijk toch nog niet de optimale gezien het grillige verloop van de nauwkeurigheid over de tijd en gezien de veel betere schattingen voor mannen.

Een vergelijking van de heterogene versie van het spaarzame demografische model (zie paragraaf 8.6) met expliciete geboorteaantallen met de twee middenvarianten van het uitgebreide demografische model (zie paragraaf 7.5) geeft enkele duidelijke verschillen te zien:

- De berekende bevolkingsaantallen vertonen voor de twee varianten van het uitgebreide demografische model een veel gladder verloop dan voor het spaarzame model.
- Tot de oudste leeftijdsgroepen geven de uitgebreide demografische modellen een betere overeenstemming met de CBS prognose dan het spaarzame model. Voor de oudste groepen is dit echter andersom. Het laatste toont weer aan, dat modellen, die gebruik maken van waargenomen sterfecijfers meestal een overschatting geven van de omvang van de oudste bevolkingsgroepen.

De conclusie op basis van deze bevindingen luidt, dat de beschreven spaarzame demografische modellen redelijk voldoen, indien de

heterogeniteit over de leeftijd binnen elk van de leeftijdsgroepen expliciet gemodelleerd wordt. De beschreven uitgebreide demografische modellen geven duidelijk betere resultaten, behalve voor de oudste leeftijdsgroepen. Het gebruik van de waargenomen sterfteaantallen, dat voor de overige leeftijdsgroepen goede resultaten geeft, leidt tot grote overschattingen voor de oudste groepen.

Literatuurlijst

- Casparie A.F., Verkleij H. (in opdracht van de Stuurgroep Toekomstscenario's Gezondheidszorg). Chronische ziekten in het jaar 2005. Deel 1. Scenario's over Diabetes Mellitus 1990-2005. Bohn, Scheltema & Holkema, Utrecht/Antwerpen, 1990
- CBS. Maandstatistiek van de bevolking. Voorburg, vanaf 1955.
- CBS. Overledenen naar doodsoorzaak, leeftijd en geslacht in het jaar .. Voorburg, vanaf 1980.
- CBS. Prognose van de bevolking van Nederland na 1980. Deel II. Modelbouw en hypothesevorming. Staatsuitgeverij, 's Gravenhage, 1984
- CBS. H Cruijssen. Kroniek. Bevolkingsprognose 1989: meer buitenlandse migratie. Maandstatistiek van de bevolking (CBS) vol 1 (1990), pp 6-38
- Flipse E., Veling E.J.M. An application of the Leslie matrix model to the population dynamics of the hooded seal, *Cystophora cristata* Erxleben. Ecological modelling, vol 24 (1984), pp 43-59
- Fritz F.J., Huppert, B., Willems, W. Stochastische matrisen. Springer Verlag, Berlin (etc.), 1979
- Gantmacher F.R. Matrisenrechnung II, Spezielle fragen und anwendungen. Deutscher Verlag der Wissenschaften, Berlin, 1959
- Golub G.H., Loan C.F. van. Matrix computations. North Oxford Academic, Oxford, 1983
- Hoogenveen R.T. en Brouwer J.J. (in opdracht van de Stuurgroep Toekomstscenario's Gezondheidszorg). Scenario's in de volksgezondheid, Inleiding in de methodiek van de STG, Uitgeverij Jan van Arkel, Utrecht, 1989
- Hoogenveen R.T. en Jager J.C. Overlevingsduuranalyse en concurrerende doodsoorzaken. Een inleiding. RIVM-rapport, nr 958606 001, 1990
- Hoogenveen R.T. en Kooiker S.E. Markov-modellen van sociale mobiliteit. RIVM-rapport, nr 958606 003, 1990
- Keilman N., Gill R. On the estimation of multidimensional demographic models with population registration data. Kwantitatieve Methoden, nr 23 (1987), 85-115

- Lancaster P. Theory of matrices. Academic Press, New York and London, 1969
- Minc H. Nonnegative matrices. John Wiley & Sons, New York (etc.), 1988
- Mode C.J. Stochastic processes in demography and their computer implementation. Springer Verlag, Berlin (etc.), 1985
- Prabhu N.U. Stochastic processes. Macmillan Company, New York, 1965
- Ross S.M. Applied probability models with optimization applications. Holden Day, San Francisco, 1970
- Schoen R. Modeling multigroup populations. Plenum Press, New York and London, 1988
- Seneta E. Non-negative matrices and Markov chains, Second edition. Spinger-Verlag, New York (etc.), 1981
- Singer B., Spilerman S. The representation of social processes by Markov models. American Journal of Sociology, vol 82 (1976), pp 1-54

Bijlage A De gebruikte CBS cijfers.

In deze bijlage worden de gebruikte CBS cijfers gepresenteerd, namelijk sterfte-, migratie-, en geboortecijfers, en de toegepaste administratieve correcties.

A.1 Sterftecijfers.**Specifieke sterftefracties voor mannen**

Jaar	Leeftijdsklassen									
	0	1	2-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39
1955	2272	192	118	68	49	72	96	96	116	148
1956	2161	178	113	61	43	72	97	103	111	152
1957	1946	196	112	65	52	74	102	98	117	142
1958	1911	202	119	60	43	73	94	97	115	138
1959	1900	221	129	72	48	82	99	98	117	162
1960	1846	168	122	61	36	71	100	104	106	146
1961	1790	170	101	62	43	75	105	100	104	153
1962	1725	176	110	59	44	72	105	108	118	153
1963	1779	152	103	63	40	72	96	103	108	158
1964	1669	135	105	59	49	85	102	102	112	160
1965	1609	140	97	56	41	80	105	98	104	166
1966	1679	152	93	62	39	92	104	94	106	156
1967	1478	156	95	58	47	96	107	104	118	152
1968	1565	118	91	60	41	93	101	100	120	161
1969	1507	131	89	54	40	91	104	94	107	159
1970	1435	106	89	57	45	101	110	95	109	154
1971	1376	116	92	53	39	102	105	92	107	154
1972	1326	123	80	52	45	106	104	97	109	158
1973	1276	141	70	49	37	95	114	89	112	152
1974	1259	132	73	44	35	111	98	80	87	148
1975	1204	115	68	38	37	82	94	79	100	134
1976	1163	113	57	38	33	82	99	82	97	140
1977	1096	77	57	42	31	93	101	84	93	136
1978	1097	102	59	41	35	87	95	76	90	140
1979	994	86	57	32	31	78	88	81	91	128
1980	970	93	41	35	26	72	96	88	90	130
1981	937	86	52	29	29	66	86	82	90	130
1982	931	84	48	29	31	66	78	81	82	136
1983	926	73	46	27	28	60	80	74	91	123
1984	958	88	40	27	27	58	77	81	91	131
1985	888	68	38	21	22	60	79	76	88	120
1986	703	196	48	23	23	50	73	80	83	116
1987	753	157	47	20	20	52	76	65	83	120
1988	703	129	49	25	21	48	74	77	87	115

Leeftijdsklassen

Jaar	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
1955	241	442	712	1141	1823	2861	4704	8027	13132	24212
1956	242	396	703	1196	1897	2894	4789	8070	13929	24302
1957	242	417	703	1102	1824	2830	4647	7747	12624	23277
1958	250	380	685	1166	1858	2773	4615	7713	13150	23483
1959	243	391	687	1213	1870	2876	4651	7688	12537	22673
1960	243	396	716	1211	1888	2993	4555	7533	12774	22834
1961	246	413	733	1204	1928	2984	4590	7427	12332	21631
1962	235	422	749	1306	2074	3169	4646	7902	13426	23024
1963	246	436	748	1276	2058	3236	4924	7964	13163	23447
1964	248	454	712	1289	2073	3282	4757	7572	11663	20540
1965	261	434	785	1309	2102	3257	4991	7764	12408	22127
1966	256	430	769	1315	2143	3244	5017	7823	12439	21876
1967	261	434	767	1308	2109	3335	5101	7673	11694	20647
1968	265	451	755	1307	2132	3400	5181	8054	12599	22719
1969	263	461	780	1294	2205	3462	5395	7963	12340	21775
1970	255	455	796	1368	2268	3542	5359	8113	12385	21898
1971	273	430	787	1315	2187	3465	5343	8146	12380	21958
1972	269	452	778	1350	2219	3548	5554	8389	12733	22376
1973	240	449	757	1256	2131	3443	5248	8203	12235	21869
1974	239	440	723	1277	2081	3412	5107	7987	11701	21094
1975	231	442	762	1297	2122	3546	5398	8253	12601	22156
1976	236	448	738	1239	2155	3512	5458	8250	12654	21458
1977	244	421	698	1207	2019	3360	5236	7805	11607	19395
1978	225	423	700	1196	2056	3376	5366	8129	12111	20606
1979	213	408	685	1155	1981	3211	5299	7953	11760	19380
1980	214	405	693	1143	1972	3224	5235	7967	11825	19832
1981	200	367	670	1137	1919	3179	5161	7982	12046	19838
1982	215	374	655	1143	1959	3202	5009	8000	11824	20501
1983	204	358	633	1116	1839	3145	5091	8049	11803	20853
1984	193	352	646	1101	1844	3108	5065	7916	11873	20645
1985	208	332	629	1107	1857	3130	4966	8040	12005	21316
1986	193	341	604	1027	1769	2937	4776	7659	11597	20519
1987	186	318	557	1017	1694	2829	4563	7333	11235	20058
1988	179	296	552	939	1649	2812	4604	7260	11410	19525

Noot: aantallen per 100.000 mannen van de gemiddelde bevolking.
Bron: CBS Maandstatistiek van de bevolking.

Specifieke sterftefracties voor vrouwen**Leeftijdsklassen**

Jaar	0	1	2-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39
1955	1725	198	82	43	33	36	48	64	87	135
1956	1621	176	88	42	34	33	46	53	86	131
1957	1479	187	98	41	31	44	48	62	88	120
1958	1531	166	80	36	29	43	46	61	75	118
1959	1459	176	89	47	27	33	38	57	79	121
1960	1443	172	76	39	23	28	45	57	69	103
1961	1300	142	64	35	27	32	38	49	65	110
1962	1326	137	75	37	27	34	40	55	79	107
1963	1364	121	63	58	28	33	35	48	74	107
1964	1284	129	72	39	29	31	36	48	69	104
1965	1269	110	64	39	25	37	46	46	64	105
1966	1251	132	65	32	29	37	40	54	72	102
1967	1188	110	63	34	29	37	38	52	74	102
1968	1145	104	65	37	29	41	42	43	65	105
1969	1129	126	67	35	28	41	41	53	66	113
1970	1106	110	56	32	26	46	46	51	69	105
1971	1008	102	56	29	29	42	37	44	68	99
1972	1006	103	56	28	29	43	42	46	64	104
1973	1023	94	53	30	26	48	44	52	70	103
1974	997	89	46	31	23	35	38	42	62	96
1975	918	83	52	27	22	35	33	43	66	95
1976	970	85	41	31	22	37	40	44	63	87
1977	799	81	39	27	23	30	39	43	67	96
1978	812	89	45	23	24	38	41	45	62	91
1979	742	79	37	23	21	31	35	45	59	84
1980	741	71	42	22	18	31	40	39	52	93
1981	720	87	40	20	19	29	33	42	54	81
1982	723	86	36	21	18	33	34	39	57	80
1983	752	75	33	19	20	29	37	41	56	80
1984	700	66	29	18	17	26	29	37	59	77
1985	714	59	30	14	14	25	30	37	58	82
1986	582	158	38	17	17	22	33	39	59	83
1987	549	137	37	14	19	25	32	38	56	73
1988	480	110	30	17	14	23	27	36	52	76

Leeftijdsklassen

Jaar	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
1955	192	303	459	742	1217	2193	4037	6985	12093	22027
1956	188	305	435	720	1243	2206	4073	7388	12614	23450
1957	192	288	435	702	1185	2111	3830	6925	11771	21280
1958	187	281	438	680	1138	2065	3754	6762	11616	21330
1959	183	276	421	651	1125	1943	3574	6501	11451	21386
1960	177	272	423	693	1080	1892	3648	6568	11472	21034
1961	172	264	428	638	1063	1914	3402	6405	11032	20091
1962	171	261	392	648	1060	1912	3501	6314	11491	21394
1963	165	270	411	616	1052	1953	3485	6307	11187	20927
1964	163	268	416	619	1053	1794	3325	5852	10211	19100
1965	163	275	403	618	1051	1868	3330	6126	10361	19880
1966	178	265	421	645	1022	1831	3322	5993	10450	20371
1967	173	267	406	609	1049	1722	3132	5725	9941	18786
1968	180	265	419	616	1002	1764	3169	5825	10269	20187
1969	181	272	426	638	1065	1821	3218	5742	10345	19691
1970	164	284	423	629	1032	1758	3211	5677	10108	19580
1971	159	278	426	606	983	1706	3162	5596	10200	19649
1972	169	273	407	618	977	1673	3048	5768	10317	19726
1973	168	284	411	606	925	1629	2901	5289	9514	18890
1974	159	266	390	575	894	1614	2819	5040	9214	18159
1975	160	255	387	567	930	1546	2816	5096	9221	18562
1976	147	250	382	563	874	1547	2723	5017	8820	18067
1977	145	258	380	562	890	1475	2550	4685	8026	16421
1978	148	265	374	540	860	1426	2574	4614	8291	16969
1979	153	238	378	570	870	1398	2466	4416	7903	15995
1980	136	230	349	526	854	1354	2422	4384	7822	15722
1981	136	226	370	528	818	1392	2349	4299	7608	15848
1982	139	228	344	539	804	1338	2328	4230	7591	15962
1983	132	220	357	544	810	1313	2226	4027	7411	15883
1984	132	217	341	529	828	1335	2227	4142	7393	15786
1985	121	211	365	532	812	1317	2295	4078	7545	15931
1986	125	206	331	532	783	1241	2099	3810	7061	15754
1987	123	209	316	502	780	1210	2039	3651	6577	15106
1988	124	203	295	496	749	1212	2060	3589	6718	14964

Noot: aantallen per 100.000 vrouwen van de gemiddelde bevolking.
Bron: CBS Maandstatistiek van de bevolking.

Voor de jaren 1980-1988 is gebruik gemaakt van de door het CBS gepresenteerde absolute sterfte-aantallen in plaats van (relatieve) sterftefracties. De uit de absolute aantallen berekende en vervolgens toegepaste sterftefracties zijn:

Leeftijdsklassen

Mannen

Jaar **0-4** **5-9** **10-14** **15-19** **20-24** **25-29** **30-34** **35-39**

1980	243	34	26	72	97	88	92	131
1981	236	28	29	66	87	83	88	137
1982	226	28	30	66	78	81	81	141
1983	221	27	27	60	81	74	91	126
1984	233	27	26	58	78	82	91	133
1985	217	21	22	60	80	76	89	122
1986	218	23	24	53	76	80	84	124
1987	223	20	21	57	74	66	84	122
1988	205	24	20	53	74	78	91	117

Jaar **40-44** **45-49** **50-54** **55-59** **60-64** **65-69** **70-74** **75-79** **80-84** **85+**

1980	217	404	690	1139	2012	3233	5257	8049	11853	19873
1981	202	369	673	1137	1948	3197	5168	7995	12056	19938
1982	217	375	659	1144	1984	3191	5042	8042	11896	20217
1983	207	362	636	1115	1871	3127	5121	8135	11984	20502
1984	197	354	650	1101	1868	3103	5092	7998	11988	20620
1985	210	337	630	1118	1851	3187	4976	8094	12057	21526
1986	215	363	638	1097	1838	3170	5035	8005	12140	21192
1987	207	341	594	1082	1759	3049	4736	7792	11818	20651
1988	194	317	583	997	1757	3007	4784	7654	12066	20251

 Noot: aantallen per 100.000 mannen van de gemiddelde bevolking.
 (Bewerkte) bron: CBS Overledenen naar doodsoorzaak statistieken.

Vrouwen

Jaar 0-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39

1980	192	22	18	32	40	40	53	94
1981	187	19	19	29	33	42	54	85
1982	179	20	17	32	34	40	56	83
1983	179	18	20	29	37	42	56	82
1984	170	18	17	26	29	38	59	79
1985	176	14	14	25	30	37	59	83
1986	182	15	16	24	34	40	60	88
1987	168	14	19	25	34	40	58	74
1988	143	17	13	24	28	36	56	77

Jaar 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85+

1980	137	229	350	524	872	1356	2433	4469	7960	16351
1981	137	227	370	525	832	1396	2353	4351	7777	16301
1982	141	228	345	537	815	1335	2349	4272	7757	16854
1983	134	222	356	541	826	1304	2244	4081	7581	16036
1984	134	218	340	529	841	1328	2239	4195	7523	16267
1985	123	213	363	535	808	1345	2299	4107	7684	16379
1986	137	215	346	558	817	1327	2220	4078	7633	16795
1987	136	218	335	530	805	1295	2177	3882	7052	15913
1988	133	217	311	514	784	1299	2145	3836	7254	16059

 Noot: aantallen per 100.000 vrouwen van de gemiddelde bevolking.
 (Bewerkte) bron: CBS Overledenen naar doodsoorzaak statistieken.

A.2 Migratiecijfers.

Absolute migratie-aantallen voor mannen.

Jaar	Immigratie						Emigratie					
	Leeftijdsklassen											
	0	15	20	30	50	65	0	15	20	30	50	65
1955	7.5	1.8	6.3	8.5	2.1	0.6	8.4	2.1	9.4	9.1	1.3	0.3
1956	6.9	2.3	8.0	8.4	1.9	0.5	9.4	2.2	10.2	9.9	1.4	0.3
1957	6.8	2.0	7.8	8.1	1.8	0.6	9.2	2.3	10.3	10.2	1.3	0.3
1958	10.7	1.8	6.5	10.9	3.0	0.9	7.4	2.0	10.1	9.2	1.2	0.2
1959	5.1	1.0	5.0	6.3	1.4	0.4	7.8	2.1	9.1	8.7	1.1	0.3
1960	6.6	1.3	6.6	7.4	1.5	0.5	8.3	2.0	9.7	9.5	1.2	0.2
1961	7.4	1.6	11.1	9.7	1.6	0.5	6.6	1.7	8.6	7.8	1.1	0.2
1962	9.2	2.1	12.4	12.3	1.9	0.5	6.0	1.6	9.8	8.2	1.1	0.2
1963	6.1	1.8	13.1	10.7	1.6	0.4	5.2	1.4	10.2	8.1	1.1	0.3
1964	6.7	2.1	17.4	15.0	1.8	0.4	5.7	1.5	11.9	9.8	1.3	0.3
1965	7.1	2.6	19.8	18.2	1.9	0.5	6.2	1.6	12.6	11.2	1.3	0.3
1966	7.4	2.4	20.9	21.0	2.0	0.4	6.2	1.6	13.8	13.5	1.4	0.4
1967	6.9	1.7	10.1	9.6	1.6	0.5	6.9	1.7	14.5	16.2	1.5	0.4
1968	7.2	2.0	12.6	13.0	1.7	0.5	6.5	1.5	11.6	11.7	1.4	0.4
1969	8.0	2.4	16.9	16.4	1.8	0.5	6.4	1.4	10.6	10.5	1.4	0.4
1970	9.5	2.5	21.5	19.5	1.8	0.5	6.3	1.4	10.9	10.8	1.5	0.5
1971	11.3	3.1	20.4	17.9	2.0	0.6	6.3	1.5	12.6	12.3	1.6	0.5
1972	11.2	3.0	13.2	11.4	2.0	0.6	6.6	1.4	12.2	13.0	1.7	0.5
1973	12.1	3.3	14.6	11.9	2.0	0.7	7.2	1.5	11.6	12.3	1.7	0.5
1974	14.0	4.0	15.3	12.5	2.2	0.7	7.2	1.5	10.3	11.2	1.8	0.6
1975	19.1	6.1	21.9	17.5	3.3	1.1	6.7	1.4	9.1	10.4	1.7	0.6
1976	12.8	3.3	12.2	11.0	2.0	0.7	7.7	1.5	9.0	11.8	2.0	0.7
1977	12.7	3.7	11.7	11.1	2.1	0.7	7.7	1.6	8.4	11.5	2.1	0.8
1978	13.7	4.5	12.7	11.7	2.1	0.7	7.9	1.7	8.2	11.6	2.2	0.9
1979	15.9	6.0	16.4	13.4	2.2	0.8	7.4	1.7	8.5	11.6	2.1	0.9
1980	17.1	6.9	17.7	13.8	2.4	0.8	7.0	1.7	9.0	11.5	2.1	0.8
1981	11.4	4.1	12.3	11.1	1.9	0.7	7.7	1.8	9.2	12.3	2.4	0.8
1982	9.8	3.0	9.8	10.4	2.0	0.7	8.7	1.9	9.0	12.8	2.5	0.8
1983	8.9	2.4	8.8	10.3	2.1	0.7	7.8	1.7	7.8	11.5	2.5	0.8
1984	8.7	2.4	9.1	10.7	2.1	0.8	7.6	1.6	7.3	11.2	2.4	0.8
1985	10.0	3.0	13.1	12.6	2.3	0.9	6.4	1.5	7.4	10.9	2.4	0.9
1986	10.6	3.8	14.4	13.8	2.4	0.8	5.9	1.4	7.2	10.5	3.0	0.9
1987	10.5	4.4	18.7	15.6	2.5	0.9	5.6	1.3	7.3	10.2	2.7	0.8
1988	10.1	4.4	16.8	14.3	2.3	0.9	5.7	1.4	7.7	10.7	3.4	0.8

Noot: aantallen * 1000. Voor de 85+ groep worden in het ontwikkelde model de im- en emigratie gelijk verondersteld. Bron: CBS Maandstatistiek van de bevolking.

Absolute migratie-aantallen voor vrouwen.

Jaar	Immigratie						Emigratie					
	Leeftijdsklassen											
	0	15	20	30	50	65	0	15	20	30	50	65
1955	6.9	1.8	5.6	8.0	2.1	0.9	7.5	1.7	8.1	8.0	1.3	0.4
1956	6.4	1.9	5.5	7.2	1.8	0.8	8.6	1.8	8.6	8.7	1.3	0.4
1957	6.3	1.7	5.5	7.1	1.7	0.7	8.5	1.7	8.8	8.4	1.2	0.3
1958	10.5	2.0	6.4	10.6	3.0	1.6	7.0	1.6	8.4	7.3	1.1	0.3
1959	4.8	1.2	4.9	5.3	1.3	0.5	7.2	1.6	7.8	7.1	1.1	0.3
1960	6.2	1.4	5.4	6.5	1.5	0.6	7.9	1.8	8.6	7.4	1.2	0.4
1961	7.0	1.6	6.0	6.6	1.6	0.6	6.4	1.5	7.3	6.0	1.1	0.3
1962	8.8	1.9	6.9	7.7	1.6	0.6	5.9	1.6	7.5	5.7	1.0	0.4
1963	5.8	1.7	6.2	6.0	1.3	0.5	4.8	1.6	7.7	5.0	0.9	0.3
1964	6.4	1.9	7.0	6.3	1.5	0.6	5.5	1.7	8.8	5.4	1.1	0.4
1965	6.8	2.2	8.3	6.8	1.6	0.6	5.8	1.9	9.6	5.7	1.1	0.4
1966	7.0	2.3	9.1	7.1	1.6	0.6	5.8	1.7	10.1	5.8	1.1	0.5
1967	6.5	2.0	8.4	6.3	1.5	0.6	6.3	1.6	10.5	6.0	1.1	0.5
1968	7.0	2.3	9.1	6.7	1.6	0.7	6.2	1.6	10.1	5.8	1.1	0.6
1969	7.7	2.5	10.4	7.4	1.6	0.7	6.1	1.6	10.2	5.8	1.2	0.5
1970	9.2	3.2	12.2	8.4	1.8	0.7	6.2	1.5	10.5	6.0	1.2	0.5
1971	10.8	3.9	13.4	9.1	1.9	0.8	6.0	1.7	11.4	6.2	1.3	0.6
1972	10.6	3.8	13.2	8.9	1.9	0.8	6.3	1.7	10.6	6.1	1.3	0.6
1973	11.3	4.0	13.0	9.0	2.1	0.8	7.2	1.8	10.9	6.9	1.4	0.6
1974	13.4	4.7	14.0	9.9	2.2	0.8	6.9	1.7	10.4	6.8	1.5	0.6
1975	18.9	6.1	16.0	12.8	3.4	1.1	6.5	1.5	8.8	6.4	1.5	0.6
1976	12.5	3.8	12.5	9.3	1.9	0.8	7.6	1.8	9.2	7.6	1.7	0.7
1977	12.5	4.2	12.6	9.9	1.9	0.8	7.6	1.8	9.0	7.9	1.8	0.8
1978	12.9	4.6	12.8	10.7	2.0	0.8	7.4	1.8	8.5	8.2	1.9	0.8
1979	15.0	5.5	14.5	11.7	2.2	0.9	7.0	1.8	8.2	8.1	1.7	0.9
1980	16.0	6.6	15.5	12.3	2.4	1.0	6.7	1.9	8.4	8.0	1.7	0.8
1981	10.9	4.2	11.5	9.6	1.8	0.8	7.3	1.9	8.8	8.5	1.8	0.8
1982	9.6	3.6	10.4	9.1	1.7	0.8	8.2	2.2	9.0	9.4	1.9	0.9
1983	8.6	3.3	9.9	9.1	1.8	0.9	7.3	2.0	8.1	8.5	1.8	0.8
1984	8.5	3.0	9.9	9.0	1.9	0.8	7.1	2.0	7.8	8.5	1.8	0.8
1985	9.7	3.7	11.1	9.9	2.0	0.9	6.1	1.8	7.5	8.0	1.7	0.8
1986	9.9	4.4	12.9	11.0	2.2	0.9	5.8	1.7	7.5	8.0	2.1	0.8
1987	10.0	5.0	13.5	11.4	2.4	1.0	5.1	1.6	7.3	7.6	1.8	0.7
1988	9.6	5.3	13.3	11.1	2.2	0.9	5.5	1.7	7.7	8.2	2.1	0.8

Noot: aantallen * 1000. Voor de 85+ groep worden in het ontwikkelde model de im- en emigratie gelijk verondersteld. Bron: CBS Maandstatistiek van de bevolking.

Migratiesaldo.

(= totaal immigratie - totaal emigratie)

Jaar Mannen Vrouwen Totaal				Jaar Mannen Vrouwen Totaal			
1955	- 3.8	- 1.7	- 5.5	1972	6.0	12.6	18.6
1956	- 5.4	- 5.8	-11.2	1973	9.8	11.4	21.2
1957	- 6.5	- 5.9	-12.4	1974	16.1	17.1	33.2
1958	3.7	8.4	12.1	1975	39.1	33.0	72.1
1959	- 9.9	- 7.1	-17.0	1976	9.3	12.2	21.5
1960	- 7.0	- 5.7	-12.7	1977	9.9	13.0	22.9
1961	5.9	0.8	6.7	1978	12.9	15.2	28.1
1962	11.5	5.4	16.9	1979	22.5	22.1	44.6
1963	7.4	1.2	8.6	1980	26.6	26.3	52.9
1964	12.9	0.8	13.7	1981	7.3	9.7	17.0
1965	16.9	1.8	18.7	1982	0.0	3.6	3.6
1966	17.2	2.7	19.9	1983	1.1	5.1	6.2
1967	-10.8	- 0.7	-11.5	1984	2.9	5.1	8.0
1968	3.9	2.0	5.9	1985	12.4	11.4	23.8
1969	15.3	4.9	20.2	1986	16.9	15.4	32.3
1970	23.9	9.6	33.5	1987	24.7	19.2	43.9
1971	20.5	12.7	33.2	1988	19.1	16.4	35.5

Noot: aantallen *1000. Cijfers betreffen totaalaantallen. (Bewerkte)
bron: CBS Maandstatistiek van de bevolking.

Veronderstelde saldo buitenlandse migratie vanaf 1989.

	Lage variant	Middenvariant	Hoge variant
1989	22	25	28
1990	15	25	35
1991	9	24	39
1992	8	23	38
1993	7	22	37
1994	6	21	36
1995	5	20	35
1996	6	20	34
1997	7	20	33
1998	8	20	32
1999	9	20	31
2000	10	20	30
2005	10	20	30
2010	10	20	30

Noot: aantallen *1000. Cijfers betreffen totaalaantallen. De cijfers tot 1988 zijn exclusief, die vanaf 1989 zijn inclusief (veronderstelde) administratieve correcties. Bron: CBS, 1990

A.3 Geboortecijfers.

Bij de beschrijving van de aantallen nieuwgeborenen wordt in het model gebruikt gemaakt van direct gegeven aantallen, van fertilitetskijfers (gepubliceerd vanaf 1967) en van de sexratio's. De sexratio's vormen de verhouding tussen de aantallen mannelijke en vrouwelijke nieuwgeborenen.

Geboorte-aantallen.

Jaar	Aantal				
1955	228.9	1965	245.2	1975	177.9
1956	231.2	1966	239.6	1976	177.1
1957	233.9	1967	238.7	1977	173.3
1958	236.9	1968	237.1	1978	175.6
1959	242.5	1969	247.6	1979	175.0
1960	239.1	1970	238.9	1980	181.3
1961	247.4	1971	227.2	1981	178.6
1962	246.2	1972	214.1	1982	172.1
1963	249.9	1973	195.0	1983	170.2
1964	250.9	1974	186.0	1984	174.4
1985	178.1				
1986	184.5				
1987	186.7				
1988	186.6				

 Noot: aantallen * 1000. Bron: CBS Maandstatistiek van de bevolking.

Fertiliteitscijfers.**Leeftijdsklassen voor moeder**

Jaar	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1967	22.2	145.2	196.0	120.1	60.2	19.2	1.7
1968	21.5	142.3	193.3	115.0	54.9	16.4	1.5
1969	22.7	146.3	198.8	115.9	53.5	15.7	1.37
1970	22.6	138.9	186.9	109.5	48.6	14.1	1.14
1971	22.4	130.1	181.3	96.6	41.2	12.0	1.02
1972	20.1	123.4	169.6	83.8	33.5	9.2	0.54
1973	17.4	111.8	152.5	70.8	27.7	7.1	0.52
1974	15.4	105.6	145.3	65.4	22.7	5.9	0.41
1975	12.7	98.4	139.9	61.2	20.2	4.7	0.38
1976	11.4	94.9	135.9	63.5	18.3	4.4	0.32
1977	10.2	88.3	135.3	64.1	17.4	3.9	0.31
1978	9.5	85.9	139.3	64.8	16.8	3.5	0.34
1979	8.9	82.5	139.3	64.3	17.3	3.3	0.32
1980	9.2	80.7	143.4	67.8	17.7	3.5	0.39
1981	9.0	75.2	140.5	65.9	18.4	3.5	0.41
1982	8.2	69.8	133.0	66.3	18.7	3.5	0.40
1983	7.6	66.0	130.1	69.7	18.4	3.2	0.41
1984	7.4	64.5	132.1	73.9	18.8	3.3	0.38
1985	5.0	52.8	130.3	89.4	23.2	3.9	0.54
1986	5.1	50.0	130.2	95.3	25.2	4.2	0.6
1987	5.2	46.6	127.5	100.7	27.9	4.3	0.5
1988	5.6	44.3	122.4	103.5	29.7	4.3	0.52

Noot: aantallen levendgeborenen per 1000 vrouwen.
Bron: CBS Maandstatistiek van de bevolking.

Sexratio's.

Jaar	Sexratio				
1955	1.053	1967	1.055	1979	1.049
1956	1.061	1968	1.047	1980	1.053
1957	1.058	1969	1.055	1981	1.043
1958	1.056	1970	1.049	1982	1.051
1959	1.051	1971	1.055	1983	1.046
1960	1.055	1972	1.046	1984	1.054
1961	1.050	1973	1.046	1985	1.045
1962	1.050	1974	1.053	1986	1.035
1963	1.058	1975	1.050	1987	1.055
1964	1.059	1976	1.056	1988	1.047
1965	1.058	1977	1.056		
1966	1.053	1978	1.051		

Noot: sexratio = aantal mannelijke nieuwgeborenen / aantal vrouwelijke nieuwgeborenen.
Bron: CBS Maandstatistiek van de bevolking.

A.4 Administratieve correcties.

Deze paragraaf bevat de administratieve correcties, in absolute aantallen vanaf 1955, en verdeeld naar leeftijd en geslacht vanaf 1985.

Administratieve correcties

1955 - 678	1967 - 1590	1979 - 1700
1956 - 323	1968 - 879	1980 - 2477
1957 256	1969 - 881	1981 - 2799
1958 17834	1970 - 938	1982 - 4318
1959 - 325	1971 182	1983 - 3425
1960 270	1972 - 1626	1984 - 3433
1961 - 90	1973 - 1982	1985 - 4300
1962 - 588	1974 - 1801	1986 - 5900
1963 - 389	1975 - 1708	1987 - 9100
1964 - 838	1976 - 3142	1988 - 8200
1965 - 1029	1977 - 2672	
1966 - 932	1978 - 1550	

Noot: totaalaantallen *1000

Bron: CBS Maandstatistiek van de bevolking.

Administratieve correcties (verdeeld naar leeftijd en geslacht)

Jaar	0-19	20-29	30-39	40-49	50+	0-19	20-29	30-39	40-49	50+
1985	-266	- 980	-1111	-427	-224	-279	-411	-365	-122	- 75
1986	-279	-1433	-1502	-712	-395	-248	-525	-512	-153	-130
1987	-682	-2741	-2216	-806	-377	-522	-735	-676	-199	-149
1988	-494	-2751	-2075	-620	-330	-428	-635	-574	-175	-123

Bron: CBS Maandstatistiek van de bevolking.

Bijlage B De belangrijkste programmateksten van het uitgebreide demografische model.

B.1 Inleiding.

In deze bijlage worden de belangrijkste programmaregels beschreven in pseudo-programmeertaal. De beschreven regels hebben betrekking op de voorbewerkingen van de invoercijfers en op de berekening van de nieuwe bevolkingsaantallen.

B.2 De gebruikte symbolen.

De toegepaste symbolen zijn:

Indices

s	geslacht, 1 mannen, 2 vrouwen
x	leeftijdsjaar
c	type indeling in leeftijdsklassen
i	leeftijdsklasse
jr	jaar

Variabelen

De variabelen, die betrekking hebben op de gegevens c.q. uitkomsten voor mannen en vrouwen en voor de verschillende leeftijdsklassen:

bev_{si} , bev_{sx}	berekende bevolkingsaantallen aan begin van jaar
BEV_{si}	gegeven bevolkingsaantallen aan begin van jaar
$mort_{si}$	gegeven c.q. bewerkte mortaliteitscijfers
$migr_{si}$	gegeven c.q. bewerkte migratiecijfers
$fert_{2i}$	gegeven fertiliteitscijfers
$corr_{si}$	gegeven c.q. bewerkte administratieve correctie-cijfers

De variabelen, die betrekking hebben op de definiëring van de verschillende indelingen in leeftijdsklassen:

inage_{ci}	eerste jaar van leeftijdsklasse i voor indeling c
length_{ci}	de breedte van klasse i
meanage_{ci}	het middelste (geheeltallige) jaar van klasse i
toew_{cx}	de klasse voor indeling c , waartoe jaar x behoort

B.3 De indeling in leeftijdsklassen.

De verschillende toegepaste indelingen in leeftijdsklassen worden als volgt vastgelegd. Aangenomen wordt, dat ieder leeftijdsjaar tot één klasse behoort. Omgekeerd bevat iedere klasse één of meer (hele)leeftijdsjaren. Voor iedere indeling (met index c) worden de klassen (met index i) vastgelegd door voor iedere klasse het eerste leeftijdsjaar (met index x) op te geven, dat tot deze klasse behoort:

$$\text{inage}_{ci} = \text{eerste leeftijdsjaar behorende tot klasse } i \text{ van indeling } c$$

De breedte van de klassen wordt vastgelegd door de eerste leeftijdsjaren van de opeenvolgende klassen van elkaar af te trekken:

$$\text{length}_{ci} = \text{inage}_{c,i+1} - \text{inage}_{ci}$$

Het middelste (geheeltallige) leeftijdsjaar van de verschillende klassen:

$$\text{meanage}_{ci} = \text{inage}_{ci} + \text{length}_{ci} \text{ div } 2;$$

De operator div is de zogenaamde integer-deling. Omgekeerd kan op basis van de opgegeven eerste leeftijdsjaren inage_{ci} voor ieder leeftijdsjaar x de klasse berekend worden, waartoe ci deze behoort.

De volgende belangrijkste leeftijdsindelingen, vastgelegd middels de opeenvolgende waarden voor de variabele inage , zijn toegepast:

(Gegeven) cijfers van sterfte en bevolkingsaantallen:

$$\mathbf{c=1} \quad 0, 1, 2, 5, 10, 15, \dots, 80, 85$$

Invoercijfers van migratie-aantallen:

$$\mathbf{c=2} \quad 0, 15, 20, 30, 50, 65$$

Invoercijfers van administratieve correcties:

c=3 0,20,30,40,50

B.4 De bewerking van de mortaliteitscijfers.

De gegeven CBS mortaliteitscijfers betreffen de jaarlijkse sterfte per 100.000 individuen (tot 1980) resp. absolute sterfte-aantallen (vanaf 1980) verdeeld naar geslacht en in leeftijdsklassen. De relatieve sterftecijfers zijn betrokken op de gemiddelde omvang van de leeftijdsgroepen over het betreffende jaar. Zowel de relatieve als de absolute sterftecijfers worden omgezet in relatieve cijfers betrokken op de omvang van de leeftijds-groepen aan het begin van het jaar. De gehanteerde indeling in leeftijdsklassen is gegeven in paragraaf A.4 en wordt genummerd als $c=1$. De fracties worden als volgt bepaald ($s=1..2, i=0..19$):

Bewerking

tot 1980

$$\text{mort}_{si} := \text{mort}_{si} * \frac{\text{BEV1}_{si} + \text{BEV2}_{si}}{2 * \text{BEV1}_{si}} / 100.000$$

vanaf 1980

$$\text{mort}_{si} := \text{mort}_{si} / \text{BEV1}_{si} / 100$$

Bij de alternatieve sterfteversie is ook vanaf 1980 gebruik gemaakt van de gegeven CBS relatieve sterftecijfers in plaats van de absolute sterfte-aantallen.

B.5 De bewerking van de migratiecijfers.

De gegeven migratiecijfers betreffen absolute aantallen over een jaar verdeeld naar geslacht en in leeftijdsklassen. De gehanteerde indeling in leeftijdsklassen is gegeven in de paragrafen A.2 en B.3 en wordt genummerd als $c=2$. Aangenomen is, dat vanaf de leeftijd van 85 jaar geen (netto) migratie meer plaatsvindt. De aantallen worden omgezet in fracties ($s=1..2$):

Bewerking

```

voor i=0..nac[2]

    hulpsi := migrsi / bev[2]si
voor i=0..nac[1]

    migrsi := hulps,toew2,inage1i

```

B.6 De bewerking van de geboortecijfers.

De geboortecijfers bestaan uit fertiliteitscijfers, geboorte-aantallen en sexratio's. De fertiliteitscijfers betreffen aantallen levendgeborenen per 1000 vrouwen van de gemiddelde bevolking voor de verschillende klassen voor de leeftijdsindeling $c=1$ (zie de paragrafen A.3 en B.3). Deze cijfers betreffen centrale maten (zie paragraaf 4.5). De geboorte-aantallen betreffen aantallen levendgeborenen (*1000). De sexratio's zijn gelijk aan de verhouding tussen de aantallen mannelijke en vrouwelijke levendgeborenen.

Van de periode vóór 1967 zijn geen fertiliteitscijfers bekend. Voor deze jaren wordt direct gebruik gemaakt van de gegeven geboorte-aantallen. Vanaf 1967 worden de aantallen nieuwgeborenen berekend met behulp van gegeven fertiliteitscijfers. Uitgaande van de geboorte-aantallen worden met behulp van de sexratio de geslachts-specifieke aantallen berekend.

Bewerking:

```

# levendgeborenen :=

Σ(i=5..11) fert2i * ( bev12i + bev22i ) / 2

# levendgeborenen[2] := # levendgeborenen / ( 1 + sexratio )

# levendgeborenen[1] := sexratio * # levendgeborenen[2]

```

De aantallen nieuwgeborenen worden berekend als de som over de aangegeven leeftijdsklassen van telkens het product van fertiliteitscijfer en gemiddeld aantal vrouwen. Dit gemiddelde aantal is gedefinieerd als het gemiddelde van het (gegeven) aantal aan het begin van een jaar en het (berekende) aantal aan het einde van het jaar.

Bij de alternatieve geboorteverisie worden ook vanaf 1967 in plaats van de fertiliteitscijfers de direct gegeven geboorte-aantallen toegepast.

B.7 De bewerking van de administratieve correcties.

De administratieve correcties zijn gegeven in de vorm van absolute aantallen (vanaf 1955) en aantallen verdeeld naar geslacht en naar leeftijd (vanaf 1985). De toegepaste leeftijdsindeling is beschreven in paragraaf A.5 en wordt genummerd als c=3. Aangenomen wordt, dat de gemiddelde verdeling tussen 1985 en 1988 ook geldig is voor de voorafgaande jaren. De verschillende bewerkingen luiden als volgt (s=1..2, i=0..4):

$$\text{corr_fract}[s,i] := \frac{\sum(\text{jr}=1985..1988) \text{corr}_{\text{jr},\text{si}}}{\sum(\text{jr}=1985..1988, \text{s}=1..2, \text{i}=0..4) \text{corr}_{\text{jr},\text{si}}}$$

B.8 De berekening van de nieuwe bevolkingsaantallen.

Met behulp van de in de vorige paragrafen bewerkte invoercijfers betreffende de relatieve demografische veranderingen wordt ieder jaar op basis van de oude bevolkingssamenstelling de nieuwe berekend. Hierbij wordt de bevolking verdeeld verondersteld naar geslacht en afzonderlijk leeftijdsjaar. De berekeningen van de **basisversie** luiden als volgt (s=1..2):

- (1) $\text{bev}_{\text{s},\text{xmax}+1} := \text{bev}_{\text{s},\text{xmax}} * (1 - \text{hmort}_{\text{s},\text{xn}})$
- (2) voor i=xn-1..0 voor x=inage[1,i+1]-1..inage[1,i] (terugwaarts)
- (3) $\text{bev}_{\text{s},\text{x}+1} := \text{bev}_{\text{s},\text{x}} * (1 - \text{hmort}_{\text{s},\text{x}} + \text{migr}_{\text{s},\text{i}})$
- (4) $\text{bev}_{\text{s},85} := \text{bev}_{\text{s},85} + \text{bev}_{\text{s},86}$
- (5) $\text{bev}[\text{s},0] := \# \text{levendgeborenen}[\text{s}] * (1 - \text{mort}_{\text{s},0} / 2)$

Regel (1)

De omvang van de 85+ groep aan het einde van een jaar is gelijk aan die aan het begin ervan minus de sterfte. De sterfte is gelijk aan de (gecorrigeerde) sterftefractie voor deze klasse maal de oude omvang. Vanwege de veroudering vormt deze groep nu de 86+ groep. Voor de berekening van de gecorrigeerde sterftefractie, zie onder.

Regel (2)

De gegeven oude bevolkingsaantallen worden overschreven door de nieuwe berekende aantallen. Om geen informatie voortijdig te verliezen gebeurt dit terugwaarts.

Regel (3)

De nieuwe bevolkingsaantallen zijn gelijk aan de oude aantallen maal een vermenigvuldigingsfactor. Deze factor is gelijk aan 1 minus de mortaliteitsfractie plus de migratiefraction. Voor de migratiefraction wordt de waarde van de variabele migr van de betreffende klasse genomen, voor de sterffracie worden de waarden van de variabele mort voor de klassen omgezet in waarden van de variabele hmort voor afzonderlijke leeftijdsjaren. Dit gebeurt middels een lineaire benadering van de leeftijdsafhankelijkheid (zie onder).

Regel (4)

De nieuwe 85+ groep bestaat uit de nieuwe 86+ groep - gevormd uit de oude 85+ groep - tezamen met de 85-jarigen - gevormd uit de oude 84-jarigen.

Regel (5)

De nieuwe groep 0-jarigen bestaat uit de nieuwgeborenen gedurende het afgelopen jaar. Wordt de verdeling hiervan lineair over het jaar verondersteld, dan staat gemiddeld de helft hiervan aan de sterfte bloot.

Op de zo berekende nieuwe bevolkingsaantallen worden administratieve correcties toegepast (s=1..2, x=0..85):

$$\text{bev}_{sx} := \text{bev}_{sx} + \text{corr_fract}_{s,\text{toew}_{3x}} * \text{corr_tot} / \text{length}_{3,\text{toew}_{3x}}$$

Bij de berekende bevolkingsaantallen worden de eventueel negatieve administratieve correcties opgeteld. De correctie voor een specifiek leeftijdsjaar vormt een fractie van de totale correctie. Deze fractie vormt het quotiënt van de fractie voor de klasse, waartoe het leeftijdsjaar behoort, en het aantal jaren, dat tot de klasse behoort. De correcties worden dus homogeen over de klassen verondersteld.

Voor de oudste leeftijdsklasse, de 85+ groep, luidt de gecorrigeerde sterffracie (s=1..2):

$$\text{hmort}_{s,85} := \text{mort}_{s,\text{nac}[1]} / \left(1 + \frac{\text{bev}_{s,84}}{2 * \text{bev}_{s,85}} \right)$$

Voor de overige leeftijdsjaren gaat de lineaire benadering van de leeftijdsafhankelijkheid van de sterffracies op basis van de klassespecifieke cijfers als volgt (s=1..2, x=0..84):

```

hmortsx :=
  indien x < meanage1i
    
$$\frac{(x - \text{meanage}_{1,i-1}) * \text{mort}_{s,i} + (\text{meanage}_{1,i} - x) * \text{mort}_{s,i-1}}{\text{meanage}_{1,i} - \text{meanage}_{1,i-1}}$$

  indien x = meanage1i
    mortsi
  indien x > meanage1i
    
$$\frac{(x - \text{meanage}_{1,i} * \text{mort}_{s,i+1} + (\text{meanage}_{1,i+1} * \text{mort}_{s,i} - \text{meanage}_{1,i+1} - \text{meanage}_{1,i}))}{\text{meanage}_{1,i+1} - \text{meanage}_{1,i}}$$


```

met:

- leeftijdsjaar x behoort tot klasse i volgens indeling c=1
- mort is de gegeven sterftefractie verdeeld naar leeftijds-klasse, hmort de berekende sterftefractie verdeeld naar leeftijdsjaar.

B.9 De verschillende invoeropties.

Bij de vooruitberekeningen van de samenstelling van de Nederlandse bevolking wordt gebruik gemaakt van aannames betreffende de sterfte, de geboorte en de migratie. Deze aannames betreffen de waarden van de bijbehorende invoergrootheden. Hierbij zijn vier opties mogelijk.

Bij de formele beschrijving van deze opties wordt gebruik gemaakt van de volgende variabelen:

a _{old} _{jr,si}	reeks van jaarlijks waargenomen waarden van variabele a (jr ≤ 1988)
a _{new} _{jr,si}	reeks van toekomstige waarden van variabele a (jr ≥ 1989)
cst	de opgegeven correctiefactor

Het gemiddelde over de laatste $1 \leq n \leq 5$ jaar.

$$\begin{aligned} \text{hulp}_{si} &:= \Sigma(\text{jr}=1989-n, 1988) \text{ a_old}_{\text{jr},si} / n \\ \text{a_new}_{\text{jr},si} &:= \frac{\text{cst} * (\text{jr}-1988) + (2010-\text{jr})}{2010-1988} \text{ hulp}_{si} && \text{jr} < 2010 \\ &:= \text{cst} * \text{hulp}_{si} && \text{jr} \geq 2000 \end{aligned}$$

De laatst gemeten waarde.

$$\begin{aligned} \text{a_new}_{\text{jr},si} &:= \frac{\text{cst} * (\text{jr}-1988) + (2010-\text{jr})}{2010-1988} \text{ a_old}_{1988,si} && \text{jr} < 2000 \\ &:= \text{cst} * \text{a_old}_{1988,si} && \text{jr} \geq 2000 \end{aligned}$$

De trend gebaseerd op de laatste $2 \leq n \leq 5$ waarnemingen.

(s,i willekeurig, jr getransleerd zdd jr=n overeenkomt met jaar 1988)

$$\begin{aligned} -x &:= \Sigma(\text{jr}) \text{ jr} && -y &:= \Sigma(\text{jr}) \text{ a_old}_{\text{jr},si} \\ -xx &:= \Sigma(\text{jr}) \text{ jr} * \text{ jr} && -xy &:= \Sigma(\text{jr}) \text{ jr} * \text{ a_old}_{\text{jr},si} \\ \text{richting} &:= \frac{n * -xy - -x * -x}{n * -xx - -x * -x} \\ \text{intercept} &:= \frac{-y - \text{richting} * -x}{n} + n * \text{richting} = \text{a_new}_{1988,si} \end{aligned}$$

$$\text{a_new}_{\text{jr}+1,si} := \text{a_new}_{\text{jr},si} + \text{mult} * \exp^{(1988-\text{jr}) * \text{demp}} * \text{richting}$$

met:

mult de richting, waarin de trend doorgetrokken wordt
demp > 0 de mate, waarin de trend afgevlakt wordt

Expliciete cijfers (in geval van migratie).

$$\begin{aligned} \text{migr_fract}_{si} &:= \Sigma(\text{jr}=1984..1988) \text{ migr}_{\text{jr},si} / \\ &\quad \Sigma(\text{jr}=1984, 1988, s, i) \text{ migr}_{\text{jr},si} \\ \text{migr}[\text{jr}][s, i] &:= \text{migratiesaldo}[\text{jr}] * \text{fractie}[s, i] \end{aligned}$$

C De belangrijkste programmateksten van het spaarzame demografische model.

C.1 Inleiding.

In dit hoofdstuk worden de belangrijkste programmateksten van het spaarzame demografische model gepresenteerd, voorzover deze afwijken van die van het uitgebreide model. De gebruikte symbolen komen grotendeels overeen. De variabele bev kan zowel betrekking hebben op de omvang van een bevolkingsgroep voor een klasse i , meestal aangeduid als bev_{si} , als op het aantal voor een leeftijdsjaar, meestal aangeduid als bev_{sj} . De nieuwe grootheden zijn:

Variabelen

$imax$ het nummer van de oudste leeftijdsklasse voor de bij de berekeningen toegepaste leeftijdsverdeling

$uitfr_{si}$ de jaarlijkse uitstroomfractie

$breedte_i$ de breedte van klasse $i = inage_{i+1} - inage_i$

C.2 De basisversie.

De wiskundige modelvergelijkingen voor de berekening van de nieuwe bevolkingsaantallen luiden:

voor $s=1..2$

(1) $bev_{s,imax} :=$

$$bev_{s,imax} * (1 - uitfr_{s,imax}) + \\ bev_{s,imax-1} * (1 - uitfr_{s,imax-1}) / breedte_{imax-1}$$

(2) voor $i:=imax-1..1$ terugwaarts

$bev_{si} :=$

$$bev_{si} * (1 - uitfr_{si}) * (1 - 1/breedte_i) + \\ bev_{s,i-1} * (1 - uitfr_{s,i-1}) / breedte_{i-1}$$

(3) $bev[s,0] :=$

$$bev_{s,0} * (1 - uitfr_{s,0}) * (1 - 1/breedte_0) + \\ \# \text{nieuwgeborenen}[s]$$

De uitstroomfracties dienen in principe geschat te worden, de aantallen nieuwgeborenen zijn gegeven.

C.3 De heterogene versie.

De heterogene versie verschilt van de basisversie door een andere aanname omtrent de verdeling van de individuen over elk van de bevolkingsgroepen. In de basisversie wordt deze verdeling homogeen verondersteld. Dit houdt in, dat voor ieder afzonderlijk leeftijdsjaar de groep een gelijk aantal individuen bevat. Dit aantal is omgekeerd evenredig met de klassebreedte. Deze breedte wordt gedefinieerd als het aantal jaren, dat tot een klasse behoort.

In de heterogene versie wordt de verdeling van de individuen lineair verondersteld. Dit houdt in, dat het aantal individuen afhankelijk van de leeftijd een rechte (lineaire) lijn volgt. De aantallen individuen voor iedere leeftijd van een klasse worden volledig vastgelegd door:

- de omvang van de bevolkingsgroep,
- het aantal individuen voor het eerste leeftijdsjaar van de klasse.

Gegeven de breedte van de klasse kan op basis van de bovengenoemde twee grootheden ook het aantal individuen voor het laatste leeftijdsjaar van de klasse bepaald worden. Het laatstgenoemde aantal vormt (afgezien van de uitstroomfractie) de groep, die aan het einde van het jaar doorstroomt naar de volgende leeftijdsgroep.

De toegepaste grootheden zijn (voor willekeurig geslacht s en klasse i):

$-x_{si}$	de som van de leeftijdsjaren van de klasse
$-y_{si}$	de som van de aantallen voor elk leeftijdsjaar = de omvang van de leeftijdsgroep
$-xx_{si}$	de som van de kwadraten van de leeftijdsjaren
$-xy_{si}$	de som van het product van leeftijdsjaar en bijbehorend aantal
n_{si}	het aantal individuen van de bevolkingsgroep

Regel (1)

De aantallen individuen in het laatste leeftijdsjaar van de klasse worden berekend als functie van de omvang van de bevolkingsgroep, de breedte van de klasse en het aantal individuen in het eerste jaar van de klasse.

Regel (2)

De oudste leeftijdsklasse vormt een vangnet, zodat de uitstroom uit deze klasse nihil is.

Regel (3)

Het nieuwe aantal individuen in de oudste leeftijdsklasse vormt de som van het overlevende deel van dezelfde klasse minus de uitstroom plus het overlevende deel van de uitstroom van de voorafgaande klasse.

Regel (4)

Opeenvolgend worden de aantallen individuen van de verschillende klassen, behalve de oudste en jongste, bepaald. Deze vormen de som van het overlevende deel van de klasse minus de uitstroom plus het overlevende deel van de uitstroom van de voorafgaande klasse.

Regel (5)

Het nieuwe aantal van de jongste leeftijdsgroep is gelijk aan de som van het overlevende oude aantal minus de uitstroom plus het aantal nieuwgeborenen.

C.4 Het aantal nieuwgeborenen.

Twee manieren zijn toegepast voor de bepaling van het aantal nieuwgeborenen. De eerste rekenwijze is gebaseerd op de door het CBS expliciet opgegeven geboorte-aantallen voor de drie varianten. De tweede rekenwijze is gebaseerd op fertiliteitscijfers.

Bij de eerste rekenwijze zijn de geboorte-aantallen tot en met het jaar 2000 direct gegeven, vervolgens met een interval van 5 jaar. Voor de tussenliggende jaren is aangenomen, dat de aantallen een rechte lijn tussen de beide eindpunten volgt. De geboorte-aantallen vormen aldus een gewogen gemiddelde van de aantallen voor het begin- en eindjaar van het interval. Bijvoorbeeld voor de jaren tussen 2000 en 2005:

$$\text{geb}_{jr} := ((jr-2000) * \text{geb}_{2005} + (2005-jr) * \text{geb}_{2000}) / 5$$

in_{si} het (gegeven) aantal individuen voor het eerste leeftijdsjaar van de groep

uit_{si} het (te berekenen) aantal individuen voor het eerste leeftijdsjaar van de groep

met:

$$-x_{si} = \Sigma(j=0..breedte_i-1) j$$

$$-xx[s,i] = \Sigma(j=0..breedte_i-1) j^2$$

$$-y[s,i] = \Sigma(j=0..breedte_i-1) bev_{s,inage_i+j} = n_{si}$$

$$-xy_{si} = \Sigma(j=0..breedte_i-1) j * bev_{s,inage_i+j}$$

De grootheden $-x_{si}$ en $-xx_{si}$ worden bepaald door de toegepaste leeftijdsverdelingen en veranderen in de loop van de tijd niet. De grootheden $-y_{si}$ en $-xy_{si}$ worden bepaald door de leeftijdsopbouw van de bevolking voor de betreffende klasse en veranderen gedurende de tijd wel. De volgende vergelijkingen beschrijven deze veranderingen:

voor $s=1,2$

(1) voor $i=0..imax-1$

$$uit_{si} := 2 * bev_{si} / breedte_i - in_{si}$$

(2) $uit_{s,imax} := 0$

(3) $bev_{s,imax} := bev_{s,imax} * (1-uitfr_{s,imax}) +$
 $uit_{s,imax-1} * (1-uitfr_{s,imax-1})$

voor $i:=imax-1..1$ terugwaarts

(4) $bev_{si} := (bev_{si} - uit_{si}) * (1-uitfr_{si}) +$
 $uit_{s,i-1} * (1-uitfr_{s,i-1})$

(5) $bev_{s0} := bev_{s0} * (bev_{s0} - uit_{s0}) * (1-uitfr_{s0}) +$
 $\#$ nieuwgeborenen[s]

Een korte toelichting bij de tekstregels:

Bij de tweede rekenwijze volgen de geboorte-aantallen uit de aantallen vrouwen voor een beperkt aantal klassen en de fertiliteitscijfers:

```
gebjr := 0;
voor i=minfert..maxfert
    gebjr := gebjr + bev2i * ferti
```

De verdeling van het aantal nieuwgeborenen over de beide geslachten volgt uit de sexratio, de verhouding tussen het aantal nieuwgeborenen van het mannelijk en dat van het vrouwelijk geslacht:

```
# nieuwgeborenen2 := 1 / (1+sexratio) * # nieuwgeborenen
# nieuwgeborenen1 := # nieuwgeborenen - # nieuwgeborenen2
```

C.5 De uitstroomfractie.

De uitstroomfracties $uitfr_{si}$ worden in principe geschat. Daarnaast zijn er versies van het model beschouwd, waarbij rekening gehouden wordt met de waargenomen fracties in het jaar 1988. Hierbij wordt een lineaire verandering van de uitstroomfracties tot het jaar 2005 verondersteld (s, i willekeurig):

```
voor jr < 2005
    uitfrsi :=
        ( (jr-1988) * uitfrlsi + (2005-jr) * uitfr1988si ) / (2005-1988)
voor jr >= 2005
    uitfrsi := uitfrlsi
met:
uitfrsi      de in het jaar jr toegepaste uitstroomfractie
uitfrlsi     de onbekende te schatten 'limiet'fractie
uitfr1988si de in het jaar 1988 waargenomen fractie
```