

Methodebeschrijving tijdreeks huishoudens



Verklaring van tekens

.	gegevens ontbreken
*	voorlopig cijfer
**	nader voorlopig cijfer
x	geheim
–	nihil
–	(indien voorkomend tussen twee getallen) tot en met
0 (0,0)	het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	een cijfer kan op logische gronden niet voorkomen
2011–2012	2011 tot en met 2012
2011/2012	het gemiddelde over de jaren 2011 tot en met 2012
2011/'12	oogstjaar, boekjaar, schooljaar enz., beginnend in 2011 en eindigend in 2012
2009/'10– 2011/'12	oogstjaar, boekjaar enz., 2009/'10 tot en met 2011/'12

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312
2492 JP Den Haag

Prepress

Centraal Bureau voor de Statistiek
Grafimedia

Omslag

Telldesign, Rotterdam

Inlichtingen

Tel. (088) 570 70 70
Fax (070) 337 59 94
Via contactformulier:
www.cbs.nl/infoservice

Bestellingen

E-mail: verkoop@cbs.nl
Fax (045) 570 62 68

Internet

www.cbs.nl

© Centraal Bureau voor de Statistiek,
Den Haag/Heerlen, 2012.
Vereenvoudiging is toegestaan,
mits het CBS als bron wordt vermeld.



Centraal Bureau voor de Statistiek

Divisie Macro-economische statistieken en publicaties
Sector Ontwikkeling en ondersteuning

*Postbus 24500
2490 HA Den Haag*

Methodebeschrijving tijdreeks huishoudens

R.H.M. van der Stegen en C.N. Harmsen

Kennisgeving:

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

Projectnummer:

MOO-

BPA-nummer:

-MOO

Datum:

7 september 2011

METHODEBESCHRIJVING TIJDREEKS HUISHOUDENS

Samenvatting: Dit rapport beschrijft de methode waarmee de tijdreeks huishoudens 1971-2010 is samengesteld uit de verschillende bronnen, namelijk de volkstelling 1971, de arbeidskrachtentelling, de enquête beroepsbevolking en de gemeentelijke basisadministratie. De gevolgde werkwijze is om eerst de breuken op detailniveau te repareren en vervolgens alle resultaten in te passen zodanig dat een consistente set van tabellen ontstaat. Het resultaat zijn tabellen naar plaats in het huishouden en naar type huishouden voor het aantal personen en het aantal huishoudens, onder andere uitgesplitst naar leeftijd, geslacht, provincie, aantal kinderen en aantal leden.

Trefwoorden: tijdreeks, huishoudens, huishoudenssamenstelling, één-ouderhuishouden, alleenstaanden, paren

1.	Inleiding.....	2
1.1	Doel.....	2
1.2	De tabellen.....	2
2.	Methodebeschrijving	4
2.1	Bronnen	4
2.2	Breuken schatten en verwijderen	5
2.2.1	Indeling van de breuken naar bronbestanden.....	6
2.2.2	Breuken schatten en verwijderen in theorie	7
2.2.3	Breuken schatten en verwijderen in praktijk.....	9
2.2.4	Samenvatting repareren van de breuken	11
2.3	Controleren en waar nodig inpassen van de resultaten	11
2.3.1	Theoretische beschrijving van de inasprocedure	11
2.3.2	Praktische implementatie van het inpassen.....	11
2.4	Samenvatting methodebeschrijving	14
	Referenties	14

1. Inleiding

In de periode tussen de Volkstelling 1971 en de introductie van de huishoudensstatistiek in 2001 die gebaseerd is op integrale gegevens uit de gemeentelijke basisadministratie (GBA) zijn achtereenvolgens de arbeidskrachtentelling (AKT), het Woningbehoefteonderzoek (WBO) en de Enquête beroepsbevolking (EBB) belangrijke bronnen geweest voor de statistiek over het aantal huishoudens en de samenstelling daarvan. Zowel de volkstellinggegevens over huishoudens in 1971 als de GBA-gegevens die vanaf 1995 beschikbaar zijn vormen een goed houvast om voor de tussenliggende jaren als ijkpunt voor een tijdreeks huishoudens te dienen. Met het realiseren van een tijdreeks huishoudens van 1971 tot heden wordt een lang gekoesterde wens gerealiseerd om de dynamiek in huishoudensvorming en -ontbinding vanuit het tijdreeksperspectief eenduidig vast te leggen.

In deze methodebeschrijving wordt beschreven welke methodes zijn gekozen en welke inhoudelijke keuzes zijn gemaakt om te komen tot de voorliggende tijdreeks huishoudens.

1.1 Doel

Het doel is om een tijdreeks van de huishoudens in Nederland samen te stellen vanaf 1971 tot en met 2010. 1971 is gekozen als begin vanwege de beschikbaarheid van data, namelijk de volkstelling 1971, bovendien hebben zich in de periode na 1971 grote veranderingen in samenstellingen van de huishoudens voorgedaan. Vóór 1970 zijn er weinig data op jaarlijkse basis beschikbaar en is de dynamiek in de huishoudensamenstelling kleiner.

1.2 De tabellen

Er worden 4 tabellen gemaakt voor publicatie op Statline en nog twee andere uitsplitsingen die nodig waren voor een consistente overgang tussen het aantal personen en het aantal huishoudens. Deze laatste worden niet op Statline gepubliceerd maar zijn op verzoek als Excel bestand leverbaar.

1. Aantal personen in particuliere huishoudens; geslacht en leeftijd, vanaf 1971, jaarlijkse stand per 1 januari
 - a. naar plaats in het huishouden
 - b. naar type huishouden
2. Aantal particuliere huishoudens; type van het huishouden, leeftijd en kindertal (0 t/m 4 en 5 of meer) vanaf 1971, jaarlijkse stand per 1 januari
3. Aantal particuliere huishoudens; type van het huishouden, leeftijd en grootte (1 t/m 6 en 7 of meer) vanaf 1971, jaarlijkse stand per 1 januari

4. Particuliere huishoudens; samenstelling en provincie vanaf 1971, jaarlijkse stand per 1 januari
 - a. naar aantal personen volgens plaats in het huishouden
 - b. naar aantal huishoudens volgens type van het huishouden

Het verschil tussen de vijfde en de zesde tabel en de voorgaande tabellen naar aantal personen is dat bijvoorbeeld een kind van 5 jaar in deze tabel in de leeftijd van de referentiepersoon in het huishouden wordt gerekend en in de vorige tabel bij de gezinnen met kinderen in de leeftijd van 0-15 jaar. Als gevolg hiervan klopt de som van het aantal personen per leeftijd in deze tabel niet met het aantal personen met die leeftijd in Nederland.

5. Aantal personen in particuliere huishoudens; type van het huishouden, leeftijd van de referentiepersoon in het huishouden en kindertal vanaf 1971, jaarlijkse stand per 1 januari
6. Aantal personen in particuliere huishoudens; type van het huishouden, leeftijd van de referentiepersoon in het huishouden en grootte vanaf 1971, jaarlijkse stand per 1 januari

De referentiepersoon in een huishouden is gedefinieerd als het lid van het huishouden ten opzichte van wie de posities van de andere leden in het huishouden worden bepaald en van wie de kenmerken eventueel ook aan het huishouden worden toegekend. Uit de leden van het huishouden wordt de referentiepersoon als volgt gekozen:

- als er een paar is binnen het huishouden: de man;
- als het paar van gelijk geslacht is: de oudste van het paar;
- in een één-ouderhuishouden: de ouder;
- in de overige huishoudens: de oudste meerderjarige man of - als deze ontbreekt - de oudste meerderjarige vrouw.

Bij 'Plaats in het huishouden' wordt onderscheid gemaakt naar de volgende 9 categorieën:

- Kind
- Alleenstaande
- Partner in ongehuwd paar zonder kinderen
- Partner in gehuwd paar zonder kinderen
- Partner in ongehuwd paar met kinderen
- Partner in gehuwd paar met kinderen
- Ouder in één-ouderhuishouden
- Referentiepersoon in overig huishouden
- Overig lid van een huishouden

Het type huishouden wordt verdeeld in:

- Alleenstaande
- Ongehuwd paar zonder kinderen
- Gehuwd paar zonder kinderen
- Ongehuwd paar met kinderen
- Gehuwd paar met kinderen
- Eén-ouderhuishoudens
- Overig

De leeftijd wordt ingedeeld in de volgende categorieën:

- 0-14 jaar
- vanaf 15 jaar in 10-jaarsgroepen, dus 15-24 jaar tot en met 65-74 jaar
- en 75+

Een uitsplitsing in 5-jaarsgroepen is ook bekeken maar dit leverde te veel groepen op met te kleine aantallen, waardoor eigenlijk alleen maar ruis zichtbaar is en geen bevolkingstrends.

1.3 Opbouw van het rapport

Na deze inleiding waarin de tabellen beschreven staan die gemaakt worden, volgt de methodebeschrijving. Deze bestaat uit de volgende delen, namelijk een beschrijving van de gebruikte data, de methode om de breuken te repareren tussen de verschillende databronnen en vervolgens de verschillende tabellen op elkaar af te stemmen.

2. Methodebeschrijving

De methodebeschrijving bestaat uit 3 delen, namelijk eerst worden de beschikbare bronnen voor de tijdreeks beschreven. Omdat de bronnen op verschillende manieren zijn samengesteld en met verschillende doelen, treden er breuken op de overgangen van de ene bron naar de andere. De methode om deze breuken te bepalen en te verwijderen wordt beschreven in het tweede deel. Als laatste worden de tijdreeksen gecontroleerd en waar nodig worden verschillen tussen de verschillende sets met tabellen geminimaliseerd door het gebruik van een inpasprocedure. Bijvoorbeeld de verhouding tussen het aantal getrouwde paren met en zonder kinderen en aantal personen in getrouwde paren met en zonder kinderen moet een plausibele waarde hebben.

2.1 Bronnen

De volgende bronnen zijn voor de tijdreeks beschikbaar:

- De volkstelling (VT) uit 1971. Dit is een integrale telling van de bevolking geweest. Daarom wordt aangenomen dat in dit bestand geen methodefout/breuk zit.
- De arbeidskrachtentelling (AKT) uit 1975, 1977, 1981, 1983 en 1985. De AKT heeft een breuk ten opzichte van de VT71. Daarnaast zit tussen 1977 en 1981 een breuk vanwege een methodewijziging, respectievelijk AKT-oud en AKT-nieuw. De AKT is ook voor een paar andere jaren beschikbaar echter zonder ophoogfactoren waardoor de informatie niet bruikbaar is.
- Het woningbehoefte-onderzoek (WBO) is een enquête die onder andere in 1981, 1985 en 1987 is uitgevoerd. De overige jaren bevatten breuken. Vanwege de grote overlap met de AKT en slechts een beperkt aantal jaren die beschikbaar zijn, is besloten deze informatie niet te gebruiken tenzij zou blijken dat de modellen anders niet oplosbaar zijn.
- De enquête beroepsbevolking (EBB) is beschikbaar tussen 1988 en 1999 zonder breuken in de EBB; na 1999 volgen breuken, de EBB-data uit deze jaren worden niet gebruikt omdat ze beter worden beschreven door de GBA.
- De Gemeentelijke basisadministratie is een integrale telling; op de oude manier bewerkt (GBA-oud) is deze beschikbaar vanaf 1995 tot en met 2010.
- De op de nieuwe manier¹ bewerkte Gemeentelijke basisadministratie (GBA-nieuw) is ook een integrale telling, beschikbaar vanaf 2006 tot en met 2010.

Er wordt aangenomen dat GBA-nieuw en VT71 de meest nauwkeurige weergave geven van de huishoudens. Dit zijn dus de uitgangspunten van de tijdreeks. Andere bronnen worden aangepast ten opzichte van deze informatie om tot een plausibel verloop van de tijdreeks te komen.

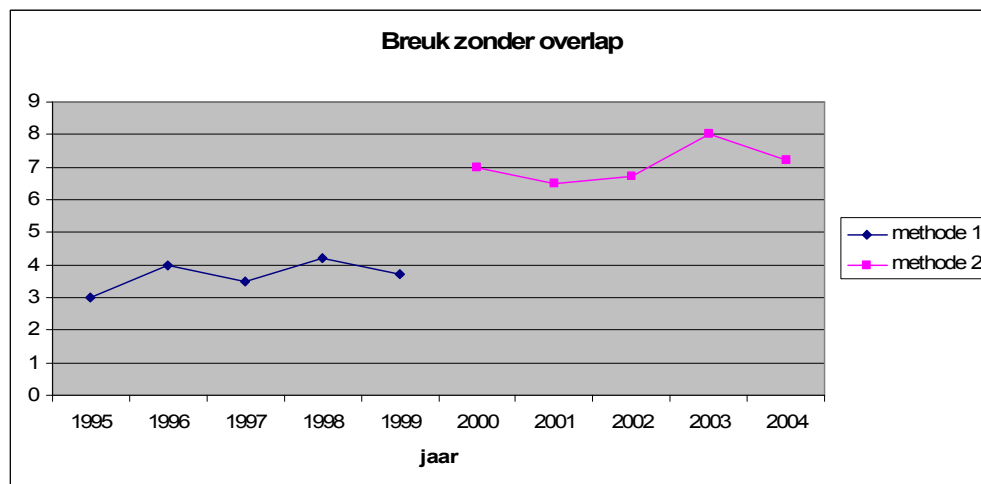
De eerste bewerkingsstap van alle bestanden is dat aan alle records de leeftijd, het type en het geslacht van de referentiepersoon van het huishouden wordt toegevoegd. Daarnaast is in de bronbestanden de classificatie van de huishoudensvariabelen in lijn gebracht met de sinds 1995 geldende classificatie. Vervolgens worden de bestanden geaggregeerd tot het laagste publicatieniveau waarbij de records in de AKT en EBB worden gewogen met hun gewichten om de tabellen zoals beschreven in paragraaf 1.2 te kunnen afleiden. Voor elke tabel, elk jaar en elke bron wordt apart geaggregeerd. Dit betekent dat er voor een aantal jaren informatie uit twee bronnen beschikbaar is die niet aan elkaar gelijk hoeft te zijn.

2.2 Breuken schatten en verwijderen

Uit de vorige paragraaf moge blijken dat er twee type breuken op te lossen zijn in de tijdreeks. Ten eerste zijn er breuken waarbij er geen periode is waarin de 2

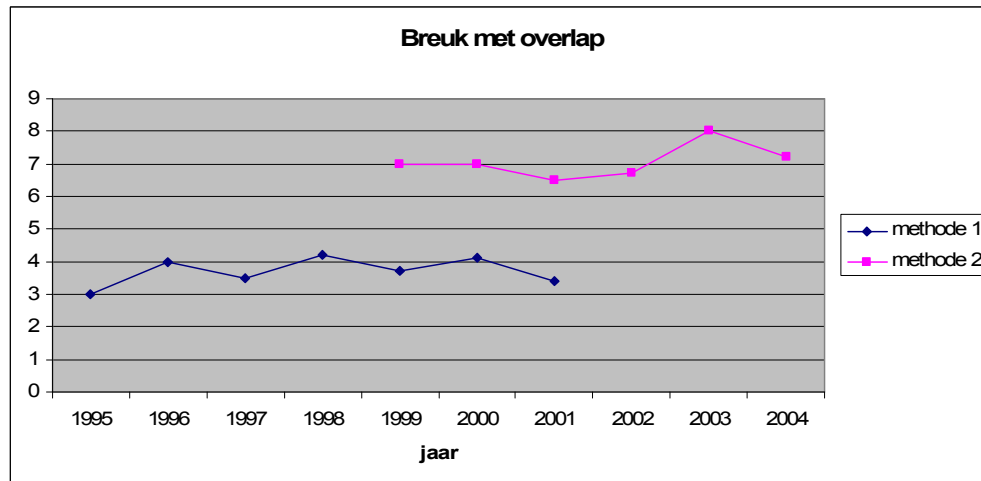
¹ Verschillen tussen de nieuwe en de oude bewerkingsmethode van de GBA zijn beschreven in Business Analyse Document Huishoudens (H. Witvliet, 2011).

bronnen/methoden tegelijk informatie geven. De ene bron geeft tot een bepaald tijdstip informatie en de volgende bron pas na dat tijdstip, de zogenaamde breuk zonder overlappende informatie (figuur 1).



Figuur 1: Schematisch voorbeeld van een breuk zonder overlappende data voor een bepaalde periode.

Ten tweede zijn er breuken waarbij voor een bepaalde periode data beschikbaar zijn uit twee (of meerdere) bronnen met ongelijke methoden, waardoor de uitkomsten verschillen, de zogenaamde breuken met overlappende data (figuur 2).



Figuur 2: Schematisch voorbeeld van een breuk met overlappende data voor dezelfde periode

2.2.1 Indeling van de breuken naar bronbestanden

In chronologische volgorde hebben we te maken met de volgende breuken in de tijdreeks huishoudens:

- De breuk tussen VT71 en AKT-oud is een breuk tussen 1971 en 1975, dus een breuk zonder overlappende data. Hierbij wordt aangenomen dat de VT71 op zichzelf geen breuk bevat.
- De breuk tussen AKT-oud en AKT-nieuw is een breuk tussen 1977 en 1981, ook zonder overlappende data.
- De breuk tussen AKT-nieuw en EBB loopt tussen 1985 en 1988, dus eveneens een breuk zonder overlappende data.
- Breuk tussen EBB en GBA-oud is een breuk met data volgens beiden methoden tussen 1995 en 1999 dus een breuk met overlappende data.
- Breuk GBA-oud en GBA-nieuw is een breuk met data tussen 2006 en 2010 dus een breuk met overlappende data. Er wordt aangenomen dat GBA-nieuw geen breuk bevat ten opzichte van de VT71.

In de volgende paragraaf wordt voor beide type breuken een methode beschreven om de breuken te kwantificeren en te verwijderen.

2.2.2 Breuken schatten en verwijderen in theorie

In de tijdreeksen zijn 2 typen breuken aanwezig, namelijk met en zonder overlappende periode. Er geldt dat de breuken in de verschillende reeksen naar plaats of type op moeten tellen naar de breuk in het totaal aantal huishoudens of het totaal aantal personen over alle typen of plaatsen in het huishouden. Echter in sommige gevallen blijkt de breuk in het totaal verwaarloosbaar klein te zijn.

2.2.2.1 Breuken met overlappende periode

Er zijn meerdere methoden beschikbaar om breuken te schatten waarbij over 1 of meerdere jaren data beschikbaar zijn die hetzelfde beschrijven maar een andere waarde hebben, omdat ze op verschillende manieren bepaald zijn. Naar mate er meer overlap is, zijn er meer mogelijkheden. Eerst worden twee gewone methoden besproken, vervolgens wordt de stochastische variant van beide methoden gegeven:

- De eenvoudigste methode voor het berekenen van de breuk is het gemiddelde berekenen van de overlappende periode volgens de ene methode en volgens de andere en daar het verschil van nemen. Dit is een zogenaamde 0^{de}-orde breuk. Er wordt geen rekening gehouden met verschillen in ontwikkeling. Meestal wordt de gehele periode van overlap genomen, maar er kunnen redenen zijn om slechts een gedeelte te nemen. Deze methode is toepasbaar vanaf 1 punt overlap, maar de toevalsfactor is dan wel erg groot als gevolg van ruis in beide manieren van bepalen van de overlappende waarden.
- Als er meer dan 1 punt overlap is, is een uitbreiding op de vorige methode het toevoegen van een ontwikkeling aan de breuk. Dit kan bijvoorbeeld een ontwikkeling in de tijd zijn of een ontwikkeling van een andere onafhankelijke variabele. In dit geval spreekt men van een eerste-orde of

hogere-orde breuk. Hiervoor wordt eerst per overlappend punt het verschil bepaald tussen de waarden die op 2 verschillende manieren berekend zijn. Met bijvoorbeeld een kleinste-kwadratenmethode kan bijvoorbeeld dan een formule worden afgeleid voor de breuk. Deze methode is toepasbaar vanaf 2 punten overlap, maar de toevalsfactor is dan wel erg groot als gevolg van ruis in beide manieren van bepalen van de overlappende waarden.

- Met stochastische tijdreeksmethoden, zoals ARMA of structurele tijdreeksmodellen is het mogelijk om een breuk te schatten waarbij informatie uit de gehele reeks wordt gebruikt, niet alleen uit de overlappende periode. Ook hier zijn er varianten waarbij een constante breuk geschat wordt of een breuk afhankelijk van een onafhankelijke variabele zoals de tijd. Deze methoden zijn nog in een onderzoekstadium en worden daarom nog niet toegepast.

Het blijkt dat de eerste en eenvoudigste methode zeer goede resultaten geeft. Daarom wordt hiervoor gekozen bij breuken met overlappende reeksen. Bij een aantal breuken blijkt de steekproef-onnauwkeurigheid in de enquêtes te groot te zijn voor de tweede methode. Daarnaast houdt de eerste methode op eenvoudige wijze rekening met de inherente relaties tussen de verschillende breuken in deelreeksen, zoals beschreven in paragraaf 2.2.2.3.

2.2.2.2 Breuken zonder overlappende periode

Voor breuken zonder overlappende periode zijn er grofweg dezelfde mogelijkheden, de gewone methoden en de stochastische methoden. Om de breuk te kunnen bepalen is eerst een voorspelling/extrapolatie nodig zodanig dat er overlap gecreëerd wordt. En vervolgens kunnen de bovenstaande methoden gebruikt worden voor het bepalen van de breuken. In tegenstelling tot de vorige paragraaf zijn de stochastische modellen in het algemeen het beste voor deze situatie.

2.2.2.3 Relaties tussen breuken

In het algemeen geldt dat de breuken die in een aantal deelreeksen worden geschat, niet hoeven op te tellen tot de breuk in de aggregaatreeks. Hierbij moeten we 2 situaties apart beschouwen, namelijk die zonder breuk in het aggregaat en die met een breuk in het aggregaat.

Zo is de totale particuliere (niet-institutionele) bevolking, het aggregaat, bekend zonder breuken. Als blijkt dat er een breuk zit in bijvoorbeeld de deelreeks van één-ouderhuishoudens, betekent dit dat deze personen verdeeld moeten worden over andere categorieën, zodanig dat zowel voor als na reparatie van de breuk de deelreeksen optellen tot het aggregaat. Hiervan wordt gebruik gemaakt bij de breukreparatie van tabel 1 uit paragraaf 1.2.

Het aantal huishoudens en het aantal personen naar type referentiepersoon hebben geen extern gegeven aggregaat (dit betreft alle tabellen behalve tabel 1 uit paragraaf 1.2). Bij de bepaling van de breuken in de deelreeksen moet dus gelden dat de breuken van de deelreeksen optellen tot de breuk in het aggregaat.

Omdat de eerste methode in paragraaf 2.2.2.1 een volledig lineaire methode is, is hier automatisch aan voldaan. Dit is een van de redenen om voor deze methode te kiezen. Echter als er geen overlappende periode is, is het extrapoleren belangrijk voor het bepalen van de breuk, daarom is voor een stochastische methode gekozen. In het artikel van van den Brakel (2008) staat beschreven hoe in dat geval de consistentie tussen de breuken gerealiseerd kan worden. Dit is geprogrammeerd in SSF-pack (Koopman 1999).

2.2.3 Breuken schatten en verwijderen in praktijk

In deze sectie wordt eerst de algemene procedure voor het verwijderen van de breuken beschreven. In het tweede deel worden de uitzonderingen die gemaakt zijn voor bepaalde tabellen weergegeven.

2.2.3.1 Algemeen deel

De volgende stappen zijn voor alle tabellen hetzelfde:

- De data wordt gegroepeerd of per leeftijd en eventueel geslacht, of per provincie of per aantal kinderen of per aantal leden van het huishouden. Elke groep wordt in een aparte Excel sheet geplaatst. Bij het aantal personen naar plaats of type in tabel 1 wordt een jaar proportioneel opgehoogd naar de gegeven particuliere (niet institutionele) bevolking als hier niet aan voldaan is.
- Vervolgens worden met de eerste methode uit sectie 2.2.2.1 de breuken tussen GBA-nieuw en GBA-oud verwijderd, waarbij een GBA-reeks ontstaat die als referentie ‘nieuw’ heeft. Dezelfde procedure wordt herhaald tussen de GBA-reeks en de EBB-reeks, waarbij deze laatste aan de GBA-reeks wordt gekoppeld.
- In het structurele tijdreeksmodel worden de data per groep ingelezen. Eigenlijk is de wens om alle groepen per tabel tegelijk te analyseren in een meerdimensionaal probleem maar dat is numeriek nog niet mogelijk. De VT71 en de gekoppelde GBA-reeks zijn de referentiewaarden die niet veranderen. Het model bepaalt de breuken van de AKT-oud en AKT-nieuw ten opzicht van de referentie. Er zijn twee situaties te onderscheiden:
 - Geen breuk in het aggregaat, zoals bij aantal personen naar plaats of type. In dit geval moet voor zowel de AKT-oud als AKT-nieuw de breuk optellen tot 0.
 - In de andere gevallen wordt een aggregaatreeks toegevoegd waarvoor geldt dat de breuken in deze reeks de som zijn van de breuken in onderliggende reeksen.

Het resultaat van de berekeningen, de uitvoer, bestaat onder andere uit de breuken en de originele trend van de tijdreeksen.

- De uitvoer van het structurele tijdreeksmodel wordt ingelezen in de Excel-sheet waarna de trend van reeksen gecorrigeerd wordt door de breuken te verwijderen. Er wordt gekozen voor de trend en niet voor de originele reeks. De trend heeft geen ontbrekende waarden, terwijl tussen 1972 en 1987 een aantal jaar ontbreken in de brondata van de originele reeks, en bovendien heeft de trendreeks minder last van ruis in de steekproeven van de AKT en EBB.

2.2.3.2 Specifieke bewerkingen

Per tabel zijn een aantal uitzonderingen gemaakt die hard worden opgelegd. Dit zijn de volgende zaken:

- In de tabel 1 ‘aantal personen in particuliere huishoudens’ geldt:
 - Onder de 15 jaar is een persoon
 - ‘kind’ of ‘overig’ in plaats in het huishouden.
 - Opgenomen in een ‘(on)gehuwd paar met kinderen’, ‘één-ouderhuishouden’ of overig huishouden volgens type van het huishouden.
 - In de groep 75+ komen de categorieën ‘kinderen’ en ‘partners in ongehuwd paar met kinderen’ niet voor. Vanwege te geringe aantal is het niet mogelijk te bepalen of de aantallen meetfouten zijn of werkelijk bestaan.
- In de tabellen 2 en 3 ‘aantal particuliere huishoudens; type van het huishouden, leeftijd en kindertal of grootte’ gelden:
 - Een referentiepersoon in een huishouden is per definitie 15 jaar of ouder, dus de categorie 0-14 komt niet voor
 - In de groep 75+ komt de categorie ‘geen partners in ongehuwd paar met kinderen’ niet voor.
 - In de tabellen naar kindertal, komen kinderen alleen voor in één-oudergezinnen en (on)gehuwde paren met kinderen
 - In de tabel naar grootte: Alleen alleenstaanden hebben per definitie 1 lid, de andere type huishoudens hebben meerdere leden. Eén-oudergezinnen, gehuwde en ongehuwde paren zonder kinderen en overige huishoudens hebben minimaal twee leden. Aan de paren zonder kinderen kunnen ook overige leden worden toegevoegd waardoor meer dan 2 leden mogelijk is. Ongehuwde en gehuwde paren met kinderen hebben minimaal 3 leden.
- ‘Huishoudens naar provincie’ worden geen aannamen gemaakt bij het repareren van de breuken.

2.2.4 Samenvatting repareren van de deelbreuken

In deze sectie is beschreven hoe de breuken gerepareerd zijn per tabel per groep (bijvoorbeeld, leeftijd, geslacht etc). Er is alleen consistentie tussen de verschillende reeksen in een groep, zogenaamde gedeeltelijke consistentie. De hierboven beschreven procedures zijn niet bruikbaar om op tabelniveau of tussen de tabellen onderling consistentie te bereiken. Volledige consistentie wordt bereikt door in te passen, zoals beschreven wordt in de volgende paragraaf.

2.3 Controleren en waar nodig inpassen van de resultaten

De methode waarbij de reeksen per groep geanalyseerd worden heeft als nadeel dat er geen consistentie is tussen de deelgroepen en de aggregaatgroep, en ook niet tussen de verschillende tabellen. De verschillen zijn vaak niet groot maar werken wel verstorend, vooral bij zeer kleine aantallen. Er is bijvoorbeeld een klein verschil in aantal personen in ongehuwd paar zonder kinderen volgens ‘type huishouden’ en volgens ‘plaats in het huishouden’, maar het aantal in ‘type huishouden’ moet wel groter zijn dan in ‘plaats in het huishouden’ omdat enkele personen van de categorie overig uit ‘plaats in het huishouden’ daaraan toegevoegd zijn. Daarom worden de verschillende resultaten ingepast. Dit heeft als bijkomend voordeel dat fouten in de brondata gedetecteerd worden. Als eerste wordt nu de mathematische procedure beschreven en vervolgens wordt uitgelegd hoe deze geïmplementeerd is.

2.3.1 Theoretische beschrijving van de inpasprocedure

Voor het inpassen zijn verschillende mogelijkheden beschikbaar. In dit geval moet de procedure kunnen omgaan met:

1. gewichten; de variabelen moeten een gewicht hebben wat aangeeft hoeveel ze mogen veranderen tijdens het inpassen. Dit is nodig omdat de getallen variëren van tientallen tot miljoenen.
2. gelijkheidsrestricties
3. ongelijkheden
4. en eventueel harde grenzen, bij voorbeeld dat alle getallen positief moeten zijn en de particuliere bevolking.

De procedure QUADPROG in MATLAB voldoet aan deze eisen en daarom wordt deze gebruikt, voor details zie de help-file van MATLAB.

2.3.2 Praktische implementatie van het inpassen

Het inpassen wordt in meerdere ronden gedaan om de rekentijd niet te groot te laten worden waarbij steeds een set van afhankelijke tabellen wordt ingepast. Dit resulteert in 4 inpassingen, waarbij zo veel mogelijk in de eerste keer wordt ingepast met de beperking dat de rekentijd niet te veel mag oplopen.

1. De eerste inpassing is de grootste voor tabel 1 en de randen van tabellen 2, 3, 5 en 6. (De uitsplitsing naar zowel leeftijd als grootte of kindertal wordt

in de tweede inpassing gedaan). Deze tabellen worden hierbij simultaan ingepast met hun onderlinge relaties.

- a. Tabel 1 heeft de volgende eisen:
 - i. Het aantal personen naar plaats in het of type van het huishouden per leeftijd en geslacht wordt gegeven door de particuliere (niet-institutionele) bevolking naar leeftijd en geslacht per jaar². Dit zijn harde randen waaraan voldaan moet worden. Tevens moet gelden dat de verschillende categorieën per leeftijd en geslacht optellen tot de aggregaten van die categorieën.
 - ii. Daarnaast is er een relatie tussen type en plaats in het huishouden. Het aantal alleenstaanden in type van het huishouden moet precies gelijk zijn aan het aantal alleenstaanden in plaats in het huishouden. Vervolgens moet gelden dat plaats in het huishouden de ondergrens geeft voor het aantal personen in (on)gehuwde paren met of zonder kinderen en in één-ouderhuishoudens.
- b. Tabellen 5 en 6 hebben de volgende restricties:
 - i. Deze tabel heeft als hard uitgangspunt de gegevens van de particuliere (niet-institutionele) bevolking. Deze eis geldt alleen voor de sommatie over alle leeftijden en kindertal of grootte.
 - ii. Plaats in het huishouden geeft het exacte aantal alleenstaanden per leeftijd. Tevens moet gelden dat de verschillende categorieën ‘type huishouden’ optellen tot de aggregaten van die categorieën naar type van het huishouden uit tabel 1. Alleen is de verdeling per type huishouden over de verschillende leeftijden is anders, omdat in tabel 1 de leeftijd van de persoon is opgenomen en in onderhavige tabel de leeftijd van de referentiepersoon.
- c. Tabellen 2 en 3 ‘aantal huishoudens naar leeftijd, kindertal of grootte’ hebben de volgende restricties:
 - i. In tegenstelling tot de voorgaande tabellen is ligt het totale aantal huishoudens niet vast. Het aantal huishoudens met alleenstaanden is in tabellen 5 en 6 bepaald omdat elk persoon per definitie ook een huishouden is. Ouders in één-ouderhuishoudens en referentiepersonen in overige

² De bevolking in particuliere huishoudens wordt voor de periode 1995-2011 gevonden in de Statline tabel: Bevolking; kerncijfers: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37296NED&D1=0,56&D2=20-60&HD=110913-0945&HDR=T&STB=G1....>

huishoudens volgens plaats in het huishouden geven ook direct het aantal huishoudens in deze categorieën. De overige categorieën kunnen worden begrensd door het maximum te bepalen. Dit wordt gegeven door het aantal personen naar type huishouden naar leeftijd referentiepersoon te delen door het minimale aantal leden van een huishouden. Dit is voor de (on)gehuwde paren met kinderen 3, voor de (on)gehuwde paren zonder kinderen en voor 1-ouderhuishoudens en overige huishoudens 2. Tevens moet gelden dat de verschillende categorieën per leeftijd, kindertal of grootte optellen tot de aggregaten van die categorieën.

- ii. De aantallen huishoudens naar grootte worden bepaald door uit tabel 4 het aantal personen te delen door de grootte van het huishouden. Voor de categorie '7 of meer leden' gaat dit niet omdat onbekend is hoeveel leden dit precies zijn. Dit is een restpost die wordt bepaald uit het verschil tussen het totaal en de optelling van de categorieën 1 t/m 6 leden.
 - iii. Bij de vermenigvuldiging van het aantal huishoudens naar kindertal met het kindertal wordt het totaal aantal kinderen per type gekregen. Dit moet kleiner zijn dan het totaal aantal kinderen volgens 'plaats in het huishouden'. Het product is kleiner omdat de groep '5 of meer kinderen' niet exact weergeeft hoeveel kinderen in deze groep aanwezig zijn.
2. In de tweede inpasronde worden de uitsplitsingen naar het aantal huishoudens naar type en leeftijd van de referentiepersoon verdeeld over kindertal en grootte ingepast. Per leeftijd of per grootte/kindertal is in de vorige inpassing bepaald, echter die inpassing voor zowel leeftijd als grootte/kindertal is nog niet bepaald. Dat gebeurt aan de hand van de randtotalen bepaald in de vorige inpassing.
 3. Het aantal personen naar plaats in het huishouden per provincie en aantal huishoudens naar type en provincie is de derde inpasgroep.
 - a. Het aantal personen per plaats in het huishouden per provincie heeft als harde rand de particuliere (niet-institutionele) bevolking per provincie³. Daarnaast worden de totalen per plaats gegeven door de totalen per plaats uit de vorige inpasprocedure.

³ Deze kan voor de periode 1995 – 2011 gevonden worden op Statline in de tabel Huishoudens; personen naar geslacht, leeftijd en regio, 1 januari:
<http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71488NED&D1=0,11&D2=0&D3=0&D4=5-16&D5=a&HD=110913-0949&HDR=G4,T&STB=G1,G2,G3>

...

- b. Het aantal huishoudens naar type huishouden per provincie heeft de volgende randvoorwaarden: het aantal huishoudens per type in Nederland uit de tabel ‘aantal huishoudens naar leeftijd referentiepersoon’ en uit de volgende categorieën in het aantal personen naar plaats in het huishouden per provincie: ouders in één-ouderhuishoudens, referentiepersonen in overig huishouden en de alleenstaanden. De som van deze personen is gelijk aan het aantal huishoudens.

2.3.3 Vergelijkbaarheid met reeds gepubliceerde resultaten.

Omdat er een aantal econometrische bewerkingen op de data zijn uitgevoerd, verschillen de tijdreeks data van eerder gepubliceerde data. Een aantal oorzaken zijn:

1. In de oude resultaten waren niet ingedeelde mensen waardoor niet voldaan wordt aan het randtotaal van de bevolking. Deze ontbrekende personen zijn wiskundig bijgeschat.
2. Omdat, bijvoorbeeld de EBB, in een aantal gevallen een behoorlijke steekproeffout heeft ten opzichte van het aantal personen, is het niet verstandig om deze resultaten zonder bewerking te gebruiken. Er is daarom gekozen om een tijdreeks te splitsen in een trend en fluctuatie. De trend wordt gebruikt.
3. Als laatste wordt de trend verschoven door het inpassen.

2.4 Samenvatting methodebeschrijving

In een aantal stappen zijn tabellen over huishoudens afgeleid uit microdata-bestanden. De methode van bewerking van de data kan per databron verschillen maar is in zijn principe voor alle tabellen gelijk. Het begint met aggregeren tot het laagste niveau waarop gepubliceerd gaat worden. Vervolgens worden de breuken gerepareerd en tot slot worden de tabellen met elkaar consistent gemaakt.

Referenties

- Brakel, Jan A. van den, Paul A. Smith Simon Compton 2008; Quality procedures for survey transitions – experiments, time series and discontinuities; Survey Research Methods (2008); volume 2; no 3; pp. 123-141
- Koopman, Siem Jan, Niel Shephard, Jurgen A. Doornik 1999; Statistical algorithms for models in state space using SsfPack 2.2; Econometrics Journal (1999); volume 2, pp. 113-166.