

# Methodebreuken

# 10

*Jan van den Brakel, Thomas Kraan en Joeri Roels*

**Statistische Methoden (10006)**



## 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
2008–2009	= 2008 tot en met 2009
2008/2009	= het gemiddelde over de jaren 2008 tot en met 2009
2008/'09	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2008 en eindigend in 2009
2006/'07–2008/'09	= oogstjaar, boekjaar enz., 2006/'07 tot en met 2008/'09

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*

TelDesign, Rotterdam

### *Inlichtingen*

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

### *Bestellingen*

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

### *Internet*

[www.cbs.nl](http://www.cbs.nl)

ISSN: 1876-0333

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2010.  
Verveelvoudiging is toegestaan, mits het CBS als bron wordt vermeld.

## Inhoudsopgave

1.	Inleiding .....	5
1.1	Algemene beschrijving en leeswijzer .....	5
1.2	Afbakening en relatie met andere thema's .....	6
1.3	Plaats in het statistisch proces .....	7
2.	Kwantificeren via herberekenen .....	8
2.1	Korte beschrijving .....	8
2.2	Toepasbaarheid.....	9
2.3	Uitgebreide beschrijving .....	9
2.3.1	Schatters van standen en ontwikkelingen.....	9
2.3.2	Herberekeningen .....	11
2.3.3	Alternatieve herberekeningsmethoden.....	11
2.3.4	Backcasting procedures.....	12
2.4	Voorbeeld: SBI-overgang voor de economische statistieken.....	15
2.4.1	De SBI voor economische statistieken.....	15
2.4.2	Stratificatie volgens de oude indeling SBI'93.....	15
2.4.3	Stratificatie volgens de nieuwe indeling, SBI2008 .....	18
2.4.4	Toepassing van de variantieschatters bij het steekproefontwerp .....	18
2.4.5	Backcasten .....	19
2.5	Kwaliteitsindicatoren.....	19
3.	Experimenten .....	21
3.1	Korte beschrijving .....	21
3.2	Toepasbaarheid.....	21
3.3	Uitgebreide beschrijving .....	22
3.3.1	Inleiding .....	22
3.3.2	Principes van proefopzetten .....	23
3.3.3	Proefopzetten .....	23
3.3.4	Steekproefomvang.....	25
3.3.5	Traditionele analysemethoden voor experimenten .....	26
3.3.6	Design-based analysemethoden voor experimenten in kanssteekproeven .....	27
3.4	Corrigeren van reeksen.....	29
3.5	Voorbeeld .....	31

3.6	Kwaliteitsindicatoren.....	32
4.	Tijdreeksmodellen.....	34
4.1	Korte beschrijving .....	34
4.2	Toepasbaarheid.....	34
4.3	Uitgebreide beschrijving .....	35
4.3.1	Univariate structurele tijdreeksmodellen .....	35
4.3.2	Multivariate structurele tijdreeksmodellen.....	36
4.3.3	State Space vorm.....	39
4.3.4	Kalmanfilter .....	41
4.3.5	Correctie.....	42
4.3.6	Alternatieve interventies .....	42
4.3.7	Transformatie van de doelvariabelen.....	43
4.3.8	Software .....	44
4.4	Voorbeeld .....	44
4.5	Eigenschappen.....	47
4.6	Kwaliteitsindicatoren.....	49
5.	Literatuur .....	51

## **1. Inleiding**

### **1.1 Algemene beschrijving en leeswijzer**

De missie van het CBS is het verzamelen, bewerken en publiceren van onbetwiste en relevante statistische informatie over allerlei aspecten van de Nederlandse samenleving. Om ontwikkelingen te kunnen beschrijven worden veel statistische processen continu in de tijd uitgevoerd of met vaste tijdsintervallen herhaald. Door de opzet van statistische processen zoveel mogelijk ongewijzigd te laten, worden tijdreeksen verkregen die de ontwikkeling van een verschijnsel door de tijd heen zo goed mogelijk beschrijven. Het blijft echter onvermijdelijk om de opzet van een onderzoek van tijd tot tijd te herzien. Dataverzamelmethodeën en vragenlijstontwerpen kunnen verouderen waardoor vanuit kwaliteitsoverwegingen aanpassingen noodzakelijk zijn. Daarnaast wordt geïnvesteerd in onderzoek naar de ontwikkeling van nieuwe waarnemingstrategieën die vooral gericht zijn op het realiseren van noodzakelijke kosten- en lastendrukvermindering. Bij voldoende kwaliteit wordt primaire waarneming vervangen door secundaire waarneming uit registers. Indien primaire waarneming noodzakelijk is, zal dit zoveel mogelijk via mixed-mode worden uitgevoerd om de kosten van de primaire dataverzameling zoveel mogelijk te reduceren.

Veranderingen in de opzet van het statistische proces kunnen de uitkomsten systematisch beïnvloeden. Dit heeft tot gevolg dat de tijdreeksen die de ontwikkeling van indicatoren beschrijven worden verstoord. Het verschijnsel dat een wijziging in de onderzoeksopzet de tijdreeks van een indicator verstoort, wordt kortweg aangeduid met de term methodebreuk. Indien op een bepaald moment veranderingen worden doorgevoerd in het statistische proces is het wenselijk dat kan worden aangegeven hoe groot het effect hiervan is op de schattingen van de belangrijkste indicatoren. Hierdoor kan de autonome ontwikkeling van een indicator worden gescheiden van het effect van de veranderingen in de onderzoeksopzet op de uitkomsten.

Voor veel gebruikers is de vergelijkbaarheid van de statistische informatie met het verleden een belangrijk onderdeel van de relevantie van de statistische output. Daarom worden in dit themarapport methoden beschreven die de nadelige effecten van een herontwerp op de continuïteit van reeksen zoveel mogelijk beperken.

De aard van de veranderingen in het herontwerp is mede bepalend voor de wijze waarop methodebreuken gekwantificeerd kunnen worden. Indien het herontwerp zich beperkt tot veranderingen in classificaties, steekproefkaders of de verwerking of de analyse van een steekproefonderzoek, dan kunnen breuken in kaart worden gebracht door dezelfde data te verwerken volgens het reguliere en het vernieuwde proces. In hoofdstuk 2 wordt een methodiek beschreven om het effect van een verandering in classificaties te kwantificeren en te corrigeren. Indien veranderingen plaats vinden in de waarneming van een steekproefonderzoek, is het doorgaans

onvermijdelijk dat een grootschalig veldexperiment moet worden uitgevoerd. Statistische methoden voor het ontwerpen en analyseren van grootschalige veldexperimenten in een regulier survey onderzoek worden beschreven in hoofdstuk 3. In dit hoofdstuk worden ook methoden beschreven om reeksen te corrigeren voor de waargenomen breuken. Ten slotte kan de omvang van methodebreuken worden geschat door het uitvoeren van een interventieanalyse met behulp van een geschikt tijdreeksmodel. Deze methodiek wordt beschreven in hoofdstuk 4.

Analyseren en corrigeren van methodebreuken is een onderwerp waarbij een groot scala van statistische methoden toegepast kan worden. In deze bijdrage wordt ingegaan op herberekenen gebruik makend van steekproefmethoden, ontwerp en analyse van proeven en tijdreeksanalyse. De scope van ontwerp en analyse van proeven is breder dan van herberekenen en het gebruik van tijdreeksmodel. Daarom is, omwille van de omvang van het document, het onderdeel ontwerp en analyse van proeven ten opzichte van de andere twee onderwerpen wat globaler behandeld. Naast deze drie methoden kunnen in specifieke situaties andere methoden meer geschikt zijn voor het kwantificeren en corrigeren van methodebreuken. Op het moment dat deze situatie in de praktijk voordoet kan de Methodenreeks hiermee uitgebreid worden. Tot slot wordt vermeld dat met name de design-based analysemethoden voor grootschalige veldexperimenten en de tijdreeksmodellen voor het schatten en corrigeren van methodebreuken nog verder door ontwikkeld kan worden. Op deze onderdelen zal deze bijdrage voor de Methodenreeks in de toekomst aangevuld kunnen worden.

## **1.2 Afbakening en relatie met andere thema's**

Methodebreuken doen zich voor in reeksen die zijn verkregen via directe schatters uit de steekproeftheorie, zoals de Horvitz-Thompsonschatting, en de gegeneraliseerde regressieschatting. In dit hoofdstuk wordt ook ingegaan op het schatten van varianties, het bepalen van de minimale steekproefomvang en de allocatie over de strata. Deze onderwerpen worden uitgebreider behandeld in de Methodenreeks over steekproeftheorie, Banning en Knottnerus (2010).

In hoofdstuk 2 wordt aangegeven dat bij beperkte steekproefomvang modelmatige schattingsmethoden gebruikt kunnen worden om de precisie van directe schatters te verbeteren. Deze technieken zijn nader uitgewerkt in de Methodenreeks over modelmatig schatten. Onder dit thema is een Methodenreeksbijdrage geschreven over synthetische schatters en kleinedomeinschatters op basis van gemengde modellen (Boonstra en Buelens, 2007). Tevens is een Methodenreeksbijdrage geschreven over kleinedomeinschatters op basis van structurele tijdreeksmodellen (Van den Brakel en Krieg, 2010). De Methodenreeksbijdrage over structurele tijdreeksmodellen vertoont grote raakvlakken met de technieken beschreven in hoofdstuk 4 van deze bijdrage, die ook gebaseerd zijn op structurele tijdreeksmodellen.

Methodebreuken ontstaan doordat veranderingen in het statische proces worden doorgevoerd. In deze bijdrage voor de Methodenreeks wordt beschreven hoe de

effecten van deze veranderingen op de uitkomsten kunnen worden gekwantificeerd. Voor een kwalitatieve interpretatie van de waargenomen effecten ontstaan relaties met onderdelen van de Methodenreeks die de onderdelen van het waarnemingsproces beschrijven. Hierbij gaat het om Methodenreeksbijdrage over benaderingsstrategieën, deelthema's responsverhogende maatregelen bij bedrijfsenquêtes (Berkenbosch, 2009), responsverhogende maatregelen bij sociale statistieken (Luiten, 2009), algemene beschrijving en uni-mode designs (Beukenhorst en Kerssemakers, 2010), mixed-mode benadering (Beukenhorst en Schouten, 2010), Methodenreeksbijdrage over vragenlijstontwikkeling (Giesen, e.a., 2010) en de Methodenreeksbijdrage over foutenbronnen (Bethlehem, 2010).

### **1.3 Plaats in het statistisch proces**

Kwantificeren en analyseren van methodebreuken kan op verschillende plaatsen effect hebben op de verwerking van microdata omdat deze bijvoorbeeld opnieuw moeten worden gecodeerd. Vervolgens grijpt deze methodiek in op het schattingsproces indien reeksen opnieuw worden berekend via standaard schattingstechnieken uit de steekproeftheorie. Indien op een macromethode wordt gecorrigeerd kan dit worden gezien als een tweede stap in het schattingsproces.

Het kwantificeren en corrigeren van methodebreuken via tijdreeksmodellen kan worden gezien als een extra analyseslag die plaats vindt na het schatten. De input voor de tijdreeksmodellen zijn immers de tijdreeksen die via het reguliere schattingsproces zijn samengesteld.

Ontwerp en analyse van veldexperimenten grijpt in op bijna het gehele statistische verwerkingsproces. Dit komt omdat via deze aanpak de oude en de nieuwe opzet van het steekproefproces, parallel aan elkaar uitgevoerd dienen te worden. Na afronden van beide processen is een aanvullende analyse noodzakelijk om hypothesen over de methodebreuken te toetsen.

## 2. Kwantificeren via herberekenen

### 2.1 Korte beschrijving

Bij veel statistieken wordt gebruik gemaakt van onderverdelingen binnen de populatie. Men kan bijvoorbeeld geïnteresseerd zijn in resultaten voor een deel van de populatie. In dat geval spreekt men van *domeinen* waarover men uitspraken wil doen. Zo is er binnen de economische statistieken een classificatie naar activiteiten van de bedrijven. De cijfers worden dan per activiteit gepubliceerd. Het is doorgaans efficiënt om de onderverdeling in domeinen waarnaar gepubliceerd wordt, ook te gebruiken als stratificatievariabele in een zogenaamd gestraficteerd steekproefontwerp. Hierdoor wordt de precisie van de schattingen per domein verhoogd. Daarnaast is het mogelijk om de steekproefomvang per domein te bepalen zodat bij de steekproeftrekking beter rekening kan worden gehouden met vooraf gespecificeerde precisie eisen voor de domeinschattingen. Daarom valt de onderverdeling van de populatie in strata voor de steekproeftrekking vaak samen met de onderverdeling in domeinen waarnaar gepubliceerd wordt; noodzakelijk is dit echter niet.

Economische of sociologische ontwikkelingen of nieuwe regelgeving kunnen een wijziging van de relevante onderverdeling in deelpopulaties waarnaar gepubliceerd wordt nodig maken. Het is dan interessant om te reconstrueren welke cijfers in eerdere jaren zouden gelden voor de nieuwe domeinindeling, uitgaande van responsgegevens die verzameld zijn volgens het toen gebruikte oude steekproefontwerp. Dit noemt men *terugleggen* of *backcasten*. Omgekeerd kan het interessant zijn te bezien wat de ontwikkeling is in de oude domeinen in jaren na de wijziging.

Steekproeven worden vaak ontworpen door te kijken voor welke steekproefaantallen per stratum de te verwachten nauwkeurigheden van de schatters (uitgedrukt in steekproefvarianties) voor relevante domeinen zo klein mogelijk zijn, gegeven een bepaalde totale steekproefomvang. De overgang op een andere indeling vraagt extra aandacht bij de berekening van de steekproefvarianties die bij het maken van zo'n ontwerp nodig zijn.

In dit hoofdstuk wordt beschreven op welke manier niveauschattingen kunnen worden gemaakt voor een herziene domeinindeling en de bijbehorende steekproefvarianties. Dit gebeurt met het oog op terugleggen en backcasten, d.w.z. het reconstrueren naar het verleden van reeksen van een herziene classificatie, uitgaande van historische responsgegevens. In dit hoofdstuk worden voornamelijk zogenaamde design-based schattingsmethoden uit de steekproeftheorie toegepast. Dit zijn schattingsmethodieken die voornamelijk gebaseerd zijn op de eigenschappen van het steekproefontwerp. Andere technieken, die meer expliciet gebruik maken van statistische modellen, zullen kort worden benoemd.



## 2.2 Toepasbaarheid

De methode is toepasbaar wanneer men voor domeinen, die zijn gebaseerd op een herziene indeling wil berekenen wat de totalen zijn als men de beschikking heeft over een set steekproefdata die zijn waargenomen onder de oude indeling. Voor deze herberekeningen is het noodzakelijk dat voor de steekproefelementen de classificatie op basis van de oude en nieuwe domeinindeling bekend is. Ook dienen insluitkansen van de steekproefelementen volgens de oorspronkelijke indeling bekend te zijn. Uiteraard moeten ook de waarden van de doelvariabelen voor deze steekproefelementen bekend zijn.

Het vereist doorgaans veel inspanning om te achterhalen hoe steekproefelementen, geclassificeerd op basis van de oude indeling, zouden worden geclassificeerd op basis van de nieuwe indeling. In veel gevallen is dit zelfs niet mogelijk zonder additionele informatie bij de betreffende steekproefelementen te verzamelen. Ook is het niet gegarandeerd dat voor het schatten van parameters op basis van de nieuwe domeinindeling voldoende steekproefelementen beschikbaar zijn. Bij de steekproeftrekking heeft men immers geen rekening kunnen houden met de minimale steekproefaantallen op basis van de nieuwe classificatie. Een te geringe steekproefmassa kan resulteren in onacceptabel grote design-varianties voor de schattingen van bepaalde domeinen.

Een alternatief is dat de cijfers op het geaggregeerde niveau van de domeinen worden vertaald naar een andere indeling. Hiervoor is een verdeelsleutel in termen van overgangsproporties nodig. Deze worden geschat op basis van een steekproef waarbij voor alle steekproefelementen de indeling volgens de oude en de nieuwe classificatie bekend is. Als alternatief kunnen ze worden afgeleid uit het steekproefkader. Deze wijze van terugleggen op macroniveau leidt doorgaans tot stabielere schatters maar het synthetische karakter van deze methodiek kan ook snel resulteren in vertekening in de uitkomsten.

## 2.3 Uitgebreide beschrijving

### 2.3.1 *Schatters van standen en ontwikkelingen*

De behandeling van herberekeningen bij een wijziging in de indeling van de populatie past in de algemene theorie over schattingen voor deelpopulaties met behulp van domeinschatters. In de volgende deelparagrafen presenteren we daarom eerst de theorie van deze domeinschatters.

#### 2.3.1.1 *De domeinschatter*

Uitgaande van de responsdata kunnen grootheden voor de populatie en deelverzamelingen van de populatie geschat worden. Er zijn verschillende schattingsmethoden. We bespreken er twee, de Horvitz-Thompsonschatter en de Hájekschatter.

De *Horvitz-Thompsonschat*ter is een design-zuivere schatter voor het populatietotaal en wordt gegeven door:

$$\hat{Y}_s = \sum_{i \in s} \frac{y_i}{\rho_i}, \quad (2.3.1)$$

waarbij  $y_i$  de respons is behorende bij steekproefelement  $i$  en  $\rho_i$  de insluitkans van dat element. De sommatie loopt dan over alle responselementen die zijn waargenomen in de steekproef  $s$ .

Wanneer bij een statistiek over deelpopulaties (domeinen) gepubliceerd gaat worden, kan ook weer de Horvitz-Thompsonschatte gebruikt worden, waarbij nu gesommeerd wordt over alle elementen die in domein  $d$  vallen. De Horvitz-Thompsonschatte voor het totaal van domein  $d$  wordt dan

$$\hat{Y}_d = \sum_{i \in s} \frac{y_i d_i^d}{\rho_i}. \quad (2.3.2)$$

In deze uitdrukking is  $d_i^d$  een indicatorvariabele die de waarde 1 aanneemt als steekproefelement  $i$  in domein  $d$  zit, en de waarde 0 in alle andere gevallen.

Het domeintotaal kan nauwkeuriger geschat worden door via een ratioschatte gebruik te maken van het bekende populatie totaal van het domein. Deze schatter staat bekend als de *Hájek schatter* en wordt gegeven door

$$\hat{Y}_d = \frac{\sum_{i \in s} \frac{y_i d_i^d}{\rho_i}}{\sum_{i \in s} \frac{d_i^d}{\rho_i}} N_d, \quad (2.3.3)$$

waarbij  $N_d$  de omvang van de populatie van domein  $d$  is. Het label  $i$  indiceert weer de steekproefelementen.

Het quotiënt in (2.3.3) bestaat uit de Horvitz-Thompsonschatte voor het populatietotaal van de doelvariabelen en de Horvitz-Thompsonschatte voor de omvang van de populatie voor domein  $d$ . Dit resulteert dus in een ratioschatte voor het domeingemiddelde, dat vervolgens wordt opgehoogd met de bekende populatieomvang ( $N_d$ ). Deze schatter is doorgaans stabielier dan de Horvitz-Thompsonschatte voor het populatietotaal omdat een over- of onderschatting voor het populatietotaal in de teller doorgaans gecompenseerd wordt door een over- of onderschatting voor de populatieomvang in de noemer. Verder kan de nauwkeurigheid van zowel de Horvitz-Thompsonschatte als de Hájekschatter verbeterd worden door gebruik te maken van meer achtergrondkenmerken waarvoor de populatieverdelingen bekend zijn via de zogenaamde gegeneraliseerde regressieschatte. Voor meer details over de domeinschatte verwijzen we naar Särndal e.a. (1992), hoofdstukken 2 en 5.7, 8, 10.

### 2.3.1.2 Berekenen van ontwikkelingen

Periode-op-periode ontwikkelingen worden bij de bedrijfseconomische statistieken vaak uitgedrukt met een groeivoet, bijvoorbeeld de Laspeyres index. Voor een domein  $d$  is deze voor periode  $t$  gedefinieerd als:

$$\hat{Q}_d^t = \frac{\hat{Y}_d^t}{\hat{Y}_d^{t-1}}. \quad (2.3.4)$$

In deze uitdrukking is  $\hat{Y}_d^t$  de schatter in domein  $d$  in periode  $t$ . In de voorbeelden zullen we voor deze schatter in sommige gevallen de Horvitz-Thompsonschatting en in andere de Hájekschatter kiezen.

Tot zover de introductie van domeinschatters en berekeningen van indices.

### 2.3.2 Herberekeningen

Indien de indeling van domeinen waarnaar gepubliceerd wordt, verandert, kan het noodzakelijk zijn om tijdreeksen waargenomen in het verleden opnieuw te berekenen voor deze nieuwe indeling. Indien het mogelijk is om de afzonderlijke elementen in de steekproef die zijn waargenomen onder de oude classificatie in te delen naar de nieuwe classificatie, dan kunnen de in de vorige paragraaf geïntroduceerde domeinschatters voor de herberekening gebruikt worden.

De domeinschatting (de Horvitz-Thompsonschatting in vergelijking (2.3.2) of de Hájekschatter in vergelijking (2.3.3)) kan immers gebruikt worden om een schatting te maken voor willekeurige domeinen. Zolang het mogelijk is om de afzonderlijke elementen te classificeren op basis van de nieuwe indeling, kunnen schattingen worden gemaakt voor de nieuwe domeinen, ook als het stratificatieschema van het steekproefontwerp gebaseerd is op de oude classificatie. Men dient zich wel te realiseren dat de oorspronkelijke steekproef niet is ontworpen voor het bepalen van domeintotalen naar de nieuwe domeinindeling. Sommige domeinen kunnen dan te weinig steekproefelementen bevatten om voldoende betrouwbare schattingen te maken.

Met de geschatte populatietotalen voor de nieuwe domeinindeling kunnen ook ontwikkelingen en groeivoeten worden geschat via formule (2.3.4).

In paragraaf 2.4 wordt een voorbeeld uitgewerkt voor de overgang op een andere activiteitenindeling voor economische statistieken. In deze paragraaf worden de uitdrukkingen gegeven voor domeinschatters voor populatietotalen en indices met de bijbehorende steekproefvarianties voor een gestratificeerd steekproefontwerp. Een andere toepassing is het berekenen van de effecten van een nieuwe indeling van het steekproefkader. Dit is uitgewerkt in Van den Brakel (2007).

### 2.3.3 Alternatieve herberekeningsmethoden

De in 2.3.2 genoemde onnauwkeurigheid als gevolg van slecht gevulde domeinen in de nieuwe indeling kan deels gerepareerd worden door gebruik te maken van

hulpinformatie die voor de hele populatie bekend is. Dit gebeurt via de gegeneraliseerde regressieschatter, zie Särndal e.a. (1992), hoofdstuk 10. Een andere, meer effectieve methode is gebruik te maken van modelgebaseerde kleine domeinschatters die uitgaan van gemengde lineaire modellen (Rao, 2003), of tijdreeksmodellen (Pfeffermann en Burck, 1990). In Van den Brakel (2009) wordt nader ingegaan op de toepassing van deze methoden in de context van het herberekenen van reeksen voor nieuwe domeinindelingen.

#### *2.3.4 Backcasting procedures*

Backcasten van tijdreeksen is het bepalen van waarden volgens een nieuwe indeling voor en uitgaande van data uit perioden toen nog met de oude indeling gewerkt werd. Dit kan op verschillende manieren. De methoden worden ingedeeld in zg. micro- en macromethoden. Bij micromethoden wordt uitgegaan van individuele responsgegevens die zo goed mogelijk in de nieuwe indeling worden gepast. Dan kunnen totalen voor nieuwe domeinen worden berekend. Bij macromethoden worden deeltotalen omgerekend naar de nieuwe indeling (en dus niet individuele responsgegevens).

##### *2.3.4.1 Benadering op microniveau*

Bij de benadering op microniveau wordt gebruik gemaakt van de Hájekschatter of de Horvitz-Thompsonschatting, toegepast op responsdata uit eerdere jaren. De schatters zijn gegeven in vergelijkingen (2.3.2) en (2.3.3). Om deze te kunnen gebruiken moet bekend zijn hoe de responsdata uit eerdere jaren, toen met de oude indeling gewerkt werd, past in de nieuwe indeling. Voor de Horvitz-Thompsonschatting is het voldoende als dit voor de responsdata bekend is. Voor de Hájekschatter moet voor de hele populatie de oude indeling bekend zijn, omdat bij deze schatter ook domeinomvangen gebruikt worden. Een probleem is in beide gevallen de mogelijk beperkte celvulling in de nieuwe domeinen. Kleindomeinschatters kunnen gebruikt worden om deze problemen te overwinnen. De hulpinformatie kan dan steekproefdata uit eerdere perioden zijn (in een tijdreeksmodel), of informatie uit nabijgelegen domeinen (via een gemengd lineair model).

Het voordeel van de benadering op microniveau is dat de zo verkregen gebackcaste cijfers nog steeds gebaseerd zijn op empirie. De noodzakelijke dubbelcoderingen van data en populatie van eerdere jaren (nieuwe en oude indeling) maken deze methode wel arbeidsintensief.

##### *2.3.4.2 Benadering op macroniveau*

Gezien de nadelen van de benadering op microniveau (domeinen met slechte vulling en de arbeidsintensiviteit van het hercoderen van individuele responsgegevens voor de nieuwe classificatie) wordt vaak een benadering op macroniveau toegepast. De benadering op macroniveau gaat ervan uit dat schattingen voor de nieuwe domeinen verkregen kunnen worden als lineaire combinaties van schattingen voor de oude

domeinen. Voor een nieuw domein  $l$  kan zo'n schatting  $\tilde{Y}_l^t$  dan geschreven worden als

$$\tilde{Y}_l^t = \sum_k b_{k,l} \hat{Y}_k^t, \quad (2.3.5)$$

met  $\hat{Y}_k^t$  een directe schatter voor het totaal van de variabele  $Y$  in het oude domein  $k$ . Het label  $t$  refereert naar de periode. De coëfficiënten  $b_{k,l}$  zijn de elementen van een zogenaamde conversiematrix, die aangeven welke fractie van het totaal  $\hat{Y}_k^t$  terecht komt in domein  $l$  van de nieuwe indeling. De informatie die gebruikt wordt om deze conversiematrix op te stellen is bepalend voor de kwaliteit van deze backcasting procedure. Omdat de coëfficiënten fracties zijn die de verdeling specificeren van  $\hat{Y}_k^t$  over de klassen van de nieuwe indeling geldt:

$$\sum_l b_{k,l} = 1.$$

De conversiematrix is op verschillende manieren samen te stellen. Meestal wordt deze voor één overgangperiode opgesteld. De conversiefactoren worden vervolgens gebruikt om reeksen in het verleden te berekenen voor de nieuwe domeinindeling via formule (2.3.5).

Overgangsmatrices worden afgeleid van een hulpvariabele, zeg  $X$ , die aangeven welk aandeel van deze hulpvariabele van het ene domein naar het andere overgaat. Dit leidt tot de volgende veelgebruikte definitie van de elementen van de conversiematrix,

$$b_{k,l} = \frac{X_{k,l}}{\sum_l X_{k,l}},$$

met  $X_{k,l}$  het totaal van variabele  $X$  dat in domein  $k$  onder de oude indeling en in domein  $l$  onder de nieuwe indeling valt. In het geval van economische statistieken kan de hulpvariabele dan bijvoorbeeld aantal bedrijven, omzet, toegevoegde waarde of aantal werknemers zijn. Dergelijke grootheden zijn vaak beschikbaar uit registers, waardoor de conversiefactoren zelf geen steekproeffouten bevatten. Deze aanpak op macroniveau heeft echter een sterk synthetisch karakter. Er wordt immers verondersteld dat de overgangen zoals die worden waargenomen voor de hulpvariabelen ook van toepassing zijn op de doelvariabelen die gebackcast moeten worden. Afhankelijk van de juistheid van deze veronderstelling wordt met deze procedure een vertekening in de gebackaste cijfers geïntroduceerd. De kwaliteit van de teruggelegde reeksen hangt sterk af van de keuze van de hulpvariabele waarmee de conversiefactoren worden bepaald.

Tot nu toe is impliciet verondersteld dat de conversiematrix tijdinvariant is en gebaseerd is op de hulpinformatie uit een bepaalde periode, bijvoorbeeld het jaar waarin over wordt gegaan op de nieuwe classificatie. Indien met een tijdinvariante conversiematrix reeksen worden gebackcast, dan wordt verondersteld dat de fracties

waarmee een variabele van de oude naar de nieuwe classificatie overgaat door de tijd heen niet verandert. Dit is een zeer sterke modelveronderstelling waar doorgaans niet aan voldaan is. Het is bijvoorbeeld zeer onwaarschijnlijk dat nieuwe bedrijfstakken door de tijd heen met dezelfde verhoudingen kunnen worden afgeleid uit bestaande reeksen. Dit is een tweede potentiële bron van vertekening die geïntroduceerd wordt met backcasten op macroniveau. Naarmate reeksen over een langere periode worden gebackcast is de houdbaarheid van deze aanname twijfelachtiger. Vanuit dit oogpunt is het beter om met de micromethode te werken of, als compromis, met tijdafhankelijke overgangsmatrices te werken. Dit wordt nader uitgewerkt in paragraaf 2.3.4.3.

Naast een hulpvariabele kan de doelvariabele zelf worden gebruikt om de elementen van de conversiematrix af te leiden, uitgaande van responswaarden voor de doelvariabele. Dit voorkomt een lastige afweging welke hulpvariabele gekozen moet worden. Element  $X_{k,l}$  is dan de schatter van het populatietotaal van de elementen die behoren tot domein  $k$  onder de oude indeling, en domein  $l$  onder de nieuwe indeling. Bij deze aanpak kunnen weer de Hájek- of Horvitz-Thompsonschaters worden gebruikt, waarin dan gesommeerd wordt over die respons elementen die in domein  $k$  naar de ene indeling en domein  $l$  naar de andere zitten. Wanneer dan de Horvitz-Thompsonschaters gebruikt wordt, heeft dit het bijkomende voordeel dat in de periode dat met twee indelingen gewerkt wordt, de gebackcaste cijfers overeenkomen met de domeinschattingen voor de nieuwe domeinindeling. Dit impliceert dat de aansluiting van de gebackcaste reeksen via formule (2.3.5) op de directe schattingen voor de nieuwe domeinindeling, continu is. In andere gevallen kan de continue aansluiting via interpolaties (zie Van den Brakel e.a., 2008) bereikt worden.

Ook voor de methode op macroniveau is een arbeidsintensieve dubbelcodering nodig, bijvoorbeeld in het overgangsjaar. Afhankelijk van hoe de overgangsmatrices worden berekend (uit registers of uit waarneming, met de Horvitz-Thompsonschaters of met de Hájekschatters) is zo'n dubbelcodering nodig voor de hele populatie of alleen voor de steekproefelementen.

#### *2.3.4.3 Benadering op macroniveau met tijdafhankelijke conversiefactoren*

Met de macromethode worden doorgaans stabiele schattingen verkregen voor de gebackcaste reeksen. Het nadeel van de benadering op macroniveau is dat bij deze procedure vertekening in de gebackcaste reeksen wordt geïntroduceerd. Een van de aannames is dat de conversiematrix (bepaald uitgaande van data uit het jaar van dubbelcodering van populatie en/of steekproef) over de jaren heen constant zou zijn. Dit is te ondervangen door de conversiematrix tijdafhankelijk te maken. Dit vereist echter dubbelcoderingen voor meer tijdvakken. Men kan dit voor een beperkt aantal perioden doen, en dan de conversiematrix interpoleren in tussenliggende jaren.

Naast het dubbel coderen van steekproefbestanden voor verschillende tijdstippen, kan ook gebruik worden gemaakt van de kwalitatieve kennis van inhoudelijk deskundigen. Op basis van deze kennis kan worden ingeschat van af welk moment

in het verleden een nieuwe bedrijfstak is ontstaan en hoe de interpolatie tussen de beschikbare conversiematrices het beste kan plaats vinden.

#### 2.3.4.4 *Indices backcasten*

Voor het backcasten van indices kan men de onderliggende reeks voor de populatietotalen backcasten. Deze worden vervolgens gebruikt om de indices voor de nieuwe indeling uit te rekenen. Er bestaan ook alternatieve methoden om indices op macroniveau te backcasten. Deze hebben als voordeel dat ze betrekkelijk eenvoudig zijn te berekenen. Zo'n methode staat uitgelegd in paragraaf 2.2.3 van Eurostat (2006).

Voor meer informatie over backcasten op macroniveau wordt verwezen naar Buiten e.a. (2008), Van den Brakel (2009), Kampen (2007), Eurostat (2006).

## 2.4 Voorbeeld: SBI-overgang voor de economische statistieken

### 2.4.1 *De SBI voor economische statistieken*

Een voorbeeld van een verandering van indeling is de modernisering van de standaard bedrijvenindeling (SBI) die in 2008 is ingegaan. De SBI is de indeling van bedrijven naar activiteit bij een aantal economische statistieken, zoals de productiestatistiek (PS), de korte termijnstatistieken (KS) en de investeringsstatistiek (IS). Bij economische statistieken is het publicatieniveau (het domein  $d$ ) de kerncel en het stratificatieniveau kerncel gekruist met grootteklasse. Voor de SBI'93 zijn de domeinen de kerncellen van die indeling weergegeven met  $k$ . Voor de SBI2008 zijn de domeinen de kerncellen van de SBI2008 weergegeven met  $l$ .

### 2.4.2 *Stratificatie volgens de oude indeling SBI'93*

In de periode van de overgang van de bedrijfsindeling kan voor het steekproefontwerp de SBI'93 als stratificatievariabele worden genomen. De strata zijn dan SBI'93-kerncellen gekruist met grootteklasse. De relevante domeintotalen naar SBI'93 worden dan uitgerekend met de Horvitz-Thompsonschatting en die naar SBI2008 met de Hájekschatter. Voor beide schatters zijn de insluitkansen van de steekproefelementen nodig. Voor een gestratificeerd steekproefontwerp zonder terugleggen, waarbij de strata gelijk zijn aan kerncel  $k$  naar SBI'93 gekruist met grootteklasse  $g$ , ( $1, \dots, G$ ), worden de eerste-orde insluitkansen van de respondenten, gegeven door

$$\pi_i = \frac{n_{g,k}}{N_{g,k}} \text{ voor } i \in U_{g,k}, \quad (2.4.1)$$

waarbij voor stratum  $U_{g,k}$   $n_{g,k}$  de netto steekproefomvang is en  $N_{g,k}$  de populatieomvang. De tweede-orde insluitkansen, nodig voor variantieschatters, worden gegeven door

$$p_{ij} = \begin{cases} \frac{n_{g,k}(n_{g,k} - 1)}{N_{g,k}(N_{g,k} - 1)} \text{ voor } i, j \hat{=} U_{g,k} \\ \frac{n_{g,k}n_{g',k'}}{N_{g,k}N_{g',k'}} \text{ voor } i \hat{=} U_{g,k}, j \hat{=} U_{g',k'} \end{cases} \quad (2.4.2)$$

Met deze insluitkansen vereenvoudigt de Horvitz-Thompsonschatteer voor het populatietotaal in periode  $t$  voor kerncel  $k$  naar SBI'93 tot de bekende schatteer voor een gestratificeerd steekproefontwerp:

$$\hat{Y}_k^t = \sum_{g=1}^G \frac{N_{g,k}}{n_{g,k}} \sum_{i=1}^{n_{g,k}} y_{i,g,k}^t, \quad (2.4.3)$$

met  $y_{i,g,k}^t$  de responswaarde in periode  $t$  van respondent  $i$  uit stratum  $g,k$ . De bijbehorende variantieschatteer voor een gestratificeerd steekproefontwerp volgt uit de variantie van de Horvitz-Thompsonschatteer door invullen van de eerste- en tweede-orde insluitkansen (2.4.2) en wordt gegeven door:

$$\hat{V}(\hat{Y}_k^t) = \sum_{g=1}^G N_{g,k} (N_{g,k} - n_{g,k}) \frac{\hat{S}_{g,k}^2}{n_{g,k}}$$

met stratumvarianties

$$\hat{S}_{g,k}^2 = \frac{1}{n_{g,k} - 1} \sum_{i=1}^{n_{g,k}} (y_{i,g,k}^t - \hat{Y}_{g,k}^t)^2$$

en geschatte stratumgemiddelden

$$\hat{Y}_{g,k}^t = \frac{1}{n_{g,k}} \sum_{i=1}^{n_{g,k}} y_{i,g,k}^t.$$

Uitdrukking (2.4.3) kan ingevuld worden in de uitdrukking voor de Laspeyres index, vgl. (2.3.4), zodat de groeivoet van periode  $t-1$  op periode  $t$  gelijk is aan:

$$\hat{Q}_k^t = \frac{\hat{Y}_k^t}{\hat{Y}_k^{t-1}}.$$

De uitdrukking voor  $\hat{Y}_k^{t-1}$  is analoog aan die in vgl. (2.4.3), met  $y_{i,g,k}^{t-1}$  in plaats van  $y_{i,g,k}^t$ .

Een variantieschatteer voor de Laspeyres index gebaseerd op de variantie voor een gestratificeerd steekproefontwerp

$$\hat{V}(\hat{Q}_k^t) = \frac{1}{(\hat{Y}_k^{t-1})^2} \sum_{g=1}^G N_{g,k} (N_{g,k} - n_{g,k}) \frac{\hat{S}_{g,k}^2}{n_{g,k}},$$

met geschatte stratumvarianties



$$\hat{S}_{g,k}^{t^2} = \frac{1}{n_{g,k} - 1} \sum_{i=1}^{n_{g,k}} (\hat{z}_{i,g,k}^t - \hat{Z}_{g,k}^t)^2, \quad (2.4.4)$$

uitgedrukt in de residuen

$$\hat{z}_{i,g,k}^t = y_{i,g,k}^t - \hat{Q}_k^t y_{i,g,k}^{t-1} \quad (2.4.5)$$

en gemiddelden van deze residuen

$$\hat{Z}_{g,k}^t = \frac{1}{n_{g,k}} \sum_{i=1}^{n_{g,k}} \hat{z}_{i,g,k}^t.$$

Deze berekening van de variantie geldt indien de steekproefelementen in opeenvolgende perioden worden waargenomen. Voor panelsteekproeven als gebruikt in de KS is dit het geval. Bij cross-sectionele steekproeven zijn de schattingen voor de populatietotalen tussen de opeenvolgende perioden onafhankelijk en kan de variantie van de index worden benaderd met een eerste-orde Taylor linearisatie.

De populatietotalen voor de kerncellen naar SBI2008 kunnen nu worden herberekend, uitgaande van de aanpak uit 2.3.2. Dit gaat met de Hájek domeinschatting (vergelijking (2.3.3)),

$$\hat{Y}_l^t = \frac{\sum_{i \in s} y_i^t \delta_i^l}{\sum_{i \in s} \pi_i} N_l = \frac{\sum_{k=1}^K \sum_{g=1}^G \frac{N_{g,k}}{n_{g,k}} \sum_{i=1}^{n_{g,k}} y_{i,g,k}^t \delta_i^l}{\sum_{k=1}^K \sum_{g=1}^G \frac{N_{g,k}}{n_{g,k}} \sum_{i=1}^{n_{g,k}} \delta_i^l} N_l. \quad (2.4.6)$$

De bijbehorende variantieschatting is

$$\hat{V}(\hat{Y}_l^t) = \sum_{k=1}^K \sum_{g=1}^G N_{g,k} (N_{g,k} - n_{g,k}) \frac{\hat{S}_{g,k}^{t^2}}{n_{g,k}},$$

met  $\hat{S}_{g,k}^{t^2}$  gedefinieerd als in (2.4.4), met nu als residuen

$$\hat{z}_{i,g,k}^t = y_{i,g,k}^t \delta_i^l - \hat{Y}_l^t \delta_i^l.$$

Deze schatters kunnen weer worden ingevuld in de uitdrukking voor de Laspeyres index. Dan geldt voor de groeivoet

$$\hat{Q}_l^t = \frac{\sum_{k=1}^K \sum_{g=1}^G \frac{N_{g,k}}{n_{g,k}} \sum_{i=1}^{n_{g,k}} y_{i,g,k}^t \delta_i^l}{\sum_{k=1}^K \sum_{g=1}^G \frac{N_{g,k}}{n_{g,k}} \sum_{i=1}^{n_{g,k}} y_{i,g,k}^{t-1} \delta_i^l}.$$

Een schatter voor de bijbehorende variantie is

$$\hat{V}(\hat{Q}_l^t) = \frac{1}{(\hat{Y}_l^{t-1})^2} \sum_{k=1}^K \sum_{g=1}^G N_{g,k} (N_{g,k} - 1) \frac{\hat{S}_{g,k}^{t^2}}{n_{g,k}},$$

met  $\hat{S}_{g,k}^{t^2}$  en  $\hat{Z}_{g,k}^t$  gedefinieerd als boven, echter nu met

$$\hat{z}_{i,g,k}^t = y_{i,g,k}^t \delta_i^l - \hat{Q}_l^t y_{i,g,k}^{t-1} \delta_i^l.$$

Wederom is er bij deze variantie schatter van uit gegaan dat de steekproefelementen in opeenvolgende perioden via een panelopzet worden waargenomen.

### 2.4.3 Stratificatie volgens de nieuwe indeling, SBI2008

Stel dat in het overgangsjaar de steekproef wordt getrokken volgens een allocatie naar SBI2008. De ophoging naar SBI2008 gaat dan weer via de schatters afgeleid voor een gestratificeerd steekproefontwerp en de ophoging voor SBI'93 domeinen kan met de Hájekschatter uitgevoerd worden. De uitdrukkingen voor deze schatters zijn, evenals die voor de bijbehorende varianties, analoog aan de schatters beschreven in paragraaf 2.4.2.

Indien in het overgangsjaar wordt gestratificeerd naar de nieuwe SBI indeling, kan het noodzakelijk zijn, ten behoeve van het bepalen van de steekproefomvang en allocatie, schattingen te maken van stratumvarianties voor de stratificatie naar de SBI2008 indeling. Daarvoor zijn alleen steekproefgegevens beschikbaar van eerdere jaren toen met een stratificatie naar SBI'93 gewerkt werd. In deze situatie moeten stratumvarianties voor de nieuwe indeling worden geschat op basis van steekproefelementen afkomstig uit verschillende strata onder de oude indeling met mogelijk ongelijke insluitkansen. In deze situatie kan gebruik worden gemaakt van de volgende variantieschatter (Van den Brakel, 2009),

$$\hat{S}_{g,l}^2 = \frac{1}{2N_{g,l}(N_{g,l} - 1)} \sum_{i=1}^{n_{g,l}} \sum_{j \neq i}^{n_{g,l}} \frac{(z_{i,g,l}^t - z_{j,g,l}^t)^2}{\pi_{ij}}.$$

De tweede-orde insluitkansen  $\pi_{ij}$  worden gegeven door vergelijking (2.4.2).

### 2.4.4 Toepassing van de variantieschatters bij het steekproefontwerp

De in de paragrafen 2.4.2 en 2.4.3 gegeven uitdrukkingen voor de varianties van de verschillende schatters van totalen en indices kunnen worden gebruikt om steekproeven te ontwerpen. Op basis van eisen aan de maximaal toelaatbare variantie voor de uitkomsten kan de minimale steekproefomvang en allocatie over de strata bepaald worden. Dit is een onderwerp dat in de Methodenreeks over steekproeftheorie behandeld wordt. Voor een toepassing in de context van de verandering van de SBI indeling wordt verwezen naar Van den Brakel (2009).

#### 2.4.5 Backcasten

In deze paragraaf wordt een voorbeeld gegeven van de SBI-overgang bij de KS. Voor de meeste kerncellen is verondersteld dat deze onder de SBI'93 één op één overgaan op de indeling onder de SBI2009. Voor de Zakelijke dienstverlening bleek deze veronderstelling niet houdbaar. Onder de SBI'93 werd de omzetontwikkeling voor Accountants, Adviesbureaus, Markt en opinieonderzoeksbureaus als één kerncel gepubliceerd. Onder de SBI2008 is deze kerncel gesplitst in drie afzonderlijke kerncellen. Bij het backcasten is in eerste instantie uitgegaan van een macromethode, die er op neer kwam dat de gezamenlijke omzetontwikkeling voor Accountants, Adviesbureaus, Markt en opinieonderzoeksbureaus ook van toepassing is op de drie afzonderlijke kerncellen. Nadere bestudering door een inhoudelijk deskundige leidde tot de conclusie dat deze methode leidt tot een onrealistisch beeld van de omzetontwikkeling voor deze drie afzonderlijke kerncellen. Omdat bij deze kerncellen de steekproefbestanden betrekkelijk eenvoudig kunnen worden dubbelgecodeerd, zijn uiteindelijk via de micromethode de reeksen voor deze drie kerncellen gebackcast.

Een tweede voorbeeld van de zakelijke dienstverlening zijn de Architecten en Ingenieursbureaus. Onder de SBI'93 werd de omzet-ontwikkeling van deze bedrijfstak als één kerncel gepubliceerd. Onder de SBI2008 is deze in twee afzonderlijke kerncellen gesplitst. De macromethode, die veronderstelt dat de gezamenlijke omzetontwikkeling voor Architecten en Ingenieursbureaus van toepassing is voor de twee afzonderlijke kerncellen, bleek ook voor deze bedrijfstak niet houdbaar. Daarom zijn de reeksen voor de afzonderlijke kerncellen herberekend via de micromethode. Een complicerend probleem bij deze kerncellen is dat een deel van de Architecten is overgegaan naar de Ingenieursbureaus. Omdat deze bedrijven niet op een eenvoudige wijze konden worden gecodeerd volgens de nieuwe classificatie, is in een vervolgstap nog een macromethode toegepast. Daarbij is op basis van het dubbelgecodeerde bestand voor 2008 de fractie van de omzet van de Architecten bepaald die overgaat naar de Ingenieursbureaus. Deze fractie is gebruikt om op macroniveau de totale omzet voor beide kerncellen te backcasten. Hieruit is vervolgens de omzetontwikkeling berekend.

## 2.5 Kwaliteitsindicatoren

Kwaliteitsmaten voor de in de vorige paragrafen beschreven schatters zijn de vertekening en de steekproefvariantie. Voor de bepaling van de steekproefvariantie van de schatters en indices konden we standaarduitdrukkingen uit de steekproeftheorie gebruiken.

Voor de kwaliteit van het backcasten op macroniveau is tevens relevant hoe goed de modelveronderstelling achter het gebruik van de conversiematrix is. De keuze van de hulpvariabele  $X$  is mede bepalend. De kwaliteit kan alleen in jaren worden getoetst waarin er een dubbelcodering is. Het gebruik van de conversiematrix voor het backcasten in tussenliggende jaren berust dan op de aanname van tijdinvariantie van de conversiematrix of de geldigheid van de interpolatie.

Het is doorgaans erg lastig om iets te zeggen over de omvang van de vertekening die optreedt bij backcasten op basis van macromethoden. Vooral de kwalitatieve kennis van inhoudelijk deskundigen is van belang om een inschatting te maken van de plausibiliteit van gebackcaste reeksen. Op basis van deze informatie kunnen realistischere tijdsafhankelijke conversiefactoren worden samengesteld.

### **3. Experimenten**

#### **3.1 Korte beschrijving**

Op het CBS worden regelmatig experimenten uitgevoerd om het effect van één of meerdere veranderingen in het surveyonderzoek op de belangrijkste uitkomsten hiervan te vergelijken. In een experiment worden de steekproefeenheden aselekt verdeeld over twee of meer deelsteekproeven. Bij iedere deelsteekproef worden data verzameld via één van de surveymethodieken. Vervolgens kunnen schattingen worden gemaakt van de belangrijkste doelvariabelen op basis van de verschillende surveymethodieken. De schattingen voor deze doelvariabelen kunnen worden vergeleken en er kan worden getoetst of de verschillende surveymethodieken leiden tot significante verschillen tussen de schattingen voor de doelvariabelen. Omdat in een experiment de steekproefeenheden aselekt verdeeld zijn over de verschillende behandelingen is het mogelijk te concluderen of waargenomen verschillen zijn veroorzaakt door de verschillen tussen de surveymethodieken.

#### **3.2 Toepasbaarheid**

Experimenten kunnen worden toegepast om de effecten van wijzigingen in het surveyproces te vergelijken. Omdat de steekproefeenheden worden gerandomiseerd over de verschillende behandelingen van het experiment, wordt ervoor gezorgd dat de verschillen tussen deze behandelingen de enige systematische verschillen tussen de deelsteekproeven zijn. Hierdoor wordt bereikt dat er een causaal verband tussen de behandelingen en de waargenomen verschillen in de doelvariabelen ontstaat.

Experimenten zijn vooral geschikt om de effecten van veranderingen in de waarneming en de dataverzameling van het surveyonderzoek te kwantificeren. Veranderingen in de waarneming en de dataverzameling leiden ertoe dat de microdata onder de verschillende methoden niet meer consistent zijn. In een dergelijke situatie is het noodzakelijk om via aparte deelsteekproeven data te verzamelen onder de verschillende behandelingen of surveymethodieken. Het uitvoeren van een experiment is tamelijk kostbaar. Een alternatief is de opgebouwde reeks te modelleren met een tijdreeksmodel en de overgang van het oude naar het nieuwe ontwerp expliciet te modelleren met behulp van een interventievariabele. Deze aanpak wordt nader uitgewerkt in hoofdstuk 4.

Indien veranderingen in het surveyonderzoek zich beperken tot de verwerkingsfase blijven de microdata onder de oude en de nieuwe methodiek consistent. In een dergelijke situatie kunnen verschillen worden bepaald door het uitvoeren van een herberekening op basis van dezelfde steekproef. Ook effecten van nieuwe classificaties van eenheden kunnen op deze wijze worden vastgesteld na toevoeging van de nieuwe classificatievariabele, zie hoofdstuk 2.

Indien een experiment naar het effect van een herontwerp van het survey-proces grootschalig opgezet wordt, is een bijkomend voordeel dat de overgang van het oude naar het nieuwe ontwerp minder riskant is. Gedurende de periode dat het experiment wordt uitgevoerd worden immers data verzameld onder de oude en de nieuwe methodiek. Als achteraf blijkt dat de nieuwe aanpak tot grote problemen leidt, wordt het risico weggenomen dat er een periode ontstaat waarover geen cijfers gepubliceerd kunnen worden. Een mooi voorbeeld van deze manier van werken is het herontwerp van het Onderzoek Verplaatsingsgedrag (OVG). In 1998 is het OVG een jaar lang via de oude en de nieuwe methode parallel aan elkaar uitgevoerd. Gedurende dit jaar is voldoende ervaring opgedaan met het nieuwe ontwerp om in 1999 zonder grote risico's over te gaan op het nieuwe ontwerp. Daarnaast zijn op basis van dit experiment de effecten op de belangrijkste doelvariabelen van het OVG gekwantificeerd.

### **3.3 Uitgebreide beschrijving**

#### *3.3.1 Inleiding*

Onder een experiment wordt een test verstaan waarbij doelbewust wordt gevarieerd met de niveau's van één of meerdere variabelen of factoren om het effect daarvan op een doelvariabele te kunnen meten. Uit deze beschrijving volgt dat bij een experiment ten minste twee verschillende behandelingen betrokken zijn waarvan het effect op een doelvariabele moet worden onderzocht. Bij veel toepassingen op het CBS worden experimenten ingebouwd in het reguliere steekproefonderzoek. Dit betekent dat de steekproef die ten behoeve van het onderzoek wordt getrokken, via een proefopzet in twee of meer deelsteekproeven wordt verdeeld. Iedere deelsteekproef wordt aan één van de behandelingen van het experiment toegewezen. De deelsteekproef die aan het reguliere onderzoek wordt toegewezen dient als controlegroep in het experiment en wordt tegelijkertijd gebruikt voor de officiële publicatiedoelstellingen van het steekproefonderzoek.

Ontwerp en analyse van experimenten behoort tot het vakgebied van de toegepaste statistiek omdat de waarnemingen die in een experiment worden verzameld meestal worden vertroebeld door verstoringen. Deze verstoringen worden veroorzaakt door bijvoorbeeld meetfouten die worden gemaakt bij de dataverzameling, variatie tussen de experimentele eenheden, en het niet onder exact gelijke omstandigheden uitvoeren van de behandelingen in het experiment. Het gevolg van deze verstoringen is dat er variatie optreedt in de waargenomen effecten indien een experiment wordt herhaald. Naarmate de bovengenoemde verstoringen groter zijn, zal het steeds moeilijker worden om eventuele effecten van de behandelingen op de doelvariabele aan te tonen. Ten gevolge van deze verstoringen wordt een zekere mate van onzekerheid geïntroduceerd bij het trekken van conclusies over de waargenomen behandelingseffecten in een experiment. Daarom zijn statistische methoden noodzakelijk om experimenten te ontwerpen en te analyseren zodanig dat op een

objectieve manier conclusies kunnen worden getrokken uit de waargenomen resultaten.

### 3.3.2 *Principes van proefopzetten*

De theorie van ontwerp en analyse van experimenten is gebaseerd op een aantal grondbeginselen. De vier belangrijkste worden hieronder kort samengevat en vormen de basis voor de proefopzetten die in deze paragraaf worden beschreven.

1) *Herhalen of repliceren*. Om aan te kunnen tonen dat een behandelingseffect statistisch significant is, is een minimum aantal herhalingen van een behandeling noodzakelijk. Naarmate het aantal herhalingen toeneemt, neemt ook de nauwkeurigheid toe waarmee een behandelingseffect wordt waargenomen. Een belangrijk onderdeel van ontwerpen van experimenten is het bepalen van de minimale proefomvang om een vooraf gespecificeerd effect met een bepaalde betrouwbaarheid te kunnen waarnemen.

2) *Randomiseren*. Dit is het aselekt toewijzen van de experimentele eenheden aan de behandelingen. Door randomiseren wordt bereikt dat verstoringen willekeurig over de behandelingen worden verdeeld. Hierdoor wordt voorkomen dat een behandeling systematisch wordt bevoordeeld of benadeeld waardoor de resultaten van het experiment worden vertekend. Door randomiseren wordt bereikt dat de waargenomen effecten kunnen worden toegeschreven aan de verschillen tussen de behandelingen in het experiment. Anders gezegd: door randomiseren wordt bereikt dat de behandelingseffecten zuiver worden geschat.

3) Lokale controle door middel van *blokken*. Dit impliceert dat experimentele eenheden worden gegroepeerd in min of meer homogene groepen, ook wel blokken genoemd. Door de behandelingen te vergelijken binnen deze blokken, kan in de analyse voor de variatie tussen de blokken worden gecontroleerd, waardoor de behandelingseffecten nauwkeuriger geschat kunnen worden.

4) Het simultaan testen van meerdere factoren in één *factoriële proef*. Het is efficiënt om verschillende factoren niet in afzonderlijke experimenten te onderzoeken, maar zoveel mogelijk simultaan binnen één experiment. Dit is efficiënt omdat minder experimentele eenheden nodig zijn en omdat de wisselwerking (interactie) tussen de verschillende factoren kan worden onderzocht.

In de volgende paragraaf wordt dieper ingegaan op drie elementaire proefopzetten, namelijk volledig gerandomiseerde proeven, gerandomiseerde blokken proeven en factoriële proeven.

### 3.3.3 *Proefopzetten*

Een volledig gerandomiseerde proef wordt in de Engelstalige literatuur aangeduid als een completely randomized design (CRD). Het doel van een CRD is om twee of meer gelijksoortige behandelingen met elkaar te vergelijken. Meer formeel gezegd, in een CRD wordt het effect van één factor op  $I \geq 2$  behandelingsniveaus

onderzocht. Bijvoorbeeld het effect van  $I$  verschillende vragenlijsten op de uitkomsten van een enquête. In een CRD worden de experimentele eenheden aselekt toegewezen aan de behandelingen in het experiment. Aan de hand van deze proefopzet wordt bereikt dat alle verstoringen aselekt verdeeld worden over de behandelingen zodat geen enkele behandeling systematisch bevoordeeld of benadeeld wordt.

De nauwkeurigheid van een CRD kan worden verbeterd door de experimentele eenheden in te delen in groepen zodanig dat de variatie tussen de experimentele eenheden binnen een groep kleiner is dan de variatie tussen de experimentele eenheden van verschillende groepen. Groepen van min of meer homogene experimentele eenheden worden *blokken* genoemd. Nadat de experimentele eenheden zijn ingedeeld in  $J$  blokken, worden binnen ieder blok de experimentele eenheden aselekt verdeeld over de  $I$  behandelingen van het experiment. Een dergelijke proefopzet wordt een gerandomiseerde blokkenproef genoemd en wordt in de Engelstalige literatuur aangeduid als een Randomized Block Design (RBD). Bij deze proefopzet vindt de vergelijking tussen de behandelingen alleen binnen de blokken plaats waardoor de variatie tussen de blokken buiten de analyse van de behandelings effecten kan worden gehouden. Hierdoor neemt de nauwkeurigheid toe waarmee de behandelingen in het experiment worden vergeleken.

Bij veel toepassingen op het CBS wordt het experiment ingebouwd in een lopend steekproefonderzoek. Hierbij bepaalt het steekproefontwerp welke potentiële variabelen gebruikt kunnen worden als een blokvariabele in het experiment. Steekprofeenheden die afkomstig zijn uit hetzelfde stratum, primaire steekprofeenheid of cluster, zijn doorgaans homogener dan steekprofeenheden die afkomstig zijn uit verschillende strata, primaire steekprofeenheden of clusters. Ook is uit de literatuur bekend dat ten gevolge van interviewereffecten, de respons van steekproefpersonen die zijn toegewezen aan de zelfde interviewer homogener zijn dan de respons van steekproefpersonen die zijn toegewezen aan verschillende interviewers. Hieruit volgt dat de nauwkeurigheid van een experiment kan worden vergroot door het experiment te ontwerpen als een RBD waarbij strata, primaire steekprofeenheden, clusters of interviewers als blokvariabelen worden gebruikt. Indien de fractie van steekprofeenheden die aan de alternatieve behandelingen wordt toegewezen relatief klein is t.o.v. de fractie die aan de reguliere steekproef wordt toegekend, kan via een RBD worden gegarandeerd dat alle strata, primaire steekprofeenheden of clusters in de kleine deelsteekproeven vertegenwoordigd zijn.

Indien interviewers als blokvariabele worden gebruikt, heeft dit tot gevolg dat iedere interviewer aan alle behandelingen van het experiment wordt toegewezen. Dit kan om praktische redenen bezwaarlijk zijn. Bijvoorbeeld, indien in een experiment het effect van een aantal verschillende vragenlijsten wordt onderzocht, bestaat de kans dat de interviewers de verschillende behandelingen door elkaar halen waardoor het experiment verstoord wordt. Indien het noodzakelijk is dat een interviewer aan één van de behandelingen wordt toegewezen, zijn doorgaans complexere proefopzetten vereist.



Bij het herontwerp van een statistisch proces wordt doorgaans meer dan één factoren veranderd. Bijvoorbeeld het effect van een nieuwe vragenlijst en het effect van een andere dataverzamelingsmethode. In dit voorbeeld zijn er twee verschillende factoren. Voor beide factoren onderscheiden we twee niveaus; een oude en een nieuwe vragenlijst en een oude en een nieuwe benaderingsstrategie. Indien het gewenst is om het effect van de afzonderlijke factoren te kwantificeren is het mogelijk om voor iedere factor een afzonderlijk experiment uit te voeren. Vanuit een statistisch oogpunt is het echter efficiënter om de verschillende factoren simultaan te testen in één experiment. Hierbij worden de niveaus van de verschillende factoren met elkaar gekruist. In het bovengenoemde voorbeeld ontstaan vier verschillende behandelingscombinaties; 1) oude vragenlijst met de oude benaderingsstrategie, 2) oude vragenlijst met de nieuwe benaderingsstrategie, 3) nieuwe vragenlijst met de oude benaderingsstrategie, 4) nieuwe vragenlijst met de nieuwe benaderingsstrategie.

Het voordeel van een factoriële proefopzet is dat minder experimentele eenheden nodig zijn om de effecten van de afzonderlijke factoren te toetsen en dat de wisselwerking tussen de factoren kan worden onderzocht. Het nadeel van een factoriële proefopzet is dat het aantal behandelingscombinaties snel erg groot wordt. Een experiment met een groot aantal behandelingscombinaties leidt doorgaans al snel tot praktische problemen bij de planning en uitvoering van het veldwerk. Om deze reden wordt in de meeste toepassingen het effect van het nieuwe steekproefontwerp vergeleken met het oude ontwerp. Dat betekent dat er slechts twee behandelingscombinaties zijn en dat alleen het netto effect van alle factoren die in het herontwerp veranderen kan worden gekwantificeerd. Dit volstaat voor het maken van consistente reeksen en is daarom voor de meeste praktische toepassingen voldoende, zie bijvoorbeeld Van den Brakel e.a. (2008) en Van den Brakel (2005).

Voor meer gedetailleerde informatie over het ontwerpen van experimenten in steekproefonderzoeken wordt verder verwezen naar Fienberg and Tanur (1987, 1988, 1989), Van den Brakel and Renssen (1998), en Van den Brakel (2008), Van den Brakel e.a. (2008).

### *3.3.4 Steekproefomvang*

Een ander belangrijk onderdeel van het ontwerpen van een experiment is het bepalen van de steekproefgrootte. Om een uitspraak te doen over de minimale steekproefomvang moet worden aangegeven welke behandelingseffecten minimaal moeten leiden tot het verwerpen van de nulhypothese. Daartoe moet vooraf worden aangegeven wat de omvang van de behandelingseffecten zijn waarin de onderzoeker geïnteresseerd is, hoe groot het betrouwbaarheidsniveau en het onderscheidingsvermogen van de toets moet zijn en een indicatie van de omvang van de variantie. Op basis van deze informatie kan de steekproefomvang worden berekend of worden bepaald aan de hand van tabellen (zogenaamde Operating Characteristic Curves). Dit zijn aparte onderwerpen binnen de theorie van ontwerp en analyse van experimenten waar hier niet nader op in wordt gegaan. Voor meer

informatie zie Montgomery (1997), hoofdstuk 3 voor een CRD en hoofdstuk 4 voor een RBD.

In veel praktische toepassingen wordt de steekproefgrootte en de verdeling over de verschillende behandelingen bepaald door het beschikbare budget en andere restricties die aan de uitvoering van het steekproefonderzoek worden gesteld. Bij veel toepassingen is het echter niet verantwoord om de data verzameld onder de alternatieve behandelingen ook te gebruiken voor de reguliere publicatiedoeleinden. Vaak worden betrouwbaarheidseisen gesteld aan de te publiceren cijfers. Dit heeft tot gevolg dat de minimale steekproefomvang van de deelsteekproef die aan het reguliere onderzoek wordt toegewezen (de controlegroep) vastligt en daardoor veel groter is dan de steekproefomvang van de overige experimentele groepen. Bij sommige toepassingen, bijvoorbeeld bij het testen van een nieuwe aanschrijfbrief of kleine beloningen op de respons, is het wel mogelijk om de data verzameld onder beide behandelingen te gebruiken in de reguliere publicaties. In dat geval kan voor een meer gebalanceerde verdeling van de experimentele eenheden over de behandelingen worden gekozen omdat dit het onderscheidingsvermogen van het experiment verhoogt.

### *3.3.5 Traditionele analysemethoden voor experimenten*

De analyse van een experiment start met het toetsen van de hypothese op behandelingseffecten. Bij deze hypothese wordt getoetst of ten minste één paar van behandelingen significant verschillend is.

Deze hypothese wordt getoetst aan de hand van een zogenaamde variantieanalyse. De manier waarop een proef is ontworpen bepaalt in sterke mate de wijze waarop een variantieanalyse moet worden uitgevoerd. Verondersteld wordt dat de waarnemingen in het experiment kunnen worden gemodelleerd via een lineair model waar de storingstermen onderling onafhankelijk en normaal verdeeld zijn. Op basis hiervan wordt een F-toets afgeleid voor het toetsen van de hypothese dat ten minste één paar van behandelingen significant verschillend is.

Een speciaal geval is het vergelijken van twee behandelingen. Doorgaans wordt een experiment met twee behandelingen geanalyseerd met behulp van de t-toets. Voor meer informatie wordt verwezen naar Montgomery (1997, hoofdstuk 2).

De analyse van een experiment is niet afgerond na het toetsen van hypothese op behandelingseffecten. Indien de nulhypothese wordt verworpen, wordt het interessant om te onderzoeken welke behandelingen specifiek van elkaar verschillen. Een voor de hand liggende manier is om voor alle behandelingscombinaties, onafhankelijk van elkaar, het verschil tussen twee behandelingen te toetsen met een t-toets. Het gevaar van een dergelijke procedure is dat de kans op het maken van een fout van de eerste soort onacceptabel groot wordt. Dat betekent dat naarmate meer behandelingen paarsgewijs met elkaar worden vergeleken, de kans steeds groter wordt dat voor een behandelingscombinatie de nulhypothese ten onrechte verworpen wordt. Dit kan worden voorkomen door

gebruik te maken van de zogenaamde simultane betrouwbaarheidsintervallen of multiële vergelijkingsmethoden. Deze procedures garanderen een minimaal betrouwbaarheidsniveau indien voor een vooraf vastgesteld aantal variabelen onafhankelijk van elkaar hypothesen worden getoetst. Dit gebeurt door het betrouwbaarheidsinterval voor iedere afzonderlijke variabele groter te kiezen naarmate het aantal variabelen dat onafhankelijk van elkaar getoetst wordt toeneemt. Een bekende procedure is het multiële betrouwbaarheidsinterval van Bonferroni. Zie Montgomery (1997, hoofdstuk 3).

Bij de analyse van een experiment is het ook van belang dat zorgvuldig wordt nagegaan of de data voldoen aan de gemaakte modelveronderstellingen, dat wil zeggen dat de residuen onderling onafhankelijk en normaal verdeeld zijn met gelijke variantie. In veel toepassingen zijn de data zeer scheef verdeeld en kunnen er uitbijters aanwezig zijn waardoor niet aan deze modelveronderstellingen voldaan wordt. Er bestaat een groot aantal technieken om deze modelveronderstellingen te evalueren. Indien op grond van een modevaluatie blijkt dat deze veronderstellingen niet gerechtvaardigd zijn, kan worden gezocht naar een geschikte transformatie van de data waardoor wel of beter aan de modelveronderstellingen wordt voldaan. Als dit niet mogelijk is kan er voor worden gekozen om de analyse uit te voeren met een parameter vrije toets (Lehmann, 1975). Een geschikte parameter vrije toets voor een CRD is de toets van Kruskal-Wallis. In geval van een RBD is de verdelingsvrijetoets van Friedman voor blokkenproef een geschikt alternatief. Voor een beschrijving wordt verwezen naar Lehmann (1975).

In veel experimenten wordt onderzocht wat het effect is van alternatieve behandelingen op de responscijfers van een onderzoek. In dat geval is de doelvariabele een dichotome of een categoriale variabele die het responsresultaat over twee of meer categorieën specificiert. In een dergelijke situatie ligt het voor de hand om de analyse uit te voeren via logistische regressie of multinomiale logistische regressie (Agrestie, 1990).

### *3.3.6 Design-based analysemethoden voor experimenten in kanssteekproeven*

De standaard analysemethoden voor experimenten, beschreven in paragraaf 3.3.5, zijn niet altijd geschikt voor het analyseren van grootschalige veldexperimenten die worden uitgevoerd om het effect van een herontwerp te kwantificeren, omdat deze methoden onvoldoende rekening houden met het steekproefontwerp. Deze experimenten worden uitgevoerd om het effect van een verandering op de doelvariabelen van een surveyonderzoek te kwantificeren.

Het uitgangspunt is een kanssteekproef die is getrokken uit de doelpopulatie conform het steekproefontwerp van het reguliere surveyonderzoek. Deze steekproef is aan de hand van een van de hiervoor beschreven proefopzetten aselect verdeeld in twee of meer deelsteekproeven.

De data verzameld in de verschillende deelsteekproeven kunnen worden gebruikt om schattingen te maken voor de doelvariabelen van het reguliere surveyonderzoek

onder de verschillende behandelingen van het experiment conform de schattingsmethodiek zoals die in het reguliere onderzoek wordt gehanteerd. Vervolgens wordt getoetst of deze schattingen significant van elkaar verschillen. Uit de doelstelling van het experiment volgt daarom dat hypothesen worden getoetst over de populatieparameters zoals die zijn gedefinieerd in het reguliere surveyproces. Dit zijn doorgaans gemiddelden, totalen of ratio's van totalen. Verder volgt uit de doelstelling dat bij de analyse rekening moet worden gehouden met de wijze waarop een steekproef uit de eindige doelpopulatie wordt getrokken en de schattingsmethodiek die in het reguliere surveyproces wordt gebruikt. In een reeks van publicaties (Van den Brakel, 2001, 2008, Van den Brakel and Renssen, 1998, 2005, en Van den Brakel and Van Berkel, 2002) is een design-based analysemethode uitgewerkt voor experimenten ingebouwd in kanssteekproeven die rekening houden met de bovengenoemde aspecten. In deze paragraaf wordt de aanpak beschreven in de context van herontwerpen. Voor meer achtergrondinformatie zie de genoemde literatuur.

Ter illustratie gaan we uit van een experiment waarbij het netto-effect van een herontwerp moet worden gekwantificeerd. Dat betekent dat onderzocht wordt wat het effect is op de doelvariabelen indien van het reguliere surveyproces wordt overgegaan op het nieuwe ontwerp. Daartoe wordt een kanssteekproef uit de doelpopulatie getrokken. Vervolgens wordt deze steekproef aselekt in twee deelsteekproeven verdeeld volgens een van de hiervoor besproken proefopzetten (CRD of een RBD). In één deelsteekproef worden data verzameld volgens het reguliere ontwerp, in de andere deelsteekproef volgens het nieuwe ontwerp. Op basis van deze twee deelsteekproeven wordt onderzocht of de populatieparameter waargenomen onder het reguliere en het nieuwe surveyproces aan elkaar gelijk zijn. Uit deze doelstelling volgt de volgende hypothese

$$\begin{aligned} H_0 : Y_R &= Y_N \\ H_1 : Y_R &\neq Y_N \end{aligned} \tag{3.3.1}$$

Hierbij zijn  $Y_R$  en  $Y_N$  de populatieparameter waargenomen onder respectievelijk het reguliere en het nieuwe surveyproces. Deze parameters zijn doorgaans gedefinieerd als gemiddelden, totalen of ratio's van twee populatietotalen. Het verschil  $Y_R - Y_N$  is de methodebreuk die ontstaat ten gevolge van de overgang van het reguliere naar het nieuwe ontwerp. Merk op dat in de traditionele design-based steekproeftheorie de kenmerken die aan de elementen in de steekproef worden waargenomen vaste waarnemingen zijn die foutloos worden waargenomen, zie bijvoorbeeld Cochran (1977). Onder dit paradigma bestaat er dus slechts één waarde voor de eindige populatieparameter en zijn  $Y_R$  en  $Y_N$  per definitie aan elkaar gelijk. Om eventuele verschillen tussen een eindige populatieparameter toe te laten ten gevolge van verschillende waarnemingsmethoden is het noodzakelijk om een meetfoutmodel te veronderstellen. Zie Van den Brakel (2008) of Van den Brakel en Renssen (2005) voor details.

De populatiegemiddelden  $Y_R$  en  $Y_N$  worden geschat aan de hand van de data waargenomen in de twee deelsteekproeven. Laat  $\hat{Y}_R$  en  $\hat{Y}_N$  de gegeneraliseerde regressieschatters zijn voor respectievelijk  $Y_R$  en  $Y_N$ . Deze schatter houdt rekening met de manier waarop een kanssteekproef uit de doelpopulatie is getrokken, de proefopzet die gebruikt is om deze steekproef in twee (of meer) deelsteekproeven te verdelen en de schattingsmethodiek zoals die in het surveyonderzoek is gedefinieerd. Vervolgens kan hypothese (3.3.1) worden getoetst met een toetsgrootheid van de vorm:

$$z = \frac{\hat{Y}_R - \hat{Y}_N}{\sqrt{\hat{Var}(\hat{Y}_R - \hat{Y}_N)}}. \quad (3.3.2)$$

In de noemer van (3.3.2) staat een schatter voor de variantie van het verschil van twee gegeneraliseerde regressieschatters waarbij wederom rekening wordt gehouden met de manier waarop een kanssteekproef uit de doelpopulatie is getrokken, de proefopzet die gebruikt is om deze steekproef in twee (of meer) deelsteekproeven te verdelen en de schattingsmethodiek zoals die in het surveyonderzoek zijn gedefinieerd. Voor grote steekproefaantallen is toetsgrootheid (3.3.2) bij benadering standaardnormaal verdeeld.

Een schatting voor de omvang van de methodebreuk wordt gegeven door  $\hat{Y}_R - \hat{Y}_N$ . In de volgende paragraaf wordt beschreven hoe dit verschil kan worden gebruikt om reeksen te corrigeren voor het waargenomen verschil.

De exacte vorm van de punt- en variantieschatters hangt af van de gekozen proefopzet en het type doelvariabelen. Voor expliciete uitdrukkingen wordt verwezen naar Van den Brakel and Renssen (2005) en Van den Brakel (2008).

De analysemethoden zijn geïmplementeerd in het softwarepakket X-tool. Dit pakket is beschikbaar als een component van Blaise. Voor een korte beschrijving van de functionaliteit van dit pakket wordt verwezen naar Van den Brakel (2008).

### 3.4 Corrigeren van reeksen

Er kan op verschillende manieren worden omgegaan met methodebreuken. Een conservatieve aanpak is om de breuk uitsluitend te kwantificeren met als doel het effect van het herontwerp te scheiden van de autonome ontwikkeling van de doelvariabelen op het moment waarop het herontwerp wordt geïmplementeerd. Met een dergelijke aanpak loopt men weinig risico omdat uitsluitend uitspraken worden gedaan over de omvang van de methodebreuken over de periode waar de reguliere en de nieuwe methodiek parallel aan elkaar zijn uitgevoerd. Aan de andere kant komt deze aanpak meestal niet tegemoet aan de wensen van veel gebruikers omdat deze geïnteresseerd zijn in ononderbroken reeksen die de ontwikkeling van doelvariabelen zo goed mogelijk in beeld brengen. In deze paragraaf worden daarom methoden voorgesteld om de uitkomsten op basis van een survey onderzoek voor en na het doorvoeren van een herontwerp vergelijkbaar te maken. Dit gebeurt door de schatting voor de methodebreuk, verkregen via een experiment, te extrapoleren in de

tijd. Dit impliceert dat verondersteld wordt dat de methodebreuk tijdonafhankelijk is.

De periode waarin de reguliere en de nieuwe aanpak via een experiment parallel aan elkaar zijn uitgevoerd, wordt aangegeven met  $T$ . Verder zijn  $\hat{Y}_{R,T}$  en  $\hat{Y}_{N,T}$  de directe schatters voor de populatieparameters  $Y_{R,T}$  en  $Y_{N,T}$  waargenomen via de reguliere en de nieuwe aanpak gedurende het paralleltraject. De meest voor de hand liggende methode om een reeks te corrigeren voor een waargenomen methodebreuk is de volgende additieve correctie:

$$\tilde{y}_{N,t} = \hat{y}_{R,t} + (\hat{y}_{N,T} - \hat{y}_{R,T}) \equiv \hat{y}_{R,t} + \hat{\Delta}_T, \text{ voor } t = 1, \dots, T-1, \quad (3.4.1)$$

met  $\hat{\Delta}_T = \hat{y}_{N,T} - \hat{y}_{R,T}$ . Model (3.4.1) impliceert dat de correctie voor de methodebreuk onafhankelijk is van de waarde van  $\hat{Y}_{R,t}$ . Dit kan resulteren in gecorrigeerde reeksen die waarden aannemen buiten de toegestane parameterruimte.

Om te voorkomen dat gecorrigeerde reeksen negatieve waarden aannemen, kan de volgende multiplicatieve correctie gebruikt worden:

$$\tilde{y}_{N,t} = \hat{y}_{R,t} \frac{\hat{y}_{N,T}}{\hat{y}_{R,T}}, \text{ voor } t = 1, \dots, T-1. \quad (3.4.2)$$

Model (3.4.2) veronderstelt dat de omvang van de correctie proportioneel is met de waarde van  $\hat{Y}_{R,t}$ . Voor continue doelvariabelen die uitsluitend waarden groter of gelijk aan nul kunnen aannemen is deze aanname doorgaans plausibeler dan een correctie die volledig onafhankelijk is van de waarde van de te corrigeren reeks.

Voor variabelen die gedefinieerd zijn als fracties of percentages zijn modellen (3.4.1) en (3.4.2) ongeschikt om reeksen te corrigeren. Fracties kunnen uitsluitend waarden aannemen in de range  $[0,1]$ . Model (3.4.2) kan nog steeds resulteren in aangepaste waarden die groter zijn dan 1 of 100%. Voor dergelijke parameters is door Van den Brakel e. a. (2008) de volgende correctie ontwikkeld:

$$\tilde{y}_{N,t} = \hat{y}_{R,t} + \gamma \hat{\Delta}_T \mathcal{D}(\hat{y}_{R,t}), \text{ voor } t = 1, \dots, T-1. \quad (3.4.3)$$

Hierbij is  $\mathcal{D}(\hat{y}_{R,t})$  een soort dempende factor die waarden aanneemt in de range  $[0,1]$  en is gedefinieerd als een functie van  $\hat{Y}_{R,t}$ , zodanig dat  $\mathcal{D}(\hat{y}_{R,t}) = 1$  indien  $\hat{y}_{R,t} = 1/2$  en  $\mathcal{D}(\hat{y}_{R,t}) = 0$  indien  $\hat{y}_{R,t} = 1$  of  $0$ . Een voor de hand liggende functie die aan deze condities voldoet is:

$$\mathcal{D}(\hat{y}_{R,t}) = 4\hat{y}_{R,t}(1 - \hat{y}_{R,t}). \quad (3.4.4)$$

Omdat  $\hat{y}_{R,t}(1 - \hat{y}_{R,t})$  de populatievariantie van een fractie is, heeft (3.4.4) de aantrekkelijke statistische interpretatie dat de factor  $\mathcal{D}(\hat{y}_{R,t})$  die de omvang van de correctie regelt, evenredig is aan de variantie van de te corrigeren doelvariabele

$\hat{Y}_{R,t}$ . De omvang van de correctie van  $\hat{Y}_{R,t}$  is daardoor afhankelijk van de precisie van  $\hat{Y}_{R,t}$ . Tot slot wordt  $\gamma$  in (3.4.3) zodanig gekozen dat de gelijkheid in (3.4.3) exact geldt in de periode  $T$ , waar het parallelle traject is uitgevoerd, dat wil zeggen  $\hat{y}_{N,T} = \hat{y}_{R,T} + \gamma \hat{\Delta}_T \delta(\hat{y}_{R,T})$ . Indien  $\gamma = 1/\delta(\hat{y}_{R,T})$  wordt ingevuld in (3.4.3) volgt dat:

$$\tilde{y}_{N,t} = \hat{y}_{R,t} + \hat{\Delta}_T \frac{\hat{y}_{R,t}(1 - \hat{y}_{R,t})}{\hat{y}_{R,T}(1 - \hat{y}_{R,T})}. \quad (3.4.5)$$

Formule (3.4.5) illustreert dat de omvang van de correctie uiteindelijk evenredig is met de ratio van de populatievarianties van  $\hat{y}_{R,t}$  en  $\hat{y}_{R,T}$ .

Benaderingen voor de varianties voor reeksen die zijn gecorrigeerd via (3.4.1), (3.4.2) en (3.4.5) zijn uitgewerkt in Van den Brakel e.a. (2008).

Tot nu toe is er van uitgegaan dat de reeks in het verleden wordt gecorrigeerd voor de waargenomen methodebreuk. In de meeste toepassingen heeft een herontwerp als doel om de kwaliteit en de efficiency van een statistiek te verbeteren. Daarom ligt het in de meeste toepassingen voor de hand om de reeks in het verleden te corrigeren en daarmee vergelijkbaar te maken met de uitkomsten zoals die worden verkregen na invoering van het herontwerp. Indien gewenst is het echter ook mogelijk om de waarnemingen verkregen onder de nieuwe methodiek, te corrigeren voor de waargenomen breuk. Daarvoor kunnen de modellen die in deze paragraaf zijn beschreven op een vergelijkbare manier worden gebruikt.

Indien reeksen die aan elkaar zijn gerelateerd onafhankelijk van elkaar worden gecorrigeerd voor waargenomen breuken, kan dat leiden tot ongewenste inconsistenties. Dit doet zich bijvoorbeeld voor indien een reeks van geschatte populatietotalen voor de hele doelpopulatie tevens wordt uitgesplitst naar deelpopulaties. De relatie tussen deze reeksen is dat de som over de totalen voor de deelpopulaties gelijk is aan het totaal voor de hele populatie. Indien deze reeksen onafhankelijk van elkaar worden gecorrigeerd via model (3.4.2) zal deze consistentie worden verstoord. Vereiste consistenties kunnen worden hersteld via een geschikte Lagrange functie. Aan de hand van deze techniek worden de schattingen minimaal aangepast zodanig dat de vereiste consistentie tussen de schattingen hersteld wordt. De vereiste aanpassingen zijn evenredig aan de variantie van de afzonderlijke schattingen verdeeld. Dat betekent dat schattingen met een grote variantie sterker worden aangepast dan schattingen met een kleinere variantie. Dit is nader uitgewerkt in Van den Brakel e.a. (2008).

### 3.5 Voorbeeld

Er zijn diverse voorbeelden waarbij aan de hand van een experiment het effect van een herontwerp op de uitkomsten van de statistiek is gekwantificeerd. Eén daarvan is het herontwerp van het Onderzoek Verplaatsingsgedrag. In 1998 is de waarnemingsstrategie van deze statistiek drastisch herzien met als doel het verhogen

van de respons. Gedurende dit jaar zijn het oude en het nieuwe ontwerp via een grootschalig veldexperiment parallel aan elkaar uitgevoerd. Op basis hiervan zijn methodebreuken voor de meeste doelvariabelen bepaald. Voor een uitgebreide beschrijving, zie CBS (2002) en Van den Brakel e. a. (2008).

### **3.6 Kwaliteitsindicatoren**

De traditionele analysemethoden zijn gebaseerd op een lineair model waarbij verondersteld wordt dat de residuen onderling onafhankelijk en normaal verdeeld zijn met gelijke variantie. Indien de analyse op deze methoden is gebaseerd is het van belang om te onderzoeken in welke mate aan deze modelveronderstellingen wordt voldaan. Dit gebeurt door het inspecteren van de (gestandaardiseerde) residuen. In Montgomery (1997) worden diverse grafische diagnostieken beschreven die kunnen worden gebruikt voor het visueel inspecteren van de residuen. De design-based analysemethoden maken geen gebruik van modelveronderstellingen en daarom spelen dergelijke modeldiagnostieken voor deze methoden nauwelijks een rol.

De belangrijkste kwantitatieve kwaliteitsindicator voor een experiment is het verschil dat bij de geplande steekproefomvang kan worden waargenomen met een vooraf gespecificeerd onderscheidingsvermogen en significantieniveau. Het significantieniveau is de kans dat de nulhypothese verworpen wordt gegeven dat de nulhypothese waar is, d.w.z. als er geen sprake is van een methodebreuk. Het onderscheidingsvermogen is de kans dat de nulhypothese verworpen wordt als de alternatieve hypothese waar is, d.w.z. als er wel sprake is van een methodebreuk.

Bij het ontwerpen van een experiment dat er op gericht is om methodebreuken vast te stellen is het van belang dat rekening wordt gehouden met het onderscheidingsvermogen. Op deze wijze wordt de kans geminimaliseerd dat na implementatie significante methodebreuken optreden terwijl deze tijdens het experiment niet tot verwerping van de nulhypothese leiden.

Een aanvullende indicator is de verhouding tussen het minimaal waarneembare verschil in het experiment en het minimaal waarneembare verschil tussen twee opeenvolgende perioden indien het onderzoek in zijn reguliere omvang wordt uitgevoerd.

Daarnaast is er een aantal kwalitatieve kwaliteitsindicatoren. Bepalend voor de interpretatie en de bruikbaarheid van de resultaten van het experiment is de mate waarin de experimentele groep is uitgevoerd conform de nieuwe surveymethodiek zoals die uiteindelijk wordt geïmplementeerd.

Voordat begonnen wordt via een grootschalig veldexperiment methodebreuken te kwantificeren, moet duidelijk zijn welke veranderingen in het statistische proces worden doorgevoerd. Het is dus van belang om onderscheid te maken tussen pilots en kleinschalige experimenten die zijn gericht op het testen van bijvoorbeeld nieuwe vragenlijsten en alternatieve benaderingsstrategieën en grootschalige veldexperimenten die zijn gericht op het kwantificeren van methodebreuken.



Tijdens de voorbereidende fase van een herontwerp moet de opzet van het nieuwe statistische proces zo goed mogelijk zijn uitgekristalliseerd. Pas als dit het geval is, kan worden onderzocht wat de effecten zijn van het herontwerp op de cijfers die met het onderzoek worden gepubliceerd. Van belang is dat wordt voorkomen dat een grootschalig veldexperiment wordt uitgevoerd als een pilot om een nieuwe veldwerkstrategie uit te proberen en tegelijkertijd gebruikt wordt om methodebreuken te kwantificeren. Het combineren van dergelijke doelstellingen is riskant omdat het risico bestaat dat uit de analyseresultaten blijkt dat de nieuwe veldwerkstrategie op onderdelen moet worden aangepast terwijl hier geen tijd en budget meer voor is omdat het grootschalige experiment ook bedoeld was om methodebreuken te kwantificeren. In deze fase zijn grote aanpassingen in het nieuwe proces nauwelijks te permitteren omdat hierdoor de methodebreuken worden beïnvloed. In een ideaal transitieproces wordt daarom expliciet onderscheid gemaakt tussen experimenten en pilots die in de voorbereidende fase plaats vinden om tot een definitieve versie van het nieuwe proces te komen en grootschalige veldexperimenten die bedoeld zijn om methodebreuken te kwantificeren.

## 4. Tijdreeksmodellen

### 4.1 Korte beschrijving

Veranderingen in het statistische proces leiden vaak tot systematische effecten in de indicatoren. Als het niet mogelijk is om een bepaalde periode de oude en de nieuwe enquête via een experiment parallel aan elkaar af te nemen via de methoden beschreven in hoofdstuk 3, kan als alternatief het effect van de overgang worden gemodelleerd via een tijdreeksmodel. Hierbij wordt het netto effect gekwantificeerd van alle veranderingen in de onderzoeksopzet op de desbetreffende doelvariabele. Dit kan via een zogeheten interventieanalyse. Een dergelijke analyse kan worden uitgevoerd door het toevoegen van een regressieterm aan een ARIMA model of met een structureel tijdreeksmodel (Harvey 1989, Durbin and Koopman 2001). In dit hoofdstuk wordt uitgegaan van het structurele tijdreeksmodel. Voor aanvullende informatie over het uitvoeren van interventieanalyses met state-space modellen voor het schatten van methodebreuken, wordt ook verwezen naar Van den Brakel e. a. (2008) en Van den Brakel en Roels (2010).

Het structurele tijdreeksmodel modelleert een tijdreeks met een *trend* component, een *seizoens*component, een *cyclische* component, een *autoregressie*component, een regressiecomponent en een storingsterm. Indien gewenst kunnen ook ARMA componenten aan het model worden toegevoegd. In een interventieanalyse wordt in de regressiecomponent een *interventiev*ariabele toegevoegd die aangeeft op welk tijdstip overgegaan wordt op het nieuwe ontwerp.

Onder de modelveronderstelling dat het tijdreeksmodel de ontwikkeling van de doelvariabele goed benadert en als op het moment dat overgegaan wordt op het nieuwe ontwerp de reële ontwikkeling van de doelvariabele niet afwijkt van de ontwikkeling zoals die beschreven wordt door het tijdreeksmodel, dan kan de regressiecoëfficiënt voor de interventievariabele worden geïnterpreteerd als de omvang van de methodebreuk ten gevolge van de overgang van het oude naar het nieuwe ontwerp.

Via een dergelijk structureel tijdreeksmodel kunnen schattingen worden gemaakt van de methodebreuken. Tevens kunnen deze schattingen gebruikt worden om de reeks te corrigeren.

### 4.2 Toepasbaarheid

De methode kan gebruikt worden indien er geen periode is geweest waarin het oude en het nieuwe ontwerp parallel zijn gedraaid, meestal vanwege budgettaire redenen. Als er voldoende observaties zijn onder het oude en het nieuwe ontwerp dan is het mogelijk om tijdreeksmodellen te gebruiken om de methodebreuken te schatten. Vervolgens kunnen de reeksen gecorrigeerd worden voor publicatie.

Als op het moment van de overgang van de oude naar de nieuwe enquête de reële ontwikkeling van de doelvariabele afwijkt van de ontwikkeling zoals die beschreven wordt door het tijdreeksmodel, dan wordt deze afwijking ontrect door het model opgevat als een onderdeel van de methodebreuk. Deze situatie kan resulteren in vertekening in reeksen die op basis van dit model zijn gecorrigeerd.

Een belangrijk aspect van het gebruik van tijdreeksmodellen voor het schatten van methodebreuken is dat bij het voortschrijden van de tijd steeds nieuwe data voor de tijdreeks beschikbaar komen. Daardoor kunnen steeds betrouwbaardere schattingen voor de methodebreuken worden gemaakt. Een consequentie is dat de schattingen voor de methodebreuken en de eventueel gecorrigeerde reeksen iedere keer gereviseerd zouden kunnen worden. Het schatten van methodebreuken via een interventieanalyse met een tijdreeksmodel werkt daarom vooral goed indien onder de oude en de nieuwe methodiek voldoende waarnemingen beschikbaar zijn.

### 4.3 Uitgebreide beschrijving

In deze paragraaf wordt eerst beschreven hoe een interventieanalyse kan worden uitgevoerd met het univariate structurele tijdreeksmodel. Vervolgens wordt de theorie uitgebreid naar multivariate modellen. Bij het univariate model wordt één tijdreeks gemodelleerd. Bij multivariate modellen worden meerdere tijdreeksen gelijktijdig gemodelleerd, waarbij (meestal) extra restricties worden opgelegd aan de regressiecoëfficiënten van de interventievariabelen om het verband tussen de variabelen in stand te houden. Voorbeelden van multivariate modellen zijn:

1. Uitsplitsingen van variabelen in subcategorieën of van populaties in deelpopulaties, waarbij de subcategorieën of deelpopulaties op ieder tijdstip moeten blijven optellen tot de hele categorie of populatie,
2. Antwoordcategorieën van meerkeuzevragen, waarbij op ieder tijdstip moet gelden dat de som van de categorieën gelijk is aan 100%.

Ten behoeve van de eenvoud wordt uitgegaan van een tijdreeksmodel dat bestaat uit een trend, een interventie en een storingsterm. Seizoenscomponenten, cyclische componenten ARMA componenten en regressiecomponenten kunnen op analoge wijze worden toegevoegd indien nodig.

#### 4.3.1 Univariate structurele tijdreeksmodellen

De tijdreeks wordt gemodelleerd aan de hand van een structureel tijdreeksmodel dat is opgebouwd uit een trend, een interventievariabele en een storingsterm:

$$y_t = L_t + \beta\delta_t + \varepsilon_t. \quad (4.3.1)$$

Hierbij is  $y_t$  de doelvariabele via het oude of het nieuwe ontwerp op tijdstip  $t$ ,  $L_t$  een stochastisch trendmodel,  $\delta_t$  een interventievariabele die de waarde nul heeft gedurende de periode dat de waarnemingen onder het oude ontwerp zijn verzameld en 1 gedurende de periode dat de waarnemingen onder het nieuwe ontwerp zijn

verzameld,  $\beta$  de regressiecoëfficiënt voor de interventievariabele en  $\varepsilon_t$  een storingsterm.

Verondersteld wordt dat de storingstermen onderling onafhankelijk en normaal verdeeld zijn, dat wil zeggen:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2). \quad (4.3.2)$$

De variantie kan eventueel omgekeerd evenredig worden gekozen met de steekproefomvang  $n_t$  van de enquête:

$$\varepsilon_t \sim N\left(0, \frac{\sigma_\varepsilon^2}{n_t}\right).$$

Dit is aan te bevelen indien de steekproefomvang over de tijd heen sterk fluctueert. Een andere mogelijkheid om rekening te houden met niet-homogene steekproefvarianties is om de geschatte steekproeffouten als priorinformatie in het tijdreeksmodel op te nemen. Deze methode is ontwikkeld door Binder and Dick (1990) voor ARMA modellen. Zie Van den Brakel en Roels (2010) voor een toepassing in de context van interventie analyse via een structureel tijdreeksmodel.

De stochastische trend kan op verschillende manieren gemodelleerd worden. Een veelvoorkomend model is het zogenaamde *smooth trend* model dat gedefinieerd is als:

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1}, \\ R_t &= R_{t-1} + \eta_t. \end{aligned} \quad (4.3.3)$$

Hierbij is  $L_t$  het niveau en  $R_t$  de stochastische hellingsparameter van het trendmodel. Verder is  $\eta_t$  een storingsterm waarvoor verondersteld wordt dat deze onderling onafhankelijk en normaal verdeeld zijn;

$$\eta_t \sim N(0, \sigma_\eta^2). \quad (4.3.4)$$

Onder de modelveronderstelling dat het trendmodel de ontwikkeling van de parameter goed benadert, kan de regressiecoëfficiënt  $\beta$  voor de interventievariabele  $\delta_t$  worden geïnterpreteerd als de omvang van de methodebreuk ten gevolge van de overgang van het oude naar het nieuwe ontwerp.

#### 4.3.2 Multivariate structurele tijdreeksmodellen

Bij veel toepassingen zijn reeksen aan elkaar gerelateerd omdat variabelen worden uitgesplitst in subcategorieën of deelpopulaties waarbij de subcategorieën of deelpopulaties op ieder tijdstip moeten blijven optellen tot de hele categorie of populatie. Reeksen zijn ook gerelateerd met elkaar indien ze betrekking hebben op antwoordcategorieën van meerkeuzevragen waarbij op ieder tijdstip moet gelden dat de som van de categorieën gelijk is aan 100%. De onderlinge relaties tussen reeksen leggen restricties op aan de modelparameters voor de methodebreuken. Hiermee kan

rekening worden gehouden door deze reeksen simultaan te modelleren in een multivariaat tijdreeksmodel. In de volgende twee deelparagrafen wordt dit behandeld. Voor meer technische details wordt verwezen naar Van den Brakel e.a. (2008) en Van den Brakel en Roels (2010).

#### 4.3.2.1 Subcategorieën of deelpopulaties

In veel toepassingen wordt een populatieparameter opgesplitst in een aantal subcategorieën. Een voorbeeld is de uitsplitsing van het totaal aantal misdrijven in de subcategorieën geweldsdelicten, vermogensdelicten, vandalisme en overige delicten. Dergelijke parameters kunnen gedefinieerd zijn als populatietotalen of als populatiegemiddelden. Er is in deze situatie sprake van een variabele die in  $K \geq 2$  subcategorieën of deelpopulaties worden uitgesplitst. Voor deze variabele geldt op ieder tijdstip dat de som over de  $K$  subcategorieën gelijk is aan het totaal. In deze situatie is het wenselijk dat de breuken in de onderliggende categorieën optellen tot de breuk voor het totaal. Een dergelijke eis moet worden vertaald naar een restrictie die opgelegd wordt aan de regressiecoëfficiënten van de interventievariabelen.

Zoals in het univariate geval, wordt de tijdreeks van iedere afzonderlijke categorie of deelpopulatie gemodelleerd met een trend, een interventievariabele die aangeeft wanneer de waarneming overgaat van het oude naar het nieuwe ontwerp, en een storingsterm:

$$y_{t,k} = L_{t,k} + \beta_k \delta_t + \varepsilon_{t,k}, \quad k = 1, \dots, K, \quad (4.3.5)$$

$$y_{t,+} = L_{t,+} + \beta_+ \delta_t + \varepsilon_{t,+}. \quad (4.3.6)$$

Vergelijking (4.3.5) bevat de modellen voor de subcategorieën, waarbij  $y_{t,k}$  de geschatte variabele is op tijdstip  $t$  voor categorie  $k$  waargenomen via het oude of het nieuwe ontwerp,  $L_{t,k}$  een stochastisch trendmodel voor categorie  $k$ ,  $\delta_t$  een interventievariabele zoals in (4.3.1),  $\beta_k$  de regressiecoëfficiënt voor de interventievariabele behorende bij categorie  $k$  en  $\varepsilon_{t,k}$  een storingsterm. Vergelijking (4.3.6) is het model voor het totaal, waarbij  $y_{t,+}$  de geschatte variabele is op tijdstip  $t$  voor de hele categorie of populatie. Verder zijn  $L_{t,+}$ ,  $\beta_+$  en  $\varepsilon_{t,+}$  op analoge wijze gedefinieerd.

Voor  $L_{t,k}$  en analoog  $L_{t,+}$  worden dezelfde trendmodellen gebruikt, zoals bijvoorbeeld de *smooth trend*, zoals gedefinieerd in (4.3.3):

$$\begin{aligned} L_{t,k} &= L_{t-1,k} + R_{t-1,k}, \\ R_{t,k} &= R_{t-1,k} + \eta_{t,k}, \quad k = 1, \dots, K. \end{aligned} \quad (4.3.7)$$

Voor de storingstermen gelden dezelfde aannames als in (4.3.2) en (4.3.4).

In de set vergelijkingen wordt de tijdreeks van het totaal (4.3.6) uitgesplitst in  $K$  reeksen (4.3.5). Dat betekent dat op ieder tijdstip geldt dat de som over de  $K$

subcategorieën of deelpopulaties optelt tot het totaal van de categorieën of populaties:

$$y_{t,+} = \sum_{k=1}^K y_{t,k}. \quad (4.3.8)$$

Vanwege restrictie (4.3.8) geldt dat de regressiecoëfficiënten van de interventievariabelen moeten voldoen aan de restrictie:

$$b_+ = \sum_{k=1}^K b_k. \quad (4.3.9)$$

De schattingen voor de coëfficiënten van de interventievariabelen  $b_+$  en  $b_k$  kunnen worden geïnterpreteerd als de schattingen voor de methodebreuken en kunnen eventueel gebruikt worden om de tijdreeksen te corrigeren. In deelparagraaf 4.3.3 wordt beschreven hoe deze restrictie in een multivariaat tijdreeksmodel wordt opgenomen.

#### 4.3.2.2 Meerkeuzevragen in antwoordcategorieën

In een andere situatie, die veel voorkomt, is er sprake van  $K \geq 2$  antwoordcategorieën van een meerkeuzevraag, waarbij geldt dat de som van de categorieën gelijk is aan 100%. Voor een dergelijke variabele kunnen  $K$  reeksen voor de afzonderlijke antwoordcategorieën worden geconstrueerd, die op ieder tijdstip de procentuele verdeling over de  $K$  categorieën specificeren.

De tijdreeks van iedere afzonderlijke categorie wordt gemodelleerd met een trend, een interventievariabele die aangeeft wanneer de waarneming overgaat van het oude naar het nieuwe ontwerp, en een storingsterm;

$$y_{t,k} = L_{t,k} + b_k d_t + e_{t,k}, \quad k = 1, \dots, K. \quad (4.3.10)$$

Hierbij is  $y_{t,k}$  het geschatte percentage op tijdstip  $t$  voor categorie  $k$  waargenomen via het oude of het nieuwe ontwerp, en  $L_{t,k}$ ,  $d_t$ ,  $b_k$  en  $e_{t,k}$  zoals gedefinieerd in (4.3.5).

Voor  $L_{t,k}$  wordt wederom vaak hetzelfde trendmodel gebruikt, bijvoorbeeld de *smooth trend* zoals gedefinieerd in (4.3.7).

Voor iedere parameter is op deze manier een  $K$ -dimensionaal structureel tijdreeksmodel gedefinieerd die de procentuele verdeling over  $K$  categorieën beschrijft. Omdat de afhankelijke variabelen optellen tot 100% geldt voor de schattingen voor de methodebreuken dat de som over de  $K$  categorieën gelijk is aan nul. Dat betekent dat de regressiecoëfficiënten van de interventievariabele moeten voldoen aan de restrictie:

$$\sum_{k=1}^K b_k = 0. \quad (4.3.11)$$

In de volgende deelparagraaf 4.3.3 wordt beschreven hoe de restricties (4.3.9) en (4.3.11) in een multivariaat tijdreeksmodel worden opgenomen.

### 4.3.3 State Space vorm

Zowel het univariate tijdreeksmodel (4.3.2.1) als het  $K$ -dimensionaal structureel tijdreeksmodel (4.3.2.2) kunnen in toestandsruimteform ("State Space") worden geschreven. Vervolgens kunnen optimale schattingen voor de modelparameters worden verkregen via het Kalmanfilter (deelparagraaf 4.3.4). Zie Harvey (1989) of Durbin en Koopman (2001) voor technische details.

Enkele definities:

$A \otimes B$  : het kronecker product van A en B. Als A een  $m \times n$  matrix is en B een  $p \times q$  matrix, dan is  $A \otimes B$  de  $mp \times nq$  blokdiagonaalmatrix:

$$\begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

$Diag(d_1, d_2, \dots)$  : een diagonaalmatrix met diagonaalelementen  $d_1, d_2, \dots$ , en buiten-diagonaalelementen gelijk aan nul.

$Blockdiag(D_1, D_2, \dots)$  : een blokdiagonaalmatrix waarbij de blokken die niet op de diagonaal staan gelijk zijn aan matrices met ieder element gelijk aan nul zijn en de blokken op de diagonaal gevormd worden door de matrices  $D_1, D_2, \dots$ .

$\mathbf{0}_p$  : een kolomvector van de orde  $p$  met elk element gelijk aan nul.

$\mathbf{1}_p$  : een kolomvector van de orde  $p$  met elk element gelijk aan 1.

$\mathbf{I}_p$  : de  $p \times p$  identiteitsmatrix.

Indien men te maken heeft met uitsplitsingen van variabelen in subcategorieën of deelpopulaties, zoals beschreven in paragraaf 4.3.2.1 kan het multivariate tijdreeksmodel op de volgende manier in toestandsruimte vorm geschreven:

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \quad (4.3.12)$$

$$\boldsymbol{\alpha}_t = \mathbf{T} \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t. \quad (4.3.13)$$

De vergelijking (4.3.12) wordt de meetvergelijking genoemd en beschrijft hoe de waargenomen tijdreeks afhangt van een vector van (niet geobserveerde) toestandsvariabelen  $\boldsymbol{\alpha}_t$  en een vector van storingstermen  $\boldsymbol{\varepsilon}_t$ . In dit geval zijn de

toestandsvariabelen het *niveau* (level) en de *helling* (slope) van het trendmodel en de regressiecoëfficiënten van de interventievariabelen. Vergelijking (4.3.13) wordt de systeemvergelijking genoemd en geeft aan hoe de toestandsvariabelen veranderen in de tijd. De vector  $\boldsymbol{\eta}_t$  bevat de storingstermen van de onderliggende stochastische processen van de toestandsvariabelen. Als de tijdreeksen in (4.3.5) en (4.3.6) gemodelleerd worden met ieder hun eigen stochastisch trendmodel met interventievariabelen samen met restrictie (4.3.8), dan zijn de matrices en vectoren in (4.3.12) en (4.3.13) als volgt gedefinieerd:

$$\begin{aligned}\mathbf{y}_t &= (y_{t,+}, y_{t,1}, \dots, y_{t,K}) \\ \boldsymbol{\alpha}_t &= (L_{t,+}, R_{t,+}, L_{t,1}, R_{t,1}, \dots, L_{t,K}, R_{t,K}, \beta_+, \beta_1, \dots, \beta_K)^T, \\ \mathbf{Z}_t &= (\mathbf{I}_{K+1} \otimes (1,0) \mid \delta_t \mathbf{I}_{K+1}), \\ \mathbf{T} &= \text{Blockdiag}(\mathbf{T}_{tr}, \mathbf{T}_{iv}), \\ \mathbf{T}_{tr} &= \mathbf{I}_{K+1} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \\ \mathbf{T}_{iv} &= \begin{pmatrix} 0 & \mathbf{1}_K^T \\ \mathbf{0}_K & \mathbf{I}_K \end{pmatrix}.\end{aligned}$$

De vectoren van storingstermen zijn gedefinieerd als

$$\begin{aligned}\boldsymbol{\varepsilon}_t &= (\varepsilon_{t,+}, \varepsilon_{t,1}, \dots, \varepsilon_{t,K})^T, \\ \boldsymbol{\eta}_t &= (0, \eta_{t,+}, 0, \eta_{t,1}, \dots, 0, \eta_{t,K}, \mathbf{0}_{K+1}^T)^T.\end{aligned}$$

Er wordt aangenomen dat

$$\begin{aligned}E(\boldsymbol{\varepsilon}_t) &= \mathbf{0}_{K+1}, \quad \text{Cov}(\boldsymbol{\varepsilon}_t) = \text{Diag}(\sigma_{\varepsilon,+}^2, \sigma_{\varepsilon,1}^2, \dots, \sigma_{\varepsilon,K}^2), \\ E(\boldsymbol{\eta}_t) &= \mathbf{0}_{3(K+1)}, \quad \text{Cov}(\boldsymbol{\eta}_t) = \text{Diag}(0, \sigma_{\eta,+}^2, 0, \sigma_{\eta,1}^2, \dots, 0, \sigma_{\eta,K}^2, \mathbf{0}_{K+1}^T).\end{aligned}$$

Indien men te maken heeft met antwoordcategorieën van meerkeuzevragen, dan wordt (4.3.10) samen met restrictie (4.3.11) in toestandsruimte vorm gelijk aan:

$$\begin{aligned}\mathbf{y}_t &= (y_{t,1}, \dots, y_{t,K}) \\ \boldsymbol{\alpha}_t &= (L_{t,1}, R_{t,1}, \dots, L_{t,K}, R_{t,K}, \beta_1, \dots, \beta_K)^T, \\ \mathbf{Z}_t &= (\mathbf{I}_K \otimes (1,0) \mid \delta_t \mathbf{I}_K), \\ \mathbf{T} &= \text{Blockdiag}(\mathbf{T}_{tr}, \mathbf{T}_{iv}), \\ \mathbf{T}_{tr} &= \mathbf{I}_K \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\end{aligned}$$



$$\mathbf{T}_{iv} = \begin{pmatrix} \mathbf{I}_{K-1} & \mathbf{0}_{K-1} \\ -\mathbf{1}_{K-1}^T & 0 \end{pmatrix}.$$

Verder geldt dan:

$$\boldsymbol{\varepsilon}_t = (\boldsymbol{e}_{t,1}, \dots, \boldsymbol{e}_{t,K})^T,$$

$$\boldsymbol{\eta}_t = (0, h_{t,1}, \dots, 0, h_{t,K}, \mathbf{0}_K^T)^T.$$

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_K, \text{Cov}(\boldsymbol{\varepsilon}_t) = \text{Diag}(\boldsymbol{s}_{e,1}^2, \dots, \boldsymbol{s}_{e,K}^2),$$

$$E(\boldsymbol{\eta}_t) = \mathbf{0}_{3K}, \text{Cov}(\boldsymbol{\eta}_t) = \text{Diag}(0, \boldsymbol{s}_{h,1}^2, \dots, 0, \boldsymbol{s}_{h,K}^2, \mathbf{0}_K^T).$$

#### 4.3.4 Kalmanfilter

In de vorige paragraaf is beschreven hoe structurele tijdreeksmodellen in de zogenaamde state-space of toestandsruimteform geschreven kunnen worden. Vervolgens kan het Kalmanfilter worden gebruikt om dit model te schatten. Zie Harvey (1989) of Durbin en Koopman (2001) voor een introductie in het schatten van toestandsruimte modellen met behulp van het Kalmanfilter.

Via het Kalmanfilter wordt voor iedere tijdstip een optimale schatting gemaakt voor de doelvariabele en de modelparameters op basis van de informatie die beschikbaar is tot en met deze periode. Dit zijn de zogenaamde gefilterde schattingen. Het Kalmanfilter is een recursief algoritme dat start aan het begin van de tijdreeks en eindigt bij de waarneming van de laatste periode. Vervolgens kunnen de gefilterde schattingen worden verbeterd met de informatie die beschikbaar is gekomen na de periode waarop de gefilterde schatting betrekking heeft. Dit is een recursief algoritme dat start bij de laatst waargenomen periode en eindigt bij het begin van de reeks. Dit proces wordt smoothen genoemd. De gefilterde schattingen voor ieder tijdstip  $t$  zijn de optimale schattingen gebaseerd op de waarnemingen tot en met periode  $t$ . De gesmoohte schattingen zijn de optimale schattingen voor periode  $t$ , gebaseerd op alle informatie uit de beschikbare tijdreeks. In dat geval wordt bij de schatting voor de model- en doelvariabelen van periode  $t$  ook gebruik gemaakt van de informatie die na dit tijdstip beschikbaar zijn gekomen. De gesmoohte schattingen zijn gebaseerd op de fixed interval smoother. Zie Harvey (1989) of Durbin en Koopman (2001) voor technische details.

Alle toestandsvariabelen van het toestandsruimte model zijn niet-stationair. De niet-stationaire toestandsvariabelen worden geïnstalleerd met een diffuse prior. Dat wil zeggen dat deze toestandsvariabelen aan het begin van de reeks een startwaarde krijgen die gelijk is aan nul met een diagonale covariantiematrix met zeer grote waarden.

#### 4.3.5 Correctie

Voor het univariate tijdreeksmodel kan  $\beta$  gebruikt worden om de tijdreeks te corrigeren. Men kan de tijdreeksen na het moment van de overgang van de oude naar de nieuwe enquête corrigeren met:

$$\tilde{y}_t = y_t - \beta.$$

Als alternatief kan men ook de tijdreeks vóór de overgang corrigeren met

$$\tilde{y}_t = y_t + \beta.$$

Voor multivariate tijdreeksmodellen geldt dat zowel de regressiecoëfficiënten als de Kalman filter schattingen voldoen aan de restricties (4.3.9) of (4.3.11). Zodoende kan men de tijdreeksen na het moment van de overgang van de oude naar de nieuwe enquête corrigeren met:

$$\tilde{y}_{t,k} = y_{t,k} - \beta_k.$$

Als alternatief kan men ook de tijdreeks vóór de overgang corrigeren met

$$\tilde{y}_{t,k} = y_{t,k} + \beta_k.$$

Een voordeel is dat de methodebreuken worden geschat voor de reeksen zoals die gepubliceerd worden met een methodiek die er voor zorgt dat voldaan is aan de consistentie-eis dat de reeksen over de subcategorieën of deelpopulaties moeten blijven optellen tot de hele categorie of populatie of in het geval van meerkeuzevragen de onderliggende categorieën optellen tot 100%.

#### 4.3.6 Alternatieve interventies

Tot nu toe is verondersteld dat het effect van een herontwerp alleen het niveau van een reeks beïnvloedt. Een dergelijk effect kan worden gemodelleerd met een zogenaamde level-interventie zoals beschreven in deelparagraaf 4.3.2. Bij een level-interventie is interventievariabele  $\delta_t$  gelijk aan 0 onder het oude ontwerp en gelijk aan 1 onder het nieuwe ontwerp. De bijbehorende regressiecoëfficiënt zorgt voor een niveauverschil in de reeks die verder tijdsinvariant is. Bij een dergelijke interventie en een correctie van de reeks op basis hiervan wordt verondersteld dat het effect van het herontwerp constant is in de tijd. Het effect van een herontwerp kan echter complexer zijn. Verder is bij het corrigeren van reeksen de aanname dat het effect constant is in de tijd onwaarschijnlijker naarmate een reeks over een langere periode gecorrigeerd wordt.

Een eenvoudige methode om de level-interventie tijdsafhankelijk te maken, is het toepassen van een transformatie op de doelvariabele. Variabelen zoals slachtofferschap kunnen geen negatieve waarden aannemen. Door de logaritmisches getransformeerde reeksen te modelleren en vervolgens te corrigeren via een level-interventie, is de correctie na toepassen van de inverse transformatie evenredig aan de waarde van de doelvariabelen. Dit voorkomt bijvoorbeeld dat gecorrigeerde reeksen die per definitie een positieve waarde hebben, na correctie negatief worden.

Een dergelijke correctie is vergelijkbaar met de multiplicatieve correctie (3.4.2) beschreven in paragraaf 3.4. In de volgende deelparagraaf wordt nader ingegaan op het transformeren van doelvariabelen.

Het is mogelijk om interventies voor een state-space model tijdsafhankelijk te maken. In Van den Brakel en Roels (2010) wordt beschreven hoe tijdsafhankelijke interventies worden verkregen door de interventie op de hellingsparameter van het trendmodel toe te passen of een interventie op het seizoenspatroon te modelleren. In de literatuur zijn meerdere alternatieve vormen van interventies beschreven, zie bijvoorbeeld Box en Tiao (1975) of Durbin en Koopman (2002, hoofdstuk 3). In het algemeen moet voorzichtig worden omgegaan met het toepassen van tijdsafhankelijke interventies. Een tijdsafhankelijke interventie is flexibeler waardoor het risico toeneemt dat een deel van de werkelijke ontwikkeling van de populatieparameter ten onrechte in de interventievariabele wordt opgenomen.

Ten slotte is het mogelijk dat een herontwerp niet alleen effect heeft op de puntschattingen, maar ook de omvang van de meetfouten of de beweeglijkheid van de reeks beïnvloedt. Dergelijke effecten kunnen worden gemodelleerd door de variantieparameter van de meetvergelijking tijdsafhankelijk te kiezen of de geschatte steekproefvariantie als priorinformatie aan het model mee te geven. Zie Van den Brakel en Roels (2010) voor technische details.

#### *4.3.7 Transformatie van de doelvariabelen*

Er zijn verschillende redenen om een transformatie toe te passen op de waargenomen tijdreeksen. In paragraaf 4.3.6 is reeds aangegeven dat een logaritmische transformatie kan worden overwogen om ervoor te zorgen dat de correctie van een reeks evenredig is aan het niveau van de doelvariabelen. Een dergelijke transformatie past goed bij variabelen die per definitie niet negatief kunnen worden omdat de gecorrigeerde reeks na terugtransformeren per definitie positief is.

Voor een set van reeksen die voor ieder tijdstip de procentuele verdeling over een aantal categorieën specificeren kan worden overwogen om de zogenaamde logratio transformatie toe te passen, Aitchison (1986). Deze transformatie houdt rekening met de eigenschap dat de som over de reeksen op ieder tijdstip gelijk is aan 100% en dat de afzonderlijke reeksen uitsluitend waarden tussen 0% en 100% kunnen aannemen. Gecorrigeerde reeksen voldoen na toepassen van de inverse transformatie altijd aan de restrictie dat de som over de gecorrigeerde reeksen gelijk is aan 100% en dat de waarden binnen het interval  $[0,100]$  liggen. Het nadeel van deze transformatie is dat de resultaten van de analyse beïnvloed worden door de keuze van de categorie die in de noemer van de logratio transformatie wordt gebruikt.

Een tweede belangrijke reden om de waargenomen reeksen te transformeren is om beter aan de modelveronderstellingen van het tijdreeksmodel te voldoen. De in dit hoofdstuk beschreven modellen gaan er vanuit dat de storingstermen in de verschillende vergelijkingen onderling onafhankelijk en normaal verdeeld zijn met

gelijke variantie. Reeksen die per definitie positief zijn, kunnen een sterke asymmetrische verdeling hebben waardoor de normaliteitveronderstelling geschonden kan worden. Door dergelijke reeksen logaritmisch te transformeren wordt naar een meer symmetrische verdeling toegewerkt waardoor de veronderstelling dat residuen normaal verdeeld zijn plausibeler wordt.

Een set van reeksen die voor ieder tijdstip de procentuele verdeling over een aantal categorieën specificeren zijn doorgaans afgeleid van een variabele die een multinomiaal verdeeld is. De hiervoor genoemde logratio transformatie kan overwogen worden indien uit een modevaluatie volgt dat de veronderstelling dat de residuen van het tijdreeksmodel bij benadering normaal verdeeld zijn ernstig geschonden wordt.

#### 4.3.8 Software

Voor univariate modellen kunnen de op het CBS aanwezige pakketten STAMP (Koopman e.a. 1999a) en EVIEWS (referentie) gebruikt worden. De schatting van de methodebreuk  $\beta$  kan geschat worden, waarna de tijdreeksen eventueel gecorrigeerd kunnen worden. De pakketten zijn redelijk eenvoudig in het gebruik.

Er is geen reguliere statistische software beschikbaar waarmee het voorgestelde multivariate structurele tijdreeksmodel kan worden geschat. Voor dergelijke toepassingen moet de gebruiker het model implementeren in de matrix programmeertaal *Ox* waarbij gebruik kan worden gemaakt van de subroutines uit Ssfpack 3.0, zie Doornik (1998) en Koopman e.a. (1999b, 2008).

## 4.4 Voorbeeld

Van 1997 tot en met 2004 is informatie over milieuedrag en maatschappelijke participatie waargenomen in het Permanent Onderzoek Leef Situatie (POLS). Het POLS was opgezet als een continu doorlopend onderzoek waarbij interviewers de elektronische vragenlijst afnemen via een persoonlijk vraaggesprek met de respondent (CAPI). Voor het POLS werd een zogenaamde basisvragenlijst afgenomen waar algemene informatie werd gevraagd over de leefsituatie van de respondent. Vervolgens werd de steekproef voor de basisvragenlijst verdeeld in deelsteekproeven die werden toegewezen aan vervolgmodes. Onder deze opzet werd de informatie over maatschappelijke participatie en milieu verzameld in twee afzonderlijke vervolgmodes. De vragen met betrekking tot milieuedrag werden waargenomen in de vervolgmodes Recht en Milieu (REM). Vragen met betrekking tot maatschappelijke participatie werden waargenomen in de module Recht en Participatie (REP).

In 2005 zijn uit bezuinigingsoverwegingen de modules REP en REM uit het POLS gestopt. De vragen met betrekking tot rechtsbescherming en veiligheid die werden waargenomen via de modules REP en REM zijn overgegaan naar de Veiligheidsmonitor (VM). De vragen over milieu uit de module REM en de vragen over participatie uit de module REP zijn samengevoegd tot een nieuwe enquête Maatschappelijke Participatie en Milieu (P&M). De elektronische vragenlijst van

deze enquête wordt door interviewers telefonisch afgenomen bij de respondenten (CATI). Verder is de enquête van een continu doorlopend onderzoek overgegaan op een waarneming in de maanden september, oktober en november. Om de interviewduur te beperken is een aantal vragen niet langer opgenomen. Het aantal respondenten verminderde van netto ruim 5.200 in 2004 naar netto 4.400 in 2005.

Veranderingen in het proces door de overgang van het POLS naar de enquête P&M hebben geleid tot verschillen in de uitkomsten, zogenaamde methodebreuken. Systematische verschillen in de schattingen van de doelvariabelen zijn het netto resultaat van de volgende factoren die gelijktijdig in de onderzoeksopzet zijn gewijzigd:

- Verandering van de waarnemingsmethodiek. De overgang van CAPI naar CATI kan op verschillende manieren een systematisch effect op de uitkomsten hebben.
- Veranderingen in de vragenlijst. Onder het POLS werden de vragen over maatschappelijke participatie en milieu waargenomen in twee aparte vervolgmodes in combinatie met vragen over rechtsbescherming en veiligheid. Deze vragen zijn overgegaan naar een nieuwe vragenlijst waardoor grote veranderingen zijn opgetreden in de context van de enquête.
- Verandering in de waarnemingsperiode. De dataverzamelingsperiode is veranderd van een continu doorlopende enquêtering naar een driemaandelijksse enquêtering in de maanden september tot en met november.

Alle variabelen met betrekking tot milieugegedrag en maatschappelijke participatie zijn gebaseerd op meerkeuzevragen. De indicatoren die op basis van deze vragen worden gepubliceerd specificeren de procentuele verdeling over de betreffende antwoordcategorieën. Daarom is het multivariaat structureel tijdreeksmodel, zoals beschreven in deelparagraaf 4.3.2.2, gebruikt om de oorspronkelijke reeksen van de onderliggende antwoordcategorieën simultaan te beschrijven. Via dit model wordt een restrictie op de geschatte methodebreuken opgelegd, die er in dit geval voor zorgt dat de som van de breuken over de antwoordcategorieën gelijk is aan nul.

In paragraaf 4.3.7 is aangegeven dat kan worden overwogen om deze reeksen te transformeren via een logratio transformatie. Via een simulatiestudie is onderzocht of deze transformatie tot betere resultaten leidt voor reeksen die worden gegenereerd uit zowel een normale als een multinomiale verdeling. Voor beide verdelingen werden gesimuleerde methodebreuken het nauwkeurigste geschat indien geen transformatie werd toegepast. Zie Van den Brakel en Roels (2010) voor een gedetailleerde beschrijving van deze simulatie en het gebruik van logratio transformaties voor een set van reeksen die de procentuele verdeling over een aantal categorieën specificeren.

Voor vrijwel alle onderzochte variabelen blijkt dat het herontwerp resulteert in een methodebreuk die significant ongelijk is aan nul. Deze breuken zijn het netto resultaat van de verandering in dataverzamelingsmethode (van CAPI naar CATI), verschillen tussen de vragenlijst en de context van het POLS en de enquête P&M en

de overgang van een continue waarneming naar een waarneming in de maanden september tot en met november van het jaar.

In deze toepassing is besloten om de P&M cijfers te corrigeren voor de waargenomen methodebreuken zodat deze aansluiten bij de cijfers van 1997 tot en met 2004 van het POLS. Hiertoe is besloten omdat de cijfers gebaseerd op het POLS betrouwbaarder zijn dan de cijfers gebaseerd op de enquête P&M. Overwegende argumenten zijn dat met de CAPI-waarneming de doelpopulatie beter bereikt wordt en dat het POLS gebaseerd is op een continue doorlopende waarneming zodat seizoenseffecten geen rol spelen.

Als voorbeeld geven we de resultaten voor de variabele “Contact met burenen”. De exacte vraagstelling betreffende het contact met de burenen luidt: “Hoe vaak hebt u contact met de burenen?” met de antwoordcategorieën:

1. 1 keer in de week of vaker,
2. 1 keer in de 2 weken,
3. minder dan 1 keer in de 2 weken,
4. nooit.

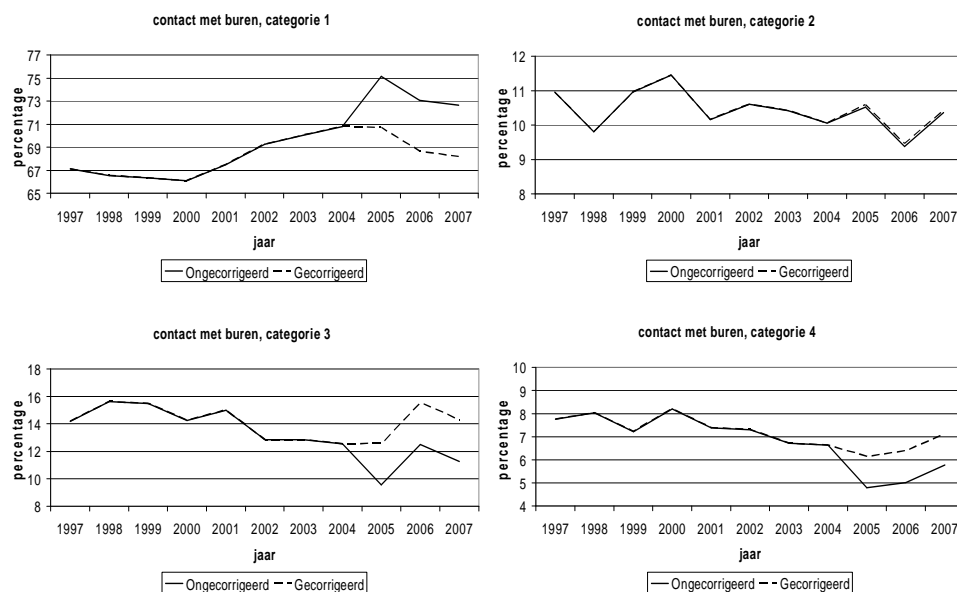
In tabel 1 zijn schattingen voor de methodebreuken gegeven op basis van een univariaat tijdreeksmodel dat geen rekening houdt met de restrictie dat de geschatte breuken moeten optellen tot nul en het multivariate model uit deelparagraaf 4.3.2.2, dat wel rekening houdt met deze restrictie. In vergelijking met het univariate model zorgt het multivariate model dat de reeksen consistent blijven omdat de som van de geschatte methodebreuken over de vier categorieën exact gelijk is aan nul. Daarnaast zijn de schattingen met het multivariate model preciezer. De interpretatie hiervan is, dat de restrictie op de parameters voor de breuken extra informatie aan het model toevoegt waardoor deze parameters preciezer kunnen worden geschat.

*Tabel 1. Schattingen methodebreuken in de reeksen over Contact met burenen (standaardfouten tussen haakjes)*

Model	Categorie							
	1		2		3		4	
Univariaat	4.40	(1.20)	-0.09	(0.59)	-3.18	(1.30)	-1.36	(0.59)
Multivariaat	4.44	(0.93)	-0.07	(0.56)	-3.01	(0.95)	-1.35	(0.56)

In figuur 1 zijn de meetresultaten en de gecorrigeerde cijfers van deze 4 antwoordcategorieën grafisch gepresenteerd. De figuur betreffende categorieën 1, 3 en 4 laten zien dat er een sterke sprong zit in de meetresultaten ten gevolge van het POLS naar de P&M in 2005. Na correctie voor de geschatte methodebreuken lijken de gecorrigeerde cijfers beter ‘in lijn’ te liggen met de ontwikkelingen in de periode 1997-2004.

Figuur 1. Tijdreeksen Contact met buren



Een ander voorbeeld waarbij structurele tijdreeksmodellen gebruikt zijn om tijdreeksen te corrigeren is te vinden in Van den Brakel e.a. (2008) waarbij uitsplitsingen van aggregaten bij slachtofferschaps-delicten behandeld worden. Dit betreft een voorbeeld waarbij een reeks van populatietotalen en een uitsplitsing daarvan in drie categorieën simultaan wordt gemodelleerd met het model beschreven in deelparagraaf 4.3.2.1.

#### 4.5 Eigenschappen

Het effect van alle veranderingen in het statistische proces kan worden geschat via een tijdreeksmodel. Het gebruik van structurele tijdreeksen is meestal interessant als (door meestal budgettaire redenen) in één keer is overgestapt op het nieuwe ontwerp. Voorwaarde is dat voldoende lange reeksen beschikbaar zijn, ook onder het nieuwe statistische proces.

In de loop van de tijd komen nieuwe data onder het nieuwe statistische proces beschikbaar. Met de nieuwe data kunnen betrouwbaardere schattingen voor de methodebreuken worden gemaakt. Een nadeel is dat de geschatte breuken gereviseerd zouden kunnen worden. Vooral in de eerste periode na de overgang op het nieuwe ontwerp kunnen de schattingen voor de omvang van de methodebreuken aan sterke revisies onderhevig zijn als een nieuwe waarneming aan de reeks wordt toegevoegd. Zie Van den Brakel e.a. (2008) voor een voorbeeld bij slachtofferschapsindicatoren. Er zal daarom moeten worden nagedacht over een praktisch hanteerbare revisiestrategie. Een mogelijkheid is om op het moment dat de omvang van de revisies klein wordt, de schattingen voor de methodebreuken vast te stellen en vanaf dat moment te gebruiken alsof ze tijdinvariant zijn.

In de publicaties kan op verschillende manieren worden omgegaan met de waargenomen methodebreuken:

1. De eerste mogelijkheid is om cijfers onder het nieuwe ontwerp te corrigeren zodat deze aansluiten bij de cijfers onder het oude ontwerp. Deze mogelijkheid ligt voor de hand als de cijfers onder het oude ontwerp betrouwbaarder zijn dan de cijfers onder het nieuwe ontwerp. De vraag bij deze aanpak is hoe lang men de correctie kan toepassen op steeds nieuw binnenkomende data.
2. De tweede mogelijkheid is om de cijfers onder het oude ontwerp te corrigeren zodat deze aansluiten bij de cijfers onder het nieuwe ontwerp. Deze mogelijkheid ligt voor de hand als de cijfers onder het nieuwe ontwerp betrouwbaarder zijn dan de cijfers onder het oude ontwerp. Een voordeel is dat er eenmalig een correctie wordt gedaan, en men daarna de nieuwe reeks kan voortzetten. Meestal is het echter zo dat er budgettaire redenen zitten achter het herontwerp zodat het waarschijnlijk is dat de nieuwe cijfers juist minder betrouwbaar zijn.
3. De derde mogelijkheid is om de reeksen zoals die zijn waargenomen onder het oude en het nieuwe ontwerp ongecorrigeerd te publiceren. In de publicatie moet duidelijk worden aangegeven dat vanaf een bepaalde datum de onderzoeksopzet gewijzigd is waardoor systematische verschuivingen in de reeksen opgetreden zouden kunnen zijn. De schattingen voor de methodebreuken kunnen bij deze reeksen worden gepubliceerd. Deze schattingen kunnen aangepast worden als er meer data onder het nieuwe ontwerp beschikbaar komen. Aan de gebruiker wordt overgelaten of en hoe de schattingen voor de methodebreuken worden gebruikt om een deel van de reeks te corrigeren. De oorspronkelijke informatie blijft behouden en de gebruiker heeft de mogelijkheid heeft om naar eigen inzicht wel of geen correctie toe te passen.

Indien voor een  $K$ -dimensionaal structureel tijdreeksmodel de breuken zijn geschat, dan kan het zijn dat voor sommige categorieën of deelpopulaties de breuken niet significant zijn en voor andere categorieën of deelpopulaties wel. Wellicht is het beter toch alle categorieën of deelpopulaties te corrigeren zodat de regressiecoëfficiënten van de interventievariabele blijven voldoen aan de restricties.

Het schatten van methodebreuken via tijdreeksmodellen is met name geschikt indien na doorvoeren van een verandering of herontwerp voldoende waarnemingen onder het nieuwe ontwerp beschikbaar zijn. Zoals aangegeven in paragraaf 4.2 en aan het begin van deze paragraaf zijn de schattingen voor de methodebreuk aan revisies onderhevig als vlak na de overgang op het nieuwe ontwerp steeds nieuwe data aan de reeks worden toegevoegd. De methode is om deze reden zeer geschikt voor schatten en repareren van breuken in lange reeksen waarbij in het verleden methodebreuken zijn opgetreden die nog niet onderzocht en gerepareerd zijn.

In zijn algemeenheid kunnen geen richtlijnen worden gegeven over hoeveel waarnemingen noodzakelijk zijn om tot een definitieve schatting van de methodebreuk te komen. Dit wordt namelijk sterk bepaald door de mate waarin de overige modelcomponenten tijdsafhankelijk zijn en de gewenste nauwkeurigheid. Naarmate bijvoorbeeld de tijdsafhankelijkheid van de trendcomponent toeneemt, is de schatting voor de methodebreuk meer gebaseerd op de lokale data rondom het



moment van het herontwerp en zullen revisies van de schattingen voor de methodebreuken sneller afnemen indien nieuwe waarden aan de reeks worden toegevoegd. Een manier om inzicht te krijgen in de betrouwbaarheid van de geschatte breuken bij een beperkt aantal waarnemingen is om een simulatie uit te voeren waarbij tijdreeksen worden gegenereerd met verschillende lengten op basis van tijdreeksmodellen die zijn geschat op de beschikbare reeks. Hoe een dergelijke simulatie kan worden uitgevoerd, is beschreven in Van den Brakel en Roels (2010).

In deelparagraaf 4.3.6 is aangegeven dat het gebruik van tijdsafhankelijke interventies het risico met zich meebrengt dat een deel van de werkelijke ontwikkeling van de doelvariabelen door het model ten onrechte wordt toegeschreven aan de interventie. Om dezelfde reden moet voorzichtig worden omgegaan met het opnemen meerdere level-interventies voor het modelleren van verschillende veranderingen in het statistische proces. Naarmate het aantal level-interventies binnen een tijdsinterval toeneemt, wordt het risico groter dat een deel van de echte ontwikkeling in de schattingen van de opeenvolgende methodebreuken terecht komt.

Bij het modelleren van lange reeksen komt het voor dat voor bepaalde perioden waarnemingen ontbreken. Een voordeel van het gebruik van state-space modellen is dat via het Kalmanfilter voorspellingen voor de ontbrekende data worden gegenereerd, Durbin en Koopman (2002, hoofdstuk 2).

#### 4.6 Kwaliteitsindicatoren

Aan het structurele tijdreeksmodel ligt een aantal modelaannamen ten grondslag. Het is van belang dat deze geëvalueerd worden.

De aanname dat voor ieder van de storingstermen  $\varepsilon_t$  en  $\eta_t$  geldt dat deze onderling onafhankelijk en normaal verdeeld zijn, leidt ertoe dat de gestandaardiseerde 1-stap voorspelfout, gedefinieerd als

$$e_t = \frac{v_t}{\sqrt{F_t}}, \quad (4.6.1)$$

ook onderling onafhankelijk en normaal verdeeld zijn. Hierbij geldt dat  $v_t = y_t - Z_t a_t$ ,  $a_t = E(\alpha_t | Y_{t-1})$ ,  $F_t = \text{var}(v_t)$  en  $Y_t = \{y_1, \dots, y_t\}$ .

Voor  $v_t$  wordt ook wel de term *innovaties* gebruikt, zodat  $e_t$  de gestandaardiseerde innovaties zijn. Zie voor uitvoerige beschrijving Durbin en Koopman (2001).

In de literatuur is een groot aantal toetsen en diagnostieken bekend om de aannemelijkheid voor de veronderstellingen voor  $e_t$  te evalueren:

- Plotjes van  $e_t$ : Bijvoorbeeld normaliteitsplot, histogram en autocorrelogram.
- Toets op normaliteit van  $e_t$ : Bijvoorbeeld de *Doornik-Hansen Test* (deze toets is een aangepaste versie van de *Bowman-Shenton test*).

- F-Toets op heteroskedasticiteit van  $e_t$ .
- Toets op autocorrelatie van  $e_t$ : *Durbin Watson Test* en de *Box Ljung Q Test*.

Met het pakket STAMP kunnen voor univariate tijdreeksmodellen deze tests uitgevoerd worden. Ook kunnen diverse plotjes gemaakt worden.

Indien er seriële correlatie in de storingstermen of de innovaties aanwezig is, kan dit een reden zijn om afhankelijk van het patroon een seizoenscomponent of een additionele cyclische component in het model op te nemen. Ook kan het noodzakelijk zijn om seriële autocorrelatie in de residuen te modelleren met een ARMA-component. Deze situatie kan vooral optreden indien steekproeven elkaar gedeeltelijk overlappen omdat ze gebaseerd zijn op een roterend panelontwerp.

Indien de veronderstelling dat de residuen en de innovaties normaal verdeeld zijn niet houdbaar is, kan hier via een transformatie van de reeksen naar toe worden gewerkt. Daarnaast bestaan er zeer geavanceerde analysemethoden voor de analyse van niet-Gausische structurele tijdreeksmodellen. Dergelijke methoden vallen buiten de scope van deze Methodenreeks. Voor meer informatie wordt verwezen naar Durbin en Koopman (2002, hoofdstuk 10 en 11).

Een eerste indicatie van de kwaliteit van de schattingen voor de methodebreuken is de bijbehorende standaardfout. De variantie of de standaardfout van de schatting voor de methodebreuk zegt echter niets over plausibiliteit van de modelveronderstelling die wordt gemaakt met de keuze voor een tijdsafhankelijke of tijdsafhankelijke interventie. Meer inzicht in de betrouwbaarheid van een schatting voor de methodebreuken kan worden verkregen door het uitvoeren van simulaties. Aan de hand van een aantal potentiële modellen waaruit een keuze gemaakt moet worden, worden tijdreeksen gegenereerd. Aan deze reeksen worden verschillende soorten methodebreuken toegevoegd en vervolgens worden deze reeksen geanalyseerd met verschillende modelspecificaties. Dit geeft een beeld hoe goed of slecht in een specifieke situatie een methodebreuk wordt waargenomen. Een uitgebreid voorbeeld van een dergelijke modevaluatie via simulaties is uitgewerkt in Van den Brakel en Roels (2010).

## 5. Literatuur

- Aitchison, J. (1986), *The statistical analysis of compositional data*. London: Chapman and Hall.
- Agresti, A. (1990), *Categorical Data Analysis*. Wiley, New-York.
- Banning, R. en Knottnerus, P. (2010), *Methodenreeks: Thema: Steekproeftheorie, Deelthema Steekproefontwerp*. Centraal Bureau voor de Statistiek, Den Haag. (in voorbereiding).
- Berkenbosch B.F. (2009), *Methodenreeks: Thema: Benaderingsstrategieën. Deelthema: Responsverhogende maatregelen bij bedrijvenquêtes*, Digidivnummer: 464527-v1. Centraal Bureau voor de Statistiek, Heerlen.
- Bethlehem J.G. (2010), *Methodenreeks: Thema: Foutenbronnen*. BPA. nr. DMV-2010-01-04-JBTM. Centraal Bureau voor de Statistiek, Leidschenvveen.
- Beukenhorst D.J. en Kerssemakers, F. (2010), *Methodenreeks: Thema: Benaderingsstrategieën. Deelthema: Algemene beschrijving en uni-mode designs*, BPA. nr. DMH-2010-01-25-DBKT-FKRS. Centraal Bureau voor de Statistiek, Heerlen.
- Beukenhorst D.J. en Schouten, J.G. (2010), *Methodenreeks: Thema: Benaderingsstrategieën. Deelthema: Mixed Mode benadering* (nog te verschijnen). Centraal Bureau voor de Statistiek, Heerlen.
- Binder, D.A. and Dick, J.P. (1990), A method for the analysis of seasonal ARIMA models. *Survey Methodology* 16, 239-253.
- Boonstra H.J. en Buelens, B. (2007), *Methodenreeks: Thema: Modelmatig schatten. Deelthema's: Synthetische schatters en Kleinedomeinschatters*, BPA. nr. DMH-2007-12-18-HBTA-BBUS. Centraal Bureau voor de Statistiek, Heerlen.
- Box, G.E.P. & Tiao, G.C. (1975), Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* 70, 70-79.
- Brakel, J.A. van den (2001), *Design and Analysis of Experiments Embedded in Complex Sample Surveys*. Proefschrift, Erasmus Universiteit Rotterdam.
- Brakel, J.A. van den (2005), *Trendbreukanalyse Veiligheidsmonitor Rijk*. TMO\_R&D\_2005\_08\_26\_JBRL, CBS-rapport, Heerlen.
- Brakel, J.A. van den (2007), *Herberekening PS-2005 naar herontwerp ABR*, CBS-nota, DMH-2007-06-05-JBRL, Centraal Bureau voor de Statistiek, Heerlen.
- Brakel, J.A. van den (2008), Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. *Journal of the Royal Statistical Society*, (Series A) 171 (3), 581-613.

- Brakel, J.A. van den (2009), *Sampling and estimation techniques for the implementation of the NACE Rev. 2 in Business Surveys*. Discussion paper 09013, Centraal Bureau voor de Statistiek, Heerlen.
- Brakel, J.A. van den and Berkel, C.A.M. van (2002), A Design-Based Analysis Procedure for Two-Treatment Experiments Embedded in Sample Surveys. *Journal of Official Statistics* 18 (2) 217-231.
- Brakel, J.A. van den en Krieg, S.. *Methodenreeks: Thema: Modelmatig schatten. Deelthema's: Kleinedomeinschatters op basis van structurele tijdreeksmodellen voor een roterend panelontwerp*. Centraal Bureau voor de Statistiek, Heerlen. (in voorbereiding).
- Brakel, J.A. van den, Smith, P. and Compton, S. (2008), Quality procedures for survey transitions – experiment, time series and discontinuities. *Journal of Survey Research Methods* 2, 123-141.
- Brakel, J.A. van den and Renssen, R.H. (1998), Design and Analysis of Experiment Embedded in Sample Surveys. *Journal of Official Statistics* 14, 277-295.
- Brakel, J.A. van den and Renssen, R.H. (2005), Analysis of experiments embedded in complex sampling designs. *Survey Methodology* 31, 23-40.
- Brakel, J.A. van den en Roels, J. (2008), *Schatten van methodebreuken bij de enquête maatschappelijke participatie en milieu via multivariate tijdreeksmodellen*. DMH-08-06-03-JBRL. CBS-nota, Heerlen.
- Brakel, J.A. van den and Roels, J. (2010), Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics*, in press.
- Buiten, G., Kampen, J.K., and Vergouw, S. (2008), *Theory on the producing of historical time series for Short-term Business Statistics in NACE Rev. 2 with an application in the industrial turnover index in the Netherlands (1995-2008)*. Research paper, BPA nr.: DMK-2008-10-02-JKPN, Statistics Netherlands.
- CBS (2002), *Trendbreuk Analyse Onderzoek Verplaatsingsgedrag*. BPA nr. H01190-02-SAH, Centraal Bureau voor de Statistiek, Heerlen.
- Cochran, W.G. (1977), *Sampling Techniques*. Wiley, New York.
- Cochran, W.G. and Cox, G.M. (1957), *Experimental Designs*. John-Wiley & Sons, New-York.
- Doornik, J.A. (1998), *Object-oriented matrix programming using Ox 2.0*. Timberlake Consultants Press, London.
- Durbin, J. and Koopman, S.J. (2001), *Time series analysis by state space methods*. Oxford University Press, Oxford.
- Eurostat (2006), Task Force on the implementation of NACE Rev. 2, *Back casting handbook*, <http://circa.europa.eu/irc/dsis/nacecpacon/info/data/en/index.htm>

- Luiten J.M.M. (2009), *Methodenreeks: Thema: Benaderingsstrategieën. Deelthema: Responsverhogende maatregelen bij sociale statistieken*. Centraal Bureau voor de Statistiek, Heerlen.
- Fienberg, S.E. & Tanur, J.M. (1987), Experimental and sampling structures: parallels diverging and meeting. *International Statistical Review* 55, 75-96.
- Fienberg, S.E. & Tanur, J.M. (1988), From the inside out and the outside in: combining experimental and sampling structures. *Canadian Journal of Statistics*, 16, 135-151.
- Fienberg S.E. & Tanur, J.M. (1989), Combining cognitive and statistical approaches to survey design. *Science* 243, 1017-1022.
- Giesen D., Meertens V., Vis-Visschers R. & Beukenhorst D.J. (2010), *Methodenreeks: Thema: Vragenlijstontwikkeling* (nog te verschijnen). Centraal Bureau voor de Statistiek, Heerlen.
- Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge.
- James, G. (2008), *Backcasting for use in Short Term Statistics*. Interim Report from the UK Office for National Statistics.
- Lehmann, E.L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*. McGrawhill, New-York.
- Kampen, J.K. (2007), *CoSBI Methodologie voor terugleggen en backcasten*. DMK-2007-05-04-JKPN, Centraal Bureau voor de Statistiek, Heerlen.
- Koopman S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (1999a), *Structural Time Series Analysis, Modelling and Prediction using STAMP*. Timberlake Consultants Press, London.
- Koopman, S.J., Shephard, N. and Doornik, J.A. (1999b), Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* 2, 113-166.
- Koopman, S.J., Shephard, N. and Doornik, J.A. (2008), *SsfPack 3.0: Statistical algorithms for models in state space form*. Timberlake Consultants Press, London.
- Montgomery, D.C. (1997), *Design and Analysis of Experiments* (5e editie). Wiley, New-York.
- Pfeffermann, D., and Burck, L. (1990), Robust small area estimation combining time series and cross-sectional data. *Survey Methodology* 16, 217-237.
- Rao, J.N.K. (2003), *Small Area Estimation*, Wiley, Hoboken, NJ.
- Roels, J en van den Brakel, J.A. (2008), *Analyse methodebreuken ten gevolge van de herziening van de enquête Maatschappelijke Participatie en Milieu*. DMK-2008-04-11-JRLS. CBS-nota, Heerlen.
- Särndal, C.-E., Swensson, B. en Wretman, J, (1992), *Model Assisted Suvey Sampling*. Springer-Verlag, New York.