

Macro-integratie

Deelthema: Inpassen

08

Reinier Bikker, Jacco Daalmans en Nino Mushkudiani

Statistische Methoden (08003)



Verklaring van tekens

.	= gegevens ontbreken
*	= 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
2005–2006	= 2005 tot en met 2006
2005/2006	= het gemiddelde over de jaren 2005 tot en met 2006
2005/'06	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2005 en eindigend in 2006
2003/'04–2005/'06	= oogstjaar, boekjaar enz., 2003/'04 tot en met 2005/'06

In geval van afronding kan het voorkomen dat de som van de totalen afwijkt van het totaal.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Prinses Beatrixlaan 428
2273 XZ Voorburg

tweede helft van 2008:

Henri Faasdreef 312
2492 JP Den Haag

Prepress

Centraal Bureau voor de Statistiek - Facilitair bedrijf

Omslag

TelDesign, Rotterdam

Inlichtingen

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

Bestellingen

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

Internet

www.cbs.nl

ISSN: 1876-0333

Inhoudsopgave

1. Inleiding op het thema.....	4
2. RAS / IPF	7
3. De methode van Stone.....	12
4. Uitbreidingen van Stone.....	20
5. Univariate Denton methode.....	25
6. Multivariabele Dentonmethode.....	30
Literatuur.....	36

1. Inleiding op het thema

1.1 Algemene beschrijving en leeswijzer

1.1.1 Beschrijving van het thema

Macro-integratie heeft twee doelen. Ten eerste wordt economische informatie samengebracht in een rekeningenstelsel, waardoor de data in onderlinge samenhang kan worden geanalyseerd. Ten tweede wordt door het onderling confronteren van statistische informatie in zo'n stelsel een nauwkeuriger schatting gemaakt van de economische werkelijkheid. Een economisch rekeningenstelsel wordt gedefinieerd door een verzameling variabelen en een verzameling relaties tussen deze variabelen. De variabelen zijn gebaseerd op een verzameling samenhangende definities van de economische grootheden. De relaties kunnen de vorm hebben van lineaire gelijkheidsrestricties, of de vorm van ratiorestricties.

Macro-integratie houdt in dat de brondata eerst moet worden "vertaald" naar de juiste definities. Vervolgens moeten grote meetfouten worden opgespoord en gecorrigeerd. In dit proces worden de variabelen, middels hun relaties, onderling geconfronteerd. De data hebben in het algemeen een zeer verschillende bron. Variabelen kunnen zijn waargenomen in externe registraties of in steekproeven, maar bij gebrek aan bron kunnen andere variabelen ook zijn gebaseerd op modelschattingen, of "expert guesses".

In het algemeen moet tijdens het inpassen zowel vertekening als steekproefruis worden opgespoord en gereduceerd. De formele methoden voor het inpassen die hier worden beschreven richten zich met name op data met ruis. Voor het detecteren en verwijderen van vertekening zijn deze methoden niet geschikt. Het onderscheid tussen vertekening en steekproefruis is in de praktijk moeilijk te maken. Meestal komt het er op neer dat de grootste verschillen met hand worden weggewerkt, en dat (de grote hoeveelheid) kleinere verschillen modelmatig kunnen worden weggesmeerd.

1.1.2 Problemen en oplossingen

Er zijn verschillende formele inpasmethoden bekend in de literatuur. Iedere methode heeft zijn eigen ontstaansgeschiedenis en er is daarom variatie in toepasbaarheid, uitlegbaarheid, en algemeenheid. Alle hier genoemde methoden kunnen worden geïmplementeerd op de computer.

De eenvoudigste methoden stammen uit een tijd waar nog niet van de rekenkracht van computers gebruik gemaakt kon worden. Een voorbeeld daarvan is de *RAS-methode*, waarmee het binnenwerk van een rechthoekige matrix kan worden aangesloten op een set randtotalen. Deze methode wordt beschreven in hoofdstuk 2.

Meer algemeen, en statistisch beter gefundeerd, zijn methoden die, op een nader te definiëren manier, ingepaste resultaten schatten uit brongegevens, zodanig dat aan bepaalde restricties wordt voldaan. Vrijwel alle bekende methoden vallen in de klasse van *gegeneraliseerde kleinste kwadraten methoden*. Kenmerkend hiervoor is dat een doelfunctie wordt geformuleerd die kwadratisch is in de storingstermen. De beste statistische schatting correspondeert met de optimale waarde van de doelfunctie. Verschillende aanvullende aannamen leiden tot specifieke modellen.

De aanname dat de variabelen onderling onafhankelijk zijn en dat alle lineaire restricties exact kloppen leidt tot de *methode van Stone*, één van de oudere (1942) en meest basale kleinste kwadratenmethoden. Deze wordt beschreven in hoofdstuk 3. Verschillende uitbreidingen zijn mogelijk op de methode Stone en twee belangrijke daarvan worden beschreven in hoofdstuk 4: modellen met “zachte” restricties en modellen met relaties in de vorm van ratio’s.

Inpassen kan ook worden uitgebreid met een tijdscomponent. Dat betekent dat data uit verschillende perioden simultaan worden ingepast. Daarbij mogen data van verschillende frequentie voorkomen, bijvoorbeeld kwartaal- en jaardata. Relaties kunnen voorkomen binnen één bepaalde tijdsperiode, maar ook tussen verschillende perioden, bijv dat de som van vier kwartalen optelt tot een jaar. Tussen variabelen die bijvoorbeeld op kwartaal- en op jaarbasis worden waargenomen kunnen op dezelfde wijze inconsistenties bestaan, als tussen variabelen van één periode onderling. De relaties tussen variabelen met een verschillende frequentie zijn lineaire gelijkheidsrestricties.

In de literatuur zijn twee veelgebruikte termen te vinden. Als het gaat om het wegsmeren van kleine verschillen, vaak ontstaan door statistische manipulaties, spreken we meestal over *benchmarking*. Waar het gaat om methoden waarmee hoogfrequente data wordt afgeleid uit laagfrequente data met behulp van indicatoren spreken we over *temporal disaggregation* methoden. De methoden die in beide gevallen toegepast kunnen worden zijn echter identiek, het verschil zit eigenlijk alleen in de eigenschappen van de hoogfrequente indicatoren.

In deze toepassingsgebieden wordt onder meer de Chow-Lin methode (1971) toegepast. Hierbij wordt een kleinste kwadraten (regressie)schatter toegepast om een hoogfrequente reeks (bijvoorbeeld een kwartaalreeks) te schatten uit een of meer gerelateerde reeksen, zodanig dat de hoogfrequente reeks aansluit op een laagfrequente reeks (bijv. jaarcijfers). Deze methode vindt de oplossing die het dichtst bij de verzameling indicatoren ligt. Als de regressierelaties echter snel veranderen in de tijd is de methode minder goed toepasbaar. De methode zal dan zogenaamde stapproblemen creëren: de verandering van het vierde kwartaal van het ene jaar op het eerste kwartaal van het daaropvolgende jaar buitenproportioneel groot is. De Chow-Lin methode wordt voorzover bekend niet op het CBS toegepast, en daarom in dit overzicht verder niet besproken.

In hoofdstuk 5 wordt de univariate *Denton methode* besproken. Deze methode sluit één hoogfrequente reeks variabelen aan op één laagfrequente reeks, zonder

stapproblemen te creëren. De laagfrequente reeksen zijn exogeen, wat wil zeggen dat deze niet worden aangepast.

De Denton methode tracht de trend van de hoogfrequente reeks zo goed mogelijk te behouden, dus niet de niveaus. De eerste verschillen (het verschil tussen twee opeenvolgende perioden) van de hoogfrequente indicator vormen de enige informatiebron over de korte-termijn bewegingen van de te schatten reeks, terwijl de niveaus en de lange-termijn trend van de laagfrequente reeks betrouwbaarder worden verondersteld dan die van de hoogfrequente indicator.

De Denton methode combineert daarom de informatie in de eerste verschillen van de hoogfrequente datareeks met de informatie over het niveau en de lange termijn trend uit de laagfrequente reeks en schat uit deze informatie een nieuwe hoogfrequente reeks, die aansluit bij de laagfrequente reeks.

Een variant van deze methode is bekend als Cholette matrix methode. Deze laatste wordt op het CBS gebruikt bij het benchmarken van de economische hoofdreeksen na seizoencorrectie.

Hoofdstuk 6 bespreekt een basale vorm van de *multivariate Denton methode*. Met deze methode kan een reeks hoogfrequente *stelsels* worden aangesloten op een reeks laagfrequente stelsels. In dit geval zijn er lineaire restricties binnen één periode en lineaire restricties tussen de verschillende frequenties. De multivariate Denton methode kan ook worden uitgebreid met zachte restricties en ratiorelaties. Een beschrijving van dit uitgebreide model wordt eind 2007 toegevoegd aan deze methodenreeks.

1.2 Afbakening en relatie met andere thema's

Van alle beschikbare methoden hebben wij gekozen voor de meest algemene en basale versie. Daarnaast beschrijven we enkele algemene uitbreidingsmogelijkheden. Waar mogelijk wordt verwezen naar de literatuur, in plaats van de materie hier te herkauwen.

Er zijn geen directe relaties met andere thema's in deze Methodenreeks.

1.3 Plaats in het statistisch proces

De inpasmethodes die hier worden beschreven zijn primair ontworpen om te worden toegepast bij de integratie van waargenomen economische grootheden in rekeningenstelsels. Het inpasproces vindt vrijwel achteraan plaats in de productieketen van de economische statistieken.

Problemen waarbij dezelfde of vergelijkbare technieken toepasbaar zijn kunnen zich echter ook op andere plaatsen voordoen.

2. RAS / IPF

2.1 Korte beschrijving

De methode RAS is voor het eerst in de econometrie toegepast door Stone. De RAS methode staat ook bekend als de IPF-methode (iterative proportional fitting), of biproportional fitting in het twee-dimensionale geval.

Het uitgangspunt is een matrix waarvan het binnenwerk niet consistent is met de randtotalen. De uitkomst is een nieuw binnenwerk dat wél consistent is met de randtotalen. De randtotalen worden niet aangepast, het binnenwerk wél. De structuur van dit binnenwerk blijft zoveel mogelijk behouden. Dat wil zeggen dat relatief kleine waarden relatief klein blijven, en vice versa.

Het algoritme is een relatief eenvoudige iteratieve procedure: beurtelings worden de rijen en kolommen kloppend gemaakt. Voor een meer specifieke beschrijving wordt verwezen naar paragraaf 2.3.

2.2 Toepasbaarheid

De RAS-methode is geschikt voor problemen waarin het doel is om cijfers in een rechthoekige matrix consistent te krijgen met randtotalen. Deze methode is alleen interessant als de data in één rechthoekige matrix kan worden gerepresenteerd, met harde randtotalen en een zacht binnenwerk. De tabel hoeft niet vierkant te zijn: het aantal rijen mag dus anders zijn dan het aantal kolommen.

Men kan geen onderscheid aanbrengen in hardheid tussen variabelen in het binnenwerk. Het is dus niet mogelijk om af te dwingen dat een bepaalde waarde uit het binnenwerk nauwelijks wordt aangepast. De aanpassing van het binnenwerk is namelijk altijd (multi)proportioneel aan de randtotalen, zodat de structuur van de tabel zo veel mogelijk wordt behouden.

Verder is het niet mogelijk om andere restricties op te leggen dan de randtotalen. Men kan dus niet afdwingen dat twee variabelen uit het binnenwerk optellen tot een derde variabele uit het binnenwerk, zelfs niet als hier in de uitgangssituatie wél aan is voldaan.

De RAS-methode is zeer geschikt bij bijvoorbeeld het updaten van de zogenaamde input-output tabellen van Nationale Rekeningen. Op het huidige tijdstip, zeg t , liggen de randtotalen vast. Deze randtotalen moeten namelijk consistent zijn met een andere statistiek: de aanbod- en gebruiktabel. Cijfers over het binnenwerk zijn slechts beschikbaar tot en met $t - 1$. Updaten houdt in dat het binnenwerk uit $t - 1$ zó wordt aangepast, dat dit consistent wordt met de randtotalen van tijdstip t en dat tevens de structuur van $t - 1$ zo goed mogelijk blijft behouden.

Een technische beschrijving van de voorwaarden van de methode staat in Bacharach (1970). Om de technische details te vermijden, wordt hieronder een globale beschrijving van de voorwaarden gegeven:

- 1) Voor iedere variabele (binnenwerk en randtotalen) zijn beginschattingen beschikbaar¹;
- 2) De broncijfers kunnen in één tabel worden gerepresenteerd;
- 3) Alle waarden zijn niet-negatief;
- 4) Rijen en kolommen met alleen nullen komen niet voor in combinatie met een rij- of kolomtotaal dat ongelijk aan nul is;
- 5) De som van de rij- en kolomtotalen is hetzelfde.

2.3 Uitgebreide beschrijving

Deze paragraaf geeft een omschrijving van het algoritme. Voor een toelichting wordt verwezen naar Bacharach (1970). De kern is dat rijen en de kolommen beurteling proportioneel worden opgehoogd tot de randtotalen.

De uitgangspositie is een $n \times m$ matrix A , en vereiste rij- en kolomtotalen: r_i en k_j .

Stap 0 Initialiseer: $t = 0$ en $A^{(t)} = A$;

Stap 1 Bereken voor iedere rij i de correctiefactor

$$c_i^{(t)} = \frac{r_i}{\sum_{j=1}^m A_{ij}^{(t)}}. \quad (2.1)$$

Deze correctiefactor is de verhouding tussen het vereiste rijtotaal en het huidige rijtotaal. Vermenigvuldig ieder binnenwerkelement in rij i met deze correctiefactor. Dit geeft een matrix $A^{(t+1)}$, met de eigenschap dat de rijen optellen tot de vereiste rijtotalen. Stop, indien ook alle kolomtotalen optellen tot de vereiste kolomtotalen (of als de discrepantie acceptabel klein is). Ga anders verder met stap 2 en neem $t = t + 1$.

Stap 2 Bereken voor iedere kolom j de correctiefactor

$$c_j^{(t)} = \frac{k_j}{\sum_{i=1}^n A_{ij}^{(t)}}. \quad (2.2)$$

Deze correctiefactor is de verhouding tussen het vereiste kolomtotaal en het huidige kolomtotaal. Vermenigvuldig ieder element in kolom j met deze correctiefactor. Dit geeft een matrix $A^{(t+1)}$, met de eigenschap dat de kolommen optellen tot de vereiste

¹ Als geen informatie beschikbaar is voor het binnenwerk van de tabel kan toch aan voorwaarde 1) worden voldaan door het binnenwerk te vullen met enen. De ingepaste cijfers zijn dan biproportioneel aan de randtotalen.

kolomtotalen. Stop, indien ook alle rijtotalen optellen tot de vereiste totalen (of als de discrepantie acceptabel klein is). Ga anders terug naar stap 1 en neem $t = t + 1$.

In stap 1 van het algoritme worden de rijen consistent gemaakt en in stap 2 wordt hetzelfde gedaan voor de kolommen. Deze keuze is willekeurig, evengoed kunnen stap 2 en stap 1 worden omgewisseld.

Dit algoritme convergeert altijd naar een oplossing, gegeven dat aan alle voorwaarden wordt voldaan, die worden genoemd in Bacharach (1970). Deze voorwaarden lijken erg op de voorwaarden 1) tot en met 5) hierboven, maar ze zijn niet precies hetzelfde. Om technische details te vermijden, worden de precieze voorwaarden niet in deze nota niet beschreven.

2.4 Voorbeeld

In tabel 1 hieronder is het binnenwerk niet consistent met de bijbehorende rij- en kolomtotalen. De RAS methode wordt toegepast om een consistente tabel te krijgen.

Tabel 1. Uitgangssituatie

2	4	12
2	4	6
9	9	18

Als eerste worden de rijen aangepast. Voor de eerste rij geldt dat het binnenwerk optelt tot zes en dat het randtotaal twee keer zo groot is. De elementen in het binnenwerk moeten derhalve met twee worden vermenigvuldigd. De tweede rij van het binnenwerk is wél consistent met het rijtotaal en wordt daarom niet aangepast. Het resultaat van het bovenstaande is tabel 2. De rijen van het binnenwerk zijn consistent met de rijtotalen, maar niet met de kolomtotalen.

Tabel 2. Eerste tussenresultaat

4	8	12
2	4	6
9	9	18

In de volgende stap van het algoritme worden de kolommen aangepast. De som van de eerste kolom is zes en het kolomtotaal is negen. De binnenwerkelementen op de eerste kolom moeten daarom worden verhoogd met een factor $9/6$, oftewel 1.5. Voor de tweede kolom geldt dat het binnenwerk optelt tot 12, terwijl het kolomtotaal 9 is. De twee elementen in het binnenwerk moeten dus worden vermenigvuldigd met $9/12$, oftewel 0.75. Dit resulteert in tabel 3. Deze tabel is geheel consistent. Het algoritme stopt. (het algoritme zou verder gaan met rijen, indien de tabel niet consistent zou zijn)

Tabel 3. Eindresultaat

6	6	12
3	3	6
9	9	18

2.5 Eigenschappen

Alle ingepaste waarden zijn niet-negatief. Een nul in het binnenwerk wordt niet aangepast. Omgekeerd kan het wel voorkomen dat een positieve waarde in het binnenwerk wordt veranderd in een nul (alleen als het rij- of kolomtotaal gelijk is aan nul).

De uitkomst van de RAS-methode toegepast op een matrix A is hetzelfde als de uitkomst van de volgende optimaliseringsprobleem:

$$\text{Min}_{A^*} \sum_{i=1}^n \sum_{j=1}^m A_{ij}^* (\ln A_{ij}^* - \ln A_{ij}), \quad (2.3)$$

$$\text{ovd} \sum_{i=1}^n A_{ij}^* = k_j, \quad j = 1, \dots, m \quad (2.4)$$

$$\sum_{j=1}^m A_{ij}^* = r_i, \quad i = 1, \dots, n \quad (2.5)$$

waar n het aantal rijen weergeeft, m het aantal kolommen, k_j een kolomtotaal en r_i een rijtotaal.

De multiplicatoren c_i en c_j in (2.1) en (2.2) hebben, afhankelijk van de praktijktoepassing, een economische interpretatie, bijvoorbeeld een substitutie-effect en een effect dat weergeeft hoeveel een bedrijfsklasse als geheel meer produceert. Meer hierover staat in Bacharach (1970).

2.6 Kwaliteitsindicatoren

De belangrijkste kwaliteitsindicator is *hoe* de cijfers zijn aangepast. Er kan hierbij naar relatieve of absolute verschillen worden gekeken. Vanwege de relaties tussen de verschillende elementen in matrix A , moeten deze verschillen in samenhang worden bekeken. Men kan bijvoorbeeld bekijken in welke mate de verhoudingen tussen binnenwerkcellen en de randtotalen veranderen bij het inpassen.

De RAS-methode tracht om deze verhoudingen zoveel mogelijk te behouden. Als een verhouding toch veel moet worden gewijzigd bij het inpassen is het raadzaam om te heroverwegen om RAS te gebruiken.

Speciale aandacht moet uitgaan naar nullen, zowel vóór als na-inpassen. De RAS-methode kan nullen creëren als een randtotaal nul is, maar kan bestaande nullen in het binnenwerk niet aanpassen. In beide gevallen moet gecontroleerd worden of de structuur van de dataset correct is.

3. De methode van Stone

3.1 Korte beschrijving

De methode van Stone past een verzameling van cijfers zo weinig mogelijk aan, zodanig dat aan bepaalde restricties wordt voldaan. In tegenstelling tot bij RAS hoeft de presentatie van deze cijfers niet in een rechthoekige tabel te zijn.

De uitkomsten van de methode worden gestuurd door betrouwbaarheidsgewichten (varianties). Deze gewichten kunnen zó worden opgesteld dat de meest betrouwbaar geachte cijfers het minst worden aangepast. De methode geeft betrouwbaarheids-gewichten van de consistente waarden als output.

Wiskundig gezien is het probleem een gewogen kwadratisch optimaliserings-probleem onder lineaire voorwaarden. Voor de oplossing van dit probleem bestaan analytische uitdrukkingen. De Lagrange Multiplier methode kan worden gebruikt voor het afleiden van de oplossing.

3.2 Toepasbaarheid

De methode is interessant in een situatie waarin een verzameling van cijfers zo weinig mogelijk moet worden aangepast, zó dat aan bepaalde restricties wordt voldaan. Een uiterst eenvoudig voorbeeld van een restrictie is dat de waarde van één variabele gelijk moet zijn aan de waarde van een andere variabele.

Er mogen zowel positieve als negatieve waarden zijn. Men kan echter niet afdwingen dat positieve cijfers positief blijven en negatieve waarden negatief.

De uitkomsten van de inpasmethode kunnen worden gestuurd door verschillende betrouwbaarheidsgewichten te gebruiken, maar dit hoeft niet. Verder staat de methode exogene variabelen toe, oftewel waarden die niet mogen worden aangepast.

De voorwaarden zijn:

- 1) Beschikbaarheid cijfers. Broncijfers en een bijbehorende relatieve (co)variantiematrix zijn beschikbaar;
- 2) Onvertekende broncijfers. Dit wil zeggen dat alle broncijfers consistent zijn met de definities. Zij zijn dus al gecorrigeerd voor fouten (non-responsfouten, meetfouten, bewerkingsfouten en conceptuele verschillen). De wiskundige methode is namelijk niet bedoeld voor de correctie van fouten. Bij het gebruik van invoerdata met fouten zullen deze fouten ook terugkomen in de resultaten;
- 3) Alleen gelijkheidsrestricties. Ongelijkheidsrestricties zoals: “omzet > 100*aantal werkzame personen” zijn dus niet mogelijk. Een speciaal geval van een ongelijkheidsrestrictie is de niet-negativiteitsrestrictie. Deze kan dus

ook niet worden gemodelleerd. Verder spreekt het voor zich dat de restricties onderling niet strijdig mogen zijn;

- 4) Alleen lineaire restricties. Niet-lineaire restricties kunnen dus niet worden gemodelleerd. Een voorbeeld van een niet-lineaire restrictie is een ratio, zoals “waarde / volume = prijs”. In hoofdstuk 4 staat een uitbreiding van deze methode die wel een bepaald type van niet-lineaire restricties toestaat (ratio’s).

3.3 Uitgebreide beschrijving

3.3.1 Methode

Aangezien het artikel van Stone (1942) erg technisch is, wordt de lezer die nog niet erg vertrouwd is met de methode, verwezen naar de bijlage van Wroe e.a. (1999). Een wiskundige afleiding van de resultaten staat in Sefton en Weale (1995).

Zoals in paragraaf 3.1 is beschreven is het probleem van Stone een kwadratisch optimaliseringsprobleem. De wiskundige formulering is

$$\text{Min}_{x^*} (x^* - x)' V^{-1} (x^* - x), \quad (3.1)$$

$$\text{Ovd} \quad Ax^* = b. \quad (3.2)$$

Hierin is x een n -dimensionale kolomvector die de data omvat. Het maakt niet uit hoe de data worden geordend binnen deze vector. De oplossing x^* is tevens een n -dimensionale kolomvector. In de doelfunctie (3.1) wordt het gewogen kwadratische verschil tussen x^* en x geminimaliseerd. Hierbij wordt gewogen met een matrix V^{-1} . Dit is de inverse van de relatieve covariantiematrix. Meer over deze matrix staat in paragraaf 3.3.2. De randvoorwaarde (3.2) dwingt af dat x^* voldoet aan een stelsel van lineaire restricties. De matrix A is de coëfficiëntenmatrix van dit stelsel en de vector b omvat de coëfficiënten aan de rechterkant van de restricties.

In het speciale geval van een diagonale covariantiematrix (alle x_i onafhankelijk), is (3.1) equivalent met:

$$\text{Min}_{x^*} \sum_{i=1}^n \frac{1}{v_{ii}} (x_i^* - x_i)^2, \quad (3.3)$$

waarbij v_{ii} de variantie weergeeft van x_i .

De uitdrukkingen van de oplossingen zijn

$$x^* = x + VA'(AVA')^{-1}(b - Ax), \quad (3.4)$$

$$V^* = V - VA'(AVA')^{-1}AV. \quad (3.5)$$

De matrix V^* is de covariantiematrix van x^* . De diagonaalelementen van deze matrix hebben de eigenschap dat ze niet groter zijn dan de overeenkomstige elementen in V .

In andere woorden: door inpassen neemt de betrouwbaarheid toe. De inverse in (3.4) kan alleen worden bepaald als de restricties onderling onafhankelijk zijn. Een coëfficiëntenmatrix van een stelsel onafhankelijke restricties kan worden verkregen door A te vegen² en de rijen met nullen weg te laten. Een alternatief voor dit vegen is om de Moore-Penrose inverse te gebruiken, zie hiervoor Knottnerus (2003, p.331).

In de Nationale Rekeningen wordt Stone vaak toegepast op zeer grote matrices. Het inverteren van een matrix in formule (3.4) kan dan rekentechnisch een probleem zijn. Bij matrices met veel nullen is het aan te raden om sparse matrices te gebruiken. Knottnerus (2003) heeft een iteratieve Kalman methode beschreven, die in theorie geschikt is voor het inverteren van zeer grote matrices.

3.3.2 De covariantiematrix

De matrix V in (3.4) omvat relatieve (co)varianties. De verhoudingen van de getallen in deze matrix bepalen welke waarden het meest worden aangepast.

Een logische vraag is hoe de matrix wordt bepaald. Idealiter staan echte (co)-varianties in deze matrix. In de praktijk zijn deze echter vaak niet beschikbaar. In praktijktoepassingen worden doorgaans schattingen van relatieve varianties gebruikt. Relatieve varianties hebben op zichzelf geen betekenis. De verhouding van relatieve varianties geeft aan hoe betrouwbaar cijfers zijn ten opzichte van elkaar.

Er zijn verschillende (ad-hoc) methoden om tot schattingen van relatieve varianties te komen.

Eén methode is om een specialist 95%-betrouwbaarheidsintervallen te laten schatten en de grootte van deze intervallen te gebruiken als benadering van varianties.

Een andere methode is om enkele categorieën onderscheiden, bijvoorbeeld: relatief onbetrouwbaar, gemiddeld betrouwbaar en relatief betrouwbaar. Binnen iedere categorie krijgen alle variabelen dezelfde gewichten (θ in (3.6) hieronder).

In de praktijk komt het vaak voor dat het gewenst is dat op grote waarden in absolute zin méér wordt ingepast dan op kleine waarden. Als dit de bedoeling is kan worden gekozen voor de onderstaande varianties:

$$Var(x_i) = \theta_i^2 x_i^2, \tag{3.6}$$

waar θ_i een parameter is die afhangt van de (categorie van) betrouwbaarheid van x_i .

Het bepalen van de juiste verhoudingen tussen de verschillende varianties is in de praktijk een ‘trial en error’ proces. Dat betekent dat op basis van enige voorkennis en eenvoudige veronderstellingen (bijv. dat varianties gelijk zijn bij gebrek aan voorkennis) voor één bepaalde verhouding wordt gekozen, en vervolgens wordt

² Vegen is een andere term voor Gauss-eliminatie (Gaussian elimination in het Engels). Voor een beschrijving van deze techniek wordt de lezer verwezen naar internet (bijv. Wikipedia) of diverse (basis) boeken over lineaire algebra.

gekeken of de resultaten naar wens zijn. Indien dit niet het geval is, worden de varianties aangepast.

In de praktijk wordt, bij gebrek aan kwantitatieve maten, doorgaans verondersteld dat alle covarianties nul zijn, oftewel dat de variabelen onderling onafhankelijk zijn.

3.3.3 Software

Voor het toepassen van de methode van Stone is *Winadjust* beschikbaar op het CBS³ zie Van Dalen (2002) voor een handleiding met details.

Winadjust is alleen geschikt voor een specifiek inpasprobleem, namelijk voor het probleem waarin het binnenwerk van een tabel consistent moet worden gemaakt met de randtotalen. In principe liggen de rij- en kolomtotalen vast en wordt alleen het binnenwerk aangepast. Door een truc toe te passen (randen naar het binnenwerk verplaatsen) is het ook mogelijk om de methode toe te passen als de rijtotalen óf de kolomtotalen niet vastliggen. Voor een beschrijving hiervan wordt de lezer verwezen naar Van Dalen (2002).

De gewichten kunnen worden ingelezen uit een bestand of automatisch worden aangemaakt. Het laatste houdt in dat voor het gewicht van een waarde x_i gekozen kan worden uit vier varianties: $|x_i|$, x_i^2 , $0.1|x_i|$ en $0.1 x_i^2$.

Het programma biedt de optie dat het programma stopt in het geval het teken van een waarde verandert (positief wordt negatief of andersom).

Behalve voor het inpassen kan Winadjust ook worden gebruikt om de resultaten consistent af te ronden op geheeltallige waarden. Hiermee wordt bedoeld dat ook ná afronding binnenwerk en randtotalen consistent zijn.

De methode van Stone is ook relatief eenvoudig te programmeren in bijvoorbeeld Matlab. Daarnaast kunnen ook zogenaamde QP-solvers worden gebruikt, zoals CPLEX en XPRESS (QP staat voor kwadratisch programmeren).

3.4 Voorbeeld

Dit voorbeeld gaat uit van sterk vereenvoudigde aanbod- en gebruiktabellen, die worden weergegeven in tabellen 4 en 5. De rijen van tabel 4 zijn gerelateerd aan het aanbod van producten en diensten en kolommen op de producerende bedrijfstakken. In tabel 5 staat in de eerste twee rijen de vraag van producten en diensten en in de eerste twee kolommen de afnemende bedrijfstakken. Het grote totaal van de gehele tabel is leeg, aangezien ervoor gekozen is om dit totaal niet op te nemen in het wiskundige model. Dit groot totaal kan namelijk direct worden afgeleid uit de overige totalen.

³ Winadjust kan worden gedownload op: http://intranet/TMO/proddienst/_pages/winadjust.htm

Er zijn slechts twee bedrijfstakken: industrie en dienstverlening en twee goederengroepen: industrieproducten en diensten. Verder is er sprake van een gesloten economie, dus er is geen buitenland. Ook is afgezien van productgebonden belastingen, subsidies, handels- en vervoersmarges, en alle andere bestedings-categorieën dan consumptie. De restricties zijn dat:

- Voor de industrie en de dienstverlening het totale aanbod gelijk is aan het totale gebruik (de kolomtotalen van tabel 4 zijn gelijk aan de eerste twee kolomtotalen van tabel 5);
- Voor industrieproducten en diensten het totale aanbod gelijk is aan het totale gebruik (rijtotalen in tabel 4 zijn gelijk aan de eerste twee rijtotalen van tabel 5).

Verder moet in tabellen 4 en 5 het binnenwerk vanzelfsprekend optellen tot de randtotalen.

Tabel 4.: Aanbod

	Industrie	Dienstverlening	Totaal
Industrieproducten	700	300	1000
Diensten	100	400	500
Totaal	800	700	

Tabel 5. Gebruik

	Industrie	Dienstverlening	Consumptie	Totaal
Industrieproducten	50	190	860	1100
Diensten	170	100	180	450
Lonen	450	350		800
Exploitatieoverschot	130	60		190
Totaal	800	700	1040	

In de uitgangssituaties wordt aan twee restricties niet voldaan: voor industrieproducten en diensten is het totale aanbod niet gelijk aan het totale gebruik (de rijtotalen van tabel 4 zijn niet consistent met de eerste twee rijtotalen van tabel 5). De varianties staan in tabellen 6 en 7; zij zijn willekeurig gekozen.

Tabel 6. Varianties: aanbodtabel

	Industrie	Dienstverlening	Totaal
Industrieproducten	100	1000	1100
Diensten	1000	100	1100
Totaal	1100	1100	X

Tabel 7. Varianties: gebruiktabel

	Industrie	Dienstverlening	Consumptie	Totaal
Industrieproducten	500	1000	1000	2500
Diensten	1000	1000	1000	3000
Lonen	700	700		1400
Exploitatieoverschot	1200	1200		2400
Totaal	3400	3000	2000	X

Merk op dat de randtotalen niet vastliggen, hun variantie is immers groter dan nul.

Met de methode van Stone worden ingepaste cijfers berekend. De ingepaste waarden in tabellen 8 en 9 voldoen aan alle restricties. Kleine verschillen in de rijoptellingen in tabel 9 komen alleen door afrondingsfouten. Tussen haakjes staan cijfers vóór de inpassing.

Tabel 8. Ingepaste aanbodtabel, afgerond

	Industrie	Dienstverlening	Totaal
Industrieproducten	705 (700)	318 (300)	1023 (1000)
Diensten	92 (100)	396 (400)	488 (500)
Totaal	797 (800)	714 (700)	1511 (1500)

Tabel 9. Ingepaste gebruiktabel, afgerond

	Industrie	Dienstverlening	Consumptie	Totaal
Industrieproducten	33 (50)	164 (190)	827 (860)	1023 (1100)
Diensten	179 (170)	118 (100)	191 (180)	488 (450)
Lonen	452 (450)	358 (350)		810 (800)
Exploitatieoverschot	133 (130)	74 (60)		207 (190)
Totaal	797 (800)	714 (700)	1017 (1040)	2527 (2540)

Ook is een covariantiematrix afgeleid voor de ingepaste cijfers. Deze covariantiematrix is niet diagonaal, er komen ook covarianties voor die niet gelijk zijn aan nul. De varianties staan in tabellen 10 en 11. Zij zijn kleiner dan in de uitgangssituatie. Tussen haakjes staan varianties vóór de inpassing.

Tabel 10. Varianties ingepaste aanbodtabel

	Industrie	Dienstverlening	Totaal
Industrieproducten	84 (100)	270 (1000)	280 (1100)
Diensten	277 (1000)	85 (100)	292 (1100)
Totaal	293 (1100)	289 (1100)	

Tabel 11. Varianties ingepaste gebruiktabel

	Industrie	Dienstverlening	Consumptie	Totaal
Industrieproducten	346 (500)	524 (1000)	463 (1000)	280 (2500)
Diensten	541 (1000)	523 (1000)	489 (1000)	292 (3000)
Lonen	415 (700)	420 (700)		519 (1400)
Exploitatieoverschot	575 (1200)	591 (1200)		667 (2400)
Totaal	293 (3400)	289 (3000)	563 (2000)	

3.5 Kwaliteitsindicatoren

De belangrijkste kwaliteitsindicator is *hoe* de cijfers zijn aangepast. Er kan naar relatieve of absolute verschillen worden gekeken, afhankelijk van de eigenschappen van de gewichten die in covariantiematrix V zijn gestopt. Vanwege de relaties tussen de verschillende variabelen in het stelsel, moeten deze verschillen in samenhang worden bekeken. Als het aantal variabelen en/of onderlinge relaties erg groot wordt, kan dit erg ingewikkeld worden. In dit geval is het soms eenvoudiger de verschillen vóór inpassing te analyseren, in plaats van de aanpassingen.

Een kwantitatieve norm voor de mate van inconsistentie van de data vóór inpassing is de waarde van doelfunctie (3.1). De som van alle inpascorrecties is groter naarmate de waarde van de doelfunctie hoger is. Als veel inpascorrecties nodig zijn kan dit erop wijzen dat de brongegevens vertekend zijn en dat dus niet aan de voorwaarde(n) van het model wordt voldaan. Dit laatste betekent dat de methode niet toegepast had mogen worden.

Een ander aspect van kwaliteit is nauwkeurigheid. De methode van Stone geeft ingepaste cijfers met minimale variantie, gegeven de variantie van de cijfers die worden ingepast. De ex-post covariantiematrix van de ingepaste waarden kunnen worden berekend met (3.5). De diagonaalelementen van deze matrix bevatten informatie over de relatieve betrouwbaarheid van de ingepaste resultaten. Vergelijking met de covariantiematrix die is gebruikt in de doelfunctie levert informatie op hoe de inpassing de variantie van de gegevens reduceert. De niet-diagonale elementen van de ex-post covariantiematrix leveren informatie over correlaties tussen variabelen die door het inpassen zijn geïntroduceerd.

Een belangrijke kwaliteitsindicator van de *implementatie* van de methode, is de vraag *of* de cijfers ingepast zijn. Hiervoor kunnen de resterende verschillen op alle

lineaire restricties worden berekend. In het algemeen zal numerieke ruis ervoor zorgen dat deze verschillen niet exact nul zijn. Zolang de verschillen beneden een bepaalde drempelwaarde blijven is er meestal geen probleem.

4. Uitbreidingen van Stone

4.1 Korte beschrijving

In de literatuur worden enkele uitbreidingen van de methode van Stone beschreven, zie bijvoorbeeld Magnus en anderen (2000) en Boonstra (2006). Dit hoofdstuk gaat over twee van deze uitbreidingen.

1) *Inpasproblemen met zachte restricties*

Een restrictie is zacht als er ‘ongeveer’ aan moet worden voldaan (aan harde restricties moet exact worden voldaan). Een voorbeeld is dat een kwartaalwaarde van een variabele ongeveer (in verwachting) gelijk moet zijn aan de waarde van het voorgaand jaar.

Zachte restricties kunnen worden opgenomen in het wiskundige inpasmodel. Daarbij kan rekening worden gehouden met verschillen in betrouwbaarheid.

2) *Inpasproblemen met restricties in de vorm van een ratio*

Een veronderstelling over een verhouding tussen twee variabelen kan wiskundig als een ratio worden gemodelleerd. Een voorbeeld van een ratiorestrictie is dat de verhouding tussen uitvoer en productie van melk ongeveer 0.2 is. Ratiorestricties kunnen hard of zacht zijn.

Ratio restricties kunnen worden opgenomen in het wiskundig inpasmodel. Daarbij kan weer rekeningen worden gehouden met verschillen in betrouwbaarheid.

Het doel van het inpasproces is hetzelfde als bij Stone: een verzameling van cijfers wordt zo weinig mogelijk aangepast om aan bepaalde restricties te voldoen. Wiskundig gezien blijft het probleem om een kwadratensom te minimaliseren onder lineaire restricties.

4.2 Toepasbaarheid

Deze methode is interessant voor inpasproblemen waarover veel vakinhoudelijke kennis bestaat, met name als die kennis is toegespitst op verhoudingen tussen variabelen. Denk bijvoorbeeld aan productie-verbruikverhoudingen, arbeidsinkomenquota, groeivoeten (t gedeeld $t - 1$) en prijzen (een prijs is een waardeverandering gedeeld door een volumeverandering). Deze kennis kan immers worden omgezet in bindende en niet-bindende (ratio-) restricties. Met behulp van ratio's kan dus een methode worden opgezet waarin aanbod en gebruiktabellen in lopende en constante prijzen simultaan worden geraamd.

De voorwaarden die staan in paragraaf 3.2 zijn, op één uitzondering na, ook hier weer van toepassing. Die ene uitzondering is dat restricties lineair moeten zijn. In dit

hoofdstuk wordt namelijk een inpasmethode beschreven die geschikt is voor niet lineaire ratio-restricties.

4.3 Uitgebreide beschrijving

Magnus en anderen (2000) hebben een inpasmethode beschreven voor een model waarin ook zachte restricties en ratio's kunnen worden meegenomen. Hun beschrijving is erg technisch van aard en kan daarom moeilijk leesbaar zijn. Een toelichting op de wiskunde staat in Boonstra (2006) en een beschrijving 'in woorden' staat in United Nations (2000).

Hieronder wordt een beschrijving gegeven van methoden voor het realiseren van de bovenstaande twee methodologische uitbreidingen.

4.3.1 Zachte randvoorwaarden

Een wiskundige formulering van een optimalisatieprobleem met harde en zachte voorwaarden is:

$$\text{Min}_{x^*} (x^* - x)' V^{-1} (x^* - x) + (A_1 x^* - b_1)' \Sigma^{-1} (A_1 x^* - b_1), \quad (4.1)$$

$$\text{Ovd } A_2 x^* = b_2. \quad (4.2)$$

Hierin is (4.2) een stelsel van harde restricties. De zachte restricties worden weergegeven in het stelsel

$$A_1 x^* \text{ met} \quad (4.3)$$

$$E(A_1 x^*) = b_1 \text{ en} \quad (4.4)$$

$$\text{Cov}(A_1 x^*) = \Sigma. \quad (4.5)$$

De uitdrukkingen voor de oplossing van (4.1) en (4.2) zijn

$$x^* = x + VA'(AVA' + \bar{\Sigma})^{-1}(b - Ax) \quad (4.6)$$

en

$$V^* = V - VA'(AVA' + \bar{\Sigma})^{-1} AV, \quad (4.7)$$

met

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \text{ en } b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \quad (4.8)$$

In (4.6) en (4.7) is $\bar{\Sigma}$ een vierkante matrix, met een dimensie die even groot is als het aantal restricties. Deze matrix heeft de vorm

$$\bar{\Sigma} = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}, \quad (4.9)$$

de bovenste rijen corresponderen met de zachte constraints. Zij omvatten de bijbehorende covariantiematrix Σ . De onderste rijen horen bij de harde constraints. Zij bevatten alleen nullen, aangezien alle (co)varianties van de harde constraints nul zijn.

4.3.2 Ratio's

Ratio's zijn niet-lineair, terwijl alle harde en zachte voorwaarden (4.2) en (4.3) lineair zijn. Voorwaarden in de vorm van een ratio kunnen echter toch aan het model worden toegevoegd als zij eerst worden gelineariseerd. Magnus en anderen (2000) en Boonstra (2006) laten zien dat de ratio x_1 / x_2 , met

$$E(x_1 / x_2) = r \text{ en} \quad (4.10)$$

$$\text{Var}(x_1 / x_2) = \sigma_R^2, \quad (4.11)$$

equivalent is met de zachte restrictie:

$$x_1 - r x_2, \text{ met} \quad (4.12)$$

$$E(x_1 - r x_2) = 0 \text{ en} \quad (4.13)$$

$$\text{Var}(x_1 - r x_2) = \sigma_R^2 (\sigma_{x_2}^2 + E(x_2)^2). \quad (4.14)$$

In paragraaf 4.3.1 hebben we gezien dat de zachte randvoorwaarden in het model kunnen worden opgenomen. Een veronderstelling die wordt gemaakt in de bovenstaande afleiding is dat R en x_2 onafhankelijk zijn.

Hetzelfde probleem wordt ook behandeld in Knottnerus (2003). Hoewel andere wiskundige methoden worden gebruikt, zijn de uitdrukkingen voor de oplossingen hetzelfde.

4.4 Voorbeeld

Dit voorbeeld sluit aan bij het voorbeeld uit paragraaf 3.4. Aanvullend op de voorwaarden die staan in paragraaf 3.4, zal ook een ratio worden gemodelleerd.

Deze ratio is

$$R = \frac{\text{Gebruik van industrieproducten door de industrie}}{\text{Totale gebruik van de industrie}}. \quad (4.15)$$

Verondersteld wordt dat verwachte waarde van deze ratio gelijk is aan 0.063: dit is ongeveer de waarde vóór de inpassing (50/800), zie tabel 5. De variantie van R wordt gelijk verondersteld aan 0.0001. Deze keuze is gebaseerd op een 95% onzekerheidsmarge ter grootte van 0.02. Tabel 9 laat zien dat, als de ratio niet in het

model wordt opgenomen, de waarde van de ratio ná de inpassing gelijk wordt aan 0.041 (= 33/797). Tabellen 12 en 13 laten zien hoe de resultaten veranderen als de ratio in het model wordt opgenomen. Tussen haakjes staan ingepaste waarden, die berekend zijn zonder de ratio-restrictie. Deze resultaten komen uit tabellen 8 en 9.

Tabel 12. Aanbod – ingepast.

	Industrie		Dienstverlening		Totaal	
Industrieproducten	705	(705)	320	(318)	1025	(1023)
Diensten	93	(92)	396	(396)	489	(488)
Totaal	798	(797)	716	(714)	1514	(1511)

Tabel 13. Gebruik – Ingepast

	Industrie		Dienstverlening		Consumptie		Totaal	
Industrieproducten	48	(33)	158	(164)	820	(827)	1025	(1023)
Diensten	174	(179)	121	(118)	193	(191)	489	(488)
Lonen	449	(452)	360	(358)			809	(810)
Exploitatieoverschot	128	(133)	77	(74)			205	(207)
Totaal	798	(797)	716	(714)	1014	(1017)	2528	(2527)

Door het toevoegen van één ratio veranderen bijna alle waarden. De meeste slechts een klein beetje. De waarde van het gebruik van industrieproducten door de industrie verandert het meest. Dit is logisch: de variabele komt namelijk voor in de teller van de ratio. De waarde van de ratio ná de inpassing is: 0.060 (= 48/798). Deze wijkt maar weinig af van 0.063, de verwachte waarde van de ratio. Het opnemen van de ratio in het inpasmodel heeft dus merkbaar invloed op de uitkomsten.

4.5 Eigenschappen

Methodologisch gezien is het mogelijk om ratio's en zachte voorwaarden te modelleren. Door deze mogelijkheid kan vakinhoudelijke kennis in het model worden meegenomen, die niet is gebaseerd op waarneming.

Inhoudelijk gezien kan men erover twisten of vakinhoudelijke kennis in een inpasmodel moet worden opgenomen.

Een argument om deze kennis wél in het model op te nemen is dat deze informatie betrouwbaar wordt geacht en daarom toegevoegde waarde heeft. Uitkomsten die in tegenspraak zijn met vakinhoudelijk inzicht zijn niet wenselijk.

Een reden om vakinhoudelijke kennis buiten het model te laten is dat deze dan niet meer kan worden gebruikt om de waargenomen cijfers op fouten te controleren. Indien de ingepaste cijfers sterk in tegenspraak zijn met de vakinhoudelijke kennis betekent dit meestal dat er fouten zijn in de waargenomen cijfers. Deze fouten

worden niet opgespoord als de vakinhoudelijke kennis een rol speelt bij de inpassing.

4.6 Kwaliteitsindicatoren

De belangrijkste kwaliteitsindicator is *hoe* de cijfers zijn aangepast. Er kan naar relatieve of absolute verschillen worden gekeken, afhankelijk van de eigenschappen van de gewichten die in covariantiematrix V zijn gestopt. Vanwege de relaties tussen de verschillende variabelen in het stelsel, moeten deze verschillen in samenhang worden bekeken. Als het aantal variabelen en/of onderlinge relaties erg groot wordt, kan dit erg ingewikkeld worden. In dit geval is het soms eenvoudiger de verschillen voor inpassing te analyseren, in plaats van de aanpassingen.

De uitbreidingen op de methode van Stone, die in dit hoofdstuk zijn beschreven, kunnen eenvoudig worden meegenomen bij de controle op de veranderingen. In dit geval gaat het om de ex-post waarden van de ratios, en de ex-post verschillen op zachte relaties.

Een kwantitatieve norm voor de mate van inconsistentie van de data vóór inpassing is de waarde van doelfunctie (3.1). De som van alle inpascorrecties is groter naarmate de waarde van de doelfunctie hoger is. Als veel inpascorrecties nodig zijn kan dit erop wijzen dat de brongegevens vertekend zijn en dat dus niet aan de voorwaarde(n) van het model wordt voldaan. Dit laatste betekent dat de methode niet toegepast had mogen worden.

Een ander aspect van kwaliteit is nauwkeurigheid. De methode van Stone geeft ingepaste cijfers met minimale variantie, gegeven de variantie van de cijfers die worden ingepast. De ex-post covariantiematrix van de ingepaste waarden kunnen worden berekend met (3.5). De diagonaalelementen van deze matrix bevatten informatie over de relatieve betrouwbaarheid van de ingepaste resultaten. Vergelijking met de covariantiematrix die is gebruikt in de doelfunctie levert informatie op hoe de inpassing de variantie van de gegevens reduceert. De niet-diagonale elementen van de ex-post covariantiematrix leveren informatie over correlaties tussen variabelen die door het inpassen zijn geïntroduceerd.

Een belangrijke kwaliteitsindicator van de *implementatie* van de methode, is de vraag *of* de cijfers ingepast zijn. Hiervoor kunnen de resterende verschillen op alle lineaire restricties worden berekend. In het algemeen zal numerieke ruis ervoor zorgen dat deze verschillen niet exact nul zijn. Zolang de verschillen beneden een bepaalde drempelwaarde blijven is er meestal geen probleem.

5. Univariate Denton methode

5.1 Korte beschrijving

Het probleem waar de Denton methode (Denton 1971) op wordt toegepast komt vaak voor bij economische tijdreeksen. Dit probleem is dat hoogfrequente tijdreeksen (bijv. kwartaalreeksen) aangesloten moeten worden bij laagfrequente reeksen (zoals jaarcijfers), die een andere bron hebben. Aansluiten betekent dat de aggregatierelatie in de tijd kloppend gemaakt moet worden. Dat wil bijvoorbeeld zeggen dat de cijfers van vier kwartalen optellen tot het jaarcijfer. Hieronder wordt, zonder verlies van algemeenheid, verondersteld dat een kwartaalreeks wordt aangesloten bij een jaarreeks. De jaarreeks is overigens exogeen: deze wordt niet aangepast. De kwartaalreeks wordt wel aangepast.

De Denton methode tracht om de zgn. eerste-orde verschillen van de kwartaalreeks te behouden, dus niet de niveaus. Deze veronderstelling heet ook wel het “movement preservation principle”. De achterliggende veronderstelling is dat de mutaties in de kwartaalreeks nauwkeuriger gemeten zijn dan de niveaus.

De optimale oplossing is die waar de eerste-orde verschillen over de hele periode van de reeks gemiddeld het minst worden aanpast. Dat betekent dat de waarde van een kwartaalcijfer niet alleen door het overeenkomstig jaarcijfer wordt bepaald, maar ook de de jaarcijfers voor en na het overeenkomstig jaar. Daardoor kan de methode voorkomen dat een grote sprong tussen het laatste kwartaal van een jaar en het eerste kwartaal van het volgende jaar wordt gecreëerd.

5.2 Toepasbaarheid

Het doel van de methode is om een kwartaalreeks aan te laten sluiten bij een jaarreeks en daarbij zoveel mogelijk de oorspronkelijke kwartaal-op-kwartaal mutaties te behouden (het movement preservation principle).

Een toepassing is bijvoorbeeld het aansluiten van een seizoensgevoelige kwartaalreeks bij een jaarreeks, waarbij de seizoenspatronen zo veel mogelijk behouden dienen te worden. De methode is met name interessant als het van belang is dat er geen grote sprongen gecreëerd worden tussen het laatste kwartaal van een jaar en het eerste kwartaal van het volgende jaar. Of de verschillen tussen kwartaalcijfers en jaarcijfers voortkomen uit steekproefruis of zijn ontstaan door statistische manipulaties, zoals bijvoorbeeld seizoencorrectie, maakt daarbij niet veel uit.

De voorwaarden van de methode zijn:

1. Broncijfers (jaarcijfers en kwartaalcijfers) zijn beschikbaar. Dit impliceert ook dat kwartaal-op-kwartaal mutaties bekend zijn.

2. Onvertekende broncijfers. Dit wil onder meer zeggen dat de definities van de broncijfers voor jaren en kwartalen overeenkomen met die van NR. Zij zijn dus al gecorrigeerd voor steekproeffouten, non-responsfouten, meetfouten, bewerkingsfouten en conceptuele verschillen.

5.3 Uitgebreide beschrijving

In deze sectie beschrijven wij de Denton methode voor het geval dat de kwartaaldata bekend zijn van een bron en jaardata van een andere bron (Denton 1971). De Denton methode wordt hieronder afgeleid uit de methode van Stone.

Stel dat de kwartaaldata als een kolomvector x gedefinieerd is en de jaardata als een kolomvector b :

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_m \end{pmatrix}, \quad (5.1)$$

waar $n = 4m$, dan willen wij dat het volgende geldt

$$\sum_{j=4(k-1)+1}^{4k} x_j = b_k, \quad k = 1, \dots, m. \quad (5.2)$$

In de uitgangssituatie wordt bijna nooit aan deze gelijkheden voldaan. Wij zoeken dus een vector x^* die op een bepaalde manier dicht bij de oorspronkelijke x ligt en voldoet aan (5.2). De methode van Stone minimaliseert de kwadratische verschillen tussen de oorspronkelijke vector x en de aangepaste vector x^* . Net als in sectie 3 krijgen we een convex optimaliseringsprobleem:

$$\min_x (x^* - x)' V^{-1} (x^* - x), \quad (5.3)$$

$$\text{ovd} \quad \sum_{j=4(k-1)+1}^{4k} x_j^* = b_k, \quad (5.4)$$

Hier is V een symmetrische, niet-singuliere matrix. Dit kwadratische probleem, zie ook (3.3), kan met behulp van Lagrange multipliers opgelost worden:

$$(x^* - x)' V^{-1} (x^* - x) - 2\lambda'(b - Ax^*), \quad (5.5)$$

met

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} \text{ en } A' = \begin{pmatrix} j & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & j & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & j \end{pmatrix}, j = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ en } O = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (5.6)$$

Hier is A' een $n \times m$ matrix. De oplossing van dit probleem is

$$x^* = x + VA'(AVA')^{-1}(b - Ax). \quad (5.7)$$

Wanneer V de identiteitsmatrix is, minimaliseren wij de som van de kwadratische verschillen:

$$\sum_{j=1}^n (x_j^* - x_j)^2. \quad (5.8)$$

Deze functie heeft als nadeel dat een discontinuïteit tussen het laatste kwartaal van een jaar en het eerste kwartaal van het volgende jaar kan ontstaan (het zogenaamde stapprobleem). Om deze discontinuïteit te vermijden, beschouwt Denton de kwadratische functie gebaseerd op de verschillen tussen de eerste-orde verschillen, $\Delta x_j^* - \Delta x_j$. Getracht wordt om de eerste-orde verschillen van alle opvolgende kwartalen zo goed mogelijk te behouden,

$$\sum_{j=1}^n (\Delta x_j^* - \Delta x_j)^2 \text{ met } \Delta x_j = x_j - x_{j-1}, \text{ en } \Delta x_1 = x_1. \quad (5.9)$$

Deze kwadratische functie kan in matrix vorm omgeschreven worden door V te kiezen als $V = D'D$, met

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}, \text{ een } (n \times n) \text{ matrix.} \quad (5.10)$$

Hier behoudt de Denton methode de beginwaarde x_1 en alle eerste orde mutaties. De waarde van het eerste kwartaal x_1 wordt behouden, aangezien Denton ervoor kiest te werken met een denkbeeldig nulde kwartaal dat niet wordt aangepast, dus $x_0^* = x_0$, zodat de eerste term in (5.9) gelijk is aan $x_1^* - x_1$. Het nadeel hiervan is dat het in de praktijk vaak niet wenselijk is om aan het niveau van x_1 vast te houden. Om dit probleem te vermijden gebruikt Cholette (1984) een enigszins gewijzigde matrix

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix}. \quad (5.11)$$

Merk op dat aantal rijen één kleiner is dan het aantal kolommen en in (5.7) moet daarom een gegeneraliseerde inverse worden berekend.

5.4 Voorbeeld

Ter illustratie van het model hebben wij in tabel 14 een kunstmatige dataset van twaalf kwartalen en drie jaartotalen gemaakt. De kwartaaldata zijn zo gekozen, dat ze uitgesproken veranderingen bevatten die de seizoenen volgen. Vervolgens moeten ze zodanig aangepast worden dat voor elk jaar de som van vier kwartalen precies gelijk is aan de jaartotalen: 300, 400 en 500. Verder hebben wij geen andere beperkingen opgelegd.

Tabel 14. Kwartaaldata voor en na inpassing.

	Originele data	Ingepaste data	Jaartotalen
Jaar 1	50 100 150 100	33 73 120 74	300
Jaar 2	50 100 150 100	36 96 155 113	400
Jaar 3	50 100 150 100	69 124 178 129	500

We passen de univariate Denton methode met de Cholette matrix toe. Het is hier eenvoudig om de inpassingsresultaten uit te rekenen van (5.7). Hierna hebben wij de resultaten afgerond. De kwartaal-op-kwartaal mutaties zijn zoveel mogelijk bewaard gebleven. De resultaten van het eerste jaar zijn lager door een lager jaartotaal en in het derde jaar hoger door het hogere jaartotaal.

5.5 Eigenschappen

In het geval dat er geen kwartaaldata beschikbaar zijn, zouden we kunnen veronderstellen dat alle kwartaalbedragen gelijk zijn. De Denton methode onder deze aanname wordt dan vergelijkbaar met degene voorgesteld door Boot, Feibes en Lisman (1967).

Een alternatief voor aritmetische verschillen zijn proportionele verschillen. Deze worden toegepast met het idee dat een grote originele waarde in absolute zin meer veranderd kan worden dan een kleine originele waarde. In paragraaf 6.3 wordt dit uitgelegd.

5.6 Kwaliteitsindicatoren

Net als in het vorige hoofdstuk, is ook hier de belangrijkste kwaliteitsindicator *hoe* de hoogfrequente reeks is aangepast. Daarbij zijn met name de gemaakte veranderingen in de eerste verschillen van belang, gegeven het uitgangspunt van de Dentonmethode. De grootte van deze veranderingen is van belang, maar vooral ook het verloop van de veranderingen in de tijd. Dit laatste kan meestal het snelst grafisch beoordeeld worden.

Een andere kwaliteitsindicator van de Denton methode, is de vraag hoe nauwkeurig de hoogfrequente reeks is aangesloten bij de laagfrequente reeks. Hiervoor kunnen de verschillen tussen deze reeksen worden berekend. In het algemeen zal zogenaamde numerieke ruis ervoor zorgen dat deze verschillen niet exact nul zijn. De kwaliteitsindicator zegt eigenlijk meer over implementatie van de Denton methode. Zolang de verschillen beneden een bepaalde drempelwaarde blijven is er meestal geen probleem.

6. Multivariabele Dentonmethode

6.1 Korte beschrijving

In deze sectie beschrijven wij de uitbreiding van de Denton methode naar meerdere variabelen.

Di Fonzo en Marini (2003 en 2005) hebben de Denton methode gecombineerd met de methode van Stone (1942), zie Sectie 3. Hun methode wordt gebruikt om multivariate data met elkaar in overeenstemming te brengen, die zowel restricties in de tijd als restricties tussen de variabelen in een periode omvatten. Een voorbeeld van een restrictie in de tijd is de jaaraansluiting, bijvoorbeeld dat vier kwartaalcijfers optellen tot een jaarcijfer. De jaaraansluiting wordt hier in principe voor iedere variabele gerealiseerd. Verder zijn er ook extra restricties die de relaties tussen de variabelen op verschillende tijdstippen beschrijven.

De bedoeling van de Denton methode is weer om de veranderingen in de hoogfrequente reeks zoveel mogelijk te behouden, zodanig dat aan alle restricties wordt voldaan.

Door verschillen in de kwaliteit van de data, zijn sommige variabelen betrouwbaarder dan andere. In het model van Di Fonzo en Marini kunnen deze verschillen niet worden gemodelleerd. Bikker en Buijtenhek (2006) hebben het model van Di Fonzo en Marini verder uitgebreid met betrouwbaarheids gewichten, waarmee verschillen in betrouwbaarheid wel kunnen worden gemodelleerd.

6.2 Toepasbaarheid

Het CBS heeft de uitbreiding van Di Fonzo en Marini met betrouwbaarheids gewichten toegepast op tijdreeksen van de Kwartaalsectorrekeningen (KSR) (zie Bikker en Buijtenhek 2006). Daarnaast kan de methode worden gebruikt voor de rebasing: het creëren van consistentie tussen kwartaalramingen en later gemaakte jaarramingen.

De methode die is beschreven is in principe niet bedoeld voor ratio's. Ratio's kunnen toch in het model worden opgenomen door een zogenaamde linearisatietechniek toe te passen (grotendeels analoog aan paragraaf 4.3.2). Daarmee kunnen ook kwartaalreeksen in lopende en constante prijzen simultaan worden aangesloten aan jaarreeksen.

De voorwaarden in paragraaf 5.2 zijn ook hier weer van toepassing.

6.3 Uitgebreide beschrijving

In deze paragraaf geven we een formele beschrijving van het inpassingsmodel van Di Fonzo en Marini. Net als in vorige paragraaf, nemen we aan dat kwartaaldata moet worden aangepast aan jaardata.

Hier gebruiken wij dezelfde notatie als in de vorige secties. De collectie kwartaaldata, vóór inpassing, noteren we met x_{ij} , waar i de verschillende tijdreeksen aanduidt ($i = 1, \dots, M$) en j de kwartalen ($j = 1, \dots, n$). (In hoofdstuk 5 was $M = 1$.) De corresponderende waarden na inpassing noteren we met x_{ij}^* . We zullen de oorspronkelijke en ingepaste kwartaaldata representeren als twee $(Mn \times 1)$ vectoren x en x^* , die de data bevatten van alle M variabelen voor alle n perioden, namelijk

$$x = (x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{M1}, \dots, x_{Mn})' \text{ en} \\ x^* = (x_{11}^*, \dots, x_{1n}^*, x_{21}^*, \dots, x_{2n}^*, \dots, x_{M1}^*, \dots, x_{Mn}^*)'. \quad (6.1)$$

De restricties in dit model zijn lineair. Zij omvatten zowel het aanpassen van de kwartaaldata aan de jaardata, als relaties tussen de verschillende variabelen. Het geheel aan restricties noteren we als

$$Ax^* = b, \quad (6.2)$$

waar A een $(k \times Mn)$ matrix is en b een $(k \times 1)$ vector die ook de jaarcijfers bevat. Het totale aantal restricties is k . De matrix A is een uitgebreide versie van de matrix A in vergelijking (5.6). Hier hebben wij dat $k > n/4$ en de eerste $n/4$ rijen van A bepalen de jaaraansluiting en daarom zijn deze rijen net zoals in A in (5.6). De laatste $k - n/4$ rijen van A bepalen de restricties tussen de variabelen in een periode of op verschillende tijdstippen.

Net zoals in hoofdstuk 3 is x^* ook hier weer de oplossing van het volgende kwadratische optimaliseringsprobleem:

$$\min_{x^*} (x - x^*)' \Omega^{-1} (x - x^*), \quad (6.3)$$

$$\text{onder voorwaarde dat } Ax^* = b, \quad (6.4)$$

waar Ω de covariantiematrix is van x . Di Fonzo and Marini (2005) beschouwen twee vormen voor Ω . Voor deze twee vormen krijgen wij twee modellen: een additief (additive first difference model, AFD) en een multiplicatief (multiplicative first difference model, PFD) model. De doelfunctie in (6.3) van het additieve model is

$$\text{AFD: } \sum_{i=1}^M \sum_{j=2}^n \frac{\left((x_{ij}^* - x_{ij}) - (x_{ij-1}^* - x_{ij-1}) \right)^2}{v_{ij}} = \sum_{i=1}^M \sum_{j=2}^n \frac{1}{v_{ij}} (\Delta x_{ij}^* - \Delta x_{ij})^2, \quad (6.5)$$

waar $\Delta x_{ij}^* = x_{ij}^* - x_{ij-1}^*$, $\Delta x_{ij} = x_{ij} - x_{ij-1}$ en v_{ij} is een variantie van de kwartaal-opkwartaal mutaties in de teller van (6.5). En de doelfunctie van het PFD-model is:

$$\text{PFD: } \sum_{i=1}^M \sum_{j=2}^n \frac{1}{v_{ij}} \left(\frac{x_{ij}^*}{x_{ij-1}^*} - \frac{x_{ij}}{x_{ij-1}} \right)^2 = \sum_{i=1}^M \sum_{j=2}^n \frac{1}{v_{ij}} (\Gamma x_{ij}^* - \Gamma x_{ij})^2, \quad (6.6)$$

waar $\Gamma x_{ij}^* = x_{ij}^* / x_{ij-1}^*$ en $\Gamma x_{ij} = x_{ij} / x_{ij-1}$.

In Bikker en Buijtenhek (2006) wordt uitgelegd dat de keuze tussen een additieve of een proportionele doelfunctie bepaalt hoe het verschil tussen de kwartaalsom en het jaarcijfer wordt weggewerkt.

Een additieve doelfunctie moet worden gebruikt als de kwartaal op kwartaal mutaties van de geschatte variabele zo dichtbij mogelijk bij de oorspronkelijke mutaties moet zijn. De model aanname voor dit geval is

$$\Delta x_{ij}^* - \Delta x_{ij} \sim (0, v_{ij}). \quad (6.7)$$

Als we geen representatieve indicator hebben voor de verandering die wij willen meten of de variabele zowel positief of negatief kan zijn, moet dan ook de additieve model gebruikt worden.

Voor de proportionele model nemen wij aan dat

$$\Gamma x_{ij}^* - \Gamma x_{ij} \sim (0, v_{ij}). \quad (6.8)$$

Wat betekent dat proporties tussen de opvolgende kwartalen behouden moet worden. Dit gebeurt vaak als de wens is om de seizoen patroon te behouden.

In matrix notatie kunnen wij (6.5) en (6.6) herschrijven als

$$1) \text{ AFD: } \Omega = (D'(X'V'VX)^{-1}D)^{-1}; \quad (6.7)$$

$$2) \text{ PFD: } \Omega = X'(D'(V'V)^{-1}D)^{-1}X, \quad (6.8)$$

waar D gelijk is aan $I_M \otimes D_n$, en D_n is een $n \times n$ matrix

$$D_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix}. \quad (6.9)$$

In de vorige sectie was D_n gelijk aan D . Aangezien we nu M variabelen hebben, hebben wij ook M kopieën nodig van de D_n matrix. Dit gebeurt door de kronecker vermenigvuldiging. Weer worden hier voor elke variabele alle eerste-orde mutaties en tevens één niveau, de waarde van de eerste kwartaal van de eerste jaar, behouden. In Bikker en Bikker en Buijtenhek (2006) worden alternatieven voor de D_n matrix beschouwd.

De matrix X is een diagonaalmatrix met de waarden uit x op de diagonaal. De V -matrix bevat indicatoren op de diagonaal die weergeven hoe betrouwbaar de kwartaal-op-kwartaal mutaties zijn ten opzichte van elkaar.

Het grote voordeel van de multivariabele Denton methode is dat additieve en proportionele modellen gecombineerd kunnen worden. Hiervoor wordt de matrix X

in twee diagonale matrices X_{add} en X_{prop} gefactoriseerd. Dit gebeurt als volgt: X_{add} bevat dezelfde variabelen als X voor de additieve variabelen en een '1' voor de proportionele, terwijl X_{prop} de oorspronkelijke waarde voor de proportionele variabelen behoudt en een '1' voor de additieve variabelen krijgt (zie Bikker en Buijtenhek (2006)). De matrix Ω kunnen wij dan herschrijven als:

$$\Omega = X_{\text{prop}} \left(D' (X_{\text{add}}' V' V X_{\text{add}})^{-1} D \right)^{-1} X_{\text{prop}}. \quad (6.10)$$

en als A volle rang heeft, dan bevat (6.2) geen overbodige restricties en de oplossing voor ons optimaliseringsprobleem is

$$x^* = x + \Omega A' (A \Omega A')^{-1} (b - Ax). \quad (6.11)$$

Een uitdrukking voor de covariantiematrix voor x^* is

$$V^* = V - \Omega A' (A \Omega A')^{-1} A V. \quad (6.12)$$

6.4 Voorbeeld

In tabel 15 hebben we een dataset met vier variabelen en twaalf kwartalen.

Tabel 15. Kwartaaldata vóór inpassing

	Jaar 1				Jaar 2				Jaar 3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
x_1	335	399	335	351	355	364	312	366	335	364	335	351
x_2	347	379	343	365	341	371	333	342	336	377	389	381
x_3	340	365	338	356	333	332	351	356	340	365	338	356
x_4	341	371	337	359	335	361	337	350	350	370	348	200

De kwartaaldata moeten zodanig aangepast worden dat de som van vier kwartalen precies gelijk is aan een jaartotaal, gegeven in tabel 16.

Tabel 16. Jaardata

	Jaar 1	Jaar 2	Jaar 3
x_1	1350	1300	1350
x_2	1350	1300	1350
x_3	1350	1350	1400
x_4	1350	1350	1400

Naast de restricties in de tijd, zijn er nog een aantal extra restricties waaraan voldaan moet worden, en wel:

$$x_{1t} = x_{2t}, \quad t = 1, \dots, 12, \text{ en} \quad (6.13)$$

$$x_{3t} = x_{4t}, \quad t = 1, \dots, 12. \quad (6.14)$$

Merk op dat de jaardata aan al deze restricties voldoet; dit moet ook, anders kan niet aan alle restricties tegelijk voldaan worden.

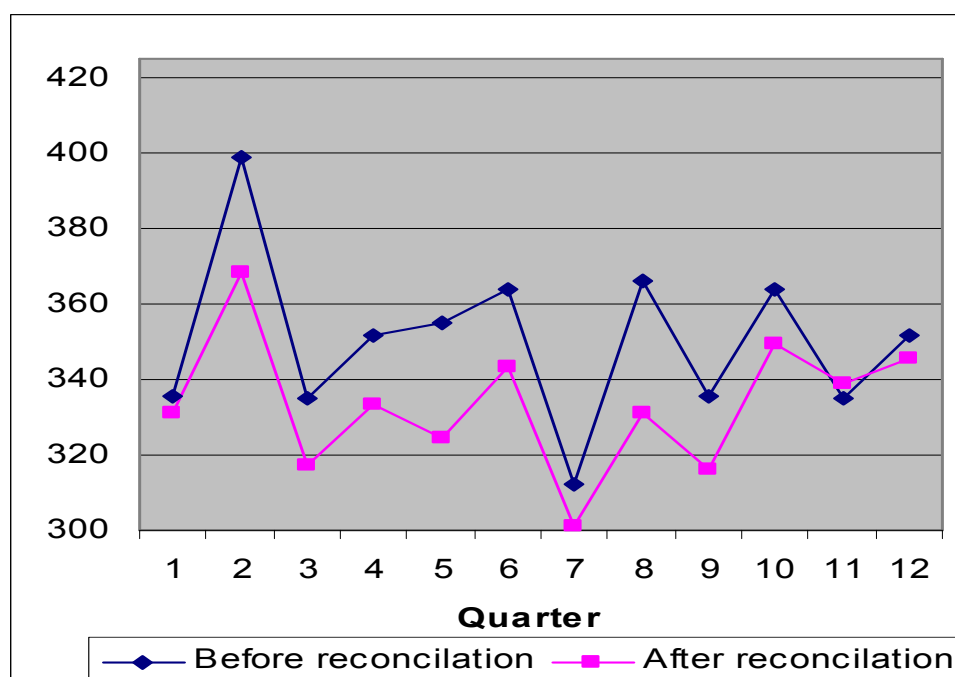
We passen de multivariate Denton methode toe. De proportionele doelfunctie wordt gebruikt voor alle variabelen. Daarbij veronderstellen we dat alle diagonaal elementen van de matrix V in (6.8) gelijk zijn aan 0.2. Dit betekent dat alle kwartaal-op-kwartaal mutaties relatief gezien even betrouwbaar zijn. De inpassingsresultaten worden gegeven in tabel 17.

Tabel 17. Ingepaste data

	Jaar 1				Jaar 2				Jaar 3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
x_1	331	369	317	333	324	343	301	331	316	349	339	346
x_2	331	369	317	333	324	343	301	331	316	349	339	346
x_3	334	355	322	339	317	332	339	362	372	402	367	259
x_4	334	355	322	339	317	332	339	362	372	402	367	259

De kwartaalcijfers van x_1 voor en na inpassing zijn weergegeven in figuur 1.

Figuur 1. Tijdreeks van x_1



De inpassing verlaagt de waarde van x_1 in alle kwartalen, behalve één. De reden hiervoor is dat de jaartotalen lager zijn dan de som van de corresponderende vier kwartaalcijfers. De originele kwartaal-op-kwartaal veranderingen zijn zoveel mogelijk bewaard gebleven. Echter, de mutatie van het vierde naar het vijfde kwartaal is opvallend anders. Vóór de inpassing is deze positief, na de inpassing negatief. De reden hiervoor is de grote daling van x_2 en de restrictie dat x_1 en x_2 gelijk moeten zijn.

6.5 Kwaliteitsindicatoren

De belangrijkste kwaliteitsindicator is weer *hoe* de hoogfrequente reeksen zijn aangepast. Gegeven het uitgangspunt van de Dentonmethode zijn daarbij met name de gemaakte veranderingen in de eerste verschillen van belang. De grootte van deze veranderingen is van belang, maar vooral ook het verloop van de veranderingen in de tijd. Dit laatste kan meestal het snelst grafisch beoordeeld worden.

Het multivariate aspect leidt er toe dat de veranderingen van verschillende reeksen in samenhang moet worden bekeken. De samenhang wordt gedefinieerd door de verzameling lineaire restricties. Als het aantal variabelen en/of onderlinge relaties erg groot wordt, kan dit erg ingewikkeld worden. In dit geval is het soms eenvoudiger de verschillen vóór inpassing te analyseren, in plaats van de aanpassingen.

Een andere belangrijke indicator wordt gevormd door de ex-post covariantiematrix. De diagonaalelementen van deze matrix bevatten informatie over de relatieve betrouwbaarheid van de ingepaste resultaten. Vergelijking met de covariantiematrix die is gebruikt in de doelfunctie levert informatie op hoe de inpassing de variantie van de gegevens reduceert. De niet-diagonale elementen van de ex-post covariantiematrix leveren informatie over correlaties tussen variabelen die door het inpassen zijn geïntroduceerd.

Een andere kwaliteitsindicator van de multivariate Denton methode, is de vraag hoe nauwkeurig de hoogfrequente reeks is aangesloten bij de laagfrequente reeks. Hiervoor kunnen de verschillen tussen deze reeksen worden berekend. In het algemeen zal zogenaamde numerieke ruis ervoor zorgen dat deze verschillen niet exact nul zijn. Zolang de verschillen beneden een bepaalde drempelwaarde blijven is er meestal geen probleem.

Literatuur

- Bacharach, M. (1970), *Biproportional matrices & input-output change*. Cambridge University Press, Cambridge.
- Bikker, R.P. and S. Buijtenhek (2006), *Alignment of Quarterly Sector Accounts to annual data*, CBS Voorburg, http://www.cbs.nl/NR/rdonlyres/D918B487-45C7-4C3C-ACD0-0E1C86E6CAFA/0/Benchmarking_QSA.pdf.
- Boonstra, H.J. (2006), *Macro-integratie Nationale Rekeningen, Een Bayesiaanse Benadering*, Nota, CBS Voorburg.
- Boot, J.C.G., W. Feibs, en J.H.C. Lisman (1967), *Further method of derivation of quarterly figures from annual data*. Applied Statistics **16** (1), 65-75.
- Cholette, P.A. (1984), *Adjusting sub-annual series to yearly benchmarks*. Survey Methodology **10** (1), 35-49.
- Chow, G.C. en A. Lin (1971), *Best Linear Unbiased Interpolation, and Extrapolation of Time Series by Related Series*. Rev. Economics and Statistics **53** (4), 372-375.
- Dalen, J. Van (2002), *Winadjust: A program for adjusting matrices to given marginal totals*. Interne Nota, CBS Voorburg.
- Denton, F.T. (1971), *Adjustment of Monthly to Quarterly Series to Annual Totals: An approach based on Quadratic Minimization*. Journal of the American Statistical Association **66** (333), 99-102.
- Harthoorn, R. en J. van Dalen (1987), *On the Adjustment of Tables With Lagrange Multipliers*, Occasional Paper NR/24, CBS Voorburg.
- Di Fonzo, T. en M. Marini (2003), *Benchmarking systems of seasonally adjusted time series according to Denton's moving preservation principle*. University of Padova, <http://www.oecd.org/dataoecd/59/19/21778574.pdf>.
- Di Fonzo, T. en M. Marini (2005), *Benchmarking a system of Time Series: Denton's movement preservation principle vs. data based procedure*. University of Padova, http://epp.eurostat.cec.eu.int/cache/ITY_PUBLIC/KS-DT-05-008/EN/KS-DT-05-008-EN.pdf.
- Knottnerus, P. (2003), *Sample Survey Theory: Some Pythagorean perspectives*. Springer-Verlag, New York.
- Magnus, J.R., J.W. van Tongeren, en A.F. de Vos (2000), *National Accounts Estimation using Indicator Ratios*, The Review of Income and Wealth **3**, 329-350, <http://center.uvt.nl/staff/magnus/paper55.pdf>.
- United Nations, Statistics Division (2000), *Handbook of National Accounting: Use of Macro Accounts in Policy Analysis*. Studies Methods, United Nations, New York.

- Sefton, J. en M.R. Weale (1995), *Reconciliation of national income and expenditure: balanced estimates for the United Kingdom, 1920-95*. Cambridge University Press, Cambridge.
- Stone, J.R.N., D.A. Champerowne en J.E. Maede (1942), *The Precision of National Income Accounting Estimates*. *Reviews of Economic Studies* **9**, 111-125.
- Van Tongeren, J.W. (1986), *Development of an Algorithm for the Compilation of National Accounts and Related Systems of Statistics*. *The Review of Income and Wealth* **32**, 25-47.
- Wroe D., P. Kenny, U. Rizki en I. Weerakoddy (1999), *Reliability and Quality Indicators for National Accounts Aggregates*. Office for National Statistics (ONS). Document CPNB 265-1 for the 33rd meeting of the GNP Committee, http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_DS_QUALITY/TAB47143266/RELIABILITY%20AND%20QUALITY%20INDICATORS%20FOR%20NATIONAL%20ACCOUNTS%20AGGREGATES.PDF.