



Discussion Paper

Criteria voor het vergelijken van reeksen prijsindexcijfers

De standpunten in dit document zijn die van de auteur(s) en komen niet noodzakelijk overeen met het beleid van Centraal Bureau voor de Statistiek

2017 | 17

**B.J.H. Lodder,
L.C.R.J. Willenborg**

Content

1. Samenvatting	3
2. Inleiding	4
2.1 Aanleiding	4
2.2 Probleemstelling	4
2.3 Aanpak en leeswijzer	4
3. Methode van onderzoek	5
3.1 Aanpak	5
3.2 Analyse van een voorbeeld	6
3.3 Tools voor berekening van modellen en kengetallen	6
4. Kwalitatieve revisieanalyse	6
4.1 Doel van de reeks en de revisie	7
4.2 Kenmerken van de data	8
4.3 Methode van dataverzameling en –analyse	8
4.4 Resultaat van de revisie	9
4.5 Vragenlijst kwalitatieve revisie	10
5. Kwantitatieve revisieanalyse	11
5.1 Inleiding	11
5.2 Reikwijdte van de revisie (scope)	11
5.3 Omvang van de revisie	11
5.4 Vertekening door de revisie	12
5.5 Mate van variatie van de revisie	12
5.6 Dynamiek van de revisie	12
5.7 Relatieve omvang van de revisie	13
5.8 Voorspelkracht van de revisie	13
5.9 Structuur van de reeksen	13
6. Procedure bij het vergelijken van reeksen	16
6.1 Inleiding	16
6.2 Toelichting per stap	16
7. Praktijkvoorbeeld: Prijsindexcijfers "Kledingzaak X"	17
7.1 Inleiding	17
7.2 Stap 1: Controle op aperte fouten	18
7.3 Stap 2: Structuur van de reeksen	18
7.4 Stap 3: Kengetallen en kwalitatieve analyse	21
8. Conclusies	24
Referenties	25
Bijlage A: Praktijkvoorbeeld "Kledingzaak X"	26
Bijlage B: Kenmerken van databronnen	32
Bijlage C: Processchema vergelijken van reeksen	34

1. Samenvatting

Het CBS heeft een methode ontwikkeld voor het berekenen van prijsindices voor de "Kledingzaak X" op basis van gegevens die met een internetrobot zijn waargenomen. Deze reeksen zijn vergeleken met reeksen die door middel van regionale waarneming zijn verkregen. Hoewel de uitkomsten een vergelijkbaar beeld laten zien, ontbreken heldere maatstaven om te beoordelen of de reeksen voldoende overeenkomen om veilig over te kunnen stappen van de ene reeks op de andere reeks. De probleemstelling waar dit stuk een antwoord op poogt te geven is: *Kunnen er algemene maatstaven of metrieken worden ontwikkeld voor het vergelijken van indexcijferreeksen op basis van verschillende bronnen en/of verschillende methoden?*

In het onderzoek hebben we ons in eerste instantie gericht op de literatuur (vanuit Eurostat) over deze thematiek. Gedurende het traject bleek dat een structurele tijdreeksanalyse mogelijk een effectievere en efficiëntere manier is om het doel te bereiken. Ook zijn kwalitatieve aspecten buitengewoon belangrijk en deze zijn eveneens onderzocht. We behandelen het thema door allereerst het theoretisch raamwerk te schetsen.

Na een inleiding (hoofdstuk 2), volgt hoofdstuk 3 met de methode van onderzoek. In hoofdstuk 4 worden de kwalitatieve aspecten behandeld. Het hoofdstuk wordt afgesloten met een vragenlijst die als checklist gebruikt kan worden. In hoofdstuk 5 worden de kwantitatieve aspecten behandeld. Allereerst de kengetallen die door Eurostat worden aanbevolen, daarna de structurele analyse die als aanvulling door ons zelf wordt voorgesteld. In hoofdstuk 6 wordt een procedure weergegeven waarbij de lengte van het totale traject mede afhangt van de complexiteit van de problematiek. In hoofdstuk 7 worden alle voorgaande elementen toegepast op een concreet voorbeeld: De prijsindexcijfers van "Kledingzaak X", zowel verkregen via regionale waarneming ("Kledingzaak X" RW) als via internet (internet). Hoewel de structurele analyse aangeeft dat er significante verschillen bestaan tussen de reeksen, moet in aanmerking genomen worden dat de kwaliteit van de "Kledingzaak X" RW reeks onvoldoende is voor een adequate vergelijking. Daarnaast is een belangrijk aspect dat de waarneming via internet veel meer waarnemingen heeft, een hogere frequentie en een meer divers assortiment van artikelen betreft. Deze kwalitatieve argumenten zijn dus in tegenspraak met de kwantitatieve structurele analyse, maar desalniettemin naar ons idee steekhoudend. Daarnaast speelt uiteraard nog een kostenoverweging een rol. Maar die is in dit onderzoek niet meegenomen, daar die niet tot de opdracht behoorde. In een conclusie worden de resultaten in grote lijnen weergegeven. In een aantal bijlagen wordt meer technische informatie gegeven over het praktijkvoorbeeld en een schema van de procedure. Er kunnen op verzoek ook tools worden geleverd waarmee snel alle benodigde analyses uitgevoerd en kengetallen berekend kunnen worden.

Keywords

CPI, tijdreeksen, prijsindices, revisie, internetwaarneming

2. Inleiding

2.1 Aanleiding

Binnen het project halvering winkelwaarneming (HWC) is een methode ontwikkeld voor het berekenen van indexcijfers voor de "Kledingzaak X" op basis van gegevens die met een internetrobot op de "Kledingzaak X"-website zijn waargenomen. "Kledingzaak X" is de aanduiding voor een specifiek concern dat in dit onderzoek is betrokken, maar wat niet met naam genoemd kan worden, omwille van de privacy. Niet alleen gaat het hier om een andere manier van waarnemen (prijswaarnemers met een beperkt aantal prijsposten, eens per maand, versus dagelijkse internetwaarneming waarbij vrijwel integraal wordt waargenomen) ook de berekening van de prijsindices is verschillend. De oude en nieuwe manier van waarnemen leveren ook andere prijsindexreeksen op voor "Kledingzaak X", en deze zijn ook weer anders dan die voor de hele COICOP kleding. Hoewel de uitkomsten een vergelijkbaar beeld laten zien, ontbreken heldere maatstaven om te beoordelen of de reeksen voldoende overeenkomen om veilig over te kunnen stappen van de ene reeks op de andere reeks¹.

2.2 Probleemstelling

De probleemstelling waar dit stuk een antwoord op poogt te geven is:

Kunnen algemene maatstaven of metrieken worden ontwikkeld voor het vergelijken van indexcijferreeksen op basis van verschillende bronnen en/of verschillende methoden?

2.3 Aanpak en leeswijzer

In dit stuk wordt getracht een aantal vergelijkingscriteria van tijdreeksen te vinden, in het bijzonder van prijzen of prijsindices, die gebruikt kunnen worden om na te gaan of de reeksen voldoende op elkaar lijken of niet. De bedoeling is niet om te komen tot criteria die blind kunnen worden toegepast en automatisch leiden tot een uitspraak of de reeksen wel, of juist niet, voldoende op elkaar lijken. Die conclusie zal een analist moeten trekken op basis van alle beschikbare informatie. De aangedragen criteria kunnen dan helpen om tot een beslissing te komen. Een deel van de informatie is overigens kwalitatief van aard, bijvoorbeeld over de wijze van waarnemen. Ook hoeven de kenmerken niet alle even zwaar te tellen. Wij gaan in dit stuk verder niet in op hoe de genoemde kenmerken te wegen. Dat is aan de experts, die er in de mo-

¹ De auteurs bedanken Sander Scholtus (CBS) voor de review van dit discussion paper.

tivatatie van hun keuze op kunnen ingaan. Bovendien kan het van geval tot geval verschillen hoe men de kenmerken weegt.

In dit stuk worden niet alleen algemene criteria besproken om prijs- en prijsindexreeksen te vergelijken. Er wordt ook een procedure voorgesteld om gegevens te verzamelen ten einde een beslissing te kunnen nemen of twee reeksen voldoende op elkaar lijken of niet.

Naast het beschrijven van algemene criteria, behandelt dit stuk ook een concreet voorbeeld van een situatie waar twee prijs(index)reeksen met elkaar vergeleken worden. Het gaat om een kwestie die op dit moment ook aan de orde is en waarover advies wordt gevraagd. Het gaat uiteindelijk om de vraag of waarneming van internetprijzen die volgens de regionale waarneming kan vervangen.

De indeling van dit stuk is als volgt. In hoofdstuk 3 wordt aangegeven welke methode gebruikt is om tot een antwoord te komen op de probleemstelling. In hoofdstuk 4 worden de kwalitatieve aspecten beschreven en in hoofdstuk 5 de kwantitatieve aspecten. De procedure voor het analyseren van reeksen wordt in hoofdstuk 6 weergegeven. Het praktijkvoorbeeld wordt behandeld in hoofdstuk 7. Tot slot worden conclusies getrokken in hoofdstuk 8. Er zijn drie bijlagen die het stuk completeren. Bijlage A geeft kengetallen voor een voorbeeld uit de kledingbranche; Bijlage B vergelijkt 3 vormen van dataverzameling voor kleding (regionaal, internet, scanner) en geeft aan hoe ze verschillen en wat ze gemeen hebben; in Bijlage C wordt een schematisch overzicht gegeven van de procedure uit hoofdstuk 6.

3. Methode van onderzoek

3.1 Aanpak

In eerste instantie is gekozen voor het gebruiken van literatuur over het reviseren van statistieken. We gebruiken hiervoor McKenzie en Gamba (2008). Dit artikel beschrijft revisies van statistieken, in het bijzonder van tijdreeksen. Het probleem dat in deze nota wordt besproken zou men als een revisieprobleem kunnen opvatten.

Gedurende het traject ontstond meer inzicht in de thematiek en vonden we een mogelijkheid om structurele tijdreeksanalyse te gebruiken. De aanleiding hiervoor is geweest dat het analysekader van McKenzie en Gamba (2008) met name beschrijvende maten geeft, terwijl in de praktijk vooral behoefte bestaat aan maatstaven die kunnen ondersteunen bij het nemen van de beslissing om wel of niet over te stappen op de latere reeks. Voor een dergelijke doelstelling zijn statistische toetsen relevant omdat ze, binnen zekere grenzen, uitsluitel geven of reeksen wel of niet van elkaar verschillen. Daarom is uiteindelijk ook besloten om de structurele tijdreeksanalyse prioriteit te geven.

Naast het kwantitatieve deel is ook veel aandacht besteedt aan factoren die niet direct in kengetallen uit te drukken zijn. Zo zijn er tal van kwalitatieve aspecten relevant bij het maken van de tijdreeksen:

- de wijze van data verzamelen,
- het soort data en de hoeveelheid data die worden verzameld,
- de fouten die bij het verzamelen kunnen optreden,
- de veronderstellingen die worden gemaakt (expliciet of stilzwijgend) ten aanzien van de data en de dataverzameling,
- de beoogde toepassing(en) van de data, bijvoorbeeld de CPI voor de kledingbranche.

Tot slot is aandacht besteed aan de prioritering. Welke zaken moeten in ieder geval gedaan worden, en in welke situaties is aanvullend onderzoek nodig?

3.2 Analyse van een voorbeeld

Aan de hand van een praktijkvoorbeeld wordt een overzicht gegeven van de kengetallen en de thematiek bij het vergelijken van reeksen (zie Bijlage A).

3.3 Tools voor berekening van modellen en kengetallen

Op verzoek van de opdrachtgever kunnen kant en klare tools worden opgeleverd waarmee op eenvoudige wijze alle noodzakelijke berekeningen worden uitgevoerd. De gebruikte programma's zijn voor methodologen bruikbaar, maar nog niet voor een eindgebruiker die geen tijd heeft om zich in technische kwesties te verdiepen.

4. Kwalitatieve revisieanalyse

De analyse van een revisie zal in eerste instantie een kwalitatief karakter hebben. In abstracte zin is een revisie het resultaat van een methode om data te gebruiken ten einde een bepaald doel te bereiken. We zullen deze aspecten apart behandelen:

- Het doel van de reeks en de revisie (waarom?)
- Wat zijn de kenmerken van de data (wat?)
- Methode van dataverzameling en -analyse (hoe?)
- Het resultaat van de revisie (waartoe?)

Deze onderwerpen worden in dit hoofdstuk verder toegelicht. Tot besluit wordt een vragenlijst gegeven die als richtsnoer kan dienen bij een kwalitatieve analyse.

4.1 Doel van de reeks en de revisie

Bij reeksen prijsindexcijfers is het doel om inzicht te krijgen in de prijzen van verkochte artikelen in een bepaalde branche, bijvoorbeeld de kledingbranche. In het concrete voorbeeld van de prijsindexcijfers bij "Kledingzaak X" is er een argument om de regionale waarneming (RW) als beter te kwalificeren dan de waarneming via internet omdat bij dit laatste aanbodprijzen worden gemeten en geen prijzen van verkochte artikelen.

Daarnaast zou men kunnen menen dat de RW-gebaseerde "Kledingzaak X"-reeksen de norm zijn waartegen de internet-reeksen voor "Kledingzaak X" worden afgezet. Het ligt voor de hand om dit te doen, want het is immers de RW-reeks die vervangen moet worden. Maar het betekent niet dat de RW-reeks per se correct is en de internet-reeks fout is als er een afwijking wordt geconstateerd ten opzichte van de RW-reeks. Men dient zich te realiseren dat de RW-reeks is gebaseerd op een betrekkelijk kleine steekproef, behept met een grote variantie en mogelijk ook nog met een bias. Over deze beide grootheden valt echter verder niet zo veel te zeggen, omdat de data waarop de RW-reeks gebaseerd is geen correct uitgevoerde steekproef is. De internetreeks is bij goede benadering nog als integraal te beschouwen (mits op een geschikt moment waargenomen). Bovendien zijn de prijsindices voor beide wijzen van dataverzameling op verschillende manieren berekend. (Zie Bijlage A).

Beter ware het als men beide reeksen kon vergelijken met de ideale, maar die is helaas niet bekend. Dan zouden eventuele manco's van beide reeksen zichtbaar worden. Het lijkt redelijk om prijzen verkregen door scannerdata dichter bij 'de ideale data' te plaatsen dan regionale waarneming of de internetwaarneming. Daarbij moeten de metadata in de scannerdata wel van voldoende detail zijn, wat niet vanzelfsprekend is. Scannerdata heeft als voordeel boven internetdata, dat het om daadwerkelijke verkochte artikelen gaat, en dat verkochte hoeveelheden artikelen er uit af te leiden zijn. Het voordeel van deze data ten opzichte van met regionale waarneming verkregen data is dat ze minder bias hebben en het om (veel) meer waarnemingen gaat. Nadeel van scannerdata voor kleding is echter wel dat de problematiek van geretourneerde artikelen ('retouren') speelt. Dit zijn verkochte artikelen die weer worden teruggezonden naar de winkel, waardoor eerder gedane verkopen feitelijk ongedaan worden gemaakt. Dit terugzenden van via internet gekochte kleding gebeurt op grote schaal. De noodzakelijke correcties op deze teniet gedane verkopen kan men pas later uitvoeren, mogelijk pas na de maand van aankoop van een dergelijk artikel.

Een doel van de revisie kan zijn dat men wil besparen op de kosten van onderzoek. Daartoe dient dan een goed beeld te worden gegeven van deze kosten, zowel van het maken van latere reeksen als van oorspronkelijke reeksen. De kosten van een eventuele migratie dienen gerekend te worden als kosten van de revisie.

Een manier om te reviseren is door dit geleidelijk te doen. De vervanging van de regionale waarneming bij de kledingbranche gebeurt niet in één keer, maar incrementeel, winkel voor winkel. Door telkens na te gaan of het effect van het vervangen van de waarneming bij één winkel wel / niet verantwoord is zou het kunnen zijn dat dat

verantwoord is omdat de veranderingen klein lijken te zijn, in vergelijking met de vorige situatie. Echter als alle veranderingen in één keer zouden worden ingevoerd kan het zijn dat dan wel veranderingen zichtbaar worden, die eerst sluipenderwijs zijn opgetreden. Als de regionale waarneming naast de internetwaarneming zou blijven bestaan voor een voldoende lange periode, zou ook het effect kunnen worden nagegaan van de toepassing van alle veranderingen in één keer in plaats van bij geleidelijke invoering.

4.2 Kenmerken van de data

Bij het analyseren van kenmerken van de data kan men denken aan zaken als:

- Welke eenheden worden waargenomen? Gaat het om micro-data en/of macro-data?
- Met welke frequentie komen data beschikbaar ?
- Wat is de populatie waarover men uitspraken wil doen?
- Over welke periode beschikt men over data, zowel de oorspronkelijke reeks als de voorgestelde latere reeks (latere reeks)? Hoeveel data ?
- Over welke periode heeft men dubbele waarneming?
- Waar wil men de reeksen voor gebruiken?
- Wat is de aard van de data die men wil vergelijken, de prijsindices of mutaties van prijsindices?
- Is er een bepaald basisjaar gebruikt in de constructie van de prijsindices.
- Hoe zijn de waargenomen artikelen geselecteerd?
- Zijn de waarnemingen gecorreleerd in de tijd?
- Kunnen de waarnemingen worden beschouwd als uitkomsten van een stochastisch proces?

Bij het vergelijken van twee prijs(index)reeksen kunnen we de reeksen sec bekijken, maar ook als onderdeel van een groter geheel. Prijsindices voor "Kledingzaak X" zijn componenten van prijsindices voor de kledingbranche en voor de hele CPI. In zo'n groter geheel (aggregaat) beschouwd kunnen verschillen die opvallen bij het bekijken van de reeksen sec, een stuk minder zijn, of zelfs geheel wegvallen. Nu zal dit snel gebeuren bij prijsindices voor één enkele winkel, omdat die een relatief bescheiden bijdrage levert aan het geheel van de kledingbranche en nog meer van de CPI als geheel.

4.3 Methode van dataverzameling en -analyse

Hoe zijn de data verzameld? Is er sprake van integrale waarneming, of van een (gestratificeerde) steekproef? Ruwweg kunnen we de wijze van dataverzameling voor kleding karakteriseren als één van de volgende methoden:

- Steekproefwaarneming door (regionale) waarnemers in winkels. De steekproef is echter geen kanssteekproef. De waargenomen prijzen betreffen die van artikelen die volgens de prijswaarnemers gangbaar zijn.
- Waarneming van aanbodprijzen op internet (internetrobot)

- Waarneming door gebruikmaking van scannerdata.

Bij scannerdata gaat het om daadwerkelijk verkochte artikelen; men meet hiermee dus marktprijzen. Nadeel bij kleding is wel dat een gedeelte van de gekochte artikelen wordt teruggebracht ('retouren'). Het verwerken van deze retouren is weliswaar mogelijk, echter zonder kennis van de verkopen die het betreft en die door de retouren teniet worden gedaan. Bij de internetwaarneming gaat het om aanbodprijzen. Bij waarneming door mensen bestaat er een gevaar van selectieve waarneming. Elke vorm van waarneming heeft zodoende voor- en nadelen. In Bijlage B worden kenmerken van de verschillende wijzen van dataverzameling in een tabel weergegeven.

Welke techniek is gebruikt bij het verzamelen van data? Gaat het om waarneming waarbij een prijswaarnemer handmatig prijzen kopieert, of betreft het scannerdata dan wel waarneming met behulp van een internetrobot (internetwaarneming)? De verwerking van de gegevens is van belang, bijvoorbeeld als de waargenomen data gefilterd zijn (omdat ze meer artikelen bevatten dan die waarin men geïnteresseerd is – zoals accessoires, schoenen, make-up e.d. naast kleding) en automatisch geclassificeerd (volgens een CBS-classificatie voor kleding).

De wijze waarop de prijs(index)reeksen berekend zijn uit de basisdata is van belang. Gaat het om artikelen die op EAN² gekoppeld worden (zoals bij de regionale waarneming) of wordt er automatisch gekoppeld op een secundaire sleutel (zoals artikelomschrijving of artikelkenmerk) aan de bovengenoemde CBS-classificatie voor kleding.

Een andere kwestie in dit verband is of men prijsindexreeksen of prijsreeksen (prijsontwikkelingen) met elkaar moet vergelijken. De prijsindexreeksen zijn uit de prijsreeksen afgeleid door in een bepaalde maand van een bepaald jaar (basis) de zaak te normeren op 1 (of 100) en van hieruit de ontwikkelingen te meten. Door te normeren knijpt men de prijsreeksen die ten grondslag liggen aan de prijsindexreeksen als het ware samen in de basis. Dit kan een geflatteerd beeld geven ten aanzien van de verschillen tussen de prijsindexreeksen. Naarmate men verder weg is van de basis kunnen de verschillen groter worden. Het kan daarom geen kwaad om naar beide typen reeksen te kijken.

4.4 Resultaat van de revisie

De oorspronkelijke reeks en de latere reeks kunnen met elkaar worden vergeleken. Dat kan al direct door visuele inspectie. De volgende vragen zijn relevant:

- Zijn er grote absolute niveauverschillen?
- Zijn er verschillen in een trendmatige ontwikkeling?
- Zijn de latere reeksen een horizontale verschuiving (vertragingseffect) van de oorspronkelijke reeks?
- Zijn er verschillen zichtbaar in seizoen patronen?

² EAN= Europese Artikel Nummering. Streepjescode voor scanners die gebruikt wordt om artikelen te identificeren en die gebruikt wordt voor kassa-afhandeling en voorraadadministratie in winkels.

- Welke verschillen zijn er in de kosten van productie van de oorspronkelijke en de latere reeks?
- Waar kan men de reeksen voor gebruiken? Wat meet de ene reeks beter / slechter dan de andere? Welke analyses kan men met de ene onderliggende data set wel maken en met de andere niet?

4.5 Vragenlijst kwalitatieve revisie

Bij het analyseren van een revisie kan men de grote lijn vasthouden door de volgende vragen te beantwoorden:

1. Wat is het doel van de reeks en de revisie?
 - Wat wil men meten met de reeks?
 - In hoeverre verschillen de oorspronkelijke reeks en de revisie van hetgeen men wil meten?
 - Wil men besparen op kosten met de revisie? Hoeveel?
2. Wat zijn de kenmerken van de data?
 - Worden er micro- en/of macro-data gebruikt? Hoeveel waarnemingen?
 - Wat is de populatie van de datareeks?
 - Wat is de periode waar het onderzoek betrekking op heeft?
 - Wat voor soort data wordt vergeleken, prijsindices of mutaties?
 - Hoe zijn de artikelen geselecteerd?
3. Hoe zijn de data verzameld?
 - Is de waarneming integraal of door middel van een (niet-probabilistische) steekproef?
 - Welke techniek wordt gebruikt (prijswaarnemer, scanner, internetrobot)?
 - Is er sprake van kunstmatige reductie van verschillen door gebruik van een basisjaar?
4. Wat is het resultaat van de revisie?
 - Verschillen in het verloop van de reeksen (niveau, trend, seizoen)?
 - Verschillen in productiekosten?
 - Welke gebruiksmogelijkheden zijn er?

In de praktijk kan een vergelijking tussen reeksen die op een geheel verschillende manier zijn waargenomen tot verkeerde conclusies leiden als bijvoorbeeld blijkt dat data verkregen via een internetrobot slecht zijn omdat de internetrobot een keer haperde, terwijl dit normaal niet voor zou komen. Daarom is het zinvol om allereerst een ruwe inschatting te maken of er geen aperte fouten zijn in de verzameling van de gegevens. Dit wordt verderop uitgewerkt in hoofdstuk 6, waarbij we een procedure weergeven voor een revisieanalyse.

5. Kwantitatieve revisieanalyse

5.1 Inleiding

We beschouwen twee tijdreeksen. de ene reeks aangeven als P_i en de andere als L_i voor $i = 1, \dots, n$. Met de eerste wordt de oorspronkelijke ('preliminary') reeks bedoeld en met de tweede de latere reeks ('later'). Zij $R_i = L_i - P_i$ de revisie, voor $i = 1, \dots, n$. Daarnaast definiëren we de mutaties van de oorspronkelijke als volgt: Mutatie $p = (P_i - P_{i-1})/P_{i-1} * 100\%$, $l = (L_i - L_{i-1})/L_{i-1} * 100\%$, en we kijken naar het verschil $r = l - p$ (dus niet naar de mutatie van R !).

Op basis van het document van McKenzie en Gamba (2008) wordt een overzicht gegeven van een aantal kengetallen die relevant zijn voor het vergelijken van reeksen. De kengetallen behoren tot de volgende thema's:

- De reikwijdte van de revisie (scope);
- De omvang van de revisie;
- De vertekening door de revisie;
- De mate van variatie van de revisie;
- De dynamiek van de revisie;
- De relatieve omvang van de revisie;
- De voorspelkracht van de revisie.

Wij voegen daar nog een thema aan toe:

- De structuur van de tijdreeks.

De maten van McKenzie & Gamba (2008+) hebben met name een beschrijvend karakter, terwijl de structurele analyse een meer toetsend karakter heeft.

5.2 Reikwijdte van de revisie (scope)

De volgende kengetallen hebben betrekking op de scope:

- De periode van analyse, bijvoorbeeld 1994Q1 – 2007Q3
- Het aantal jaren en de periode van dubbele waarneming
- N , het aantal waarnemingen (macro-data en/of micro-data)

5.3 Omvang van de revisie

De volgende kengetallen hebben betrekking op de omvang van de revisie:

Het (rekenkundige) gemiddelde van de absolute waarden van de revisies per maand:

$$MAR = \frac{1}{n} \sum_{i=1}^n |R_i|.$$

De mediaan van de absolute waarden van de revisies: $|Me| = Me|R_i|$

5.4 Vertekening door de revisie

Het is mogelijk dat de revisie $R_i = L_i - P_i$ normaal verdeeld is rondom nul, maar het is ook mogelijk dat er sprake is van een systematische vertekening, waardoor meer punten boven nul liggen of juist onder nul. De volgende kengetallen drukken de mate van vertekening uit:

- Het gemiddelde van de revisies: $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$.
- De mediaan van de revisie: $Me = Me(R_i)$.
- Percentage
 - positieve revisie: $PPos_{R_i} = 100 \frac{1}{n} \sum_{i=1}^n I_i$ met $I_i = 1$ als $R_i > 0$.
 - negatieve revisie $PNeg_{R_i} = 100 \frac{1}{n} \sum_{i=1}^n I_i$ met $I_i = 1$ als $R_i < 0$.
 - niet-nul revisie $PNN_{R_i} = 100 \frac{1}{n} \sum_{i=1}^n I_i$ met $I_i = 1$ als $R_i \neq 0$.
- De aangepaste t-waarde voor significantie van de gemiddelde revisie:
 $T = \bar{R} / \sqrt{\text{var}(\bar{R})}$, waarbij $\text{var}(\bar{R})$ gelijk is aan de HAC standaarddeviatie schatting (Heteroskedasticiteit en Autocorrelatie Consistent, zie McKenzie & Gamba, 2008+).
- De verdelingsvrije toetsgrootte voor significantie van de verschillen:
 $T = PPos_{R_i} * n = 100 \sum_{i=1}^n I_i$ m, onder H_0 heeft deze een binomiale verdeling.

5.5 Mate van variatie van de revisie

Maatstaven voor de variatie in revisie worden door de volgende kengetallen tot uitdrukking gebracht:

- Standaard deviatie van de revisie: $SDR = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \bar{R})^2}$.
- Wortel van de gemiddelde kwadratische fout van de revisie: $RMSR = \sqrt{\frac{1}{n} \sum_{i=1}^n R_i^2}$.
- Kwartielafwijking: $QDEV = Q3 - Q1$. Dit kan als maat voor de variatie dienen. Hierbij zijn Q_1 respectievelijk Q_3 het eerste en derde kwartiel. Hoe kleiner deze waarde hoe dichter de revisies bij de mediaan liggen.
- Minimum revisie.
- Maximum revisie.
- Hieruit volgt de range: $Range = Max\ revisie - Min\ revisie$.
- Scheefheid (skewness): $SKEW = \frac{3(\bar{R} - Me)}{SDR}$.

5.6 Dynamiek van de revisie

Welke ontwikkelingen in de tijd zijn zichtbaar? Vertonen de reeksen dezelfde seizoenpatronen? Daarbij kan worden gekeken naar de reeksen zelf, de eerste afgeleide (de groei) of de tweede afgeleide. (1e orde differenties: $x_i - x_{i-1}$, voor $i =$

1, ..., n) of de 'tweede afgeleide' reeks (2e orde differenties: $x_i - 2x_{i-1} + x_{i-2}$ voor $i = 1, \dots, n$). Het volgende kengetal wordt door McKenzie en Gamba genoemd: Percentage observaties waar het teken van L_i en dat van P_i gelijk zijn: Percentage [$sign(later) = sign(oorspronkelijk)$.]

5.7 Relatieve omvang van de revisie

Hoe ingrijpend een revisie is kan goed worden weergegeven door de relatieve omvang van de revisie te bepalen met de volgende kengetallen:

- Gemiddelde absolute waarde van de eerste reeks: $MAP = \frac{1}{n} \sum_{i=1}^n |P_i|$
- Relatieve gemiddelde absolute revisie: $RMAR = \frac{\sum_{i=1}^n |R_i|}{\sum_{i=1}^n |P_i|} = MAR/MAP$

5.8 Voorspelkracht van de revisie

De voorspelkracht van de 'Preliminary' reeks voor de 'Later' reeks wordt bepaald door de volgende kengetallen:

- Correlatie tussen revisie en eerdere schatting: $\rho_{R_i P_i} = \frac{\sum_{i=1}^n (P_i - \bar{P})(R_i - \bar{R})}{(n-1)\hat{\sigma}_P \hat{\sigma}_R}$ met de sigma's als SDR in 5.5 voor de R_i .
- Correlatie tussen revisie en latere schatting: $\rho_{R_i L_i} = \frac{\sum_{i=1}^n (L_i - \bar{L})(R_i - \bar{R})}{(n-1)\hat{\sigma}_L \hat{\sigma}_R}$ met de sigma's als in SDR in 5.5 voor de L_i .
- Seriële correlatie van de revisies: $\rho_{R_i R_{i-1}} = \frac{\sum_{i=1}^n (R_{i-1} - \bar{R})(R_i - \bar{R})}{(n-1)\hat{\sigma}_{R_i} \hat{\sigma}_{R_{i-1}}}$.
- Decompositie van de gemiddelde kwadratische revisie: $MSR = \bar{R}^2 + (\sigma_P - \rho\sigma_L)^2 + (1 - \rho^2)\sigma_L^2$ (met $\rho = \rho_{L_i P_i}$)

We kunnen nu definiëren:

- $UM = \frac{\bar{R}^2}{MSR}$, hetgeen de proportie van MSR die niet gelijk aan nul is ten gevolge van gemiddelde revisie; ook bekend onder: gemiddelde fout.
- $UR = \frac{(\sigma_P - \rho\sigma_L)^2}{MSR}$. Als we een lineair regressiemodel gebruiken voor de eerdere en latere schattingen, zeg $L_i = \alpha + \beta P_i + u_i$, dan is UR het deel van MSR dat een gevolg is van de omstandigheid dat de richtingscoëfficiënt verschillend van 1 is.
- $UD = \frac{(1-\rho^2)\sigma_L^2}{MSR}$, het storingsdeel van MSR, dwz het deel van de MSR dat niet veroorzaakt wordt door een systematisch verschil tussen vroegere en latere schattingen.

5.9 Structuur van de reeksen

Het is ook mogelijk om een model te schatten waarbij men onderzoekt of de structuur van reeksen systematisch van elkaar verschillen. Indien de oorspronkelijke reeks exact gelijk is aan de latere reeks, dan geldt:

$$L_t = P_t.$$

Indien de reeksen alleen van niveau verschillen, dan geldt:

$$L_t = P_t + \delta.$$

Indien de reeksen alleen een toevallige schommeling verschillen, dan geldt:

$$L_t = P_t + \varepsilon_t,$$

met $\varepsilon_t \sim N(0, \sigma^2)$.

Indien de latere reeks een bepaalde fractie is van de oorspronkelijke reeks dan geldt

$$L_t = \alpha * P_t = F_t.$$

Indien de latere reeks een trendmatig verschil in de tijd verschilt van de oorspronkelijke reeks, dan geldt:

$$L_t = \beta * t + P_t = T_t + P_t.$$

Verschilt de latere reeks een seizoenspatroon, dan geldt:

$$L_t = \sum_i D_i * \gamma_i + P_t = S_t + P_t.$$

Is er sprake van een horizontale verschuiving, een vertraging in de tijd, met één periode dan geldt:

$$L_t = P_{t-1}.$$

Is de vertraging complexer, door een relatie met meerdere oorspronkelijke reeksen, dan geldt:

$$L_t = \sum_{i=1}^t \theta_i * P_{t-i}.$$

In het geval dat $\theta_i = \omega^i$ dan heeft de vertraging een “distributed lag” structuur (Abraham & Ledolter, 1983, p. 70), en dan kan de relatie ook worden geschreven als:

$$L_t = \omega * (L_{t-1} + P_{t-1}) = V_t.$$

In het algemeen is het zinvol om al deze mogelijke relaties te testen in een multivariate regressie. Door middel van een t-toets wordt getoetst of de parameters significant van nul verschillen.

$$L_t = \omega * (L_{t-1} + P_{t-1}) + \alpha * P_t + \sum_i D_i * \gamma_i + \beta * t + \delta + \varepsilon_t, \text{ of}$$

$$L_t = V_t + F_t + S_t + T_t + \delta + \varepsilon_t.$$

Of, in termen van de revisie :

$$R_t = \omega * (L_{t-1} + P_{t-1}) + (\alpha - 1) * P_t + \sum_i D_i * \gamma_i + \beta * t + \delta + \varepsilon_t, \text{ of}$$

$$R_t = V_t + F_t - P_t S_t + T_t + \delta + \varepsilon_t.$$

met $\varepsilon_t \sim N(0, \sigma^2)$.

Dan gelden de volgende uitspraken:

- Als de parameter β niet significant van nul verschilt, is de trendmatige ontwikkeling van de reeksen hetzelfde.
- Als de parameter δ niet significant van nul verschilt, is het niveau van de reeksen hetzelfde.
- Als de parameter γ niet significant van nul verschilt, is het seizoenspatroon van de reeksen hetzelfde.
- Als de parameter α significant niet van 1 verschilt dan is de fractie gelijk aan 1. Indien de parameter ω niet significant van 0 verschilt is er geen “distributed lag” vertragingstructuur.

Opgemerkt dient te worden dat er voor het seizoenspatroon gebruik kan worden gemaakt van dummy's voor maanden en / of kwartalen, of trigonometrische functies.

Het is inzichtelijk om de verschillende componenten te berekenen als een deel van het totale effect, het totale verschil door de revisie. Uitgaande van de volgende componenten:

$$R_t = V_t + F_t - P_t S_t + T_t + \delta + \varepsilon_t.$$

kan de fractie van bijvoorbeeld de seizoen component in de totale revisie berekend worden als:

$$Totaal_t = \{ |V_t| + |F_t - P_t| + |S_t| + |T_t| + |\delta| + |\varepsilon_t| \}$$

$$f_{S_t} = |S_t| / Totaal_t * 100\%.$$

Op dezelfde manier kunnen ook andere fracties worden gedefinieerd. De som van deze fracties is gelijk aan 100%. Alleen als alle cijfers positief zijn mogen de symbolen voor de absolute waarden worden weggelaten. Dat is in de praktijk meestal niet het geval.

6. Procedure bij het vergelijken van reeksen

6.1 Inleiding

In de vorige hoofdstukken zijn een groot aantal aspecten en kwantitatieve analyses beschreven. In de praktijk is het nodig om prioriteiten te stellen om niet te verdwalen in de veelheid van details. In dit hoofdstuk wordt daarom een procedure gegeven die men in praktijkgevallen kan volgen.

We kiezen voor de volgende aanpak:

Stap 1) Een controle op aperte fouten in de dataverzameling.

Bevat de dataverzameling aperte fouten?

Ja => Stop de analyse, en zorg voor data zonder aperte fouten.
Nee => Concludeer dat de structuur van de reeksen geanalyseerd kan worden.

Stap 2) Maak een analyse van de structuur van de reeksen.

Zijn er significante verschillen tussen de reeksen ?

Nee => Concludeer dat er geen verschillen zijn en het onderzoek is klaar.
Ja => Concludeer dat aanvullend onderzoek nodig is en ga naar stap 3.

Stap 3) Bereken de kengetallen uit paragraaf 5.2 tot en met 5.8. Betrek daarbij de relevante kwalitatieve overwegingen (hoofdstuk 4).

Welke conclusie kan worden getrokken over de reeksen?

In Bijlage C wordt de procedure schematisch weergegeven.

6.2 Toelichting per stap

Stap 1) Men dient zich ervan te vergewissen dat er bij de dataverzameling voor beide reeksen geen onregelmatigheden hebben voorgedaan, zoals een haperende robot bij internetdata, waardoor een selectief deel van de data verzameld kan zijn dat zich manifesteert in een reeks die tijdelijk afwijkend gedrag kan vertonen, in feite door een fout. We gaan ervan uit voor de volgende stap dat dat bij de dataverzameling geen

problemen zijn aan te wijzen.³ Anders moet men wellicht besluiten om de data die niet betrouwbaar zijn (dat kan één of meerdere maanden betreffen) weg te laten, in de analyse én in de berekeningen voor de CPI.

Stap 2) Het analyseren van de structuur van de reeksen wordt gedaan met een modelschatting. Daarvoor is een programma beschikbaar. Zie paragraaf 5.9. Daarmee wordt een verschil tussen de reeksen gekwantificeerd en de aard van het verschil vastgesteld. Indien de reeksen voldoende op elkaar lijken, doordat bijvoorbeeld de reeksen slechts een witte ruis (storingsterm) verschillen dan is men klaar en kan er geconcludeerd worden dat er geen significante verschillen zijn. Indien de variantie van de storingsterm groot is dient gezocht te worden naar een model waarbij de variantie relatief klein is.

Stap 3) Indien er wel significante verschillen zijn tussen de reeksen worden de kengetallen uitgerekend die worden genoemd in de paragrafen 5.2 tot en met 5.8. Vervolgens betreft men de kwalitatieve aspecten erbij zoals die worden genoemd in hoofdstuk 4. Op grond van een algehele beoordeling van alle kwalitatieve en kwantitatieve aspecten volgt dan een eindoordeel.

7. Praktijkvoorbeeld: Prijsindexcijfers "Kledingzaak X"

7.1 Inleiding

Het gaat hier om twee prijsindexreeksen met "Kledingzaak X"-data. De ene is gebaseerd op data uit de regionale waarneming (RW), de andere op data verzameld via de internetsite van "Kledingzaak X".

De revisie betreft dan geen verbetering van eenzelfde basisproces (bijvoorbeeld regionale data) maar een meer radicale verandering waarbij de hele waarneming anders wordt: de prijs(index)reeksen volgens de huidige methode (gebaseerd op regionale waarneming en 'prijspaartjes'⁴ wordt vervangen door een latere methode (gebaseerd op internetdata en groepsindices⁵). De perioden waarover de gegevens worden verzameld blijven gelijk, namelijk maanden.

³ Indien dat niet zo is dient men per geval te beslissen wat met de data te doen. In een extreem geval kan men die helemaal weglaten (gedurende een bepaalde periode). Voor een specifieke winkel heeft men dan gedurende die periode geen data.

⁴ Hiermee is bedoeld dat alleen prijzen van artikelen worden vergeleken die aan een gedetailleerde specificatie voldoen.

⁵ Merk op dat twee zaken anders zijn: de wijze van dataverzameling en de methode om uit de waargenomen gegevens prijsindices te berekenen.

7.2 Stap 1: Controle op aperte fouten

Na overleg met een expert op het gebied van de "Kledingzaak X"-internetrobot, mogen we concluderen dat haperingen bij de robot wel degelijk aan de orde zijn geweest. Dit is echter ook door de technici geconstateerd. Zij hebben mogelijk voldoende maatregelen genomen om de kwaliteit van de cijfers te verbeteren. Als men met behulp van visuele inspectie de data beoordeelt, dan is de conclusie gerechtvaardigd dat de data plausibel zijn. De internetreeksen vertonen soortelijke patronen als de reeksen van de kledingbranche als geheel.

Een ander probleem is dat op basis van de visuele inspectie de data voor "Kledingzaak X" die via de regionale waarneming zijn verkregen ("Kledingzaak X" RW) mogelijk niet plausibel zijn omdat er veel afwijkingen zijn van de cijfers van de kledingbranche als geheel. Dat dit niet eerder is geconstateerd heeft mogelijk te maken met het feit dat deze reeks normaliter niet afzonderlijk wordt geproduceerd.

Op grond van de visuele inspectie van de cijfers, concluderen we dat er twijfels zijn of de data goed genoeg zijn voor een adequate vergelijking. We vinden de twijfels nog niet doorslaggevend om het onderzoeksproces te stoppen. Na afronding van de vervolgstappen wordt bij de eindconclusie deze informatie wel meegenomen.

Daarnaast is er nog een andere reden om de vervolgstappen te maken. Het praktijkvoorbeeld dient ter illustratie van de werking van een procedure. Het doorlopen van alle stappen is daarom nuttig vanuit educatief oogpunt.

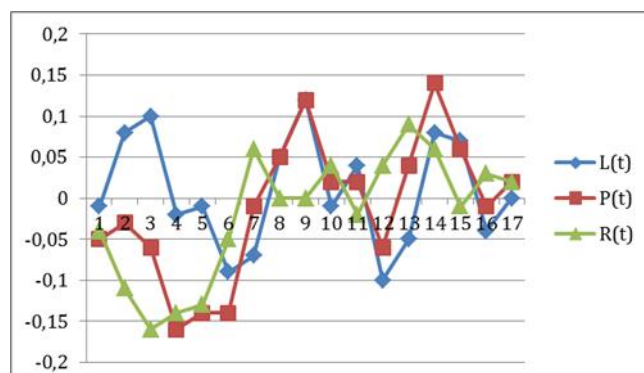
7.3 Stap 2: Structuur van de reeksen

De reeksen voor "Kledingzaak X" (RW en internet) zijn met elkaar vergeleken. Aangezien de reeksen in het eerste waarnemingspunt kunstmatig op 100 zijn gezet, is het verschil uiteraard gelijk aan nul in dat waarnemingspunt. Ook in de daaropvolgende punten zullen de afwijkingen beperkt zijn. Voor een goede analyse is het daarom beter om te werken met de procentuele mutaties. De revisiereeks wordt gedefinieerd als het verschil tussen de procentuele mutaties van de oorspronkelijke en de latere reeks. In de volgende analyses wordt per type kleding een analyse van de structuur gemaakt.

Herenkleding

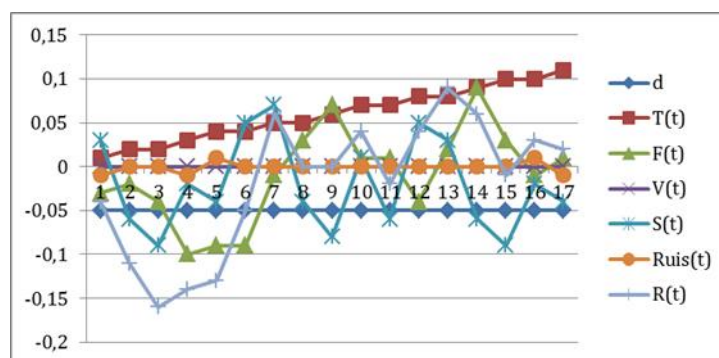
Voor de herenkleding (Coicop 031210) is de structuur van de reeks ontleed. Allereerst kijken we in hoeverre de latere reeks gelijk is aan de oorspronkelijke reeks. Zoals in de grafiek duidelijk te zien is, is dat niet het geval. De revisiereeks $R(t)$ verschilt behoorlijk van nul.

7.3.1 De oorspronkelijke - (P) , latere - (L) en revisiereeks (R) bij herenkleding.



Daarom is de revisiereeks $R(t)$ ontleed in een aantal componenten. De componenten worden in de volgende grafiek in beeld gebracht. De “horizontale” verschuiving (“distributed lag”) is niet aanwezig. Dat is ook goed te zien in de reeksen zelf, aangezien de toppen en dalen ongeveer op dezelfde momenten in de tijd plaats vinden. Wel is er sprake van een niveauverschil, een trendmatig verschil en een fractionele verschuiving. Ook is er een seizoen component aanwezig. De ruis op de schatting is heel klein.

7.3.2 Structuuranalyse bij herenkleding.

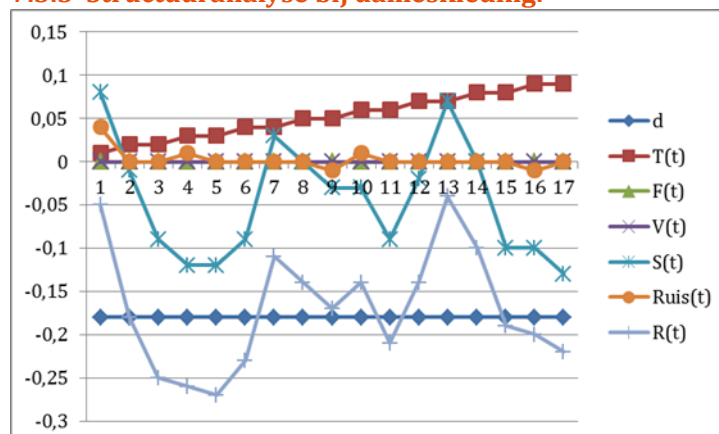


Toelichting: d = niveau, $T(t)$ = trend; $F(t)$ = fractioneel; $V(t)$ = horizontale verschuiving; $S(t)$ = seizoenspatroon, $Ruis(t)$ = storingsterm, $R(t) = L(t) - P(t)$. Op de y-as staan absolute verschillen.

Dameskleding

Bij dameskleding (Coicop 031220) is er een niveauverschil, een trendmatig verschil en een verschil in seizoenseffecten. In de volgende grafiek wordt de structuur getoond.

7.3.3 Structuuranalyse bij dameskleding.

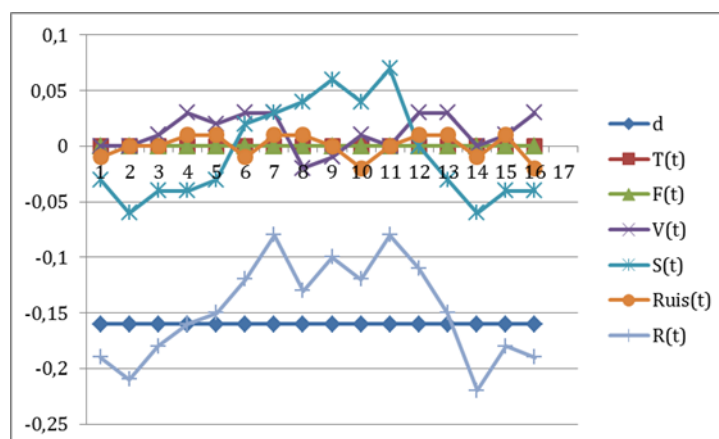


Toelichting: d = niveau , $T(t)$ = trend; $F(t)$ = fractioneel ; $V(t)$ = horizontale verschuiving; $S(t)$ = seizoenspatroon, $Ruis(t)$ = storingsterm, $R(t) = L(t) - P(t)$. Op de y-as staan absolute verschillen.

Kinderkleding

Bij kinderkleding (Coicop 031230) is er met name een groot niveauverschil. Daarnaast is er een seizoenseffect en een "horizontale" verschuiving. Er is geen trendmatig verschil of een fractionele component.

7.3.4 Structuuranalyse bij kinderkleding.

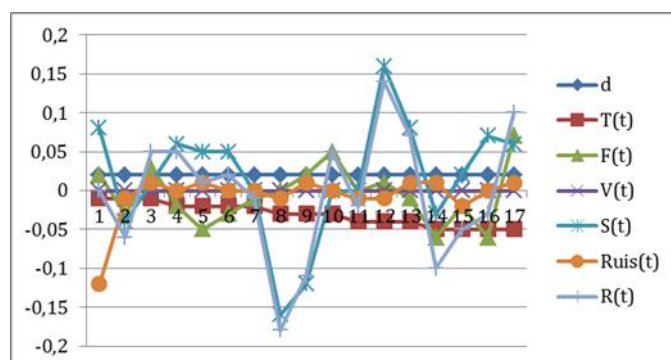


Toelichting: d = niveau , $T(t)$ = trend; $F(t)$ = fractioneel ; $V(t)$ = horizontale verschuiving; $S(t)$ = seizoenspatroon, $Ruis(t)$ = storingsterm, $R(t) = L(t) - P(t)$.

Babykleding

Bij babykleding (Coicop 031230) zijn een groot aantal componenten in de revisiereeks te onderscheiden, waaronder een sterke trendmatige ontwikkeling.

7.3.5 Structuuranalyse bij babykleding



Toelichting: d = niveau, $T(t)$ = trend; $F(t)$ = fractioneel; $V(t)$ = horizontale verschuiving; $S(t)$ = seizoenspatroon, $Ruis(t)$ = storingsterm, $R(t) = L(t) - P(t)$. Op de y-as staan absolute verschillen.

Op grond van deze analyse kan geconcludeerd worden dat er significante verschillen aanwezig zijn. Geen enkele reeks is louter met een witte ruis adequaat te modelleren.

7.4 Stap 3: Kengetallen en kwalitatieve analyse

Aangezien de structurele analyse aangeeft dat er significante verschillen zijn, is het nodig om een uitgebreidere analyse toe te passen. We beginnen met de bespreking van de kengetallen, daarna volgt een kwalitatieve analyse.

Stap 3a: Kengetallen analyse

Op basis van de kwantitatieve scores op de kengetallen, is vervolgens een waardering tussen 1 en 10 toegekend. Dat is grotendeels gebaseerd op een onderlinge vergelijking en gevoel. Hoe groter de impact van de revisie, hoe hoger de score, met 10 als maximum. Vervolgens is een weging gemaakt, waarbij elk thema even zwaar weegt. Daarmee is de totale impact van de revisie door een eindscore weergegeven. Uiteraard mogen aan dergelijke getallen niet teveel waarde worden gehecht, ze geven slechts een indicatie. De kengetallen zijn berekend op basis van de prijsindexcijfers en op basis van de -mutaties (de procentuele ontwikkeling per maand). Allereerst worden de samenvattende scores weergegeven voor de prijsindexcijfers. Daarna voor de mutaties van de prijsindexcijfers. Aangezien de prijsindexcijfers in het eerste waarnemingspunt kunstmatig op 100 zijn gesteld, mogen we er van uitgaan dat de scores over de prijsmutaties relevanter zijn.

Prijsindexcijfers

7.4.1 Samenvattende scores bij de thema's op basis van de kengetallen bij de prijsindexcijfers

Thema	D_ "Kledingzaak X"			
	31210	31220	31230	31240
Omvang van de revisie	3	3	3	6
Vertekening door de revisie	5	3	3	3
De mate van variatie van de revisie	3	3	3	3
De dynamiek van de revisie	6	8	8	5
De relatieve omvang van de revisie	7	7	4	10
De voorspelkracht van de revisie (10-V)	8	9	9	3
Totaal impact revisie (gemiddelde)	7	7	6	8

De samenvattende scores geven aan dat er tussen de soorten kleding (heren-, dames-, kinder- en babykleding) geen grote verschillen bestaan wat de revisie betreft. Wel zijn er onderlinge verschillen. Zo is de relatieve omvang van de revisie bij babykleding (Coicop 31240) groot, maar de voorspelkracht is redelijke goed (lage score).

Prijsmutaties

In tabel 2 worden de scores voor de prijsmutaties weergegeven. De samenvattende scores geven aan dat er tussen de soorten kleding (heren-, dames-, kinder- en babykleding) geen grote verschillen bestaan wat de revisie betreft. Hetzelfde patroon als bij de prijsindexcijfers is zichtbaar.

7.4.2 Samenvattende scores bij de thema's op basis van de kengetallen bij de prijsmutaties

Thema	D_ "Kledingzaak X"			
	31210	31220	31230	31240
Omvang van de revisie	3	3	3	6
Vertekening door de revisie	5	3	3	3
De mate van variatie van de revisie	3	3	3	3
De dynamiek van de revisie	6	8	8	5
De relatieve omvang van de revisie	7	7	4	10
De voorspelkracht van de revisie (10-V)	8	9	9	3
Totaal impact revisie (gemiddelde)	5	6	5	5

De kwantitatieve analyse ondersteunt enigszins de conclusie dat de revisie impact heeft gehad. De gemiddelde scores liggen rond de 5-7, op een schaal van 1-10.

Stap 3b: Kwalitatieve analyse

Opvallend is ook dat de reeks voor "Kledingzaak X" RW veelal afwijkt van alle andere reeksen. We vermoeden dat de kwaliteit van deze reeks beperkt is, mogelijk door beperkingen in de constructie van deze reeks. Dit is ook onderkend door de afdeling zelf.

Een ander belangrijk punt is de kwestie of de regionale waarneming in beginsel beter is dan de waarneming via internet. Op grond van de aspecten die in Bijlage B worden genoemd, kan gesteld worden dat nadelen van de regionale waarneming zijn: de lage frequentie, het lage aantal waarnemingen bij de achterliggende data (dus niet de reeks zelf), en het beperkte aantal artikelen. De conclusie dat er verschillen zijn tussen de waarnemingsmethoden, maar dat de internetwaarneming daar positief bij afsteekt, lijkt het meest gerechtvaardigd.

1) Wat is het doel van de reeks en de revisie?

- Wat wil men meten met de reeks? Antw.: De prijs van verkochte artikelen in de kledingbranche.
- In hoeverre verschillen de oorspronkelijke en de revisie van hetgeen men wil meten? Antw.: De oorspronkelijke reeks is mogelijk vertekend door het beperkte selectieve aantal waarnemingen. De latere reeks is mogelijk vertekend omdat aanbodprijzen worden waargenomen.

2) Wat zijn de kenmerken van de data?

- Worden er micro- en/of macro-data gebruikt? Antw.: Beide reeksen zijn macro-reeksen, gebaseerd op micro-data.
- Wat is de periode waar het onderzoek betrekking op heeft? Antw.: December 2012 – juni 2014.
- Wat voor soort data wordt vergeleken, prijsindices of mutaties? Antw.: Het is beter om naar prijsmutaties te kijken dan naar prijsindices, vanwege het feit dat de eerste waarneming voor beide reeksen op 100 is gesteld ($R=L-P=0$).
- Hoe zijn de artikelen geselecteerd? Antw.: Bij de RW door middel van een (niet-random) steekproef van artikelen. Deze bestaat uit een lijst van nauwkeurig omschreven artikelen opgesteld door een consumptieanalist. Bij internet wordt integraal waargenomen (wat op de site staat als de robot wordt geactiveerd).

3) Hoe zijn de data verzameld?

- Is de waarneming integraal of door middel van een steekproef? Antw.: De RW-reeks is op een beperkte set, maandelijkse waarnemingen gebaseerd, van een beperkt aantal kledingartikelen. De "Kledingzaak X"-internetreeks daarentegen is op de dagelijkse waarneming van alle artikelen op de website van "Kledingzaak X" gedurende de periode van waarneming. De prijzen van de artikelen op opeenvolgende dagen zijn wel sterk gecorreleerd. Door de omvang van de onderliggende data set zijn met de "Kledingzaak X"-internet data veel meer analyses te verrichten dan die gebaseerd op de RW. Zo zou men reeksen kunnen maken over allerlei kleding-groepen. Die zijn voor "Kledingzaak X" doorgaans 'goed gevuld', zodat gemiddelde prijzen voor deze groepen berekend

kunnen worden. Voor een vergelijking van de waarnemingsmethoden wordt ook verwezen naar bijlage B.

- Welke techniek wordt gebruikt (prijswaarnemer, scanner, internetrobot)?
Antw.: Regionale waarneming door een prijswaarnemer, "Kledingzaak X" internet door een internetrobot.
- Is er sprake van kunstmatige reductie van verschillen door gebruik van een basisjaar? Antw.: Ja. Als basisperiode is het eerste waarnemingspunt gekozen, december 2012. Dat reduceert de verschillen in de reeksen.

4) Wat is het resultaat van de revisie?

- Verschillen in het verloop van de reeksen (niveau, trend, seizoen)? Antw.: De reeksen vertonen een verschil in verloop. Dit kan door visuele inspectie van de reeksen (bijlage A) worden gezien. Het wordt bevestigd door de structurele analyse uit stap 2.

8. Conclusies

Bij het vergelijken van twee prijsindexreeksen spelen diverse aspecten een rol. Op basis van een kwalitatieve en kwantitatieve analyse is het mogelijk om een goed inzicht in de reeksen te krijgen. Er is een procedure ontworpen waarbij men stapsgewijs inzicht krijgt in de te vergelijken reeksen.

Met behulp van de procedure is het praktijkvoorbeeld van de Kledingzaak "Kledingzaak X" geanalyseerd. In stap 1 is daartoe bekeken of er aperte fouten zijn. Er blijkt twijfel te bestaan over met name de kwaliteit van de "Kledingzaak X" RW-reeks. Er is toch besloten om ook vervolgstappen uit te voeren. In stap 2 is een structuuranalyse gemaakt van het verschil in procentuele groei van de oorspronkelijke en de latere reeks. Er blijken significante verschillen te bestaan bij alle reeksen. In stap 3 is de verzameling kengetallen uitgerekend en is een kwalitatieve analyse uitgevoerd, aan de hand van de vragenlijst uit hoofdstuk 4. Opvallend is dat de reeks voor "Kledingzaak X" RW veelal afwijkt van alle andere reeksen. We vermoeden dat de kwaliteit van deze reeks beperkt is, mogelijk door beperkingen in de constructie van deze reeks.

Een ander belangrijk punt is de kwestie of de regionale waarneming in beginsel beter is dan de waarneming via internet. Op grond van de aspecten die in Bijlage B worden genoemd, kan gesteld worden dat nadelen van de regionale waarneming zijn: de lage frequentie, het lage aantal waarnemingen bij de achterliggende data (dus niet de reeks zelf), en het beperkte aantal artikelen. De conclusie dat er verschillen zijn tussen de waarnemingsmethoden, maar dat de internetwaarneming daar positief bij afsteekt, lijkt het meest gerechtvaardigd.

Alles in ogenschouw nemend concluderen we dat het, voor het praktijkvoorbeeld "Kledingzaak X", verantwoord is de regionale waarneming te vervangen door waarneming met een internetrobot. Er is geen reden om aan te nemen dat dit tot onbe-

trouwbare uitkomsten zou kunnen leiden. Er zijn weliswaar verschillen met de regionale waarneming. Maar deze wijkt zelf ook af van de ideale waarneming. De huidige manier om voor "Kledingzaak X"-internetdata prijsindices te berekenen (de groepsmethode) heeft beperkingen, maar die kunnen in de toekomst ondervangen worden. Dit is bijvoorbeeld mogelijk door homogenere groepen te maken. Indien echter in de toekomst blijkt dat het CBS van "Kledingzaak X" scannerdata kan krijgen (met voldoende gedetailleerde artikelinformatie) is dit een mogelijk nog aantrekkelijker data bron.

Referenties

Abraham, B. & Ledolter, J. (1983). *Statistical methods for forecasting*, John Wiley & Sons, New York.

McKenzie, R. & Gamba, M. (2008+). *Interpreting the results of revision analysis: recommended summary statistics*. Contribution to the OECD/Eurostat task force on Performing revisions analysis for sub-annual economic statistics.

Bijlage A: Praktijkvoorbeeld "Kledingzaak X"

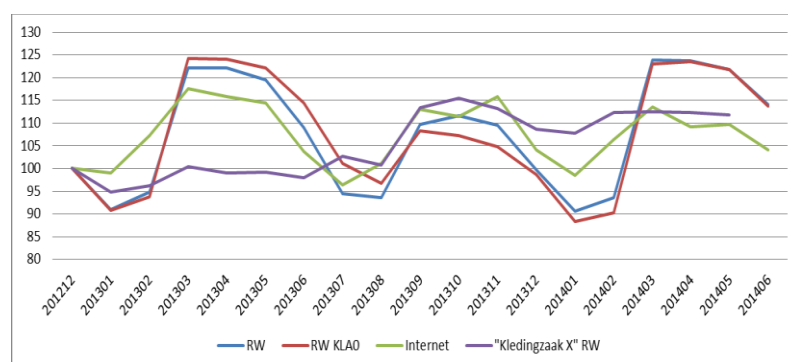
A.1 Prijsindexcijfers

A.1.1 Kengetallen bij de verschillen (R) tussen de oorspronkelijke en latere reeks.

Thema	D_ "Kledingzaak X"			
	31210	31220	31230	31240
<i>Omvang van de revisie</i>				
MAR	6,12	19,60	14,97	7,72
Mediaan absoluut	2,22	18,78	12,78	19,56
<i>Vertekening door de revisie</i>				
Mediaan Revisie	-2,22	18,78	12,78	19,56
Gemiddelde revisie (MEAN)	2,27	20,76	15,85	5,41
Aantal positieve	9	17	17	12
Aantal negatieve	8	0	0	5
Totaal N	17	17	17	17
Aantal niet-nul	17	17	17	17
PPOS	53%	100%	100%	71%
PNEG	47%	0%	0%	29%
PNN	100%	100%	100%	100%
varR	7,50	5,78	2,35	7,24
aangepaste t-waarde	0,83	8,64	10,34	2,01
toets op normaliteit (Shapiro-Wilk test)	normaal	normaal	normaal	normaal
tekentoets	Verwerp H0 niet	verwerp H0	verwerp H0	Verwerp H0 niet
<i>Mate van variatie van de revisie</i>				
SDP / SDL / SDR	8,35	9,01	5,30	9,01
RMSR	8,90	22,74	16,76	10,75
Q1	2,05	13,25	11,53	3,38
Q3	9,25	27,46	20,82	11,12
QDEV	7,20	14,21	9,29	7,74
Minimum	0,00	0,00	0,00	0,00
Maximum	17,23	34,29	24,02	21,64
Range	17,23	34,29	24,02	21,64
SKEW = 3(mean - Mediaan)/SDR	1,61	0,66	1,74	-4,71

<i>Dynamiek van de revisie</i>				
% sign (Later) =sign(earlier)	65%	82%	76%	53%
<i>Relatieve omvang van de revisie</i>				
MAP	105,81	100,64	94,13	95,84
MAR	6,12	19,60	14,97	7,72
RMAR	0,06	0,19	0,16	0,08
<i>Voorspelkracht van de revisie</i>				
Correlatie (P,R)	-0,18	0,46	0,64	0,12
Correlatie (L,R)	0,57	0,77	0,46	0,88
Seriële correlatie (R,R-1)	0,74	0,52	0,54	0,61
<i>Correlatie (L,P)</i>				
UMT	5,16	430,79	251,24	29,32
URT	117,64	34,14	14,80	4,83
UDT	40,09	78,20	25,98	78,86
MSR	163,17	543,73	292,73	113,26
<i>UM</i>				
UM	3%	79%	86%	26%
<i>UR</i>				
UR	72%	6%	5%	4%
<i>UD</i>				
UD	25%	14%	9%	70%
Totaal	100%	100%	100%	100%

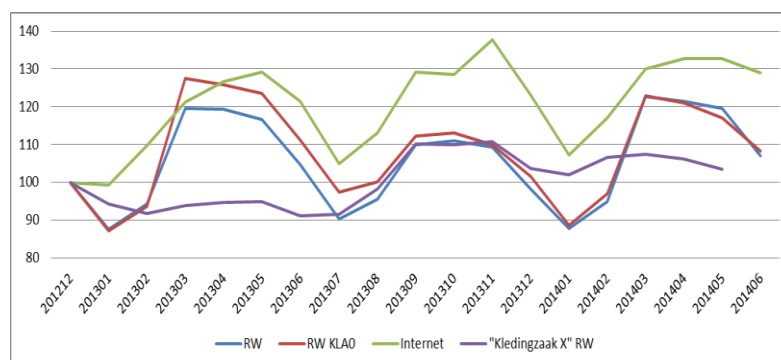
A.1.2 Prijsindexcijfers Coicop 031210 (Herenkleding).



Toelichting op de symbolen:

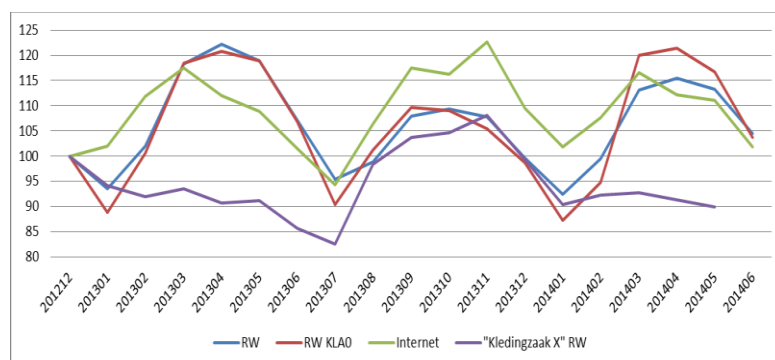
- RW = regionale waarneming totale kledingbranche
- RW KLA0 = regionale waarneming kleding algemeen
- "Kledingzaak X" RW = regionale waarneming "Kledingzaak X"
- Internet = internetwaarneming "Kledingzaak X"
- Op de y-as staan: de waarden van de index. In het basisjaar staan deze op 100

A.1.2 Prijsindexcijfers Coicop 031220 (Dameskleding).



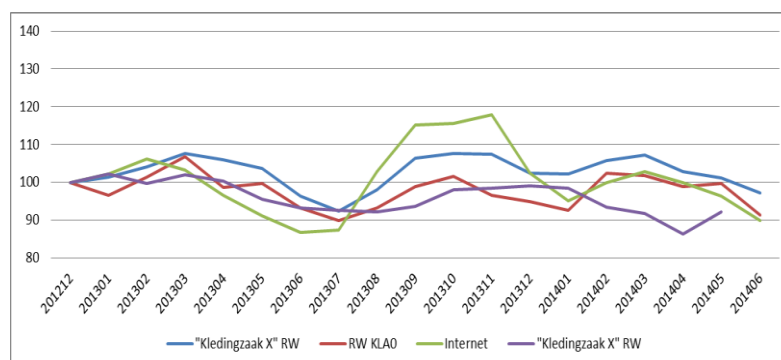
Toelichting op de symbolen: zie C2.1

A.1.3 Prijsindexcijfers voor Coicop 031230 (Kinderkleding).



Toelichting op de symbolen: zie C2.1

A1.4 Prijsindexcijfers voor Coicop 031240 (Babykleding) .



Toelichting op de symbolen: zie C2.1

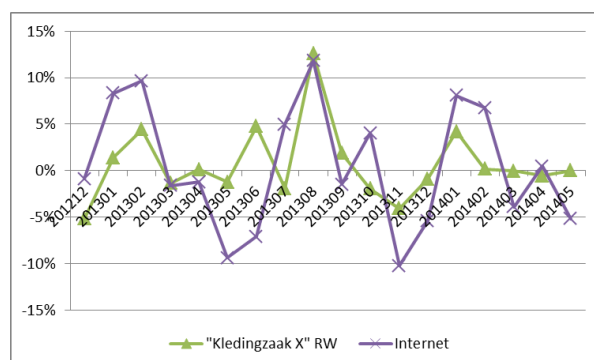
A.2 Mutaties

A.2.1 Kengetallen bij de mutaties van de verschillen (R) tussen de oorspronkelijke en de latere reeks.

Thema	D_ "Kledingzaak X" _Percentage groei			
	31210	31220	31230	31240
<i>Omvang van de revisie</i>				
MAR	0,05	0,05	0,04	0,07
Mediaan absoluut	0,02	0,01	0,04	0,08
<i>Vertekening door de revisie</i>				
Mediaan Revisie	-0,02	0,01	0,02	0,03
Gemiddelde revisie (MEAN)	0,00	0,01	0,01	0,00
Aantal positieve	8	12	9	9
Aantal negatieve	10	6	9	9
Totaal N	18	18	18	18
Aantal niet-nul	17	17	17	17
PPOS	44%	67%	50%	50%
PNEG	56%	33%	50%	50%
PNN	94%	94%	94%	94%
varR	0,003	0,004	0,004	0,006
aangepaste t-waarde	0,19	-0,83	-0,58	-0,03
toets op normaliteit (Shapiro-Wilk test)	normaal	normaal	normaal	normaal
Tekentoets	verwerp NIET	verwerp NIET	verwerp NIET	verwerp NIET
<i>Mate van variatie van de revisie</i>				
SDP / SDL / SDR	0,06	0,07	0,05	0,08
RMSR	0,06	0,07	0,05	0,08
Q1	0,01	0,02	0,02	0,02
Q3	0,07	0,08	0,05	0,10
QDEV	0,05	0,06	0,03	0,08
Minimum	0,00	0,00	0,01	0,01
Maximum	0,12	0,14	0,12	0,18
Range	0,12	0,14	0,11	0,17
SKEW = 3(mean - Mediaan)/SDR	0,95	0,24	-0,48	-1,12
<i>Dynamiek van de revisie</i>				
% sign (Later) = sign(earlier)	65%	82%	76%	53%
<i>Relatieve omvang van de revisie</i>				
MAP	0,03	0,03	0,04	0,03
MAR	0,05	0,05	0,04	0,07

RMAR	1,85	1,81	0,93	2,50
<i>Voorspelkracht van de revisie</i>				
Correlatie (P,R)	-0,36	-0,50	0,10	0,23
Correlatie (L,R)	0,42	0,29	0,34	0,28
Seriële correlatie (R,R-1)	-0,18	0,04	0,25	-0,21
Correlatie (L,P)	-0,10	0,33	0,68	0,52
UMT	0,0	0,0	0,0	0,0
URT	176,6	76,1	17,0	0,1
UDT	42,9	107,8	28,8	61,4
MSR	219,4	183,5	46,4	62,0
UM	0%	0%	0%	0%
UR	80%	41%	37%	0%
UD	20%	58%	62%	99%
Totaal	100%	100%	100%	100%

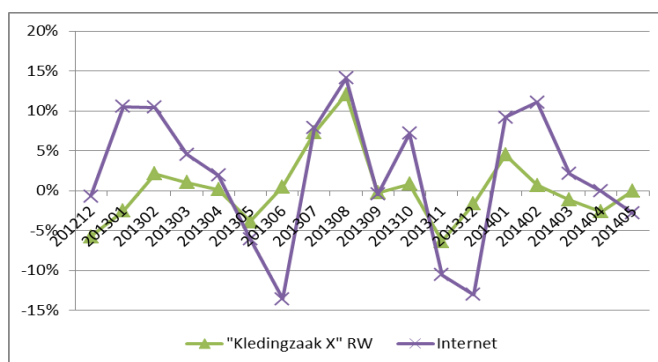
A2.2 Mutaties Coicop 031210 (Herenkleding).



Toelichting op de symbolen:

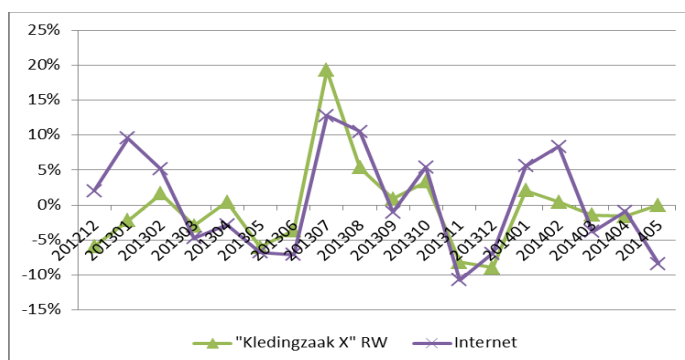
- "Kledingzaak X" RW = regionale waarneming "Kledingzaak X"
- Internet = internetwaarneming "Kledingzaak X"
- Op de y-as staan:
- De waarden van de mutaties; 0,05 is 5%

A2.3 Mutaties Coicop 031220 (Dameskleding).



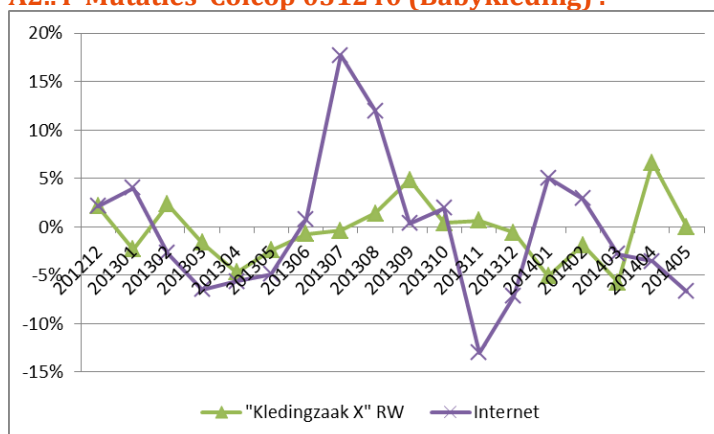
Toelichting op de symbolen: zie C3.1

A.2.4 Mutaties Coicop 031230 (Kinderkleding).



Toelichting op de symbolen: zie C3.1

A2..4 Mutaties Coicop 031240 (Babykleding) .



Toelichting op de symbolen: zie C3.1

Bijlage B: Kenmerken van databronnen

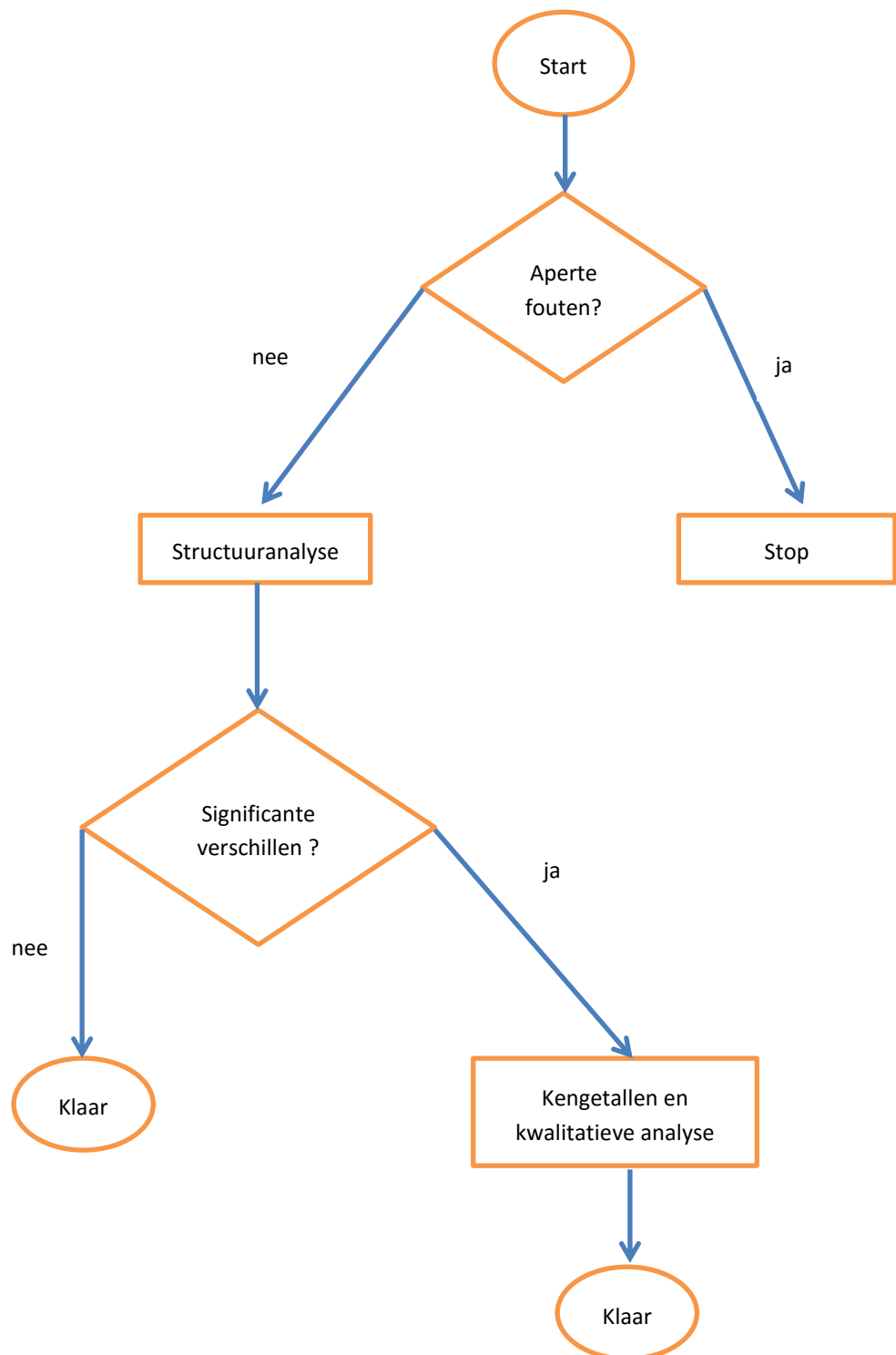
In deze bijlage zijn kenmerken, positief en negatief verzameld, met betrekking tot drie typen data die gebruikt worden, resp. zouden kunnen worden, voor de CPI: RW-data, Internetdata en Scannerdata. Zie onderstaande tabel.

B.1.1 Kenmerken van prijsdata voor drie waarneemmodi.

Kenmerken	RW	Internet	Scanner
Frequentie waarneming	1 x per maand	Dagelijks	Dagelijks, doorgaans samengenomen voor een hele maand
Aantal waarnemingen / winkel (orde van grootte)	Enkele tientallen, 1x per maand (klein deel van het aanbod)	Hele aanbod op de site op het moment van waarnemen	Alles wat verkocht is in die maand
Aantal verschillende artikelen	Beperkte selectie van kledingstukken	Alle kledingstukken op de site op moment van waarnemen	Alle verkochte artikelen die maand
Steekproef / integrale waarneming	Steekproef	Integraal	Integraal
Wel / geen hoeveelheden (mbt omzet) waargenomen	Geen	Geen	Wel
Kwaliteit datamateriaal	Hoog	Hoog bij webwinkels. Minder bij zoeksites (door verkeerd automatisch classificeren van de producten op zoeksite)	Hoog
Kwaliteit waarneming	Fouten mogelijk bij selectie artikelen (oa bij opvolgers), bij kopiëren en verwerken van prijzen door interviewers. Kwaliteit moeilijk controleerbaar	Goed	Goed

Aantekeningen specifiek voor kleding	Omschrijving kan te ruim zijn waardoor nog vele artikelen voldoen (met ruime prijs-range)	Ook slecht / niet verkopende artikelen worden gebruikt	Retouren (voor zover niet binnen dezelfde maand van verkoop)
Soort prijzen	Aanbod	Aanbod	Verkoop
Zelfde artikelen volgen	Enquêtrice zoekt ze bij omschrijving of maakt match ter vervanging	Web-id's	EANs
Problemen (algemeen)	Afhankelijk van beschikbaarheid interviewers, openingstijden van (fysieke) winkels en medewerking van winkels (ivm tolereren van prijswaarnemers)	Aanbodprijzen; geen gewichten; tijdstip van waarneming van belang voor hoeveelheid data	Detail informatie metadata soms te grof, te veel geaggregeerd (over vestigingen heen); maar kwestie van onderhandelen
Risico's	Ziekte van interviewers; interviewers die waarneming niet goed uitvoeren conform instructie	Website verandert (eventueel radicaal)	Scannerdata worden te laat of niet (meer) aangeleverd
Voor- en nadelen	Intelligentie bij waarneming (voor- én nadeel!)		
Kosten	Relatief hoog	Onderhoud robot - software. Deze kosten liggen tussen die voor RW en scannerdata in	Relatief laag; mits in staat om big data te verwerken (indien maximaal detail geleverd wordt)
Verbeterpunten	Alleen in te zetten waar geen elektronische data beschikbaar zijn	Selectie van artikelen maken	Basis bestanden zien te krijgen met maximale informatie erin

Bijlage C: Processchema vergelijken van reeksen



Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopig cijfer
**	Nader voorlopige cijfer
2015–2016	2015 tot en met 2016
2015/2016	Het gemiddelde over de jaren 2015 tot en met 2016
2015/'16	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2015 en eindigend in 2016
2013/'14–2015/'16	Oogstjaar, boekjaar, enz., 2013/'14 tot en met 2015/'16

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
www.cbs.nl

Vormgeving: Centraal Bureau voor de Statistiek, Studio BCO
Ontwerp: Edenspiekermann

Inlichtingen
Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbsl.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2017.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.