

# Seizoencorrectie

# 10

*Marcel van Velzen, Roberto Wekker en Pim Ouwehand*

**Statistische Methoden (10007)**



## Verklaring van tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
**	= nader voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2008–2009	= 2008 tot en met 2009
2008/2009	= het gemiddelde over de jaren 2008 tot en met 2009
2008/'09	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2008 en eindigend in 2009
2006/'07–2008/'09	= oogstjaar, boekjaar enz., 2006/'07 tot en met 2008/'09

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

## Colofon

### *Uitgever*

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

### *Prepress*

Centraal Bureau voor de Statistiek - Grafimedia

### *Omslag*

TelDesign, Rotterdam

### *Inlichtingen*

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

### *Bestellingen*

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

### *Internet*

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

ISSN: 1876-0333

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

## **Inhoudsopgave**

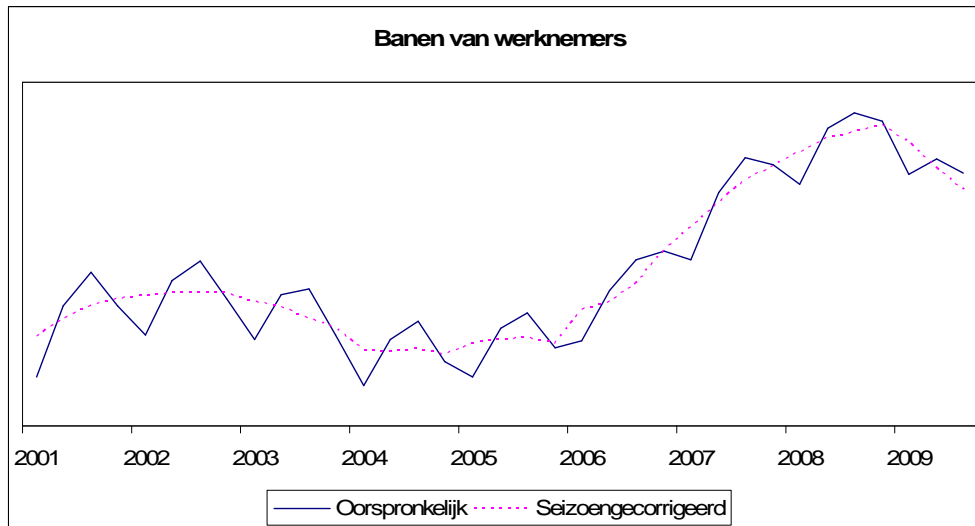
1. Inleiding op het thema.....	4
2. Seizoencorrectie met X-12-ARIMA .....	8
3. Regressie: kalendereffecten en uitbijters.....	22
4. Overige aspecten .....	29
5. Stappenplan seizoencorrectie .....	31
6. Literatuur.....	57

## 1. Inleiding op het thema

### 1.1.1 Beschrijving van het thema

Bij seizoencorrectie willen we tijdreeksen voor zogenaamde seizoenpatronen corrigeren. Dit zijn met vaste (bijv. jaarlijkse) regelmaat terugkerende op- en neergaande bewegingen in een tijdreeks. Door hiervoor te corrigeren zijn gegevens beter door de tijd heen te vergelijken.

Veel statistieken op het CBS worden immers meerdere keren per jaar gemaakt, en bieden daardoor de mogelijkheid bepaalde ontwikkelingen van periode tot periode te volgen. Echter, deze statistieken worden beïnvloed door jaarlijks terugkerende patronen in de gegevens. Voorbeelden hiervan zijn hogere detailhandelsomzetten in december en lager energieverbruik in de zomermaanden. Als we de gegevens voor deze effecten corrigeren, zijn ze beter te vergelijken tussen opeenvolgende periodes. Zo kan bijvoorbeeld blijken dat de decemberomzet nauwelijks verschilt van die van november als men de typische feestdageaankopen buiten beschouwing laat. Door seizoencorrectie uit te voeren worden trendmatige ontwikkelingen in de data beter zichtbaar. In figuur 1 staat ter illustratie de reeks 'Banen van werknemers' weergegeven, samen met de seizoengecorrigeerde reeks.



Figuur 1. Reeks 'Banen van werknemers' en seizoengecorrigeerde reeks

Seizoenpatronen zijn dus jaarlijks terugkerende variaties in de gegevens. Er valt echter ook te denken aan andere frequenties dan jaarpatronen. Zo kan bijvoorbeeld de energieconsumptie ook een weekpatroon vertonen, en zelfs een dagpatroon. Op het CBS hebben we echter alleen met maand- en kwartaalgegevens te maken en corrigeren we dus alleen voor jaarpatronen. Een bijzondere statistiek is echter die

van de omzetten bij de supermarkten, die op vierweekebasis gemeten wordt, waardoor er in feite 13 perioden per jaar ontstaan.

Om seizoencorrectie toe kunnen passen moet een patroon met redelijke nauwkeurigheid geschat kunnen worden, zodat het uit de gegevens gefilterd kan worden. Dit betekent dat er voldoende historische gegevens beschikbaar moeten zijn om het patroon nauwkeurig te schatten. In de praktijk zijn tijdreeksen soms echter relatief kort. Ook kunnen we te maken hebben met ontbrekende gegevens, trendbreuken, of andere verschuivingen binnen de reeks.

Naast correctie voor seizoenpatronen corrigeren we op het CBS ook voor kalendereffecten. Dit zijn aan de kalender gerelateerde gebeurtenissen die meestal een onregelmatig terugkerend patroon hebben, maar wel de tijdreeks beïnvloeden. Als er bijvoorbeeld minder werkdagen in een maand zitten, kan de omzet in een sector daardoor lager uitvallen. Ook heeft bijvoorbeeld het aantal zaterdagen in een maand invloed op de omzet. Daarnaast valt te denken aan feestdagen, schoolvakanties, en het schrikkeljaareffect. Door hiervoor te corrigeren worden gegevens beter door de tijd heen te vergelijken.

### *1.1.2 Problemen en oplossingen*

De voor seizoencorrectie ontwikkelde methoden en software zijn sterk met elkaar verbonden. Er zijn drie belangrijke aanpakken beschikbaar voor seizoencorrectie.

- Het US Census Bureau heeft X-12-ARIMA ontwikkeld als seizoencorrectiemethode en -software. X-12-ARIMA is feitelijk een niet-modelmatige aanpak. De methode is gebaseerd op het iteratief schatten van de seizoencomponent van een tijdreeks middels een aantal rekenrondes. Met behulp van ARIMA-modellen wordt weliswaar de reeks doorgetrokken naar de toekomst om beter schattingen te maken, maar de X-12-procedure is op empirische regels gebaseerd.
- De Bank van Spanje heeft de aanpak en gelijknamig softwarepakket TRAMO-SEATS ontwikkeld met daarin modules voor tijdreeksanalyse en seizoencorrectie. De methode probeert tijdreeksen te modelleren met behulp van ARIMA-modellen en zo seizoeneffecten te schatten.
- Een derde aanpak is die met Structurele Tijdreeksmodellen (STM). Deze modellen en de bijbehorende software zijn voor algemene tijdreeksanalyse bedoeld en dus niet specifiek voor seizoencorrectie ontwikkeld. Deze modellen modelleren elk van de componenten van een tijdreeks afzonderlijk, en kunnen dus ook de seizoencomponent afzonderlijk schatten. Een bekend pakket hiervoor is STAMP.

De meest gebruikte aanpakken zijn echter X-12-ARIMA en TRAMO-SEATS. Binnen de 27 Europese statistische bureaus gebruiken er 10 de X-12-aanpak en 14 de TRAMO-SEATS-aanpak. De overige hebben een afwijkende aanpak. Op het CBS wordt X-12-ARIMA gebruikt. In hoofdstuk 3 wordt deze nader besproken.

Voor een uitgebreidere beschrijving van de X-12-ARIMA-methode verwijzen we naar de Syllabus Seizoencorrectie met X-12-ARIMA (Vollebregt, 2002).

In de praktijk kan met X-12-ARIMA en TRAMO-SEATS dezelfde kwaliteit van seizoencorrectie behaald worden, als men althans reguliere reeksen beschouwt zonder uitzonderlijke effecten. De Eurostat-richtlijnen (Eurostat, 2009) verkiezen dan ook niet een van beide methoden boven de ander.

Binnen Eurostat heeft men Demetra ontwikkeld als schil om TRAMO-SEATS en X-12-ARIMA heen. De opvolger van Demetra is Demetra+, dat ook structurele tijdreeksmodellen aankan.

Op het CBS is Vivaldi (Boset, 2000) ontwikkeld als schil om X-12-ARIMA heen, om de gebruiksvriendelijkheid te vergroten en grotere aantallen reeksen tegelijk aan te kunnen. Inmiddels is X-12-ARIMA zelf ook gebruiksvriendelijker geworden.

Op het CBS is gekozen om te werken met X-12-ARIMA, en zijn een aantal richtlijnen opgesteld voor toepassing van seizoencorrectie binnen het CBS (Booleman, 2003). Binnen Europees verband zijn de Eurostat-richtlijnen (Eurostat, 2009) opgesteld. Dit zijn aanbevelingen en dus geen verplichtingen.

Binnen het CBS wordt voor een aantal statistieken seizoencorrectie toegepast. In Ouwehand en Kraan (2009) wordt een overzicht gegeven van de statistieken waarvoor dit geldt, en welke methoden men daarvoor gebruikt. Het blijkt dat binnen het CBS seizoencorrectie op een aantal belangrijke punten op dezelfde wijze wordt aangepakt, maar op een lager detailniveau er op uiteenlopende wijzen gehandeld wordt. Dit is voor een deel echter onvermijdelijk, aangezien elke statistiek zijn eigen specifieke aanpak vereist. Ook houdt men zich grotendeels aan de CBS-richtlijnen en Eurostat-richtlijnen. Waar dat niet gebeurt is vaak een bewuste keuze gemaakt om hier vanaf te wijken.

Naast de correctie voor het seizoenpatroon zelf, omvat seizoencorrectie ook nog de correctie voor kalendereffecten. Dit zijn gebeurtenissen die ook variaties in de gegevens veroorzaken, maar te verklaren zijn uit kalenderkenmerken. Voorbeelden hiervan zijn de correctie voor het aantal werkdagen en koopdagen in een periode, het aantal dagen in een maand, het optreden van feestdagen en brugdagen, en het schrikkeljaareffect. Seizoencorrectie kan pas goed gedaan worden als de gegevens eerst voor kalendereffecten gecorrigeerd worden. Dit gebeurt vooraf aan de daadwerkelijke seizoencorrectie, maar kan ook plaatsvinden zonder vervolgens de rest van de seizoencorrectieprocedure toe te passen. Deze correctie worden besproken in hoofdstuk 4.

## **1.2 Plaats in het statistisch proces**

Seizoencorrectie vindt plaats aan het eind van het statistisch proces. Het wordt toegepast zodra de feitelijke statistiekproductie is afgerond. Men heeft dan de ongecorrigeerde cijfers. Voor sommige statistieken zijn dit ook te publiceren cijfers. Men kan vervolgens ook nog seizoengecorrigeerde cijfers berekenen, waardoor het beter mogelijk wordt om een reeks door de tijd heen te analyseren. Voor andere

reeksen vormen deze seizoengecorrigeerde reeks de enige te publiceren cijfers. In de toekomst zou het wenselijk kunnen zijn om al eerder in het statistische proces, tijdens de analysefase, aan seizoencorrectie te doen

### 1.3 Definities

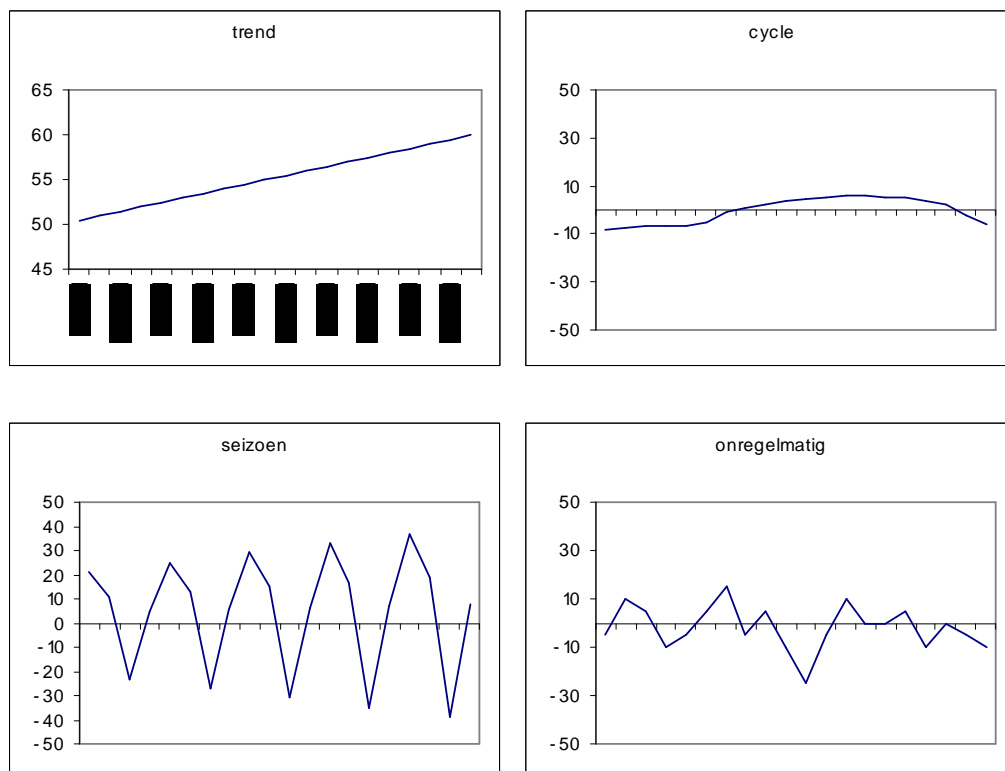
Begrip	Omschrijving
ARIMA-model	Autoregressive Integrated Moving Average model
Decompositie	Ontleden van een tijdreeks in zijn componenten (kan additief of multiplicatief zijn)
Demetra / Demetra+	Door Eurostat ontwikkelde schil om X-12-ARIMA en TRAMO-SEATS
Filter	Reeks gewichten voor waarden van de tijdreeks, benodigd om het seizoenpatroon in de reeks te identificeren.
Kalendereffect	Aan de kalender gerelateerde gebeurtenissen die meestal een onregelmatig terugkerend patroon hebben, maar wel de tijdreeks beïnvloeden
Seizoenpatroon	Een aantal met vaste (bijv. jaarlijkse) regelmaat terugkerende op- en neergaande bewegingen in een tijdreeks
STM	Structurele Tijdreeksmodellen
Tijdreeks	Gegevens over een bepaalde variabele, geordend in de tijd
TRAMO-SEATS	Door Bank van Spanje ontwikkelde methode en software voor tijdreeksanalyse en seizoencorrectie
Vivaldi	Door CBS ontwikkelde schil om X-12-ARIMA
X-12-ARIMA	Door US Census Bureau ontwikkelde methode en software voor seizoencorrectie

## 2. Seizoencorrectie met X-12-ARIMA

In dit hoofdstuk behandelen we de op het CBS gebruikte aanpak voor seizoencorrectie, namelijk die met het pakket X-12-ARIMA. In paragraaf 2.1 en 2.2 behandelen we eerst twee belangrijke basisprincipes achter deze aanpak, namelijk het ontleden van een tijdreeks in componenten, en het gebruik van filters. In 2.3 volgt dan een uitleg over ARIMA-modellen en de X-12-ARIMA-aanpak. We besluiten in 2.4 met kwaliteitsindicatoren.

### 2.1 Ontleden van een tijdreeks in zijn componenten

Onder seizoencorrectie wordt verstaan het corrigeren van een tijdreeks voor invloeden die jaarlijks op vaste tijdstippen terugkeren met een bepaalde intensiteit. Om dit te doen ontleden we een tijdreeks in zijn afzonderlijke componenten, zodat we de seizoencomponent eruit kunnen filteren. Echter, de afzonderlijke componenten zijn niet waarneembaar, men ziet immers alleen de reeks als totaal. Seizoencorrectie is dus tot op zekere hoogte subjectief, er is niet één manier om deze decompositie te doen.



*Figuur 2. Decompositie van een reeks in vier componenten*

In figuur 2 staan de componenten van een tijdreeks geïllustreerd. Naast het seizoen kunnen in de tijdreeksen nog drie componenten worden onderscheiden:



De (lange termijn) *trend* wordt gedefinieerd als de ontwikkeling op zeer lange termijn.

De *conjunctuur* of business cycle (in figuur 2 aangegeven als de cycle) is een periodiek terugkerende golfbeweging bovenop de langetermijn-trendontwikkeling. Deze cyclus varieert van twee tot tien jaar en langer. Indien er sprake is van een periodiek terugkerend conjunctuurpatroon zal het aantal jaren waarover de trendontwikkeling wordt bepaald aanzienlijk groter moeten zijn dan het aantal jaren waarover het conjunctuurpatroon zich uitstrekt.

De *onregelmatige component* bestaat uit fluctuaties in de tijdreeks veroorzaakt door toevallige niet-systematische factoren. Deze restcomponent bevat het gedeelte van de gegevens dat niet tot één van de andere componenten behoort.

De trend en conjunctuur worden door X-12-ARIMA samengenomen en de trend-cycle genoemd. Men bepaalt de trend-cycle, die een goed beeld geeft van zowel de middellange als de lange termijnontwikkeling, door verwijdering van de seizoencomponent en de onregelmatige component.

Er wordt verondersteld dat een kalender- en werkdaggecorrigeerde tijdreeks ( $Y$ ) kan worden ontbonden in vier componenten. De trendcomponent ( $T$ ), de conjunctuur- of cycluscomponent ( $C$ ), de seizoencomponent ( $S$ ) en de onregelmatige component ( $I$ ). Verder worden er twee methoden van ontbinding in componenten verondersteld. Bij de additieve ontbinding wordt ervan uitgegaan dat de seizoeneffecten in grootte constant zijn, onafhankelijk van het verloop van de trend. Bij de multiplicatieve ontbinding wordt ervan uitgegaan dat de seizoeneffecten lineair toenemen met het trendniveau van de data.

De additieve ontbinding geeft in formulevorm

$$Y_t = T_t + C_t + S_t + I_t, \quad (2.1)$$

waarbij  $t$  het tijdstip aangeeft waarop de data betrekking hebben (maand, kwartaal).

De multiplicatieve ontbinding is van de vorm

$$Y_t = T_t \times C_t \times S_t \times I_t. \quad (2.2)$$

Voor iedere tijdreeks moet dus worden vastgesteld of de ontbinding in componenten het best kan worden beschreven door een additieve dan wel een multiplicatieve ontbinding. De eenvoudigste manier om te kiezen voor een additieve of een multiplicatieve decompositie is X-12-ARIMA zelf automatisch voor één van beide decomposities te laten kiezen. Hiertoe maakt X-12-ARIMA gebruik van het Akaike Information Criterium (AIC). Dit criterium is een variant van het maximum likelihood criterium. Voor meer informatie over dit criterium zie Akaike (1973). In de meeste gevallen zal een tijdreeks echter een multiplicatieve ontbinding hebben, aangezien in het algemeen de grootte van de uitslagen van een economische tijdreeks evenredig zullen afhangen van de waarden van al zijn componenten. De multiplicatieve ontbinding wordt additief gemaakt door aan beide zijden de logaritme te nemen. Formule 2.2 wordt dan

$$\log(Y_t) = \log(T_t) + \log(C_t) + \log(S_t) + \log(I_t). \quad (2.3)$$

Het is daarom voldoende de werking van het X-12-ARIMA-programma uit te leggen aan de hand van de additieve ontbinding.

## 2.2 Filters

De filters zoals die in X-12 worden toegepast bepalen steeds een (gewogen) voortschrijdend gemiddelde. Hierbij wordt een datapunt steeds vervangen door een combinatie van omliggende datapunten. Deze filters gebruiken geen gecompliceerde concepten of modellen en zijn daardoor redelijk eenvoudig en robuust in gebruik. Het is mogelijk filters te construeren die goede eigenschappen hebben wat betreft het behoud van de trend en verwijdering van de seizoencomponent. Het onderdeel van X-12 dat uit het toepassen van symmetrische filters bestaat wordt historisch vaak het X-11 gedeelte genoemd. De combinatie van het gebruik van symmetrische filters plus regressie en ARIMA-modellering wordt gezamenlijk met X-12 aangeduid. Regressie en ARIMA-modellering worden later behandeld.

### Werking van X-11

Het rekenschema van X-11 bestaat uit drie soortgelijke rekenrondes. Iedere rekenronde maakt gebruik van de resultaten uit de vorige rekenronde en geeft een verdere verfijning van deze resultaten. In iedere rekenronde worden de trend-cycle, het seizoen en de onregelmatige component geschat. In iedere rekenronde wordt eerst een eerste schatting van de trend-cycle en het seizoen gegeven, gevolgd door werkelijke schattingen van de trend-cycle, het seizoen en de onregelmatige component voor de betreffende rekenronde. De onregelmatige component wordt gebruikt om extreme waarden te detecteren. De gevonden extreme waarden worden uit de oorspronkelijke reeks verwijderd en de zo ontstane reeks wordt gebruikt als uitgangspunt voor de volgende rekenronde. In de volgende rekenronde worden dezelfde rekenstappen herhaald. Onder aanname dat door het verwijderen van de extreme waarden de schattingen voor de onregelmatige component in opeenvolgende rekenrondes steeds verfijnder worden, worden de resultaten per rekenronde ook steeds beter. We zullen de werking van X-11 bespreken aan de hand van een additief filter aangezien zoals is aangetoond een multiplicatieve reeks additief te maken is door het nemen van een logaritme.

#### *Bepalen van de trend en seizoencomponent*

De eerste stap in de berekening bestaat uit het filteren van de trend-cycle uit de reeks. Dit wordt gedaan met behulp van een filter of voortschrijdend gemiddelde (in het Engels: moving average). Het bestaat voor kwartaalreeksen uit een 2x4 voortschrijdend gewogen gemiddelde

$$TC_t = \frac{1}{8}Y_{t-2} + \frac{1}{4}Y_{t-1} + \frac{1}{4}Y_t + \frac{1}{4}Y_{t+1} + \frac{1}{8}Y_{t+2}, \quad (2.4)$$

waarbij  $t$  het kwartaal en  $Y$  de tijdreeks waarden aangeeft. De gewichten van het filter zijn zo bepaald dat alle kwartalen even zwaar meetellen. Aangezien de

kwartalen  $t-2$  en  $t+2$  dezelfde kwartalen (van een verschillend jaar) aanduiden, worden de gewichten twee maal zo klein genomen als de gewichten van de andere kwartalen. Verder moeten de gewichten optellen tot 1. Het filter kan ook als volgt geschreven worden als  $1/8[1 \ 2 \ 2 \ 2 \ 1]$ . Aangezien het hier een symmetrisch filter betreft is het voldoende dit te schrijven als  $1/8[1 \ 2 \ [2]]$ , waarbij het getal tussen rechte haken de centrale waarden aangeeft en ook meteen aangeeft dat het hier een symmetrisch filter betreft. Met de naamgeving  $pxq$  filter wordt het resulterende filter bedoeld dat ontstaat door een voortschrijdend gemiddelde te nemen van een filter van orde  $p$  waarvan alle coëfficiënten  $1/p$  zijn, gevolgd door een voortschrijdend gemiddelde van orde  $q$  van een filter waarvan alle coëfficiënten  $1/q$  zijn. In het bovenstaande  $2 \times 4$  filter krijgen we dan inderdaad het resulterende filter  $1/8[1 \ 2 \ 2 \ 2 \ 1]$  uit de convolutie van deze twee eenvoudige filters:

$$\begin{aligned} \text{Conv}\left(\left[\frac{1}{2}, \frac{1}{2}\right], \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]\right) &= \left[\frac{1}{2} \times \frac{1}{4}, \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4}, \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4}, \frac{1}{2} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{4}, \frac{1}{2} \times \frac{1}{4}\right] \\ &= \left[\frac{1}{8}, \frac{2}{8}, \frac{2}{8}, \frac{2}{8}, \frac{1}{8}\right] \end{aligned}$$

Voor maandreeksen bestaat het uit een  $2 \times 12$  voorschrijdend gewogen gemiddelde

$$TC_t = \frac{1}{24}Y_{t-6} + \frac{1}{12}Y_{t-5} + \dots + \frac{1}{12}Y_t + \dots + \frac{1}{12}Y_{t+5} + \frac{1}{24}Y_{t+6}, \quad (2.5)$$

waarbij  $t$  nu de maand aanduidt. Ofwel  $M_{2 \times 12} = 1/24[1 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 1] = 1/24[1 \ 2 \ 2 \ 2 \ 2 \ 2 \ [2]]$ . Aangezien de randen van de filters een overeenkomstige periode aangeven, worden deze maar voor de helft meegeteld t.o.v. de andere perioden.

Door de zojuist bepaalde trend af te trekken van de oorspronkelijke reeks blijft een reeks over bestaande uit de onregelmatige component en de seizoencomponent (S+I). Er wordt nu een eerste schatting van de seizoencomponent gemaakt via een  $3 \times 3$  filter. Uitgeschreven voor een maandreeks heeft dit filter de vorm

$$\left\{ \begin{array}{l} 1/9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\ 2/9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\ 3/9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\ 2/9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\ 1/9 \end{array} \right\}$$

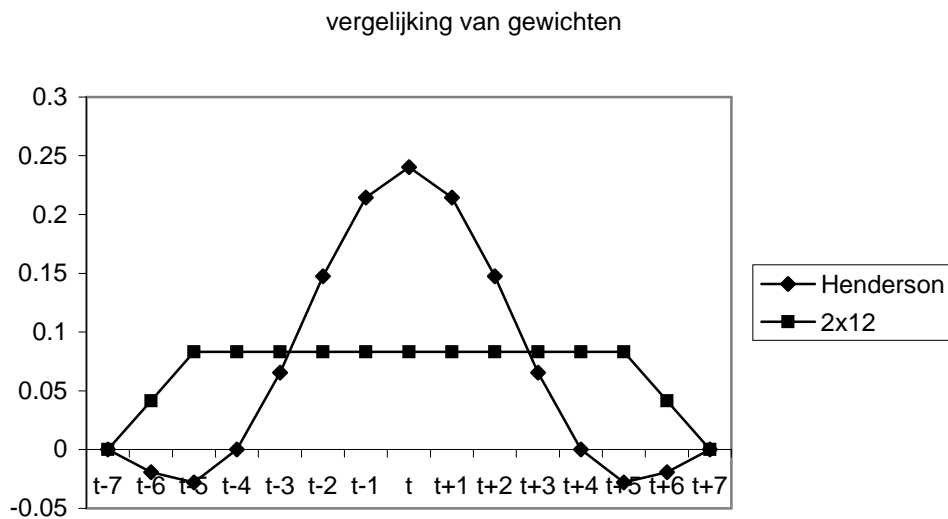
en is dus 49 maanden lang, maar alleen de overeenkomstige perioden worden meegenomen in de berekening. Ook hier worden de randen minder hard meegeteld dan de maanden van de direct omliggende jaren.

Door nu ook de seizoencomponent af te trekken van de trendvrije data blijft alleen de onregelmatige component over. De onregelmatige component wordt vervolgens ontdaan van extreme waarden op ongeveer de volgende wijze: Ieder datapunt waarvan de onregelmatige component groter is dan 2.5 keer de standaardafwijking worden beschouwd als extreem, krijgt gewicht 0 en telt straks niet meer mee voor de berekening. Vervolgens wordt opnieuw de standaardafwijking bepaald. Wederom tellen punten waarvan de onregelmatige component groter is dan 2,5 maal de standaard afwijking niet meer mee. Alle punten waarvan de onregelmatige

component kleiner is dan 1.5 keer de standaardafwijking worden beschouwd als niet afwijkend en krijgen gewicht 1. Tussen 1.5 en 2.5 keer de standaardafwijking wordt lineair een gewicht tussen 1 en 0 toegekend.

*Bepalen van de trend van de seizoengecorrigeerde reeks*

Er is een eerste schatting gemaakt van de trend-cycle en seizoencomponent, en extreme waarden zijn verwijderd. Door nu de seizoencomponent van de voor extreme waarden gecorrigeerde waarnemingen af te trekken ontstaat een seizoengecorrigeerde reeks waarvan opnieuw de trend bepaald zal gaan worden. Het bepalen van de trend uit een seizoengecorrigeerde reeks wordt nu niet meer gedaan met het voorgaande 2x12 filter maar via een zogenaamd Henderson-filter dat gladdere coëfficiënten heeft omdat aangenomen wordt dat nu ook de trend een gladder verloop heeft. De voornaamste speciale eis die aan de filtercoëfficiënten wordt gesteld bij de constructie van het Henderson-filter is dat kwadratische polynomen onveranderd blijven na toepassing van het Henderson-filter (Ladiray en Quinneville, 2001). Dit leidt ertoe dat de coëfficiënten van het Henderson-filter een wat gladdere reeks vormen. In onderstaande figuur zien we een vergelijking tussen de gewichten van het Henderson-filter en het 2x12-trendfilter: Hierbij is voor de vergelijkbaarheid een Henderson-filter gekozen dat dezelfde lengte heeft als het 2x12-trendfilter (hoewel dat niet per se zo hoeft te zijn).

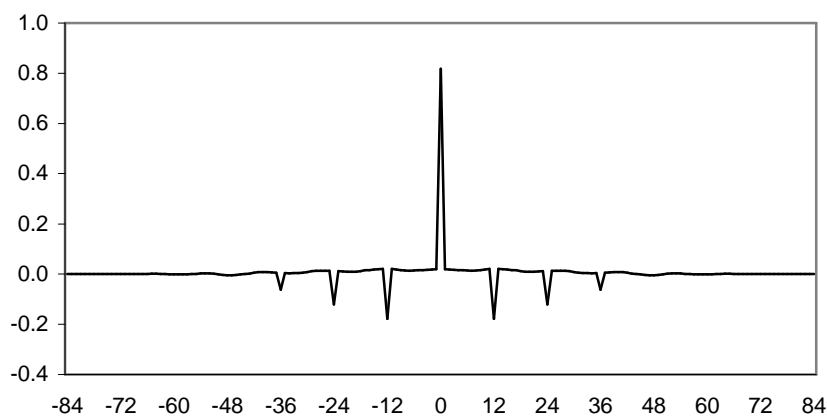


*Figuur 3. Gewichten voor het Henderson-filter en voor het 2x12-filter*

Standaard kiest X-11 zelf een geschikt Henderson-filter. X-11 maakt deze keuze op basis van de grootte van de fluctuaties in de onregelmatige component ten opzichte van de fluctuaties in de trend-cycle (Vollebregt, 2002, paragraaf 4.6).

Bovenstaande filters worden op overeenkomstige wijze door X-11 in 3 rekenrondes een aantal malen toegepast (Vollebregt, 2002, blz. 16) op de data. Dit leidt tot steeds betere schattingen van trend-cycle, seizoencomponent en onregelmatige component. Hiervoor gebruikt X-12 standaardwaarden voor de filtercoëfficiënten, waarbij het

voor de gebruiker eventueel mogelijk is de filtercoëfficiënten aan te passen. Wanneer we de diverse correcties voor extreme waarden buiten beschouwing laten, doet X-11 niets anders dan het successievelijk toepassen van diverse filters. Het na elkaar toepassen van een aantal kleine filters is hetzelfde als het in één keer toepassen van een groot filter. Als we bijvoorbeeld een 3x3-seizoenfilter voor de eerste schattingen, een 3x5-seizoenfilter voor de tweede schattingen en een Henderson-filter van lengte 13 voor het schatten van de trend-cycle toepassen, en correcties voor extreme waarden buiten beschouwing laten, heeft het uiteindelijke filter een lengte van 169. De gewichten van dit filter staan weergegeven in figuur 4. Merk op dat de gewichten voor  $-40$  en na  $+40$  nagenoeg verwaarloosbaar zijn. In werkelijkheid voert X-11 wel correcties uit voor extreme waarden. Daarom dient de onderstaande figuur slechts ter illustratie.



*Figuur 4. Gewichten voor het uiteindelijke X-11-filter*

### 2.3 ARIMA-modellen

Bij de X-11 methode wordt gebruik gemaakt van symmetrische filters en het is daarom nodig om de data aan de randen te extrapoleren om geen datapunten te verliezen na het filteren.

We geven de voorkeur aan symmetrische filters, omdat deze over het algemeen leiden tot stabielere seizoenpatronen. De extrapolaties van de tijdreeksen worden gemaakt met behulp van ARIMA-modellen. ARIMA staat voor *autoregressive integrated moving average*. ARIMA-modellen en zijn een uitgebreide klasse van modellen die vaak worden gebruikt om economische tijdreeksen te beschrijven. De materie is echter vrij complex en daarom zal alleen de essentie van de modellen worden besproken. Details zijn te vinden in Box en Jenkins (1970).

Het basisprincipe achter ARIMA-modellering is de stelling van Wold. Deze stelling zegt dat iedere stationaire tijdreeks geschreven kan worden als de som van een deterministische component en een oneindige reeks van ongecorreleerde stochasten

$e_t$ :

$$Y_t = m + \sum_{i=0}^{\infty} y_i e_{t-i}, \quad (2.6)$$

waarbij  $m$  het gemiddelde van de reeks is. Het stochastische deel ( $\sum_{i=0}^{\infty} y_i e_{t-i}$ ) is in feite een oneindig lang voortschrijdend gemiddelde van een witte-ruisproces  $e_t$ . Het probleem is echter dat het aantal stochasten en dus ook het aantal parameters oneindig groot is. Een benadering kan gegeven worden door slechts een beperkt aantal parameters  $y_i$  in het model op te nemen. We spreken dan van een moving average model, of kort MA-model. Een MA( $q$ )-model is een model met  $q$  parameters.

Bij ingewikkelde modellen zouden we in een MA-model nog steeds een groot aantal parameters nodig hebben. Een model met veel parameters is moeilijk te schatten. In dat geval kunnen we gebruik maken van een zogeheten autoregressief model, of AR-model. Hierbij gebruiken we dat bovenstaande stationaire tijdreeks ook als volgt geschreven kan worden

$$Y_t = c + e_t + j_1 Y_{t-1} + j_2 Y_{t-2} + \dots \quad (2.7)$$

Hier is  $c$  een constante die afhangt van het gemiddelde van de reeks en het is mogelijk de parameters  $j_i$  uit te drukken in de parameters  $y_i$  en omgekeerd. Deze reeks drukt dus uit dat iedere term in de reeks ook kan worden uitgedrukt in zijn voorgaande termen plus een stochast  $e_t$ . De waarden van een tijdreeks zijn immers in de tijd gecorreleerd. Zo zal bijvoorbeeld de productie of consumptie van augustus mede afhangen van die van juli.

Een model met een oneindig aantal termen wordt nu benaderd door een beperkt aantal AR- en MA-termen. Wanneer we autoregressieve en moving-average-termen combineren in één model spreken we van een ARMA-model. Een ARMA-model met  $p$  autoregressietermen en  $q$  moving-average-termen schrijven we kort als een ARMA( $p,q$ )-model en kan worden geschreven als.

$$Y_t = c + j_1 Y_{t-1} + \dots + j_p Y_{t-p} + e_t + q_1 e_{t-1} + \dots + q_q e_{t-q}. \quad (2.8)$$

We willen de tot nu toe beschreven ARMA-modellen echter ook kunnen gebruiken om niet-stationaire reeksen te beschrijven. Als er bijvoorbeeld sprake is van een lineaire trend ( $y=At+B$ ) in plaats van een constante trend, voeren we daartoe de volgende transformatie op de reeks uit:

$$Z_t = Y_t - Y_{t-1}. \quad (2.9)$$

Er volgt nu (door invullen van  $Y_t = At + B$ ) dat we de niet-stationaire reeks  $Y$  getransformeerd hebben tot een stationaire reeks  $Z$  met een constante trend  $A$ . We noemen deze transformatie een differentie. We kunnen de stationaire reeks  $Z$  nu beschrijven met een ARMA-model. Als we de reeks die beschreven wordt door het

ARMA-model terug transformeren, krijgen we de oorspronkelijke reeks  $Y$ . We noemen dit terug transformeren integratie. We hebben hierna een modelmatige beschrijving van  $Y$  en noemen een dergelijk model een ARIMA-model. De I in ARIMA staat voor integrated.

Bij trends van hogere orde dient er meerdere keren gedifferentieerd en geïntegreerd te worden. In de praktijk worden meestal niet meer dan twee differenties genomen. Een ARIMA-model met  $p$  autoregressietermen,  $d$  differenties en  $q$  moving-average- termen noteren we als ARIMA(p,d,q)-model.

Met het ARIMA(p,d,q)-model kunnen we een goede beschrijving geven van een grote hoeveelheid tijdreeksen. Met behulp van deze beschrijving kunnen we een voorwaartse en achterwaartse extrapolatie van de reeks maken waardoor in X-11 symmetrische filters gebruikt kunnen worden en ook de begin- en eindpunten van de reeks behouden kunnen worden.

ARIMA is dus ontworpen om de samenhang tussen opeenvolgende waarnemingen vast te stellen waardoor voorspellingen mogelijk zijn. Deze methode is door het Statistisch Bureau van Canada in X-11 geïmplementeerd waardoor het mogelijk werd om trend- en seizoenfilters symmetrisch toe te passen. Dit is de X-11 ARIMA-methode. Door vervolgens ook nog andere effecten in de tijdreeksen, waaronder kalenderinvloeden en uitbijters, vooraf te modelleren wordt de kwaliteit van geraamde seizoenpatronen vergroot. Deze speciale effecten worden met regressie gemodelleerd. Deze simultane toepassing van regressie en ARIMA is door het Amerikaanse Censusbureau ontworpen en staat bekend als X-12 RegARIMA.

### ARIMA-modellering van seizoeneffecten

De tijdreeksen die we met X-12-ARIMA willen analyseren zijn tijdreeksen met een sterke seizoencomponent. Om goede voorspellingen te maken zal deze seizoencomponent dus goed gemodelleerd moeten worden door het ARIMA-model. Daartoe drukken we de  $Y_t$  niet uit in de MA-terminen op tijdstip  $t, t-1, t-2, \dots$ , maar in de MA-terminen op tijdstip  $t, t-12, t-24, \dots$ ,

$$Y_t = \mu + e_t + \Theta_1 e_{t-12} + \Theta_2 e_{t-24} + \dots + \Theta_Q e_{t-12Q} \quad (2.10)$$

Merk op dat we de kleine  $\theta$  vervangen hebben door een hoofdletter  $\Theta$  en de kleine  $q$  door een grote  $Q$ . In het algemeen is het zo dat kleine letters gebruikt worden voor de gewone parameters van het ARIMA-model, en hoofdletters voor de seizoenparameters van het ARIMA-model. De geïntroduceerde seizoen-MA-terminen worden toegevoegd aan het al bestaande ARIMA-model.

Hetzelfde kunnen we doen voor autoregressietermen

$$Y_t = c + \Phi_1 Y_{t-12} + \Phi_2 Y_{t-24} + \dots + \Phi_P e_{t-12P} + e_t \quad (2.11)$$

en voor differenties:

$$Z_t = Y_t - Y_{t-12} \quad (2.12)$$

We combineren nu alle termen met elkaar in een ARIMA(p,d,q)(P,D,Q)-model, waarin P het aantal seizoen-autoregressietermen is, D het aantal seizoendifferenties en Q het aantal seizoen-moving-average-termen. Met dit model zijn we in staat om zowel de trend als het seizoen in een gegeven tijdreeks op een goede manier te beschrijven.

Zowel de trend als de seizoeneffecten zijn gemodelleerd. De seizoeneffecten zitten echter impliciet in het gegeven model. Het model geeft ons nog geen expliciete beschrijving van de trend-cycle, het seizoen en de onregelmatige component. Het is weliswaar mogelijk om deze expliciete beschrijving aan de hand van het ARIMA-model te bepalen, maar deze beschrijving komt er niet vanzelf uit rollen. Het is alles behalve eenvoudig om een gegeven ARIMA-model te vertalen in drie componenten. Bij X-12-ARIMA wordt dan ook niet voor deze aanpak gekozen. X-12-ARIMA gebruikt het ARIMA-model slechts om te extrapoleren, en rekent daarna volgens het bekende rekenschema verder. TRAMO/SEATS bepaalt de decompositie wel aan de hand van het berekende ARIMA-model. Daarbij is TRAMO het programma dat het meest geschikte ARIMA-model zoekt en SEATS het programma dat de feitelijke decompositie doet. De wiskundige onderbouwing van TRAMO/SEATS is daarmee sterker, maar de resultaten in de praktijk zijn niet beter of slechter.

### **Keuze van het ARIMA-model in X-12-ARIMA**

Binnen X-12-ARIMA kan men ofwel handmatig kiezen voor een bepaald ARIMA-model, ofwel X-12-ARIMA zelf een geschikt ARIMA-model laten zoeken. Wanneer men zelf het ARIMA-model specificeert, moet in ieder geval het type model gespecificeerd worden door het opgeven van de parameters  $p$ ,  $d$ ,  $q$  en  $P$ ,  $D$  en  $Q$ . De modelparameters worden vervolgens door X-12-ARIMA zelf geschat tenzij men deze zelf expliciet opgeeft.

De meeste gebruikers van X-12-ARIMA zullen bij de analyse van een reeks niet zelf een ARIMA-model specificeren, maar X-12-ARIMA automatisch laten kiezen voor een geschikt model. In versie 0.3 van X-12 is er ook de mogelijkheid om de software via de optie 'automodel' een geschikt ARIMA-model te laten vinden (zie paragraaf 2.3), onderscheiden van het bestaande 'pick model'.

Een tussenoplossing is 'pick model'. Hierbij rekent X-12-ARIMA vijf standaardmodellen door en kiest daarvan degene die het meest geschikt is voor de gegeven reeks. Dit zijn de modellen met de 'non seasonal' (p,d,q) parameters gelijk aan (0 1 1), (0 1 2), (2 1 0), (0 2 2) of (2 1 2), terwijl de 'seasonal' (P,D,Q) parameters steeds gelijk zijn aan (0 1 1). Deze vijf modellen zijn geschikt om voor bijna elke tijdreeks goede voorspellingen te produceren. Om tot één van de vijf modellen te komen worden alle vijf modellen doorgerekend. Ieder model wordt goedgekeurd of afgekeurd aan de hand van een tweetal voorwaarden.

Ten eerste wordt het ARIMA-model toegepast op het begin van de reeks (alle waarnemingen behalve de laatste drie jaar), om extrapolaties te maken voor de laatste drie jaar van de reeks. De extrapolaties van de laatste drie jaren worden vergeleken met de werkelijk waarden. Wanneer de extrapolaties te veel verschillen



van de werkelijke waarnemingen geeft het model geen goede beschrijving van de reeks en wordt het verworpen.

De tweede voorwaarde is dat de niet-gemodelleerde componenten van de reeks ongecorrleerd moeten zijn. Het toetsen of dit inderdaad het geval is gebeurt door middel van het berekenen van de zogenaamde Box-Ljung-grootheid. Wanneer de nulhypothese dat er geen autocorrelatie is wordt verworpen, wordt het model afgewezen.

## 2.4 Kwaliteitsindicatoren

X-12-ARIMA geeft de kwaliteit van de seizoencorrectie weer via een elftal kwaliteitsmaten, M1 tot en met M11, die beschrijven in hoeverre de seizoendecompositie geslaagd is. M1, M2, M3, M5 en M6 meten de grootte van de onregelmatige component ten opzichte van de andere componenten. M3 en M5 meten de grootte van de onregelmatige component ten opzichte van de trend. M4 meet de autocorrelatie binnen de onregelmatige component. M7 meet in hoeverre het seizoeneffect identificeerbaar is. M8 tot en met M11 meten in hoeverre het seizoenpatroon verandert. Aan de hand van deze beoordeling kunnen we besluiten om alsnog een aantal parameters bij te stellen. We kunnen de kwaliteitsmaten beschouwen als rapportcijfers, met dat verschil dat de 11 kwaliteitsmaten van X-12-ARIMA variëren van 0 tot 3, en dat een waarde tussen 0 en 1 als voldoende wordt beschouwd. Hoe kleiner de waarde van een kwaliteitsmaat, des te beter scoort de seizoendecompositie op het betreffende aspect. Aan het eind van tabel F3 wordt nog een eindcijfer gegeven, dat een gewogen gemiddelde is van de 11 getallen. Dit gemiddelde wordt aangeduid met Q. Ook hiervoor geldt dat het varieert van 0 tot 3 en dat een getal tussen 0 en 1 als voldoende wordt beschouwd. De gewichten staan in tabel 1. Bij reeksen van korter dan 6 jaar of reeksen met een stabiel seizoengemiddelde worden de maten M8, M9, M10 en M11 niet berekend. In dat geval gelden alternatieve gewichten, die eveneens in de tabel staan vermeld.

Tabel 1. Gewichten voor de bepaling van Q

<i>Maat</i>	<i>Gewicht</i>	<i>Gewicht</i> <i>(korte reeks)</i>		<i>Maat</i>	<i>Gewicht</i>	<i>Gewicht</i> <i>(korte reeks)</i>
<i>M1</i>	13	17		<i>M7</i>	16	30
<i>M2</i>	13	17		<i>M8</i>	7	
<i>M3</i>	10	10		<i>M9</i>	7	
<i>M4</i>	5	5		<i>M10</i>	4	
<i>M5</i>	11	11		<i>M11</i>	4	
<i>M6</i>	10	10				

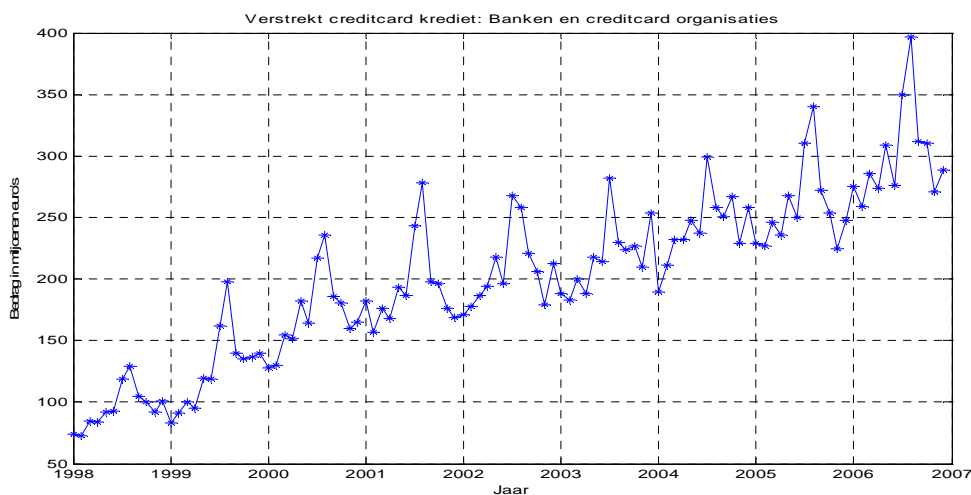
Zoals we in de tabel kunnen zien krijgt M7 het grootste gewicht voor de berekening van het gewogen gemiddelde Q. In het algemeen zullen we een seizoencorrectie met

een M7 van groter dan 1 niet accepteren, ook al heeft Q een waarde van minder dan 1.

Als voorbeeld worden de M en Q waarden van de maandstatistiek van het verstrekt creditcard krediet getoond. Bij de statistiek van het verstrekt creditcard krediet is er sprake van een duidelijk seizoenpatroon (figuur 5). De pieken in juli en augustus hangen samen met de extra uitgaven tijdens de vakantietijd. Na seizoencorrectie geeft X-12-ARIMA de volgende M en Q waarden:

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	Q
0.26	0.14	0.46	0.29	0.67	0.21	0.35	0.38	0.33	0.43	0.43	0.38

Alle M waarden zijn duidelijk kleiner dan 1, wat wil zeggen dat de seizoencorrectie op alle punten als geslaagd mag worden beschouwd. Vervolgens wordt via een gewogen gemiddelde van de M waarden (zonder M2) de Q waarde berekend, welke eveneens duidelijk kleiner is dan 1. De seizoencorrectie wordt vervolgens door X-12-ARIMA geaccepteerd.



Figuur 5. Maandstatistiek van het verstrekt creditcard krediet

Hieronder worden de genoemde kwaliteitsmaten verder toegelicht. De 11 kwaliteitsmaten worden uitgebreider besproken in [1], waarvan we nu een overzicht geven. De berekening en verantwoording van alle M waarden wordt in detail beschreven in Lothian en Morry (1978). Een goed overzicht wordt gegeven in Ladiray en Quenneville (2001, blz. 176-182) en in Vollebregt (2002, blz. 59-68), waarvan we hieronder een zeer beknopte samenvatting geven.

**M1: bijdrage van de onregelmatige component aan fluctuaties**

M1 meet de relatieve bijdrage van de onregelmatige component aan de veranderingen binnen de reeks. Wanneer deze bijdrage groot is, betekent dat dat de onregelmatige component veel meer fluctuaties in de reeks veroorzaakt dan de seizoencomponent. De seizoencomponent en de onregelmatige component kunnen in een dergelijk geval niet goed van elkaar onderscheiden worden.

**M2: bijdrage van de onregelmatige component in de stationaire reeks**

Net als M1 meet ook M2 de bijdrage van de onregelmatige component aan de totale variantie binnen de reeks. Ook wanneer M2 groot is is de onregelmatige component relatief groot. De berekening van M2 verschilt echter ten opzichte van de berekening van M1.

**M3: verhouding van de onregelmatige component tot de trend**

Om een goede seizoendecompositie te kunnen bepalen, is het van belang dat de fluctuaties in de onregelmatige component niet te groot zijn ten opzichte van de fluctuaties in de trend. M3 meet de verhouding tussen de fluctuaties in deze beide componenten.

**M4: samenhang binnen de onregelmatige component**

Eén van de belangrijkste aannamen voor de onregelmatige component is dat er geen samenhang tussen opeenvolgende waarnemingen is. Wanneer er een sterke samenhang is tussen opeenvolgende waarnemingen is deze component immers helemaal niet zo onregelmatig. M4 meet daarom de samenhang binnen de onregelmatige component.

**M5: aantal maanden voor cyclische dominantie (MCD)**

Net als M3 meet ook M5 de veranderingen in de onregelmatige component ten opzichte van de veranderingen in de trend-cycle.

**M6: verhouding van de onregelmatige component tot het seizoen**

Tijdens de eerste twee rekenronden wordt voor de berekening van de seizoencomponent gebruik gemaakt van een 3x5-filter. M6 controleert of het 3x5 filter wel geschikt is voor de gegeven reeks. Een grote waarde van M6 betekent dat de verhouding van de onregelmatige component en de seizoencomponent te klein of te groot is voor het filter.

### **M7: identificeerbaarheid van het seizoenpatroon**

Zoals we in tabel 1 hebben kunnen zien is M7 de belangrijkste kwaliteitsmaat voor de seizoencorrectie. Wanneer M7 groter is dan 1 mogen we de seizoencorrectie in principe niet als zodanig accepteren. M7 geeft aan in hoeverre het seizoeneffect in de reeks identificeerbaar is. Wanneer het seizoeneffect slecht identificeerbaar is, is de absolute fout in de uiteindelijke seizoencomponent groot.

### **M8 tot M11: verandering van het seizoenpatroon over de jaren**

M8 tot en met M11 meten in hoeverre het seizoenpatroon in de reeks aan verandering onderhevig is. Wanneer het seizoenpatroon sterk verandert zijn de seizoenfilters van X-12-ARIMA niet in staat het seizoenpatroon goed te schatten en is de fout in de schattingen groot. Met name wanneer het seizoenpatroon in de eindjaren sterk verandert is het probleem groot, omdat dan de fout in de meest recente schattingen groot is. Het zijn juist de meest recente schattingen waarin gebruikers van de reeks geïnteresseerd zijn. Het seizoenpatroon kan op twee verschillende manieren veranderen. Ten eerste kunnen er min of meer willekeurige fluctuaties in het seizoenpatroon voorkomen. Ten tweede kan er een systematische stijging of daling zijn. M8 en M10 meten de willekeurige fluctuaties in het seizoenpatroon. M9 en M11 meten de systematische stijging of daling in het seizoenpatroon. Hierbij worden M8 en M9 over de gehele reeks berekend. M10 en M11 worden berekend op basis van de meest recente jaren.

Voor elke kwaliteitsmaat geven we een korte beschrijving en een aantal hints om de betreffende waarde omlaag te brengen. Tussen haakjes staat de tabel uit het out-bestand dat gebruikt is om de kwaliteitsmaat te berekenen.

*Tabel 2. Hints om M-waarden omlaag te brengen*

M1 (uit F2B)	Bijdrage van de onregelmatige component aan veranderingen binnen de reeks <ul style="list-style-type: none"><li>• Corrigeer voor uitbijters, werkdageneffecten e.d.</li></ul>
M2 (uit F2F)	Bijdrage van de onregelmatige component in de stationair gemaakte reeks <ul style="list-style-type: none"><li>• Corrigeer voor uitbijters, werkdageneffecten e.d.</li></ul>
M3 (uit F2H)	Verhouding van de onregelmatige component tot de trend (I/C-ratio) M3 is per definitie groot bij een vlakke trend. <ul style="list-style-type: none"><li>• Corrigeer voor uitbijters, werkdageneffecten e.d.</li></ul>
M4 (uit F2D)	Samenhang binnen de onregelmatige component <ul style="list-style-type: none"><li>• Gebruik kortere filters.</li></ul>
M5 (uit F2E)	Aantal maanden voor cyclische dominantie (MCD) M5 is per definitie groot bij een vlakke trend. <ul style="list-style-type: none"><li>• Corrigeer voor uitbijters, werkdageneffecten e.d.</li></ul>
M6	Verhouding van de onregelmatige component tot het seizoen (I/S-ratio) <ul style="list-style-type: none"><li>• Kies een 3x3-seizoenfilter voor een I/S-ratio (tabel F2H) kleiner dan 1.5.</li></ul>

(uit F2H)	<ul style="list-style-type: none"> <li>• Kies een 3x9-seizoenfilter voor een I/S-ratio (tabel F2H) groter dan 6.5.</li> </ul>
M7 (uit F2I)	<p>Identificeerbaarheid van het seizoenpatroon</p> <p>Wanneer het niet lukt M7 kleiner te krijgen dan 1 is de reeks niet geschikt voor correctie</p> <ul style="list-style-type: none"> <li>• Corrigeer voor uitbijters, werkdageneffecten e.d.</li> </ul>
M8	<p>Willekeurige veranderingen in het seizoenpatroon over alle jaren</p> <ul style="list-style-type: none"> <li>• Gebruik langer seizoenfilter</li> <li>• Bij seizoenbreuk: splits de reeks of definieer zelf voorbewerkingsfactoren</li> </ul>
M9	<p>Systematische verandering in het seizoenpatroon over alle jaren</p> <ul style="list-style-type: none"> <li>• Gebruik langer seizoenfilter</li> <li>• Bij seizoenbreuk: splits de reeks of definieer zelf voorbewerkingsfactoren</li> </ul>
M10	<p>Willekeurige veranderingen in het seizoenpatroon in de laatste drie jaar</p> <ul style="list-style-type: none"> <li>• Gebruik langer seizoenfilter</li> <li>• Pas het ARIMA-model aan</li> <li>• Bij seizoenbreuk: splits de reeks of definieer zelf voorbewerkingsfactoren</li> </ul>
M11	<p>Systematische verandering in het seizoenpatroon in de laatste drie jaar</p> <ul style="list-style-type: none"> <li>• Gebruik langer seizoenfilter</li> <li>• Pas het ARIMA-model aan</li> <li>• Bij seizoenbreuk: splits de reeks of definieer zelf voorbewerkingsfactoren</li> </ul>

### 3. Regressie: kalendereffecten en uitbijters

Onder het corrigeren voor kalendereffecten wordt verstaan het corrigeren voor

- werkdagen en koopdagen
- feestdagen en bruidagen
- lengte van de maand
- schrikkeljaareffect

Er wordt voor kalendereffecten gecorrigeerd om de resultaten tussen verschillende maanden of kwartalen beter vergelijkbaar te maken. Een goed voorbeeld hiervan is de lengte van de maand. Januari heeft 31 dagen terwijl februari er 28 of 29 heeft. Als bijvoorbeeld de productie op alle dagen van de maand hetzelfde is, zal er toch een verschil gemeten worden tussen de hoeveelheid geproduceerde goederen in januari en februari. Dit verschil, veroorzaakt door het verschil in het aantal dagen in de maand, is conjunctureel niet interessant en er wordt daarom voor gecorrigeerd. Afgezien van schrikkeljaareffecten zijn de lengte van de maanden ieder jaar hetzelfde. Het is dus een gewoon seizoeneffect dat niet van de andere seizoeneffecten te onderscheiden is en er wordt dus al bij de gewone seizoencorrectie voor gecorrigeerd.

Zeer veel reeksen vertonen niet alleen een jaarlijks terugkerend patroon, maar ook een wekelijks terugkerend patroon. Denk bijvoorbeeld aan de omzet van winkels, die aan het eind van de week meer verkopen dan aan het begin van de week en op zondag vaak gesloten zijn. Een maand of een kwartaal bestaat meestal niet precies uit een geheel aantal weken. Het probleem is dan dat de ene maand een extra vrijdag en zaterdag heeft, terwijl de andere maand misschien een extra maandag en dinsdag heeft. Op vrijdag en zaterdag wordt er meer verkocht en dus is de totale omzet van de ene maand groter dan die van de andere maand. Ook deze werkdageffecten zijn vanuit conjunctureel oogpunt niet interessant en er wordt dus ook voor deze effecten gecorrigeerd.

Bij kalendercorrectie is het essentieel dat men bij het presenteren van de data in detail aangeeft voor welke effecten is gecorrigeerd, daar er zeer veel verschillende manieren bestaan om hiervoor te corrigeren. Voor feestdagcorrectie is er een belangrijk verschil tussen Kerst en Pasen. Het effect dat optreedt rond Kerst treedt ieder jaar op hetzelfde tijdstip op en is daardoor niet te onderscheiden van andere seizoeneffecten. Pasen valt echter niet ieder jaar op hetzelfde tijdstip. Daardoor kunnen eventuele paaseffecten in tegenstelling tot kersteffecten niet altijd door een normale seizoencorrectie worden opgevangen. X-12-ARIMA biedt hiervoor de mogelijkheid om te corrigeren voor paaseffecten met behulp van de regressievariabele die Pasen modelleert. Bij het corrigeren voor paaseffecten wordt aangenomen dat het niveau van de tijdreeks gedurende een vast aantal dagen vóór Pasen op een andere hoogte ligt. Dit aantal dagen wordt door de gebruiker zelf

gespecificeerd. Vervolgens wordt per jaar vastgesteld welke fractie van dit aantal vaste dagen in de maand maart valt en welke fractie in de maand april. Deze fractie wordt dan ingevuld voor de regressievariabele die Pasen modelleert. Voor de andere maanden is deze variabele gelijk aan nul. Een andere mogelijkheid is om het effect van Pasen als een normaal seizoeneffect te berekenen aangezien Pasen bijna altijd in de maand april valt. De maanden waarin Pasen in maart valt worden dan met een hulpvariabele of regressor gecorrigeerd.

Regressievariabelen om werkdageffecten of schrikkeljaareffecten te modelleren zijn standaard beschikbaar in X-12-ARIMA. Het is echter altijd mogelijk dat we een regressievariabele nodig hebben die niet standaard in X-12-ARIMA voorkomt, bijvoorbeeld voor een correctie voor een specifiek effect dat binnen een bepaalde tijdreeks optreedt. We kunnen bijvoorbeeld denken aan de correctie voor de Nederlandse feestdagen. Daarom geeft X-12-ARIMA ook de mogelijkheid om zelf regressievariabelen te definiëren. Zo kan men bijvoorbeeld corrigeren voor het aantal vorstdagen in de bouw. Door deze correctie zal het seizoenpatroon vervolgens beter te bepalen zijn.

De regressiemethode wordt ook gebruikt om te corrigeren voor trendbreuken en andere soorten uitbijters in een tijdreeks.

Voor het corrigeren van kalendereffecten zijn in X-12-ARIMA regressiemethoden en ARIMA-methoden samengebracht. Een regressiemodel is goed in staat om onregelmatige effecten zoals werkdagpatronen en breuken in de trend te bepalen. De onderlinge samenhang tussen de waarnemingen wordt door het regressiemodel niet goed beschreven omdat het model veronderstelt dat de waarnemingen in de tijd onafhankelijk van elkaar zijn. ARIMA-modellen zijn in staat de onderlinge samenhang goed te modelleren en wordt vooral gebruikt voor de seizoencorrectie en voor extrapolatie van de tijdreeksdata om symmetrische trend- en seizoenfilters mogelijk te maken. Door de regressiemodellen met ARIMA-modellen te combineren kunnen we gebruik maken van beide modellen. Het regressiedeel van het model beschrijft de bijzondere effecten, het ARIMA-deel beschrijft de onderlinge samenhang. We noemen een dergelijk gecombineerd model ook wel een RegARIMA-model (zie paragraaf 2.3).

### *Combineren van regressie en ARIMA-modellering*

Het combineren van de twee soorten modellen gaat als volgt. Zoals beschreven bij het onderdeel seizoencorrectie wordt een tijdreeks geëxtrapolleerd via ARIMA-modellering om symmetrische trend- en seizoenfilters mogelijk te maken. We nemen dus al aan dat de tijdreeks zonder werkdag effecten door een ARIMA-model beschreven kan worden. De tijdreeks  $Y_t$  gecorrigeerd voor werkdageffecten

$$Z_t = Y_t - (\beta_0 + \beta_1 X_{1t} + \dots + \beta_p X_{pt}) \quad (3.1)$$

is dus een tijdreeks  $Z_t$  welke via een ARIMA-model beschreven kan worden. Hierin staan de  $\beta$ 's voor de parameters van het regressiemodel en de  $X$ 'en voor de regressievariabelen. Met behulp van een ARIMA(p,d,q)(P,D,Q)-model kunnen we

een goede beschrijving krijgen van de variabele  $Z_t$  en daarmee, in combinatie met het regressiemodel, een goede beschrijving van de oorspronkelijke tijdreeks  $Y_t$ . Zowel de parameters van het ARIMA-model als de parameters van het regressiemodel worden geschat, vandaar de naam RegARIMA.

Bovenstaande wijze van modelleren laat zien dat het ook voor het bepalen van een goede seizoendecompositie nodig is te corrigeren voor werkdageffecten, los van de vraag of men de uiteindelijke reeks voor werkdageffecten wil corrigeren of niet.

### *Werkdagcorrectie*

Er zullen twee veel gebruikte methoden worden besproken om te corrigeren voor werkdageffecten en het schrikkeljaareffect. Men kan X-12-ARIMA zelf laten kiezen voor één van de beide manieren.

Het intuïtief eenvoudigste model om werkdageneffecten te modelleren is als volgt:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_7 X_{7t} + \varepsilon. \quad (3.2)$$

In dit model is er voor elke werkdag een regressievariabele  $X_i$ , die gelijk is aan het aantal keer dat de betreffende werkdag in de maand voorkomt.  $X_{1t}$  is dus het aantal maandagen in maand  $t$ ,  $X_{2t}$  het aantal dinsdagen en  $X_{7t}$  het aantal zondagen. Er zijn in dit model 7 parameters om te schatten. Merk op dat deze 7 parameters vast zijn voor de gehele periode. We veronderstellen dus dat de weekdageffecten niet veranderen over de tijd. Er moet nu nog worden gecorrigeerd voor het feit dat februari niet ieder jaar even veel dagen bevat (schrikkeljaareffect). Bij deze methode wordt dit gedaan via een transformatie. Waarnemingen voor februari worden daartoe vermenigvuldigd met een factor  $28.25/m_t$ . Daarbij is 28.25 de gemiddelde lengte van de maand februari, en  $m_t$  de lengte van de gegeven maand februari.

In het bovenstaande model is impliciet de lengte van de maand meegenomen. Immers, de lengte van de maand is de som van de regressievariabelen  $X_1 \dots X_7$ . Het aantal dagen dat een maand bevat is echter een gewoon seizoeneffect dat al wordt gecorrigeerd in de normale seizoencorrectie. In die zin bevat het bovenstaande regressiemodel dus redundante informatie.

We kunnen de 7 genoemde variabelen daarom vervangen door 6 contrastvariabelen. Dit geeft het volgende model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_6 X_{6t} + \varepsilon. \quad (3.3)$$

Hierin is  $X_{1t}$  het verschil tussen het aantal maandagen en het aantal zondagen,  $X_{2t}$  het verschil tussen het aantal dinsdagen en het aantal zondagen en  $X_{6t}$  het verschil tussen het aantal zaterdagdagen en het aantal zondagen. Het aantal dagen in de maand wordt nu niet meer gemodelleerd, maar verder bevat het nieuwe model precies even veel informatie als het oude. Om de uitvoer beter interpreteerbaar te maken, rekent X-12-ARIMA de zes contrastvariabelen om naar zeven coëfficiënten voor de afzonderlijke dagen van de week. Dit staat weergegeven in tabel 3. Deze getallen geven de afwijking t.o.v. de gemiddelde dag, en tellen op tot nul.



Tabel 3. *Werkdagen in de uitvoer van X-12-ARIMA*

Regression Model			
Variable	Parameter Estimate	Standard Error	t-value
Trading Day			
Mon	-0.0021	0.00671	-0.31
Tue	0.0241	0.00664	3.62
Wed	0.0153	0.00659	2.32
Thu	0.0024	0.00651	0.37
Fri	0.0092	0.00658	1.39
Sat	-0.0213	0.00691	-3.09
*Sun (derived)	-0.0275	0.00689	-3.99

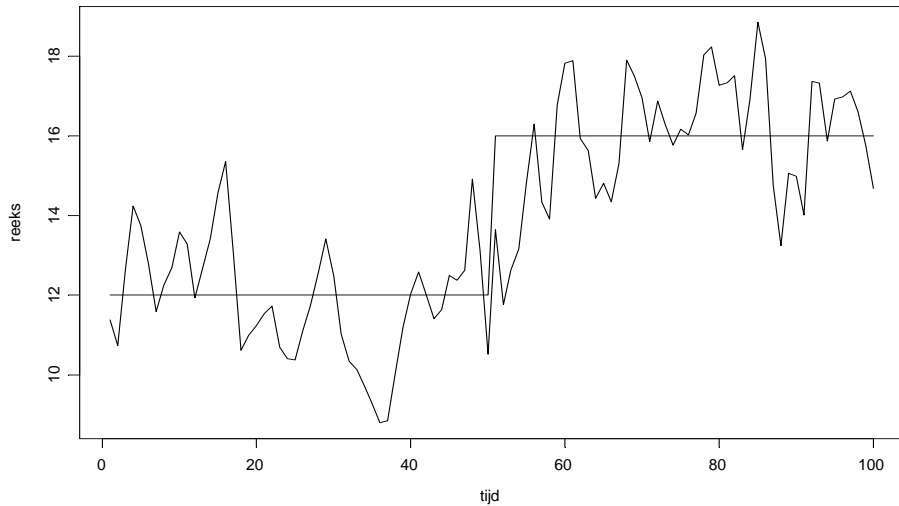
\*For full trading-day and stable seasonal effects, the derived parameter estimate is obtained indirectly as minus the sum of the directly estimated parameters that define the effect.

Er moet nu alleen nog worden gecorrigeerd voor het schrikkeljaareffect. We kunnen daarvoor een extra dummy-variabele *lpyear* (schrikkeljaar) aan het model toevoegen. Daarmee gebruikt ook deze methode van werkdagcorrectie 7 regressievariabelen.

We hebben nu dus twee methoden om te corrigeren voor werkdageneffecten en het schrikkeljaareffect, maar de gebruiker moet één van de twee methoden in X-12-ARIMA instellen. De methode waarbij zondag als contrastvariabele wordt gebruikt voor de andere weekdays leidt echter tot een betere werkdagcorrectie en is ook de standaardmethode van het CBS.

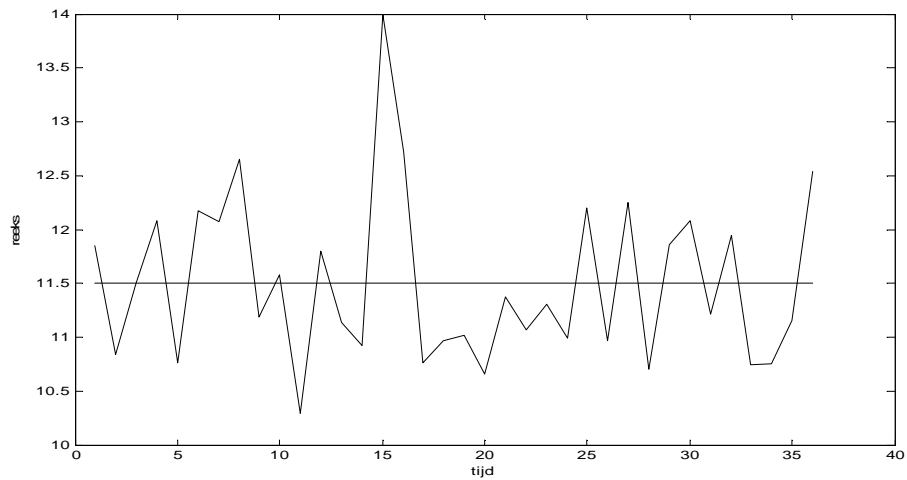
### *Uitbijters*

De regressiemethode wordt ook gebruikt om te corrigeren voor trendbreuken en andere uitbijters in een tijdreeks. Een uitbijter in een tijdreeks kan ernstige gevolgen hebben voor het schatten van de seizoeneffecten. Het is in X-12-Arima mogelijk uitbijters automatisch te laten detecteren. Bekende uitbijters kunnen echter ook door de gebruiker worden gespecificeerd. Er worden drie typen uitbijters onderscheiden. Eén hiervan is de uitbijter die wordt veroorzaakt door een trendbreuk (zie figuur 6). Een trendbreuk kan bijvoorbeeld ontstaan als gevolg van een verandering in wetgeving zoals bijvoorbeeld een BTW-verhoging. Hierbij zal het gemiddelde niveau van de cijferreeks voor altijd veranderen. Trendbreuken worden uitgebreid behandeld in het thema *Methodebreuken van de Methodenreeks* (Van den Brakel e.a., 2010).



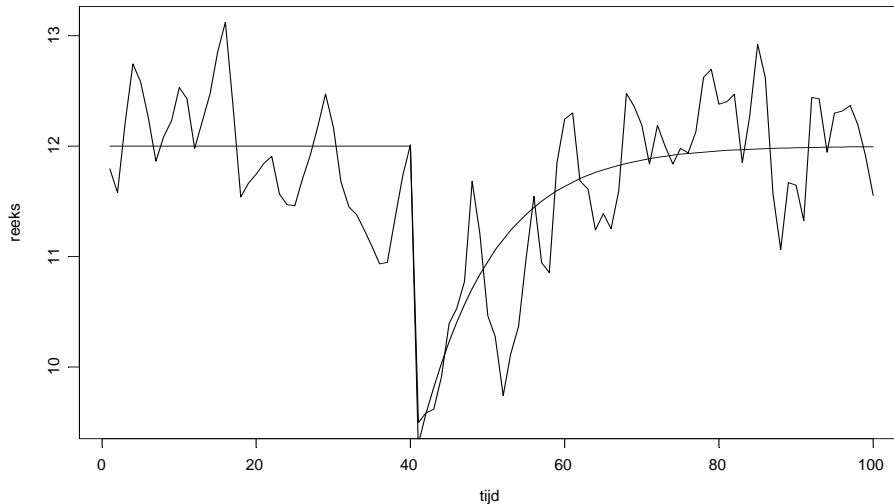
*Figuur 6. Een reeks met een trendbreuk*

Daarnaast zijn er nog twee typen uitbijters te modelleren, namelijk de additieve uitbijter en de tijdelijke niveauverandering. Een additieve uitbijter is bijvoorbeeld een plotselinge piek in het aardgasverbruik bij een extreem strenge winter (zie figuur 7). Een dergelijke uitbijter is slechts op één plaats in de reeks te zien.



*Figuur 7. Een reeks met een additieve uitbijter*

Een tijdelijke niveauverandering is een effect dat plotseling optreedt en dan vervolgens langzaam wegebt. (zie figuur 8).



*Figuur 8. Een reeks met een tijdelijke niveauverandering*

Wanneer we weten op welk tijdstip een uitbijter in een cijferreeks voorkomt kunnen we de bijbehorende regressievariabele zelf specificeren. Aangezien X-12-ARIMA de mogelijkheid biedt veel typen uitbijters automatisch te detecteren is het zelf specificeren van uitbijters, afgezien van speciale gevallen, meestal niet nodig. Het automatisch detecteren van uitbijters gebeurt door op elk tijdstip en voor elk type uitbijter een t-waarde te bepalen d.m.v. een t-toets. Een uitbijter is significant als de t-waarde groter is dan een vooraf opgegeven waarde. Als er meerdere significante uitbijters zijn, wordt eerst de meest significante uitbijter aan het model toegevoegd. Vervolgens worden de t-waarden van andere uitbijters opnieuw uitgerekend en wordt weer de meest significante uitbijter aan het model toegevoegd. Dit proces herhaalt zich tot er geen significante uitbijters meer over zijn.

Zoals gezegd bemoeilijken uitbijters het bepalen van een seizoencomponent en maken ook geen onderdeel uit van de seizoencomponent. Daarom worden uitbijters na de seizoencorrectie weer toegevoegd aan de seizoengecorrigeerde data. Een uitbijter kan namelijk belangrijke informatie geven over de data, bijvoorbeeld het effect van een staking of een uitzonderlijk hete zomer op een tijdreeks.

### *Kwaliteitsindicatoren*

Om de kwaliteit van de kalendercorrecties of uitbijters te beoordelen wordt voor alle regressievariabelen een t-toets gedaan. Een regressievariabele of uitbijter is significant als de bijbehorende absolute t-waarde groter is dan een vooraf bepaalde kritieke waarde. Wanneer er geen sprake is van een significante regressievariabele ligt de t-waarde dicht bij 0. Voor kalendereffecten is de t-waarde significant bij  $\geq \pm 1.96$ . Men kan bij uitbijters ervoor kiezen de kritieke waarde voor de t-toets zelf in te stellen. Wanneer er geen waarde wordt ingesteld wordt een standaardwaarde genomen. Deze is afhankelijk van de lengte van de reeks en varieert van 3.65 voor

een reeks van 48 waarnemingen tot 4.08 voor een reeks van 360 waarnemingen. Een voorbeeld van t-waarden is te zien in de laatste kolom van tabel 3. Men ziet dat als de t-waarde dicht bij 0 ligt de standaardfout (kolom 3) relatief groot is t.o.v. de geschatte parameter (kolom 2).

#### *Fijngeregelde uitbijterdetectie*

De kritieke T-waarde om het aantal uitbijters te detecteren kan ingesteld worden aan de hand van het aantal waarnemingen (zie de X-12 manual). Wanneer er visueel duidelijk nog uitbijters aanwezig zijn dient de drempelwaarde met 0,1 verlaagd te worden. Dit totdat de uiteindelijke drempelwaarde die is waarbij niet teveel uitbijters gedetecteerd worden. Het verlagen van de drempelwaarde kan bovendien tot een verbetering van het totale regressiemodel leiden. Belangrijk is dat als er eenmaal uitbijters gedetecteerd zijn, ze naderhand in de tijdreeks significant blijven. Dit kan via de “add one” methode van uitbijterdetectie ingesteld worden. Voor grootschalige productie is fijngeregelde uitbijterdetectie (automatisch en handmatig) echter nogal arbeidsintensief. Bij seizoencorrectie van macro-economische indicatoren is fijngeregelde uitbijterdetectie eigenlijk een must.

## 4. Overige aspecten

### 4.1 Pre-treatment

De voorbereiding of pre-treatment van de data heeft tot doel een betere bepaling van de seizoencomponent mogelijk te maken. Enkele van deze voorbereidingen zijn reeds besproken. Zowel werkdagcorrectie, feestdagcorrectie als uitbijtercorrectie behoren tot de pre-treatment fase van de databewerking. Beide maken een betere seizoencorrectie mogelijk en moeten worden uitgevoerd ook als men in de seizoengecorrigeerde data de werkdageffecten en uitbijters weer teruglegt.

Het is echter ook mogelijk dat andere bekende niet-seizoengebonden effecten een rol spelen in de data, waardoor het lastig wordt de seizoencomponent goed te schatten. Hierbij kan men denken aan een ongewoon hete zomer, het effect van vakantiespreiding en bruggen. Al deze effecten kan men corrigeren door het specificeren van een regressievariabele. Natuurlijk wordt er vervolgens alleen voor deze effecten gecorrigeerd als de bijbehorende regressievariabelen significant zijn (een hoge t-waarde hebben).

Het is daarom altijd van belang in de pre-treatment fase de tijdreeks te plotten. Hierbij kan men dan een eerste indruk krijgen van vreemde waarden in de data zoals nulwaarden of uitzonderlijk hoge waarden, trendbreuken, etc. Ook kan men hier nagaan of er eigenlijk wel sprake is van een trend-cycle, een seizoencomponent of een duidelijke onregelmatige component. Eventueel kan er ook naar het spectrum van de data gekeken worden of naar autocorrelaties. Natuurlijk is een visuele inspectie van een groot aantal tijdreeksen arbeidsintensief en staat het een geautomatiseerde seizoencorrectie in de weg. Toch is het een belangrijk onderdeel van seizoencorrectie, zeker voor de belangrijkste tijdreeksen.

### 4.2 Revisies

Er zijn twee hoofdredenen waarom er op gezette tijden revisies van seizoengecorrigeerde data plaatsvinden. Ten eerste kunnen revisies van seizoengecorrigeerde data plaatsvinden omdat er vernieuwde informatie beschikbaar is gekomen in de ruwe niet seizoengecorrigeerde data. Een probleem hierbij is dat één nieuwe waarneming gevolgen kan hebben voor jaren van seizoengecorrigeerde data. Ten tweede kunnen er nieuwe parameter instellingen worden toegepast bij het seizoencorrigeren van de data. Dit kan gevolgen hebben voor alle seizoengecorrigeerde data. Veranderen van reeds gepubliceerde cijfers kan voor gebruikers verwarrend werken. Er moet dus een balans gevonden worden tussen de nauwkeurigheid en tijdigheid van de seizoengecorrigeerde data en de stabiliteit van de seizoengecorrigeerde data over de tijd. Om dit te bereiken moet er een revisiestrategie worden ontwikkeld voor de tijdreeksen die gepubliceerd worden. Hierbij moet gekeken worden naar de behoefte van gebruikers en de beschikbaarheid van bronmateriaal. Vooral de frequentie en de grootte van de

revisies moeten van te voren worden vastgelegd. Ook moet duidelijk zijn welk deel van de seizoengecorrigeerde data door een revisie wordt beïnvloed en welk deel buiten schot blijft.

Bij de “current adjustment” methode worden de seizoencorrectieparameters op vaste tijden vastgelegd en daar tussen in vastgehouden. Bij de “concurrent adjustment”-methode worden seizoencorrectieparameters elke keer als er nieuwe data beschikbaar zijn aangepast. De eerste methode leidt tot minder revisies maar zal in het algemeen minder nauwkeurige resultaten geven. De tweede methode leidt tot betere seizoencorrectie maar zal tussendoor vaker leiden tot aanpassingen van de seizoengecorrigeerde data. In het algemeen wordt aangeraden ten minste één keer per jaar de seizoencorrectie methoden en parameters te herzien en indien nodig aan te passen.

### **4.3 Samengestelde reeksen**

Sommige tijdreeksen zijn samengesteld (geaggregeerd) uit onderliggende tijdreeksen. Als voorbeeld kunnen we uitgaan van een reeks (C) die is opgebouwd uit twee onderliggende tijdreeksen (A en B). We gaan er dus vanuit dat de ongecorrigeerde samengestelde reeks (C) wordt verkregen via  $C=A+B$ . Voor het seizoencorrigeren zijn er nu verschillende mogelijkheden.

Bij de directe methode worden alle reeksen (A, B en C) afzonderlijk (direct) seizoengecorrigeerd. Het gevolg hiervan is dat de som van de direct seizoengecorrigeerde onderliggende reeksen over het algemeen niet precies gelijk zal zijn aan de direct seizoengecorrigeerde samengestelde reeks (A direct gecorrigeerd + B direct gecorrigeerd  $\neq$  C direct gecorrigeerd). Dit verschil kan bij publicatie van alle seizoengecorrigeerde reeksen door de gebruiker natuurlijk eenvoudig worden vastgesteld door de seizoengecorrigeerde deelreeksen bij elkaar op te tellen en dan te vergelijken met de seizoengecorrigeerde samengestelde reeks.

Bij de indirecte methode worden alleen de onderliggende reeksen (A en B) seizoengecorrigeerd en vervolgens opgeteld (A direct gecorrigeerd + B direct gecorrigeerd = C indirect gecorrigeerd). Wat nu kan voorkomen is dat de directe seizoencorrectie van de samengestelde reeks (C direct gecorrigeerd) duidelijk afwijkt van de indirect seizoengecorrigeerde reeks (C gecorrigeerd  $\neq$  C indirect gecorrigeerd).

Binnen het CBS wordt vooral de directe methode gebruikt maar hier en daar ook de indirecte methode. De mogelijke afwijkingen bij de directe methode worden dan bewust geaccepteerd.

## 5. Stappenplan seizoencorrectie

Seizoencorrectie met X-12 verloopt in een bepaalde volgorde. Eerst worden cijferreeksen door de analist geïnspecteerd, dan wordt met X-12 de tijdreeks voorbehandeld waarna het seizoenpatroon geëxtraheerd kan worden. Stapsgewijs ontleedt X-12 tijdreekscomponenten. Het doel van dit hoofdstuk is om deze ontleding inzichtelijk te maken door dit te beschrijven en vervolgens met het uitwerken van een voorbeeld dit concreter te maken. Eerst wordt het ontledingsproces beschreven aan de hand van twee schema's en de opeenvolgende stappen. Dit alles wordt met een voorbeeld gedemonstreerd.

### 5.1 Het splitsen van een cijferreeks in componenten

Tabel 4 geeft het ontledingsproces van een cijferreeks met X-12-ARIMA schematisch weer vanaf de invoer tot de uitvoer. De stap "Enhanced X-11" is de feitelijke seizoencorrectie, en is een verdere ontwikkeling van de methode X-11.

Tabel 4. Schema methode seizoencorrectie X-12 software

Invoer			Oorspronkelijke tijdreekscijfers Y
Verwerking X-12	Transform	<b>Decompositie</b> •multiplicatief •additief	$Y = S * TC * I$ $Y = S + TC + I$
	RegARIMA	<b>Regressie</b>  en  <b>ARIMA</b> •trendfilters •seizoenfilters	modelleren extraheren voorspellen: •kalendereffecten •speciale effecten •voorspellingen •uitbijters
	Enhanced X-11	<b>Filtering</b> Voortschrijdende gemiddelden	Extraheren van: •seizoenpatroon •trendcyclus •witte ruis
Uitvoer			•Correctiefactoren •Gemiddelde dagcijfers •Seizoenvrije cijfers •Trend-cyclus waarden • Witte ruiswaarden •Voorspelde waarden Y

Tabel 5 laat zien welke tijdreekscomponenten aanwezig kunnen zijn in een cijferreeks en met welke frequentie ze zich voordoen, te beginnen met de component met de hoogste frequentie tot aan de component met de traagste frequentie (zie paragraaf 2.1 voor een gedetailleerde bespreking).

*Tabel 5. Schema Tijdreekscomponenten en hun frequentie*

<b>Hoofdcomponenten</b>	<b>Subcomponenten</b>	<b>Frequentie</b>
Toevalscomponent (irregular)	Witte ruis	hoog
Incidentele component (outlier; specificities)	Eenmalige verandering Tijdelijke verandering Breukverandering	Incidenteel Incidenteel Incidenteel
Kalendercomponent (calendar)	Weekdagen Feestdagen Schrikkeljaar	Jaarlijks niet systematisch Jaarlijks, vast en schuivend Vierjaarlijks
Seizoencomponent (Seasonality)	Seizoenpatroon Maandlengte Kwartaallengte	Jaarlijks, systematisch
Cycluscomponent (Business cycle)	Conjunctuur	Middellange termijn 2 á 10 jaar, periodiek, niet systematisch
Trendcomponent (Trend)	Lange termijn beweging	Oneindig

In de praktijk van seizoencorrectie met X-12 zijn er globaal vier stappen te onderscheiden:

- 1 Kennis van de statistische tijdreeks,
- 2 Grafische inspectie van de tijdreeks,
- 3 X-12 - Voorbehandeling (RegARIMA),
- 4 X-12 - Seizoencorrectie (Enhanced X-11).

De in stap 3 en 4 gebruikte diverse statistische maten en diagnostieken zullen uitgebreider besproken bij het uitwerken van een voorbeeld. Elk van deze vier



stappen wordt nu in het kort beschreven. Voor een gedetailleerde beschrijving van X-12 kan de syllabus van Vollebregt geraadpleegd worden.

## **5.2 Kennis van de statistische tijdreeks**

De analist moet de betreffende tijdreeks kennen of informatie hierover verwerven. Dit is niet onbelangrijk voor een succesvolle seizoencorrectie. Wanneer er bijvoorbeeld breuken zijn in de statistische waarneming als gevolg van een definitiewijziging, doordat een BTW-verandering, of stakingen, extreme weersomstandigheden, of varkenspest, enz. de data hebben vertekend, kan bij de berekening van het seizoenpatroon hiermee rekening worden gehouden. Verder zal hij ook moeten weten of de betreffende statistiek mede wordt bepaald door kalenderinvloeden zoals werkdageffecten of koopdageffecten en of dit door vakantiespreiding, brugdagen of koopdagenpatroon in de loop der jaren beïnvloed is. Ook de lengte van de werkweek, vijfdaags, zesdaags of volcontinu kan relevant zijn voor hoe het werkdagenpatroon wordt berekend. Over het algemeen geldt dat in economische tijdreeksen kalenderfactoren aanwezig zijn, terwijl dit in arbeidsmarktstatistieken zelden of nooit zal zijn.

## **5.3 Grafische inspectie van de tijdreeks**

Door grafisch naar een statistische cijferreeks te kijken krijgt de analist een eerste indruk van aanwezige componenten en in combinatie met wat hij weet van de tijdreeks zelf. Is er een duidelijk jaarlijks seizoenpatroon aanwezig zoals typisch is voor de horeca, toerisme, landbouw? Of heeft de reeks zichtbaar een conjuncturele beweging zoals kenmerkend is voor veel arbeidsmarktstatistieken of het consumentenvertrouwen. Of wordt het zicht op het seizoenpatroon of de conjuncturele beweging vertroebeld door grote of relatief veel toevallige schommelingen zoals kenmerkend is voor productie- en omzetreksen? Zijn er verder ook uitbijters te zien al dan niet gebaseerd op voorkennis over de cijferreeks als gevolg van een strenge winter, een staking, mond en klauwzeer, enz.? En als er een seizoenpatroon visueel te zien is, wat voor seizoendecompositie lijkt voor de hand te liggen, een additieve of een multiplicatieve? De meeste economische cijferreeksen hebben een multiplicatief seizoenpatroon. Conjunctuurtestuitkomsten of cijfers met negatieve waarden of uitkomsten dienen additief ontleed te worden.

## **5.4 Voorbehandeling met X-12**

Na stap 1 en 2 begint de feitelijke seizoencorrectie met X-12.

Eén van de dingen die als eerste bepaald wordt is hoe het seizoenpatroon in een cijferreeks berekend moet worden, of multiplicatief of additief. Hoewel informatie uit stap 1 en 2 aangewend dient te worden heeft het de voorkeur om in dit stadium eerst default instellingen te gebruiken en deze in volgende sessies met X-12 fijn te regelen.

Stel dat het een economische cijferreeks betreft, bijvoorbeeld omzet of productie, dan zal in de eerste run

- 1) de decompositie voor het seizoenpatroon op ‘automatisch’ worden ingesteld,
- 2) voor de aanwezigheid van kalenderinvloeden 6 wekdageffecten en het schrikkeljaareffect worden aangevinkt,
- 3) de detectie van uitbijters aangevinkt worden en
- 4) voor de seizoencorrectie X-11 ingesteld worden.

Hoewel het in dit stadium dus nog niet gaat om seizoencorrectie in enge zin, is dit laatste nu al nodig. Op basis van een voorlopig seizoengecorrigeerde cijferreeks moet namelijk vastgesteld worden of X-12 op basis van automatische instellingen inderdaad alle (bekende) aanwezige uitbijters heeft verwijderd. Als dit namelijk niet gebeurt dan zal hierdoor het seizoenpatroon in stap vier niet optimaal gekwantificeerd worden. Er kan bijvoorbeeld een verkeerde decompositie plaats vinden en/of elementen van niet verwijderde uitbijters zorgen er voor dat bijvoorbeeld het werkdagenpatroon maar ook het seizoenpatroon niet scherp berekend worden. Gaat het om een arbeidsmarktstatistiek dan begint standaard voorcorrectie zonder kalendercorrectie. Zouden er toch kalendereffecten in de reeks aanwezig zijn dan geeft X-12 dit aan met een zogenaamd “warning error”. Uitkomsten van stemmingstatistieken zoals het producenten- en consumentenvertrouwen worden nooit gecorrigeerd voor kalenderinvloeden.

In volgende sessies van X-12 kan alle informatie worden verwerkt welke voortvloeit uit voorgaande run, de kennis die men heeft van de cijferreeks en de visuele inspectie. Een dergelijk robuuste causale voorcorrectie van kalenderfactoren en speciale effecten met de regressiemodule van X-12 is essentieel om een zo scherp mogelijk beeld te krijgen van de onderliggende ontwikkeling in de zogenaamde jaar op jaar mutaties van de oorspronkelijke statistische maandcijfers. Met deze gemiddelde dagcijfers kan het seizoenpatroon vervolgens netter worden berekend door X-12.

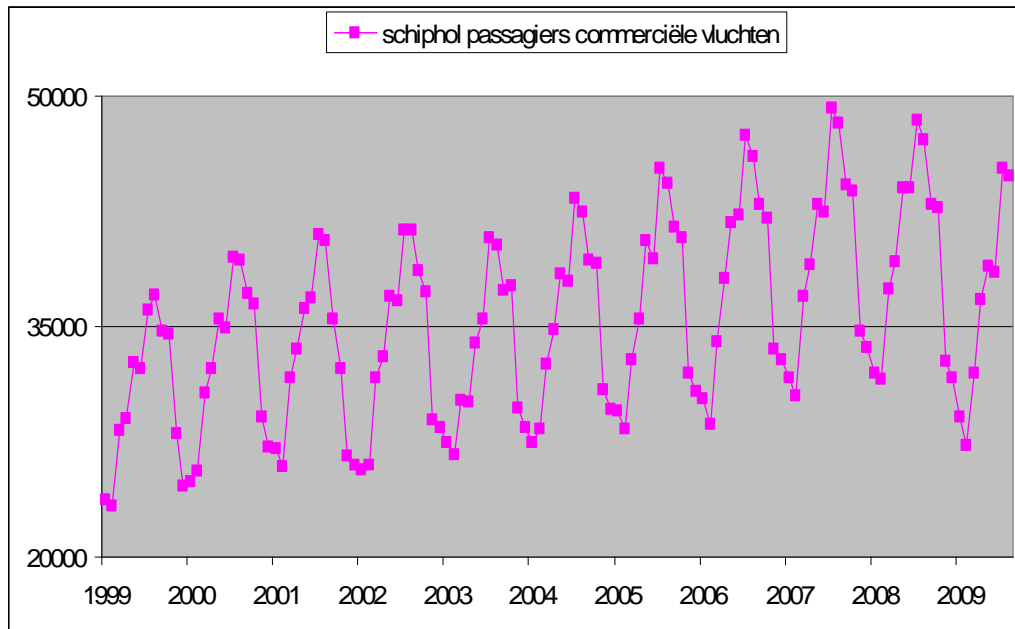
### **5.5 Seizoencorrectie met X-12**

Nadat de cijferreeks doeltreffend is voorbehandeld, dat wil zeggen gecorrigeerd is voor alle aanwezige speciale effecten (uitbijters, kalenderinvloeden), kan de reeks voor aanwezige seizoeninvloeden worden gecorrigeerd. Hiervoor gebruikt X-12 trendfilters en seizoenfilters in de vorm van voortschrijdende gemiddelden. Om deze filters symmetrisch aan te wenden wordt de reeks geëxtrapoleerd met een ARIMA-model welke tijdens de voorbehandeling als geschikt hiervoor is bevonden.

Het seizoenpatroon wordt in drie iteratieve rekenrondes via een multiplicatieve of additieve decompositie berekend (zie Vollebregt, 2002, bladzijde 81 en verder voor meer details). Doordat in elke rekenronde nog aanwezig extreme waarden worden geïdentificeerd en vervangen, kan uiteindelijk een zo stabiel mogelijk seizoenpatroon worden berekend.

## 5.6 Praktijkvoorbeeld seizoencorrectie met X-12

Als voorbeeld van seizoencorrectie met X-12 wordt gebruikt de ‘passagiersaantallen commerciële vluchten Schiphol’ (zie figuur 9). Dit zal conform de genoemde vier stappen worden uitgewerkt. Met tabellen en figuren worden correctie-instellingen, diagnostieken en uitvoer van X-12 weergegeven, van een run met automatische instellingen tot een run aangevuld met fijn geregelde correctiespecificaties.



Figuur 9. Passagiers Schiphol commerciële vluchten (x100)

### 5.6.1 Stap 1 Passagiersvervoer Schiphol - voorkennis

Angst voor een millenniumbug in de gecomputeriseerde apparatuur, verschillende terroristische aanslagen (2001 V.S; 2002 Bali), de Sars-uitbraak en oorlog in Irak begin 2003, brandstofprijzen die de pan uitreizen, bekend is dat één of meer van deze factoren het commerciële luchtvervoer van passagiers zo nu en dan sterk hebben vertekend. Bekend is ook dat het passagiersvervoer Schiphol in de jaren negentig zeer fors groeide, twee-digitcijfers waardoor een multiplicatieve decompositie waarschijnlijk is. Mogelijk kan er ook sprake zijn van een vliegdagenpatroon. Zaterdag is vaak een populaire dag om te vliegen voor vakantie reizen, maar juist minder voor zakenreizen. Een extra zaterdag in de maand kan dus per saldo leiden tot een groter of kleiner aantal reizigers dan in andere maanden. Gelet op het zeer grote aantal vluchten per dag c.q. vervoerde passagiersaantallen zal er ook zeker sprake zijn van een schrikkeljaareffect.

### 5.6.2 Stap 2 Passagiersvervoer Schiphol - visuele inspectie

Een zeer sterk seizoenpatroon is duidelijk te zien; waarschijnlijk is dit multiplicatief, gelet op het toenemen van de seizoenuitslagen met het (overwegend) stijgend niveau. Nauwgezet inspecterend voor een millenniumeffect is te zien dat alleen december 1999 er een dieptepunt is terwijl dit steevast in de maand januari ligt. De

cijfers zijn ook in 2001 en 2002 vertekend door de aanslagen van september 2001 en vanaf het voorjaar van 2003 door de uitbraak van Sars en de oorlog in Irak. Voorts kunnen eind 2008 en 2009 buitengewoon forse (conjuncturele) dalingen worden waargenomen.

### 5.6.3 *Stap 3 Passagiersvervoer Schiphol – pre-treatment met X-12*

In tabel 6 zijn de default correctie-instellingen van de eerste, voorlopige run met X-12 vermeld.

Tabel 7 geeft aan dat op basis van een default run een additieve berekening van het seizoenpatroon moet plaats vinden. Het regressie-ARIMA-model wordt in tabel 8 vermeld. Daarin is het vliegdagenpatroon over het algemeen niet significant. Een maandag heeft een significante T-value van 1,96 ( T-value  $\geq \pm 1.96$ ). De overige weekdays zijn bijna tot niet significant. De waarde voor zaterdag is bijna significant, het negatieve teken van een extra zaterdag geeft aan dat zaterdag ongunstig is voor passagiersvervoer door de lucht. Het schrikkeljaareffect is zoals verwacht inderdaad significant. Verder geeft tabel 8 ook aan dat in de Schipholreeks de aanslagen van 2001 de passagiersaantallen significant hebben beïnvloed ( $\geq \pm -3,6$ ) en dat dit van langere duur is geweest (Temporary Change). Wel is problematisch dat X-12 hiervoor 2 direct op elkaar volgende “tijdelijke veranderingen” toepast. In april 2009 is bovendien significant ( $\geq \pm -3,3$ ) een eenmalige uitbijter (AO) gedetecteerd wat kan samenhangen met de sterk verslechterde wereldconjunctuur. Het ARIMA-model voor het symmetrisch toepassen van trend en seizoenfilters bij het berekenen van seizoenpatroon heeft parameters (0,1,0)(1,1,0). De angst voor een millenniumbug en de uitbraak van Sars in combinatie met de oorlog in Irak zijn echter niet automatisch gedetecteerd, terwijl die de cijfers wel degelijk hebben beïnvloed.

*Tabel 6. Default X-12 setup*

```

Line #
1: SERIES{
2:           title = "000000001schiphol default"
3:           Format = "DateValue"
4:           Period = 12
5:           File = "F:\Output\schipholpassagiers.DAT"
6:           name = "schiph"
7:           precision = 0
8:           decimals = 1}
10: TRANSFORM{
11:          Function = auto
12:          adjust = none
13:          savelog=atr }
15: FORCE{
16:          lambda = 0.0
17:          usefcst = yes}
19: OUTLIER{
20:          types = ( ao ls tc )
21:          critical = (3.3,3.9,3.6)
22:          method = addone}
24: REGRESSION{
25:          Variables =(
26:          tdnolpyear
27:          lpyear
28:          )
29:          aictest=(tdnolpyear)}
31: AUTOMDL{
32:          acceptdefault = no
33:          checkmu = yes
37:          maxdiff = (2 1)
38:          maxorder = (2 1)
39:          savelog=automodel}
41: X11{
42:          seasonalma = msr
43:          final=(ao tc ls)
44:          savelog=(m7 m10 m11 msr)}

```

*Tabel 7. Default X-12 decompositie*

```

Likelihood statistics for model fit to untransformed series.
Likelihood Statistics
-----
Number of observations (nobs)                128
Effective number of observations (nefobs)    115
Number of parameters estimated (np)         10
Log likelihood (L)                          -949.7035
AIC                                          1919.4069
AICC (F-corrected-AIC)                    1921.5223
Hannan Quinn                              1930.5485
BIC                                         1946.8563
-----

Likelihood statistics for model fit to log transformed series.
Likelihood Statistics
-----
Number of observations (nobs)                128
Effective number of observations (nefobs)    115
Number of parameters estimated (np)         10
Log likelihood                             253.6996
Transformation Adjustment                   -1204.6425
Adjusted Log likelihood (L)                -950.9430
AIC                                          1921.8860
AICC (F-corrected-AIC)                    1924.0014
Hannan Quinn                              1933.0275
BIC                                         1949.3353
-----

Additive seasonal adjustment will be performed.

```

Tabel 8. Default X-12 Regressie- en ARIMA-model

```

Automatic model choice : (0 1 0)(1 1 0)
Regression Model
-----
Variable                Parameter      Standard      t-value
                        Estimate        Error
-----
Trading Day
  Mon                   187.3104      95.78496      1.96
  Tue                   -175.5784     92.09313     -1.91
  Wed                   -95.2893      91.11718     -1.05
  Thu                   16.0564       91.28630      0.18
  Fri                   41.4512       86.45563      0.48
  Sat                   -148.5821     86.39017     -1.72
  *Sun (derived)       174.6318      89.61313      1.95

Leap Year                750.2473      271.17710      2.77

Automatically Identified Outliers
  TC2001.Sep           -2540.6085     581.18759     -4.37
  TC2001.Oct           -3471.7462     581.42101     -5.97
  AO2009.Apr           2777.0059      555.27658      5.00
-----
*For full trading-day and stable seasonal effects, the derived
parameter estimate is obtained indirectly as minus the sum
of the directly estimated parameters that define the effect.

Chi-squared Tests for Groups of Regressors
-----
Regression Effect        df      Chi-Square      P-Value
-----
Trading Day                6          24.97          0.00
Combined Trading Day and Leap Year Regressors
                          7          34.11          0.00

ARIMA Model: (0 1 0)(1 1 0)
Nonseasonal differences: 1
Seasonal differences: 1

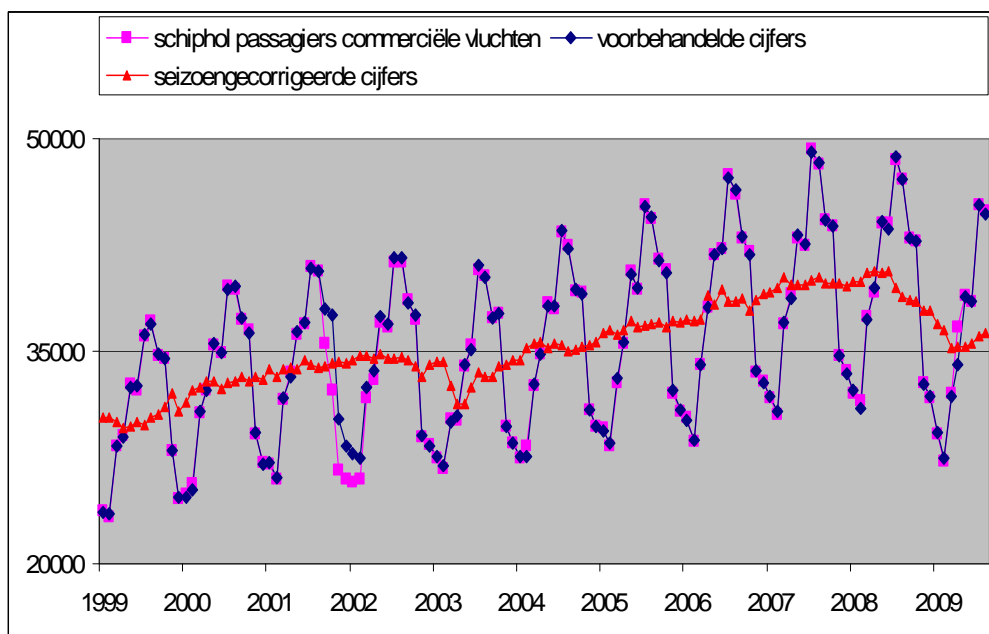
Parameter                Estimate      Standard
                        Estimate      Errors
-----
Seasonal AR
  Lag 12                 -0.4224      0.08955

Variance                 0.54785E+06
SE of Var                 0.72249E+05
-----

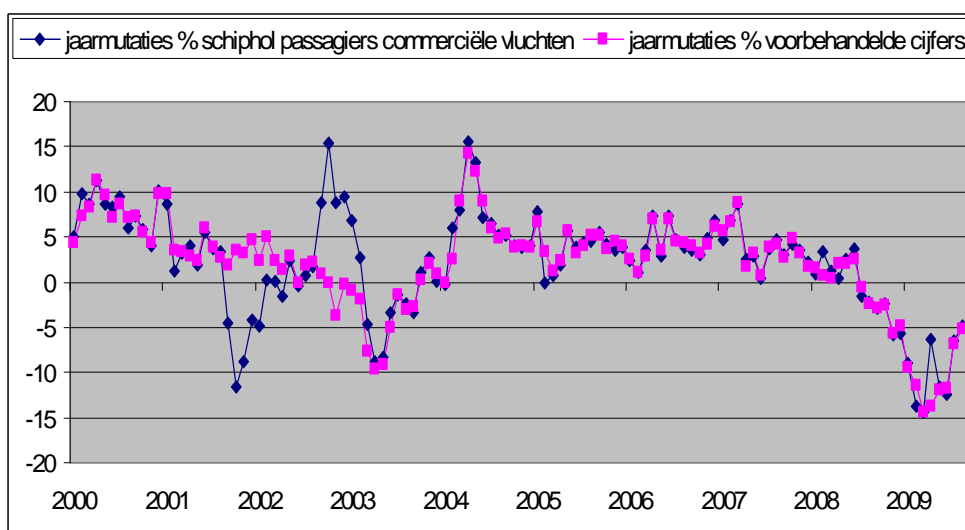
Likelihood Statistics
-----
Number of observations (nobs)                128
Effective number of observations (nefobs)     115
Number of parameters estimated (np)           12
Log likelihood (L)                           -924.1483
AIC                                             1872.2965
AICC (F-corrected-AIC)                       1875.3554
Hannan Quinn                                  1885.6664
BIC                                             1905.2357

```

In de “voorbehandelde cijfers” in figuur 10 is te zien dat de oorspronkelijke cijfers vanaf september 2001 tot begin 2002 opwaarts zijn aangepast voor de aanslagen in 2001. Hiermee rekening houdend laten seizoengecorrigeerde cijfers een beloop zien als wanneer dit niet voorgevallen zou zijn. Opvallend is de daling in de voorlopig seizoengecorrigeerde reeks begin 2003 (figuur 11). X-12 heeft dus voor de Sars uitbraak en Irak oorlog de cijfers niet automatisch gecorrigeerd. Derhalve zal *handmatig* een uitbijterdetectie hiervoor worden ingebracht.



*Figuur 10. Schiphol passagiers: voorlopig voorbehandelde en seizoengecorrigeerde reeks (x100)*



*Figuur 11. Voorlopige voorcorrecties oorspronkelijke %-jaarmutaties*

Het verder analyseren en modelleren van de oorspronkelijke cijfers leidt uiteindelijk tot een definitieve setup (tabel 9) van voorbehandeling ofwel pre-treatment. In dit geval wordt de tijdreeks nu inderdaad multiplicatief ontleed (tabel 10). Er is gecorrigeerd voor een vliegdagenpatroon en uitbijters waarvan sommige automatisch gedetecteerd zijn en andere user-defined ofwel handmatig ingebracht zijn (tabel 11). Bovendien is in dit voorbeeld volgens CBS praktijk het RegARIMA-model (ARIMA-model, vliegdagenpatroon, schrikkeljaareffect, uitbijters) bepaald op basis van een periode met complete jaren, modelspan genoemd. Toch blijkt er nog een residuele piek in het vliegdagenpatroon te bestaan (tabel 12; Spectrum G1 output X-12). X-12-ARIMA waarschuwt daarvoor in het `err`-bestand. Wanneer er

dan in het spectrum duidelijke pieken bij de trading day frequenties zichtbaar zijn (tabel 18). is er nog een werkdagenpatroon in de seizoengecorrigeerde reeks of de onregelmatige component zichtbaar. Als dit niet bij de pre-treatmentfase opgelost kan worden dan kan dit in de volgende stap van seizoencorrectie verholpen worden. De resultaten van de finale pre-treatment zijn in de volgende tabellen en grafieken weergegeven. De kalendercorrectie, hoewel significant, leidt echter niet tot een scherper beeld in de oorspronkelijke jaarmutaties. Zonder kalendercorrectie is de pre-treatment echter niet optimaal.

*Tabel 9. Finale X-12 pre-treatment setup*

```

1: SERIES{
3:           Format = "DateValue"
4:           Period = 12
5:           Modelspan = (,2008.12)
6:           File = "F:\Output\schipholpassagiers.DAT"
7:           name = "schiph"
8:           precision = 0
9:           decimals = 1}
11: TRANSFORM{
12:          Function = auto
13:          adjust = none
14:          savelog=atr}
16: FORCE{
17:          lambda = 0.0
18:          usefcst = yes}
20: OUTLIER{
21:          types = ( ao ls tc )
22:          critical = (3.0,3.9,3.6)
23:          method = addone}
25: REGRESSION{
26:          Variables =(
27:          tdnolpyear
28:          lpyear
29:          tc1999.12 ao2001.09 30: tc2001.10  ao2003.03  31:
tc2003.04 )
33:          aictest=(tdnolpyear)
34:          tcrate=(0.6) }
36: PICKMDL{
37:          file = J:\ImplSTS\Seizoen\VivX-12\cbst\Vivaldi0.6X-
120.3\Vivaldi\X-12\X-12a.mdl"
38:          mode = fcst
39:          method = best
40:          Identify = first }
42: X11{
43:          final=(ao tc ls)
44:          savelog=(m7 m10 m11 msr)}

```

*Tabel 10. Finale X-12 decompositiemethode*

```

Likelihood statistics for model fit to untransformed series.
Likelihood Statistics
Log likelihood (L)                -837.9286
AIC                               1705.8571
AICC (F-corrected-AIC)           1711.1318
Hannan Quinn                      1722.1100
BIC                               1745.9495
-----
Likelihood statistics for model fit to log transformed series.
Likelihood Statistics
Log likelihood                     289.9098
Transformation Adjustment         -1120.7069
Adjusted Log likelihood (L)       -830.7971
AIC                               1691.5942
AICC (F-corrected-AIC)           1696.8689
Hannan Quinn                      1707.8471
BIC                               1731.6866
-----
Multiplicative seasonal adjustment will be performed.

```



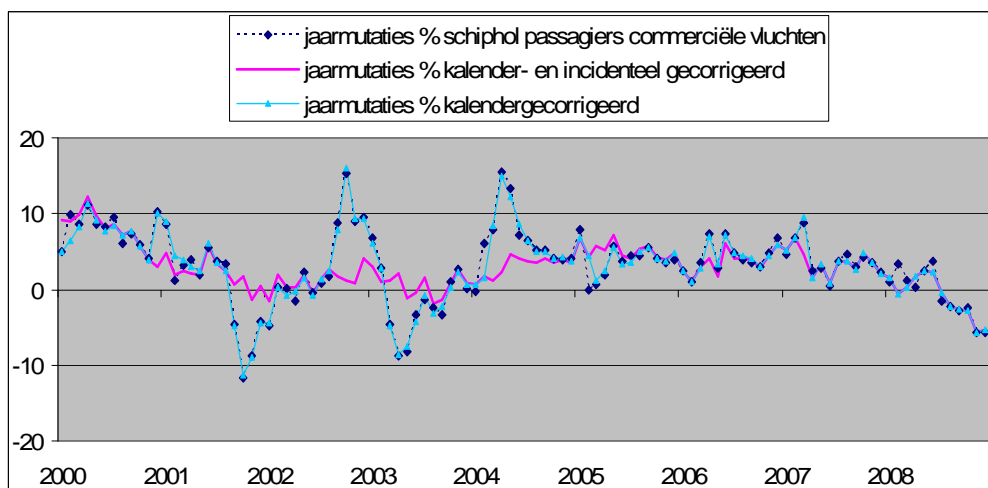
*Tabel 11. Finale X-12 pre-treatment: Regressie- en ARIMA-model*

Regression Model			
Variable	Parameter Estimate	Standard Error	t-value
Trading Day			
Mon	0.0030	0.00235	1.29
Tue	-0.0054	0.00235	-2.29
Wed	0.0015	0.00231	0.66
Thu	-0.0020	0.00233	-0.84
Fri	0.0026	0.00229	1.15
Sat	-0.0065	0.00229	-2.85
*Sun (derived)	0.0066	0.00227	2.93
Leap Year	0.0366	0.00685	5.35
TC1999.12	-0.0666	0.01333	-5.00
AO2001.09	-0.0578	0.01205	-4.80
TC2001.10	-0.1354	0.01338	-10.12
AO2003.03	-0.0692	0.01200	-5.77
TC2003.04	-0.1170	0.01341	-8.73
Automatically Identified Outliers			
TC2005.Mar	-0.0443	0.01226	-3.62
AO2006.Mar	-0.0465	0.01103	-4.21
*For full trading-day and stable seasonal effects, the derived parameter estimate is obtained indirectly as minus the sum of the directly estimated parameters that define the effect.			
Chi-squared Tests for Groups of Regressors			
Regression Effect	df	Chi-Square	P-Value
Trading Day	6	22.68	0.00
Combined Trading Day and Leap Year Regressors	7	56.96	0.00
ARIMA Model: (0 1 1)(0 1 1)			
Nonseasonal differences: 1			
Seasonal differences: 1			
Parameter	Estimate	Standard Errors	
Nonseasonal MA			
Lag 1	0.3372	0.09053	
Seasonal MA			
Lag 12	0.5331	0.08183	
Variance	0.20242E-03		
SE of Var	0.27675E-04		
Likelihood Statistics			
Number of observations (nobs)			120
Effective number of observations (nefobs)			107
Number of parameters estimated (np)			17
Log likelihood			301.1329
Transformation Adjustment			-1120.7069
Adjusted Log likelihood (L)			-819.5740
AIC			1673.1480
AICC (F-corrected-AIC)			1680.0244
Hannan Quinn			1691.5680
BIC			1718.5861

Tabel 12. *weekdag- en schrikkeljaarfactoren*

F 4. Multiplicative Trading Day Component Factors: Day of Week and Leap Year Factors							
Day of Week Component for regARIMA Trading Day Factors:							
	months starting on:						
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
31-day months	99.92	99.42	100.22	99.42	100.28	100.32	100.43
30-day months	99.77	99.62	99.96	100.07	99.61	100.01	100.97
Leap year Feb.	*****	99.46	*****	*****	100.26	*****	100.67
Nonseasonal component of length of month effect ("Leap Year" factors):							
31-day months	100.00						
30-day months	100.00						
Leap year Feb.	102.78						
Non-Leap Feb.	99.09						
Visually significant residual trading day peaks have been found in the spectral plots of the following series starting in 2001.Sep: differenced, transformed seasonally adjusted series (Table E2) (2 Trading Day peak(s))							

Door rekening te houden met de aanslagen in september 2001, de Sars uitbraak en Irak oorlog in 2003 is de onderliggende ontwikkeling scherp in beeld gebracht). De correctie voor deze twee speciale effecten laat zien waar het onderliggende dieptepunt in de oorspronkelijke jaarmutaties ligt, niet in 2001 of begin 2003 maar in de tweede helft van 2003 (figuur 12). Na 2000 had Schiphol namelijk ook last van de neergaande internationale conjunctuur.



Figuur 12. *Oorspronkelijke, kalender- en incidenteel gecorrigeerde %-jaarmutaties*

#### 5.6.4 Stap 4 Passagiersvervoer Schiphol - seizoencorrectie

Op basis van deze verfijnde finale pre-treatment is met default instellingen al een voorlopige seizoencorrectie gedaan. Net als bij de voorbereiding zijn hiervoor een aantal diagnostieken beschikbaar. Er zijn diverse toetsen en maten die aangeven of er sprake is van een (stabiel) seizoenpatroon in de jaren van de betrokken periode (between years) en binnen het jaar (between months) en wat de kwaliteit van de seizoencorrectie is. Spectra geven aan of er nog sprake is van een residuele seizoen- of werkdagenpatroon (zie Vollebregt, 2002) en ook zijn er revisiestatistieken. De seizoencorrectie van de passagiersaantallen met defaultinstellingen blijkt al goed te

zijn. De X-12-tabellen D.8.A, F2B, F3, de spectragrafieken G en revisietabellen R kunnen hiervoor geraadpleegd worden .

In het beschouwde voorbeeld zijn er geen problemen met de seizoencorrectie en de revisies van voor het eerst gecorrigeerde cijfers zijn naderhand over het algemeen kleiner dan 1% (tabel 18). In het spectrum G0 (tabel 16) treden in de voorbereekte reeks sterke pieken op wat duidelijk wijst op de aanwezigheid van een seizoenpatroon. In hoeverre maandelijkse veranderingen aan een seizoeninvloed kunnen worden toegeschreven kan worden afgelezen in tabel F2B van de X-12 uitvoer (tabel 13). Zoals opgemerkt zijn er in het spectrum G1 nog pieken in de trading day frequenties (tabel 18). Om dit probleem te verhelpen worden er meer extremen in de ruiscomponent vervangen met als resultaat dat er op de betreffende TD 0.348 frequentie in het spectrum geen piek meer is (tabel 19, 20 en 21).

*Tabel 13. Relatieve bijdrage componenten aan oorspronkelijke verandering*

F 2.B: Relative contributions to the variance of the percent change in the components of the original series								
Span in months	E3	D12	D10	A2	D18	TOTAL	RATIO (X100)	
	I	C	S	P	TD&H			
1	0.07	0.24	98.26	0.68	0.75	100.00	102.31	
3	0.01	0.33	99.13	0.43	0.10	100.00	99.66	
6	0.00	0.57	99.03	0.35	0.05	100.00	99.46	
12	0.26	80.93	0.13	16.30	2.38	100.00	114.00	

*Tabel 14. F-toetsen voor het bestaan van een seizoenpatroon*

D 8.A F-tests for seasonality					
Test for the presence of seasonality assuming stability.					
	Sum of	Dgrs.of	Mean		
	Squares	Freedom	Square	F-Value	
Between months	28872.2217	11	2624.74743	1717.959**	
Residual	177.2281	116	1.52783		
Total	29049.4498	127			
**Seasonality present at the 0.1 per cent level.					
Nonparametric Test for the Presence of Seasonality Assuming Stability					
	Kruskal-Wallis	Degrees of	Probability		
	Statistic	Freedom	Level		
	124.3913	11	0.000%		
Seasonality present at the one percent level.					
Moving Seasonality Test					
	Sum of	Dgrs.of	Mean		
	Squares	Freedom	Square	F-value	
Between Years	8.7906	9	0.976734	0.955	
Error	101.2136	99	1.022359		
No evidence of moving seasonality at the five percent level.					
COMBINED TEST FOR THE PRESENCE OF IDENTIFIABLE SEASONALITY					
IDENTIFIABLE SEASONALITY PRESENT					

*Tabel 15. Kwaliteitsmaten seizoencorrectie*

F 3. Monitoring and Quality Assessment Statistics	
All the measures below are in the range from 0 to 3 with an acceptance region from 0 to 1.	
1. The relative contribution of the irregular over three months span (from Table F 2.B).	M1 = 0.012
2. The relative contribution of the irregular component to the stationary portion of the variance (from Table F 2.F).	M2 = 0.012
3. The amount of month to month change in the irregular component as compared to the amount of month to month change in the trend-cycle (from Table F2.H).	M3 = 0.208
4. The amount of autocorrelation in the irregular as described by the average duration of run (Table F 2.D).	M4 = 0.000
5. The number of months it takes the change in the trend-cycle to surpass the amount of change in the irregular (from Table F 2.E).	M5 = 0.431
6. The amount of year to year change in the irregular as compared to the amount of year to year change in the seasonal (from Table F 2.H).	M6 = 0.130
7. The amount of moving seasonality present relative to the amount of stable seasonality (from Table F 2.I).	M7 = 0.054
8. The size of the fluctuations in the seasonal component throughout the whole series.	M8 = 0.103
9. The average linear movement in the seasonal component throughout the whole series.	M9 = 0.083
10. Same as 8, calculated for recent years only.	M10 = 0.113
11. Same as 9, calculated for recent years only.	M11 = 0.100
*** ACCEPTED *** at the level 0.11	

*Tabel 16. Statistiek gemiddelde revisies per kalendermaand*

R 1.S Summary statistics :	
average absolute percent revisions of the seasonal adjustments	
Date	Conc - Final
----	-----
<b>Months:</b>	
Jan	0.17
Feb	1.11
Mar	0.80
Apr	0.28
May	0.56
Jun	0.33
Jul	0.53
Aug	0.65
Sep	0.36
Oct	0.84
Nov	0.33
Dec	0.54
<b>Years:</b>	
2006	0.54
2007	0.39
2008	0.78



Tabel 18. Logspectrum van de seizoengecorrigeerde reeks

G 1 10*LOG(SPECTRUM) of the differenced, transformed Seasonally Adjusted Data (Table E2).						
Spectrum estimated from 2001.Sep to 2009.Aug.						
+++++++I	+++++++I	+++++++I	+++++++I	+++++++I	+++++++I	+++++++I
-34.37I			*			I -34.37
I			*			I
I			*			I
I			*			I
-35.43I			*		*	I -35.43
I			*		**	I
I			*		**	I
I			*	T	**	I
-36.49I			**	T	**	I -36.49
I			**	T	**	I
I	*		**	T	**	I
I	*	*	**	T	T **	I
-37.55I	**	*	**	T	T **	I -37.55
I	**	*	**	T	T **	I
I	**	*	**	T	T **	I
I	***	*	**	T	T **	I
I	***	**	*	T	T **	I
-38.60I	***	**	*	**	T	T ** I -38.60
I	***	**	*	**	T	T ** I
I	***	**	**	****	T *	T ** I
I	***	**	**	****	T *	T ** I
I	***	**	**	****	T *	T ** I
-39.66I	*****	**	**	****	T **	*T *** I -39.66
I	*****	**	**	****	T* **	*T *** I
I	*****	**	**	****	T* **	*T *** I
I	*****	**	**	****	T* **	*T *** I
-40.72I	*****	**	**	****	T* **	*T *** I -40.72
I	*****	**	**	****	T* **	*T *** I
I	*****	**	**	****	T* **	*T *** I
I	*****	**	**	****	T* **	*T *** I
-41.78I	*****	**	**	****	T* **	*T ***SI -41.78
I	*****	**	**	****	T* **	*T ***SI
I	*****	**	**	****	T* **	*T ***SI
I	*****	**	**	****	T* **	*T ***SI
-42.83I	*****	*	**	****	T* **	*T ***SI -42.83
I	*****	*	**	****	T* **	*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
-43.89I	*****	S**	**	****	*T* **	S*T ***SI -43.89
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
-44.95I	*****	S**	**	****	*T* **	S*T ***SI -44.95
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
I	*****	S**	**	****	*T* **	S*T ***SI
-46.01I	*****	S**	**	****	S*T* **	S*T ***SI -46.01
I	*****	S**	**	****	S*T* **	S*T ***SI
I	*****	S**	**	****	S*T* **	S*T ***SI
I	*****	S**	**	****	S*T* **	S*T ***SI
-47.06I	*****	S**	**	****	S*T* **	S*T ***SI -47.06
I	*****	S**	**	****	S*T* **	S*T ***SI
I	*****	S**	**	****	S*T* **	S*T ***SI
I	*****	S**	**	****	S*T* **	S*T ***SI
-48.12I	*****	S**	**	****	S*T* **	S*T ***SI -48.12
I	*****	S**	**	****	S*T* **	S*T ***SI
+++++++I	+++++++I	+++++++I	+++++++I	+++++++I	+++++++I	+++++++I
S=SEASONAL FREQUENCIES, T=TRADING DAY FREQUENCIES						
S-frequency	1/12	2/12	3/12	4/12	5/12	6/12
Period	12 months	6 months	4 months	3 months	2.4months	2months
TD frequency				0.348	0.432	

*Tabel 19. Finale correctiesetup passagiersvervoer Schiphol*

```

1: SERIES{
2:           title = "000000001schiphol finaal seizoencorrectie"
3:           Format = "DateValue"
4:           Period = 12
5:           Modelspan = (,2008.12)
6:           File = "F:\Output\schipholpassagiers.DAT"
7:           name = "schiph"
8:           precision = 0
9:           decimals = 1}
11: TRANSFORM{
12:           Function = log
13:           adjust = none
14:           savelog=atr}
16: FORCE{
17:           lambda = 0.0
18:           usefcst = yes}
20: OUTLIER{
21:           types = ( ao tc )
22:           critical = (3.0,3.9,3.6)
23:           method = addone}
25: REGRESSION{
26:           Variables =(
27:           tdnolpyear
28:           lpyear
29:           tcl999.12 ao2001.09tc2001.10 ao2003.03tc2003.04)
33:           aictest=(tdnolpyear)
           tcrate=(0.6)}
36: PICKMDL{
37:           file = "J:\ImplSTS\Seizoen\VivX-12\cbst\Vivaldi0.6X-120.3\Vivaldi\X-
12\X-12a.mdl"
38:           mode = fcst
39:           method = best
40:           Identify = first}
42: ESTIMATE{
45: FORECAST{
47:           Save = fct}
49: X11{
50:           mode = mult
51:           seasonalma = msr
52:           sigmalim=(0.8 1.6)
           final=(ao tc ls)
           savelog=(m7 m10 m11 msr)}
58: HISTORY{
           save=(sar)
           start=2006.01}

```

*Tabel 20. Decompositie vliegdagenpatroon en schrikkeljaarfactor*

F 4. Multiplicative Trading Day Component Factors:

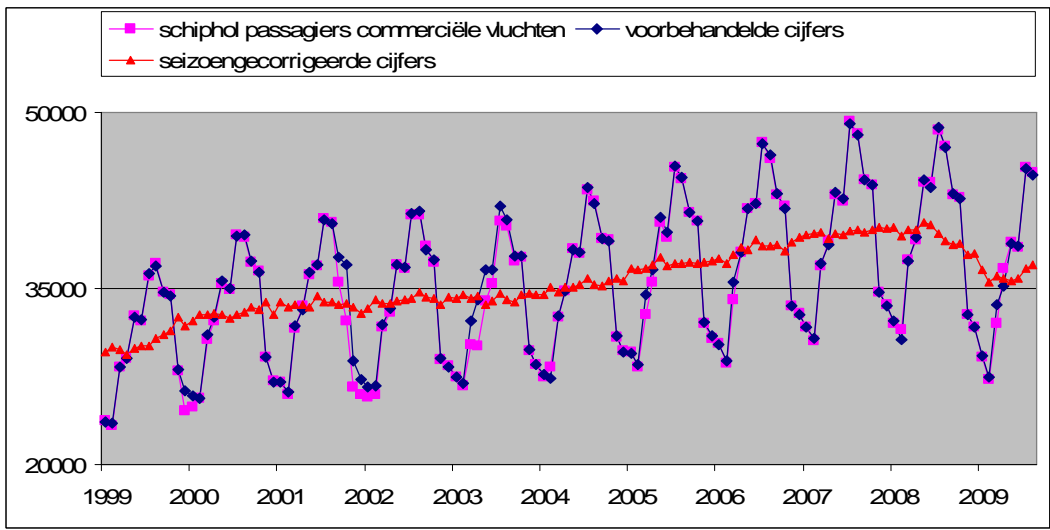
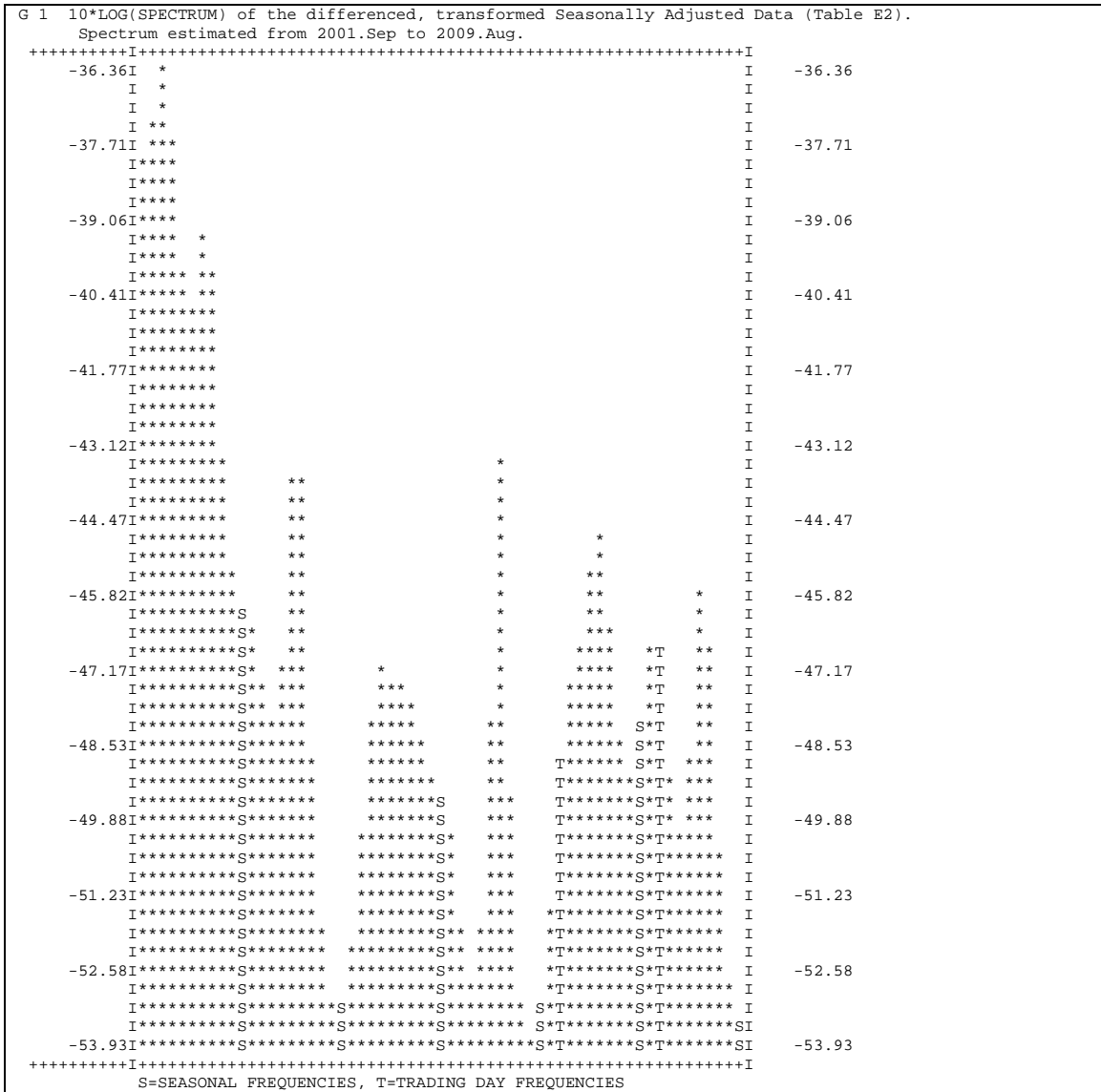
Day of Week Component for regARIMA Trading Day Factors:

	months starting on:						
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
31-day months	99.92	99.42	100.22	99.42	100.28	100.32	100.43
30-day months	99.77	99.62	99.96	100.07	99.61	100.01	100.97
Leap year Feb.	*****	99.46	*****	*****	100.26	*****	100.67

Nonseasonal component of length of month effect ("Leap Year" factors):

31-day months	100.00
30-day months	100.00
Leap year Feb.	102.78
Non-Leap Feb.	99.09

Tabel 21. Logspectrum van de seizoengecorrigeerde reeks



Figuur 13. Schiphol passagiers: definitief voorberekt en seizoengecorrigeerd (x100)



### 5.6.5 Het gekozen ARIMA-model (0,1,1) (0,1,1)

In dit deel wordt aan de hand van plots van ACF (Auto Correlation Function) en PACF (Partial Auto Correlation Function) in grote lijnen hoe X-12 het model (0,1,1)(0,1,1) heeft geselecteerd (dit model staat niet geheel toevallig ook bekend als het 'airline model'). Voor de passagiersaantallen 1999-2008 heeft dit model de kleinste voorspelfout. De eerste en belangrijkste stap bij het identificeren van het ARIMA-model is of een cijferreeks gedifferentieerd moet worden of niet en vervolgens, als er gedifferentieerd moet worden, hoeveel differenties hiervoor gebruikt moeten worden om de reeks stationair te maken. Er mag geen onder- of overdifferentiëring zijn. In X-12 kunnen plots van autocorrelaties van de residuen aangeven hoeveel differenties nodig zijn voor het 'seasonal'- en 'non seasonal'-deel.

In tabel 22 zijn de differentiatie specificaties voor de passagiersaantallen Schiphol vermeld. Voor het 'non seasonal'-deel wordt via de optie 'identify' gevraagd om eerst geen differentie toe te passen, vervolgens 1 en 2 differenties te gebruiken en ten slotte voor het 'seasonal'-deel geen en 1 differentie.

Tabel 22. Identify

```
60: IDENTIFY{
61:           diff=(0,2)
62: sdiff=(1,2)
63: maxlag=24
64: print=(none +acf +acfplot +pacf +pacfplot rgc)}
```

In tabel 23 zijn de residuen op de eerste lag en tot na de 10<sup>de</sup> lag sterk positief gecorreleerd aan elkaar. Verder zijn er ook op de 12<sup>de</sup> en 24<sup>ste</sup> lag grote pieken wat wijst op de aanwezigheid van een seizoenpatroon. Dit geeft aan dat de cijferreeks niet stationair is en gedifferentieerd moet worden voor zowel het 'seasonal'- als het 'non seasonal'-deel. Nu is het zaak om het juiste aantal differenties te vinden. Dat wordt onderzocht met 1 en 2 differenties voor het 'non seasonal' en het 'seasonal' deel. Meer dan 1 differentie gebruiken is over het algemeen niet nodig. TRAMO heeft in het meest ingewikkelde model 2 differenties voor het 'non seasonal'-deel (3,2,3). In de "pick model"-verzameling van X-12 is er een ARIMA-model met 2 differenties (0,2,2). Voor het 'seasonal'-deel wordt afgeraden om meer dan 1 differentie te gebruiken.

Tabel 23. Differencing: None

Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	0.80	0.54	0.19	-0.17	-0.36	-0.47	-0.36	-0.16	0.18	0.51	0.71	0.84
SE	0.09	0.14	0.15	0.16	0.16	0.16	0.18	0.18	0.18	0.18	0.20	0.22
Q	79.53115	15.56120	5.05123	12.54139	19.97167	26.83184	33.187	39.69192	46.04226	52.79295	59.15391	65.17
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	0.66	0.42	0.11	-0.21	-0.38	-0.48	-0.38	-0.20	0.11	0.41	0.59	0.70
SE	0.24	0.26	0.26	0.26	0.26	0.27	0.28	0.28	0.28	0.28	0.29	0.30
Q	451.21475	60477.16483	44504.26536	85557.54563	39565.21590	04642.32717	21					
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												0.804
2												0.539
3												0.189
4												-0.166
5												-0.359
6												-0.466
7												-0.357
8												-0.160
9												0.182
10												0.511
11												0.713
12												0.842
13												0.662
14												0.420
15												0.106
16												-0.211
17												-0.383
18												-0.477
19												-0.378
20												-0.200
21												0.111
22												0.408
23												0.589
24												0.701

In tabel 24 zijn de residuen opgenomen van 1 differentie. Een differentie van de eerste orde is adequaat om de reeks stationair te maken als op de eerste lag de autocorrelatie van de residuen niet gelijk aan 0.0 of negatief is en als geen aanhoudende positieve autocorrelaties meer zijn na het aantal lags van 10 of meer. Uit de plot blijkt dat dit het geval is en voldoet dus 1 differentie om het 'non seasonal'-deel stationair te maken

Tabel 24. Differencing: Nonseasonal Order=1

Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	0.20	0.32	0.02	-0.43	-0.25	-0.60	-0.25	-0.45	0.03	0.29	0.19	0.88
SE	0.09	0.10	0.10	0.10	0.12	0.12	0.14	0.15	0.16	0.16	0.16	0.17
Q	4.88	17.55	17.62	40.52	48.27	93.81	101.61	127.47	127.58	138.80	143.82	248.13
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.027	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	0.18	0.29	0.02	-0.38	-0.23	-0.53	-0.23	-0.41	0.03	0.26	0.18	0.78
SE	0.20	0.20	0.21	0.21	0.21	0.21	0.22	0.23	0.23	0.23	0.23	0.24
Q	252.56	263.91	263.96	284.33	291.65	332.03	339.35	363.60	363.71	373.40	378.06	469.63
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												0.200
2												0.321
3												0.024
4												-0.428
5												-0.248
6												-0.598
7												-0.246
8												-0.446
9												0.029
10												0.291
11												0.194
12												0.880
13												0.181
14												0.288
15												0.019
16												-0.382
17												-0.228
18												-0.532
19												-0.225
20												-0.408
21												0.028
22												0.255
23												0.176
24												0.777

Maar in de plot van tabel 24 is ook te zien dat er op de seizoenlags 12 en 24 nog steeds grote (afnemende) pieken voorkomen. Dit komt omdat er nog geen differentie voor 'seasonal' deel is gebruikt. In tabel 25 zijn er 2 differenties gebruikt voor het 'non seasonal' deel. Op de eerste lag is er sprake van een grote negatieve autocorrelatie, meer dan -0.5. Dit geeft aan dat 2 differenties teveel is, er is sprake van overdifferentiatie. Geconcludeerd kan worden dat X-12 voor het 'non seasonal'-deel terecht voor 1 differentie gekozen heeft om de reeks stationair te maken (p,d,q=0,1,1). X-12 heeft ook voor het stationair maken van het 'seasonal'-deel, aangegeven door de pieken op de seizoenlags, 1 differentie gekozen deel. Dit wordt aangetoond met behulp van tabel 26 en 27.

*Table 25. Differencing: Nonseasonal Order=2*

Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	-0.58	0.26	0.10	-0.39	0.33	-0.44	0.35	-0.42	0.13	0.22	-0.49	0.87
SE	0.09	0.12	0.12	0.12	0.13	0.14	0.15	0.16	0.17	0.17	0.17	0.18
Q	40.57	49.11	50.25	69.58	83.56	108.12	123.44	146.45	148.79	155.25	186.99	287.35
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	-0.51	0.24	0.08	-0.35	0.29	-0.38	0.31	-0.39	0.13	0.19	-0.42	0.76
SE	0.22	0.23	0.23	0.23	0.23	0.23	0.24	0.24	0.25	0.25	0.25	0.26
Q	321.98	329.67	330.59	347.22	358.96	379.92	393.47	415.23	417.77	423.04	449.95	536.60
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												-0.579
2												0.265
3												0.096
4												-0.395
5												0.334
6												-0.441
7												0.346
8												-0.423
9												0.134
10												0.222
11												-0.490
12												0.867
13												-0.507
14												0.238
15												0.082
16												-0.346
17												0.289
18												-0.385
19												0.308
20												-0.388
21												0.132
22												0.189
23												-0.425
24												0.758

De plot in tabel 26 laat duidelijk zien dat de pieken op de seizoenlags volledig verdwenen zijn bij 1 differentie. Tabel 27 waar nog gerekend is met 2 seasonal differenties laat een grote negatieve autocorrelatie zien op de seizoenlag van 12, en dit ligt buiten de bandbreedte van ongecorrleerde residuen. Ten slotte geeft tabel 28 een plot van de residuen bij 1 differentie voor zowel het seasonal als het non seasonal deel, zoals door X-12 bevonden is voor de passagiersaantallen. Er is geen sprake van dat er in de residuen nog een patroon van autocorrelatie bestaat. De negatieve autocorrelatie op de eerste lag is kleiner dan -0.5, dus er is geen sprake van overdifferentiëring.

Tabel 26. Differencing: Seasonal Order=1

Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	0.76	0.64	0.54	0.43	0.35	0.29	0.28	0.26	0.18	0.10	0.05	-0.07
SE	0.10	0.14	0.17	0.18	0.19	0.20	0.20	0.20	0.21	0.21	0.21	0.21
Q	64.28110	52143.41164	41164.64178	73188.72197	91205.72209	69210.94211	31211.84					
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	-0.06	-0.02	0.00	-0.02	-0.05	-0.09	-0.16	-0.26	-0.27	-0.20	-0.18	-0.16
SE	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.22	0.22	0.22
Q	212.26212	31212.31212	39212.39212	74213.82217	26226.48236	36242.17246	75250.24					
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												0.761
2												0.642
3												0.539
4												0.431
5												0.350
6												0.293
7												0.279
8												0.256
9												0.182
10												0.102
11												0.055
12						XX						-0.066
13						X						-0.058
14						X						-0.020
15												0.004
16						X						-0.024
17						X						-0.052
18						XX						-0.091
19						XXXX						-0.161
20						XXXXXXXX						-0.261
21						XXXXXXXX						-0.269
22						XXXXXX						-0.205
23						XXXXXX						-0.181
24						XXXX						-0.157

*Table 27. Differencing: Non seasonal Order=1, Seasonal Order=2*

Differencing: Nonseasonal Order=1, Seasonal Order=2												
Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	-0.33	-0.01	-0.02	-0.12	-0.03	-0.02	-0.03	0.23	-0.01	-0.11	0.31	-0.54
SE	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.13
Q	10.45	10.46	10.50	11.88	11.98	12.02	12.09	17.90	17.91	19.13	29.96	62.90
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.001	0.005	0.015	0.018	0.035	0.062	0.098	0.022	0.036	0.039	0.002	0.000

Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	0.03	0.05	0.10	0.05	-0.01	0.02	0.13	-0.23	-0.04	0.10	-0.09	0.05
SE	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16
Q	63.03	63.31	64.37	64.65	64.66	64.69	66.66	73.16	73.36	74.64	75.77	76.07
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1						XXXXXXXXX						-0.326
2						.						-0.011
3						.						-0.019
4						.	XXX					-0.117
5						.	X					-0.031
6						.						-0.019
7						.	X					-0.025
8						.		XXXXXX				0.234
9						.						-0.013
10						.	XXX					-0.106
11						.		XXXXXXXXX				0.314
12	- - - -					XXXXXXXXXXXXXXXXX						-0.545
13						.	X					0.034
14						.	X					0.050
15						.	XX					0.096
16						.	X					0.049
17						.						-0.009
18						.						0.017
19						.		XXX				0.127
20						.	XXXXXX					-0.230
21						.	X					-0.040
22						.		XXX				0.100
23						.	XX					-0.094
24	- - - -					.	X					0.048

Tabel 28. Differencing: Non seasonal Order=1, Seasonal Order=1

Sample Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	-0.33	-0.02	0.00	-0.07	0.00	-0.03	0.02	0.12	0.01	-0.06	0.19	-0.28
SE	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Q	11.84	11.86	11.87	12.37	12.37	12.47	12.52	14.36	14.36	14.83	19.17	29.05
DF	1	2	3	4	5	6	7	8	9	10	11	12
P	0.001	0.003	0.008	0.015	0.030	0.052	0.085	0.073	0.110	0.138	0.058	0.004

Lag	13	14	15	16	17	18	19	20	21	22	23	24
ACF	-0.07	0.10	0.09	0.02	-0.01	0.02	0.15	-0.24	-0.03	0.08	0.00	-0.08
SE	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13
Q	29.62	30.76	31.71	31.74	31.76	31.81	34.96	43.00	43.11	44.00	44.00	44.95
DF	13	14	15	16	17	18	19	20	21	22	23	24
P	0.005	0.006	0.007	0.011	0.016	0.023	0.014	0.002	0.003	0.004	0.005	0.006

The P-values approximate the probability of observing a Q-value at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Sample Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												-0.328
2						XXXXXXXX						-0.015
3						.						0.002
4						. XX						-0.067
5						.						-0.002
6						. X						-0.030
7						.						0.020
8						.	XXX					0.125
9						.						0.007
10						. XX						-0.063
11						.	XXXXX					0.189
12						XXXXXXXX						-0.284
13						. XX						-0.068
14						.	XX					0.096
15						.	XX					0.086
16						.						0.016
17						.						-0.010
18						.	X					0.021
19						.	XXXX					0.154
20						XXXXXX						-0.245
21						.	X					-0.029
22						.	XX					0.081
23						.						0.002
24						. XX						-0.082

De volgende stap in het identificeren van het ARIMA-model is het aantal AR en MA termen die nodig zijn om nog eventuele resterende autocorrelatie in de gedifferentieerde reeks te verwijderen. De PACF plot (tabel 29) van de gedifferentieerde reeks geeft een indicatie voor het aantal AR en MA termen. Een AR-term is nodig als op de 1<sup>ste</sup> lag sprake is van een grote piek en/of van een abrupte afname in het pieken-niveau. Van beide is geen sprake in de PACF plot, dus er zijn geen AR termen nodig. Hoewel in principe rest-autocorrelatie met een AR term verwijderd kan worden, is het aanwenden van een MA een betere oplossing. Dan moet naar de ACF plot van de gedifferentieerde reeks gekeken worden. Als er een negatieve piek is op de 1<sup>ste</sup> en/of op de 12<sup>de</sup> lag, een aanwijzing dat de reeks iets overgedifferentieerd is, dan is er sprake van rest-autocorrelatie. Dit is in tabel 27 inderdaad het geval. Hier kan de resterende autocorrelatie liever met een MA dan met AR term gecorrigeerd worden. Hoeveel MA termen nodig zijn wordt bepaald door de lag waarbij sprake is van een snelle afname van de autocorrelatie, in het pieken-niveau dus. De plot geeft aan dat dit op de 2<sup>de</sup> lag is van het 'seasonal' en 'non-seasonal'-deel, dus er is maar 1 MA term nodig. Dit komt exact overeen met het airlinemodel gekozen voor de passagiersaantallen door X-12: (p,d,q =0,1,1) en (P,D,Q= 0,1,1). In het model zijn er geen AR termen opgenomen, wel 1 MA term.

*Table 29. Differencing: Non seasonal Order=1, Seasonal Order=1*

Sample Partial Autocorrelations of the Residuals												
Lag	1	2	3	4	5	6	7	8	9	10	11	12
PACF	-0.26	-0.08	-0.07	-0.14	-0.03	-0.08	-0.06	0.09	0.10	-0.02	0.19	-0.25
SE	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Lag	13	14	15	16	17	18	19	20	21	22	23	24
PACF	-0.21	-0.02	0.09	0.07	0.05	0.07	0.22	-0.08	-0.09	0.00	0.01	-0.17
SE	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Lag	25	26	27	28	29	30	31	32	33	34	35	36
PACF	0.04	0.06	-0.09	0.01	0.05	0.08	0.03	-0.10	0.00	0.09	0.10	-0.06
SE	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10

Sample Partial Autocorrelations of the Residuals												
	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	
1												-0.257
2												-0.080
3												-0.074
4												-0.135
5												-0.029
6												-0.083
7												-0.059
8												0.085
9												0.104
10												-0.019
11												0.191
12												-0.252
13												-0.214
14												-0.024
15												0.090
16												0.068
17												0.051
18												0.067
19												0.221
20												-0.085
21												-0.088
22												0.005
23												0.009
24												-0.172



## 6. Literatuur<sup>1</sup>

- Akaike, H. (1973), *Information Theory and an Extension of the Likelihood Principle*. Second International Symposium on Information Theory, Budapest. Akademia Kiado, Boedapest, pp. 267-287.
- Bell, W. R. and S. C. Hillmer (1983), Modeling time series with calendar variation. *Journal of the American Statistical Association* 78, 526–534.
- Booleman, M. (2003), *CBS-standpunten over seizoencorrectie*. Interne CBS-memo.
- Boset, C. (2000), *Vivaldi beschrijving*. Intern CBS-rapport.
- Box, G.E.P. and Jenkins, G.M (1970), *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco.
- Brakel, J.A. van den, Kraan, T.C. en Roels, J. (2010), *Methodebreuken*. Rapport Methodenreeks, CBS, Heerlen.
- Eurostat (2009), *ESS guidelines on seasonal adjustment*.
- Ladiray, D., Quenneville, B. (2001), *Seasonal Adjustment with the X-11 method*. Springer-Verlag, New York.
- Lothian, J. and Morry. M. (1978), *A set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*. Working Paper 78-10.
- Ouwehand, P. en T. Kraan (2009), *Inventarisatie seizoencorrectie – een overzicht van tijdreeksen en methoden binnen het CBS*. Intern CBS-rapport, DMHV-2009-08-04-POWD-TKRN.
- U.S. Census Bureau (2007), *X-12-ARIMA Reference Manual*, Version 0.3, section 5.5. Time Series Staff Statistical Research Division, U.S. Bureau of the Census, Washington D.C. ([http://www.census.gov/srd/www/x12a/x12down\\_pc.html](http://www.census.gov/srd/www/x12a/x12down_pc.html)).
- Vollebregt M, (2002), *Syllabus bij de cursus Seizoencorrectie met X-12\_ARIMA en Vivaldi*. Intern CBS-rapport.

---

<sup>1</sup> Een aantal van deze documenten zijn te vinden op <http://cbsh1sps/sites/TmoMeth/Seizoencorrectie>.