



Centraal Bureau voor de Statistiek

Divisie Macro-economische statistieken en publicaties (MSP)
Sector Prijzen, conjunctuur en programmazaken (MPP)

CORRECTIE VOOR FEEST- EN BRUGDAGEN

Michiel Jansen

1. Inleiding

Seizoencorrecties worden binnen het CBS uitgevoerd met het programma X-12-ARIMA. Naast correcties voor seizoenen, biedt X-12-ARIMA ook de mogelijkheid correcties te doen voor andere effecten, zoals werkdagenpatronen, vorstverlet, vakantiespreiding en feest- en brugdagen. Dit gebeurt met behulp van regressiemodellen. Een zelfde effect kan meestal op verschillende manieren gemodelleerd worden. Eerder werd onderzoek gedaan naar methoden om vakantiespreiding te modelleren.¹ Het onderzoek heeft een aantal inzichten opgeleverd die toegepast kunnen worden in andere regressiemodellen. In dit verslag worden de inzichten gebruikt om de meest geschikte methode voor het corrigeren van feest- en brugdagen te introduceren.

Tot nu toe wordt op het CBS een feest- en brugdagcorrectie toegepast, waarbij in het regressiemodel voor iedere maand een absoluut aantal feest- en brugdagen is opgenomen. Door de regressor worden cijfers van maanden waarin een feest- of brugdag is gemodelleerd, naar boven toe bijgesteld. Het effect is sterker naarmate er meer feestdagen in een maand zijn. Feitelijk is dit wat we willen met een feest- en brugdagregressor. Echter, door de correctie wordt het niveau van de reeks omhooggetrokken. Dit geeft problemen bij het rechtektrekken van de cijfers in de gecorrigeerde reeks aan de jaartotalen van de oorspronkelijke reeks, hetgeen voor

¹ Michel Vollebregt, Correctie voor vakantiespreiding. Interne CBS-nota, 2003.

veel Eurostat statistieken een vereiste is. De verschillen in jaartotalen tussen de oorspronkelijke en de werkdaggecorrigeerde reeks kunnen behoorlijk groot zijn. Dit betekent dat na afloop van een verslagjaar door de rechtekking alle werkdaggecorrigeerde cijfers aanzienlijk omlaag gebracht worden. Voor lopende jaren kan echter geen rechtekking plaatsvinden omdat er nog geen jaartotaal bekend is. In de berekening van jaar-op-jaar mutaties levert dit problemen op omdat na afloop van een verslagjaar alle mutaties van dat jaar in één keer bijgesteld worden. Grote bijstellingen dienen echter zoveel mogelijk voorkomen te worden.

De bijstellingsproblemen hebben tot het inzicht geleid dat regressiemodellen idealiter gebaseerd moeten zijn op gemiddelden. Regressievariabelen dienen te bestaan uit een verzameling positieve en negatieve getallen waarvan de som bij benadering op nul uitkomt. Hierdoor blijft het niveau van de reeks intact en zijn verschillen in jaartotalen tussen de oorspronkelijke en de werkdaggecorrigeerde reeks minimaal. Bijstellingen als gevolg van rechtekkingen blijven dientengevolge eveneens beperkt.

Vollebregt heeft in zijn onderzoek naar vakantiespreidingcorrectie aangetoond dat methoden om gemiddeldes te berekenen nagenoeg tot dezelfde uitkomsten leiden. Hij beveelt aan de meest eenvoudige methode te gebruiken, namelijk op basis van een structureel gemiddelde waarbij we zoveel mogelijk beschikbare informatie gebruiken. Door het gebruik van een structureel gemiddelde veranderen de regressievariabelen niet door de tijd. In dit verslag wordt uiteengezet hoe het structureel gemiddelde voor de feest- en brugdagregressor berekend is. Daarbij is de feest- en brugdagdefinitie identiek aan het tot nog toe gebruikte regressiemodel op basis van het absolute aantal vakantiedagen.

2. Feest- en brugdagen

De volgende feest- en brugdagdefinities zijn gehanteerd. Feestdagen worden gemodelleerd als ze niet op zaterdag of zondag vallen. Hierdoor wordt Hemelvaartdag altijd gemodelleerd (valt immers altijd op een donderdag) en eerste paasdag nooit. Brugdagen worden alleen gemodelleerd als een feestdag door de week valt en de dag ervoor of erna een maandag of vrijdag is. Zo wordt de vrijdag na Hemelvaart altijd als brugdag berekend.

We kennen de volgende feestdagen die door de week vallen:

Feestdag	Datum
Nieuwjaarsdag	1 januari
Tweede paasdag	Varieert
Koninginnedag	30 april
Hemelvaartdag	Varieert
Tweede Pinksterdag	Varieert
Eerste kerstdag	25 december
Tweede kerstdag	26 december

Voor brugdagen kan een soortgelijk overzicht gemaakt worden:

Bijbehorende feestdag	Datum brugdag
Nieuwjaarsdag op dinsdag of donderdag	31 december resp. 2 januari
Hemelvaart	Varieert
Koninginnedag op dinsdag of donderdag	29 april resp. 1 mei
Eerste kerstdag op dinsdag of woensdag	24 december resp. 27 december

3. De berekening van het structureel gemiddelde

Voor alle genoemde feest- en brugdagen kunnen we een gemiddelde per maand berekenen. Het absolute aantal feestdagen in de maand wordt dan verrekend met het gemiddelde. De uitkomst wordt opgenomen in het regressiemodel. Voor iedere feest- of brugdag geldt een andere berekening.

Een tweetal opmerkingen is van belang:

- Een kalendercyclus duurt in totaal 28 jaar. Dit wil zeggen dat een feestdag die op een vaste datum valt, in elke periode van 28 jaar met dezelfde frequentie op alle zeven weekdays is gevallen. De berekening is als volgt: elke vier jaar is er een schrikkeljaar en je hebt zeven dagen per week. Met één dagverschuiving per jaar kom je dan op het kleinste gemene veelvoud van $4 \times 7 = 28$ uit. Met niet-schrikkeljaren op hele eeuwen wordt dan geen rekening gehouden.
- De duur van een volledige (Gregoriaanse) paascyclus is precies 5,7 miljoen jaar. Na 5,7 miljoen jaren is de hele cyclus voltooid en herhaalt de cyclus zich. Pasen, Pinksteren, en Hemelvaart zijn feestdagen die allemaal direct afhankelijk zijn van de paascyclus.

Met de kalendercycli in het achterhoofd kunnen we een soort eeuwigdurende gemiddelde voor alle feest- en brugdagen berekenen. Daarbij is bekend dat de feest-

en brugdagen die we meenemen nooit in februari, juli, augustus, september, oktober en november vallen. Voor deze maanden is het gemiddelde nul evenals de waarde van deze maand in het regressiemodel. Voor de overige maanden hebben we de volgende gemiddelde frequenties berekend:

Maand	Gemiddelde feestdag	Gemiddelde brugdag	Totaal
Januari	0,71429	0,14286	0,85715
Maart	0,20000	0,00000	0,20000
April	1,51429	0,14286	1,65715
Mei	1,55905	1,07816	2,63721
Juni	0,44095	0,06470	0,50565
December	1,42857	0,42857	1,85714

Een precieze berekening van deze aantallen staat in bijlage 1.

4. Resultaten

De nieuwe regressoren zijn getoetst op een aantal belangrijke reeksen van de productie-index industrie. We beschouwen de jaartotalen voor 2002 volgens de oude (met een modellering op grond van het werkelijke aantal feestdagen) en de nieuwe methode van modellering. In kolom 2 en 3 worden respectievelijk de gebruikte regressor en de bijbehorende t-waarde gegeven. De resultaten zijn als volgt:

Tabel: jaartotalen 2002 volgens oude en nieuwe methode van modellering

Sbi	Rgr ²	t-wrd	2002	Oud		Nieuw	
			Oorspr	Gm dag	Seizoen	Gm dag	Seizoen
15-37	Fb	-7,87	1170,2	1186,3	1186,4	1173,5	1173,7
15+16	F	-3,36	1204,9	1214,7	1215,3	1205,8	1206,3
17-19	Fb	-4,87	1077,9	1100,4	1100,6	1084,2	1084,8
21+22	Fb	-6,26	1144,1	1164,3	1144,8	1147,9	1144,2
23-25	F	-4,97	1247,2	1258,8	1258,8	1250,1	1250,2
27-35	Fb	-7,53	1125,4	1145,8	1145,9	1128,6	1128,6
20+26+36+37	F	-26,88	1158,4	1192,0	1191,8	1158,7	1158,5

² Toegepaste regressievariabele(n). F = feestdagen, FB = feest/brugdagen

Uit bovenstaande tabel blijkt dat de jaartotalen volgens de nieuwe modellering meestal dichterbij het oorspronkelijke jaartotaal liggen dan volgens de oude modellering. Bij de oude methode zijn de relatieve afwijkingen van de jaartotalen groter naarmate de bijbehorende t-waarde hoger is. Zo wordt sbi 20+26+36+37 na afloop van het jaar 2002 door de rechtekking met gemiddeld 3 procent bijgesteld. Via de nieuwe methode zouden de bijstellingen door rechtekking na afloop van het jaar 2002 veel beperkter zijn geweest en daarmee ook de aanpassingen van de jaar-op-jaar mutaties van 2002. Eenzelfde overzicht kan ook worden gemaakt voor het lopende jaar. Voor 2003 zijn cijfers tot en met oktober beschikbaar en de resultaten zijn als volgt:

Tabel: jaartotalen 2003 (tot en met oktober) volgens oude en nieuwe methode van modellering

Sbi	Rgr	t-wrd	Oorspr	Oud		Nieuw	
				Gm dag	Seizoen	Gm dag	Seizoen
15-37	Fb	-7,87	953,3	963,0	966,8	952,2	955,5
15+16	F	-3,36	975,2	980,3	993,0	974,1	985,8
17-19	Fb	-4,87	862,2	876,3	873,3	864,3	858,2
21+22	Fb	-6,26	920,1	931,2	932,9	918	930,4
23-25	F	-4,97	1078,0	1084,8	1078,0	1077,0	1068,8
27-35	Fb	-7,53	901,9	915,48	917,72	900,9	904,5
20+26+36+37	F	-26,88	923,2	944,7	944,6	922,9	921,4

Ook hier liggen de jaartotalen van de werkdag- en seizoen gecorrigeerde cijfers volgens de nieuwe methodiek veel dichterbij de oorspronkelijke jaartotalen dan volgens de oude methodiek. Bijstellingen door rechtekking na afloop van 2003 zullen volgens de nieuwe methodiek dan ook een stuk kleiner zijn.

5. Conclusie

De nieuwe methode van modelleren gebaseerd op een structureel gemiddelde levert betere resultaten op dan de oude methode, welke gebaseerd was op het absolute aantal vakantiedagen. De bijstellingen als gevolg van de rechtekking aan het einde van het jaar zullen kleiner zijn, zowel voor de werkdage gecorrigeerde als voor de seizoen gecorrigeerde cijfers. We raden daarom aan de oude feest- en brugdag regressor niet meer te gebruiken en te vervangen door de nieuwe regressor, gebaseerd op een structureel gemiddelde.

Bijlage 1

Hieronder worden de berekeningen gegeven van de structurele gemiddeldes volgens de nieuwe methode van modellering. Het werkelijke aantal feest- of brugdagen wordt vervolgens verrekend met deze gemiddeldes. Overigens wordt brugdagcorrectie altijd toegepast in combinatie met feestdagcorrectie, terwijl feestdagcorrectie ook afzonderlijk kan plaatsvinden. Het structureel gemiddelde voor de feest/brugdagregressor wordt berekend als som van beide gemiddeldes.

Feestdagen

Januari: 5/7 (nieuwjaarsdag)

Februari: 0

Maart: 1.140.000/5.700.000 (tweede paasdag)

April: 4.560.000/5.700.000 (tweede paasdag) + 5/7 (koninginnedag)

Mei: 3.417.625/5.700.000 (tweede pinksterdag) + 5.468.950/5.700.000 (hemelvaart)

Juni: 2.282.375/5.700.000 (tweede pinksterdag) + 231.050/5.700.000 (hemelvaart)

Juli tot en met november: 0

December: 10/7 (eerste en tweede kerstdag)

Brugdagen

Januari: 1/7 (nieuwjaarsdag op donderdag)

Februari + maart: 0

April: 1/7 (koninginnedag op dinsdag)

Mei: 1/7 (koninginnedag op donderdag) + 5.331.200/5.700.000 (dag na hemelvaart)

Juni: 368.800/5.700.000 (dag na hemelvaart)

Juli tot en met november: 0

December: 1/7 (nieuwjaarsdag op dinsdag) + 2/7 (eerste kerstdag op dinsdag of woensdag)

Paasdagencyclus³

<u>Maand</u>	<u>Dag</u>	<u>Frequentie</u>	<u>Maand</u>	<u>Dag</u>	<u>Frequentie</u>
Maart	21	0	April	09	186200
Maart	22	27550	April	10	192850
Maart	23	54150	April	11	186200
Maart	24	81225	April	12	192850
Maart	25	110200	April	13	189525
Maart	26	133000	April	14	189525
Maart	27	165300	April	15	192850
Maart	28	186200	April	16	186200
Maart	29	192850	April	17	192850
Maart	30	189525	April	18	197400
Maart	31	189525	April	19	220400
April	01	192850	April	20	189525
April	02	186200	April	21	162450
April	03	192850	April	22	137750
April	04	186200	April	23	106400
April	05	192850	April	24	82650
April	06	189525	April	25	42000
April	07	189525	April	26	0
April	08	192850			

Toelichting: de paasdagencyclus is gebaseerd op 5,7 miljoen jaren. In het overzicht staan de frequenties van Eerste paasdag vermeld voor alle mogelijke data waarop Pasen kan vallen.

Eerste paasdag valt op de eerste zondag ná de eerste volle maan óp of ná het begin van de lente. Pinksteren is 7 weken na Pasen. Hemelvaart is 10 dagen voor Pinksteren.

³ Bron: <http://members.lycos.nl/bouwzelf/paasdata.htm>

*Voorbeeld berekening regressievariabelen voor gecombineerde feest- en
brugdagcorrectie, 1990*

	Aantal feest/brugdagen	Structureel gemiddelde	Regressievariabele
Jan	1	0,85714	0,14286
Feb	0	0	0,00000
Mrt	0	0,20000	-0,20000
Apr	2	1,65715	0,34285
Mei	2	2,63721	-0,63721
Jun	1	0,50565	0,49435
Jul	0	0	0,00000
Aug	0	0	0,00000
Sep	0	0	0,00000
Okt	0	0	0,00000
Nov	0	0	0,00000
Dec	4	1,85714	2,14286